

# UC Santa Cruz

## UC Santa Cruz Electronic Theses and Dissertations

### Title

An experimental investigation of phonetic naturalness

### Permalink

<https://escholarship.org/uc/item/94x407sb>

### Author

Greenwood, Anna

### Publication Date

2016

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
SANTA CRUZ

**AN EXPERIMENTAL INVESTIGATION OF PHONETIC  
NATURALNESS**

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

DOCTOR OF PHILOSOPHY

in

LINGUISTICS

by

**Anna Greenwood**

June 2016

The Dissertation of Anna Greenwood  
is approved:

---

Professor Grant McGuire, Chair

---

Professor Jaye Padgett, Chair

---

Professor Ryan Bennett

---

Dean Tyrus Miller  
Vice Provost and Dean of Graduate Studies

Copyright © by  
Anna Greenwood  
2016

# Table of Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>Abstract</b>	<b>x</b>
<b>Dedication</b>	<b>xii</b>
<b>Acknowledgments</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Channel bias and universal learning bias . . . . .	3
1.1.1 Examples of naturalness effects in three phonological domains . .	4
1.2 An experimental framework . . . . .	6
<b>2 The channel bias approach</b>	<b>10</b>
2.1 A channel bias explanation . . . . .	10
2.1.1 Weight-sensitive stress . . . . .	10
2.1.2 Final devoicing . . . . .	13
2.1.3 Coda sonority . . . . .	14
2.2 Phonologization of phonetically natural patterns . . . . .	17
2.3 Summary . . . . .	20
<b>3 The universal learning bias approach</b>	<b>21</b>
3.1 A closer look at cognitive biases . . . . .	21
3.2 A universal learning bias explanation . . . . .	23
3.2.1 Weight-sensitive stress . . . . .	24
3.2.2 Final devoicing . . . . .	25
3.2.3 Coda sonority . . . . .	27
3.3 Interim summary . . . . .	29
3.4 The role of universal learning bias in phonology . . . . .	31
3.4.1 Do UG biases in phonology exist? . . . . .	34

<b>4</b>	<b>Naturalness in the lab</b>	<b>35</b>
4.1	Defining naturalness . . . . .	37
4.2	Review of naturalness experiments . . . . .	42
4.2.1	Simplicity experiments are not naturalness experiments . . . . .	43
4.2.2	Uncontrolled experiments do not capture naturalness effects . . . . .	46
4.3	Experimental rationale . . . . .	52
<b>5</b>	<b>Experiment: Stress learning and perception</b>	<b>56</b>
5.1	Experiment 1: The Stress Learning Experiment . . . . .	58
5.1.1	Methodology . . . . .	60
5.1.2	Results . . . . .	68
5.1.3	Discussion . . . . .	72
5.2	Experiment 2: The Stress Perception Experiment . . . . .	74
5.2.1	Methodology . . . . .	77
5.2.2	Results . . . . .	80
5.2.3	Discussion . . . . .	85
5.2.4	Universal learning bias shapes simplicity effects . . . . .	87
<b>6</b>	<b>Experiment: Final devoicing</b>	<b>92</b>
6.1	Methodology . . . . .	95
6.1.1	Participants . . . . .	95
6.1.2	Stimuli . . . . .	96
6.1.3	Procedure . . . . .	108
6.1.4	Hypothesis . . . . .	110
6.2	Results . . . . .	110
6.2.1	Novel phase . . . . .	110
6.2.2	Identification phase . . . . .	116
6.3	Discussion . . . . .	118
<b>7</b>	<b>Experiment: Coda sonority</b>	<b>124</b>
7.1	Introduction . . . . .	124
7.2	Methodology . . . . .	126
7.2.1	Participants . . . . .	126
7.2.2	Stimuli . . . . .	126
7.2.3	Procedure . . . . .	130
7.2.4	Hypothesis . . . . .	131
7.3	Results . . . . .	131
7.3.1	Novel phase . . . . .	131
7.3.2	Identification phase . . . . .	135
7.3.3	Correlation between ID performance and Novel performance . . . . .	140
7.3.4	Individual subject variance . . . . .	142
7.4	Discussion . . . . .	147

<b>8 Conclusion</b>	<b>149</b>
8.1 Implications . . . . .	153

# List of Figures

2.1	Auditory nerve response to [da] stimulus (from Gordon 2005) . . . . .	11
3.1	Stress attraction in LHL words . . . . .	24
3.2	No stress attraction in HLH words . . . . .	25
3.3	Final devoicing tableau (UR with a final voiced segment) . . . . .	26
3.4	Final devoicing tableau (UR with a final voiceless segment) . . . . .	26
3.5	Coda sonority tableau (CODACOND $\gg$ MAX <sub>IO</sub> ) . . . . .	28
3.6	Coda hierarchy tableau: Obstruent codas are deleted . . . . .	29
3.7	Coda hierarchy tableau: Nasal codas are permitted . . . . .	29
3.8	‘All shapes are black’: A pattern with one predictable dimension (color)	31
3.9	Hierarchy of simplicity (from Pertsova 2012) . . . . .	32
5.1	Proportion of correct trials by PATTERN and BLOCK ( <i>y</i> -axis begins at 0.50) . . . . .	69
5.2	d primes by PATTERN and BLOCK . . . . .	70
5.3	Effect of performance in Exposure on performance in Novel . . . . .	72
5.4	Example of two trials from the Stress Perception Experiment . . . . .	79
5.5	Proportion of correct trials in each pattern . . . . .	81
5.6	Accuracy in each pattern on initially stressed vs. peninitially stressed words . . . . .	82
5.7	Accuracy in each pattern on default-stressed vs. stress attractor words .	84
6.1	Hypothetical results: Only universal learning bias . . . . .	93
6.2	Hypothetical results: Only channel bias . . . . .	94
6.3	Spectrogram of ‘rooze’ [ɹuz] in CAREFUL speech (full voicing) . . . . .	104
6.4	Spectrogram of ‘rooze’ [ɹuz] in CASUAL speech (under 50% voiced) . . .	104
6.5	Proportion of voicing during the obstruent in each level of CLARITY . .	105
6.6	Vowel duration preceding each obstruent in each level of CLARITY . . .	106
6.7	Correlation between proportion voicing and vowel duration in the CARE- FUL stimuli . . . . .	107
6.8	Correlation between proportion voicing and vowel duration in the CA- SUAL stimuli . . . . .	107
6.9	Obstruent duration in each level of CLARITY . . . . .	108

6.10	Performance in the Novel phase by PATTERN and CLARITY . . . . .	111
6.11	d prime in the Novel phase by PATTERN and CLARITY . . . . .	112
6.12	Proportion of accepted words by final obstruent type . . . . .	114
6.13	Performance in the ID phase by PATTERN and CLARITY . . . . .	116
6.14	ID accuracy based on proportion of obstruent voicing (VOICING only) .	117
7.1	Performance in the Novel phase by PATTERN and CLARITY . . . . .	132
7.2	d prime in the Novel phase by PATTERN and CLARITY . . . . .	133
7.3	Performance in the ID phase by PATTERN and CLARITY . . . . .	138
7.4	Proportion of ID trials containing deletion errors . . . . .	140
7.5	Correlation between subjects' ID performance and Novel performance .	142
7.6	CAREFUL subject averages in the Novel phase (left) and the ID phase .	144
7.7	CASUAL subject averages in the Novel phase (left) and the ID phase . .	145



# List of Tables

1.1	Examples of phonetically grounded patterns and their channel vs. UG bias explanations . . . . .	7
5.1	The three artificial stress patterns in the Stress Learning Experiment . .	59
5.2	Examples of stimuli belonging to each pattern . . . . .	61
5.3	Numeric specifications of duration and pitch for the Stress Learning Experiment stimuli . . . . .	63
5.5	Summary of the Stress Learning Experiment procedure . . . . .	66
5.6	Average accuracy and d prime in each pattern . . . . .	69
5.7	Linear regression (Intercept = ATTESTED:Exposure) . . . . .	71
5.8	Syllable durations in default-stressed words . . . . .	76
5.9	Syllable durations in words with initial stress by stress attractor . . . .	76
5.10	Syllable durations in words with peninitial stress . . . . .	77
5.11	Sample table for calculating difference scores . . . . .	83
5.12	Syllable durations in COMPLEX words with a stress attractor . . . . .	88
6.1	DEVOICING stimuli presented in the Exposure phase . . . . .	97
6.2	VOICING stimuli presented in the Exposure phase . . . . .	98
6.3	Stimuli presented in the Novel phase in all conditions . . . . .	100
6.4	Average accuracy and d prime in the Novel phase . . . . .	112
6.5	Linear regression of Novel task (Intercept = DEVOICING:CAREFUL) . . .	113
6.6	Post hoc: Effect of PATTERN in CAREFUL groups . . . . .	113
6.7	Post hoc: Effect of PATTERN in CASUAL groups . . . . .	113
6.8	Linear regression of ID task (Intercept = DEVOICING:CAREFUL) . . . .	117
7.1	Stimuli presented in the Exposure phase . . . . .	128
7.2	Stimuli presented in the Novel phase in all conditions . . . . .	129
7.3	Average accuracy and d prime in the Novel phase . . . . .	133
7.4	Linear regression of Novel task (Intercept = SONORANT CODA:CAREFUL) .	134
7.5	Post hoc: Effect of PATTERN in CAREFUL groups . . . . .	134
7.6	Post hoc: Effect of PATTERN in CASUAL groups . . . . .	134
7.7	Average accuracy, deletion errors, and insertion errors in the ID phase .	137
7.8	Linear regression of ID task (Intercept = SONORANT CODA:CAREFUL) .	138

7.9	Post hoc: Effect of PATTERN in CAREFUL groups . . . . .	139
7.10	Post hoc: Effect of PATTERN in CASUAL groups . . . . .	139
7.11	Linear regression of deletions (Intercept = SONORANT CODA:CAREFUL)	139
7.12	Post hoc: Effect of PATTERN in CAREFUL groups . . . . .	140
7.13	Post hoc: Effect of PATTERN in CASUAL groups . . . . .	141

## Abstract

An experimental investigation of phonetic naturalness

by

Anna Greenwood

This dissertation begins with the observation of a typological asymmetry within phonological patterns related to phonetic naturalness. Patterns that are rooted within existing tendencies of perception and/or production – in other words, patterns that are phonetically “natural” – are common in phonological typology and seen in a variety of unrelated languages. Phonetically “unnatural” patterns, which are not based in these perceptual and/or production tendencies or which actively work against them, are either typologically rare or unattested. This observation has been noted in numerous works in phonology, and the debate over the nature of this asymmetry has spanned the past few decades. This dissertation makes a contribution to this debate.

There are two major arguments as to what underlies this observed typological asymmetry. One argument suggests that the bias against unnatural patterns is due to constraints on pattern learning. Learners have an innate knowledge of phonetic factors, and this knowledge predisposes them towards learning natural patterns and away from learning unnatural patterns. I refer to this argument as the *universal learning bias* account. The other argument suggests that the bias against unnatural patterns is rooted in perception and production. Since unnatural patterns work against phonetic tendencies, they are more difficult to perceive and/or produce than their natural counterparts and are more likely to be filtered out of typology through the forces of sound change. Under the *channel bias* account, the asymmetry can be explained by systematic errors in transmission that veer in the direction of natural patterns.

This dissertation explores predictions of the channel bias account using experimental methodology. A common practice within laboratory phonology is to test for naturalness-based asymmetries using artificial grammar experiments. Participants are assigned to learn one of two artificial patterns – a natural pattern or its unnatural equivalent – and then are tested on how well they learned their pattern. The results of these experiments, however, are often quite mixed: some experiments find that the

natural pattern was learned more successfully, while others find no performance differences between the two patterns. This has led researchers to question how robust these naturalness effects truly are, and whether or not they can reliably be recreated in the lab.

I argue that the channel bias account holds a solution to this issue: namely, that proper presentation of experimental stimuli matters. Simply presenting stimuli to participants in slow, hyperarticulated speech is not enough. In order to reliably see an effect of naturalness, the phonetic differences between the patterns must be encoded into the design and presentation of the stimuli. This dissertation focuses on pattern pairs where the asymmetry is based in perception, hypothesizing that the natural pattern will be learned more successfully than the unnatural pattern if and only if the stimuli of the natural pattern are also *perceived* more successfully. In this way, the cross-experimental inconsistencies are easily predicted by the channel bias account, while the universal learning bias account fails to provide an explanation.

This dissertation uses experimental methods to argue for the channel bias account in three different phonological domains. In the domain of weight-sensitive stress, an artificial pattern that was both natural and formally simple was learned more successfully than two unattested patterns: one which was unnatural but formally simple, and one which was complex but natural. Perceptual difficulty could explain the poor performance within the unnatural pattern, but not within the complex pattern. Following this, I propose a 2x2 factorial template for testing for naturalness in the lab, wherein participants are assigned to learn one of two patterns (natural or unnatural) in one of two voice qualities (careful or casual). It is predicted that the group learning the unnatural pattern in casual speech will struggle to both perceive and learn their pattern; in careful speech, the unnatural pattern should be both perceived and learned with relative ease. Two experiments in the domains of final devoicing and coda sonority demonstrate evidence for the channel bias account using this experimental design.

To my parents,  
Anita and Tim Greenwood,  
for a lifetime of support and encouragement.

## Acknowledgments

One of the truly remarkable elements about studying linguistics at UC Santa Cruz is being surrounded by vivacious, caring, and intelligent individuals. My colleagues, both students and faculty, have been a consistent wealth of intellectual and emotional support for me over the years, and my success in the program is due in no small part to this community. I would like to take this time to thank some individuals whose presence in my life has helped me get this far.

I would not be here today had it not been for the excellent foundation in linguistics I received during my undergraduate degree at New York University. Maria Gouskova in particular dedicated much of her time towards my academic success. She was the first person to encourage me to write a senior thesis, and it was under her guidance that I learned to think beyond the scope of my coursework and ask the larger, implicational questions. I am extremely grateful for the enormous impact Maria has had on my life.

My advisors and dissertation co-chairs, Jaye Padgett and Grant McGuire, have been extremely generous with their time and assistance. The three of us have spent many an hour cooped up in one of their offices, vigorously hashing out the particulars of a statistical result or a theoretical implication. Jaye and Grant cultivated an atmosphere of judgment-free academic curiosity and, in times of elevated stress, provided me with the encouragement I needed to take on my challenges. Jaye, thank you for our productive weekly meetings and for supporting me during the most difficult moments of my career. Grant, thank you for your guidance in my experiments and for our entertaining conversations on the walk back from Phonetic Analysis lectures.

When asked for my opinion on Ryan Bennett joining my dissertation committee, I responded with an immediate yes. Ryan was a fifth-year student in my first year of graduate school, and I have admired his work and his diligence ever since. Although situated across the country, Ryan has been an invaluable asset to this project, so often providing me with his expertise about experimental analysis (and, on numerous occasions, generously helping me debug my R codes from a distance!).

I would like to extend my gratitude to all of the individuals who have provided me with their input and assistance: in particular, Adrian Brasoveanu, Nameera Akhtar, Elliott Moreton, Keith Johnson, Douglas Bonett, Wendell Kimper, Ruben van de Vijver,

Dinah Baer-Henney, Matt Wagers, and Karl DeVries. Mitchell Owen and Paul Willis had a hand in the early development stages of this project, when the experimental design was little more than a scribble on a napkin. Thanks are also due to my experimental participants, my speakers, and to Evgenia Altukhova for helping out with the Coda Sonority Experiment stimuli. For financial support, I thank the Linguistics Department at UCSC for the numerous Summer Research Awards and the Institute for Humanities Research for a Summer Dissertation Fellowship.

I have had the great fortune of making some incredible, curious, wonderfully weird friends during my time here. This includes, but is not limited to: Jenny Bellik, Sara Cantor, Scarlett Clothier-Goldschmidt, Myjah Conant, Tommy Denby, Karen Duek, Peter Fabian, Nico Feria, Judith Fiedler, Wendell Kimper, Katia Kravtchenko, Eric Lopez, Rae Lundquist, Chelsea Miller, Adam Morgan, Maho Morimoto, Marks [sic] Norris, Jeremy O'Brien, Jason Ostrove, Deniz Rudin, Allan Schwade, Andreas Walker, and Paul "Pwillis" Willis. I am exceedingly grateful to Clara Sherley-Appel, my partner in crime, the April to my Helen, for all she has done for me: for all the rides to the airport, for all the movie marathons and *unfathomably* deep conversations about the human experience, for lending a patient and supportive ear through the most difficult times (and then distracting me with immature jokes). My friendship with Taylor Bell runs deep: some of my happiest memories in grad school have been spent with her, listening to our Party Mix on Spotify as we cook our way through complicated recipes and speak to each other in the Meerkat voice. Anie Thompson has continuously opened up her home and office to me, supplying me with a nigh-endless repertoire of loose-leaf teas, and she has always been enthusiastic to chat with me about Harry Potter head-canons. You all have kept me sane and happy, even when the task seemed impossible. Indeed, "perhaps the real grad school is the friends I've made along the way".

Finally, I want to thank Arielle Chavkin, Andy Massie, and my parents for the endless love they have shown me and their persistent encouragement to keep going. Thanks especially to Mitchell Owen, who has actively supported me throughout this process, shown genuine interest in my work, and counteracted many moments of panic by wrapping me in blankets and cuddles. These people have been such an incredibly positive force on my life. Thank you so much for everything you have given me.

# Chapter 1

## Introduction

A widely observed phenomenon in cross-linguistic typology is that certain phonological patterns are common across a variety of unrelated languages, while other patterns are either rare or unattested. These typological asymmetries are not random but are constrained by a number of factors. One factor that affects the typological frequency of a phonological pattern is its connection to phonetics (Javkin 1977; Jaeger 1978; Hyman 1976, 2001; Ohala 1981, 1993; Lindblom 1986; Lindblom & Maddieson 1988; Archangeli & Pulleyblank 1994, especially chapter 3; Maddieson 1996; Hume & Johnson 2001; Hayes, Kirchner & Steriade 2004; Garrett & Johnson 2012; see Blevins 2004 for discussion and an overview of the literature). Many of the common patterns have phonetic underpinnings that serve to facilitate the perception and/or production of that pattern. An example of this is utterance-final devoicing, wherein the state of the vocal tract towards the end of an utterance makes voicing more challenging and leads to phonetic devoicing of otherwise voiced obstruents (Ohala 1983). When these phonetically grounded patterns are then solidified as part of the grammar, a phonetic tendency transforms into a phonological regularity. Patterns of this sort are considered to be phonetically *natural*.

The *unnatural* counterpart to a natural pattern is one whose phonetic correlates work against perception or production. Utterance-final voicing, the counterpart of devoicing, is considered unnatural because voicing at the end of an utterance is more difficult to produce than voicelessness. I offer the statement in (1) as a definition of the terms “natural” and “unnatural” as they will be used in this dissertation.



- (1) A phonological pattern is natural if and only if it is phonetically grounded – that is, if and only if it is based in phonetic tendencies that facilitate perception and/or production.

In phonological typology, there is an apparent advantage of natural patterns over unnatural patterns. Natural patterns are more likely to survive across generations of speakers than their unnatural counterparts (Blevins 2004; Moreton 2008). Because of this, natural patterns end up being more typologically frequent than unnatural patterns. Some papers use typological frequency as a metric for how natural a pattern is, or cite that ‘natural’ constraints reflect both phonetic realities and typological frequency (Hayes, Siptár, Zuraw & Londe 2009; Carpenter 2010; Hayes & White 2013). I instead see high frequency as a *result* of a pattern’s naturalness, not as evidence of its naturalness.

While the typological tendencies that can be explained by naturalness have been widely observed, there is quite a bit that is unknown about the precise nature of these tendencies. Namely, there is disagreement as to whether the advantage of natural patterns is only based in perception and/or production, or whether there is an additional *learning* advantage. In addition, many experiments have attempted to recreate these typological tendencies in the lab, but not all have been successful – an observation that is rather baffling at first glance (Moreton & Pater 2012). The central goal of this dissertation is to investigate what is driving these naturalness-based tendencies. This is explored through an experimental framework. I am simultaneously interested in how experimental research can demonstrate the effects of perception in shaping naturalness in typology, and how determining the driving forces behind naturalness can help improve the way naturalness is studied in the lab.

I will be using the term *naturalness effects* to refer to the observed tendency for natural patterns to be at some sort of an advantage over their unnatural counterpart. This dissertation focuses on two types of naturalness effects. *Naturalness effects in typology* refers to the cross-linguistic observation that natural patterns are typologically common, while their unnatural equivalents are typologically rare or unattested. Unless stated otherwise, the term “naturalness effects” should be assumed to refer to typology. The second type, *naturalness effects in the lab*, refers to any circumstance in which a phonetically natural pattern is learned more successfully than its unnatural equivalent in

a laboratory setting. This type of effect is distinct from naturalness effects in typology, as it probes for evidence of language biases in a short period of time. Although naturalness effects in typology and in the lab are distinct processes, each can be used to provide insight about the other.

## 1.1 Channel bias and universal learning bias

There are two dominant theories as to what underlies naturalness effects (Ohala & Kawasaki 1984; Wilson 2006; Zuraw 2007; Moreton 2008, 2009). One theory states that the bias against unnatural patterns is cognitive in nature. That is, the human brain is wired to learn certain types of patterns more successfully than others, and unnatural patterns are a subset of patterns that humans are predisposed to struggle with. Learners can be predisposed towards certain patterns and away from others based on their own experience. There is, however, a certain amount of mental architecture that is shared amongst humans, making certain patterns universally more learnable than others (e.g., Pinker 1979; Mitchell 1982). It is this shared mental architecture that this dissertation is concerned with: the *universal learning biases* that, under this account, underlie the observed naturalness asymmetries in typology and in the lab.

Throughout this dissertation, I use the term “universal learning bias” to refer to all types of innate cognitive biases that are shared amongst humans. Universal Grammar (UG) is *one* type of universal learning bias that is specific to language. Theories of UG assume that humans are equipped with mental architecture that is specific to language data. From a UG perspective, unnatural patterns are cognitively challenging because they are *marked*. The principles of UG work to either allow both marked and unmarked structures to surface in a language (e.g., final voicing contrasts) or to eliminate the marked structure entirely (e.g., uniform final devoicing). Systems in which only the marked forms surface, but not the unmarked forms, are unnatural because they actively work against the principles of UG (Chomsky & Halle 1968). If the cognitive system is not configured to learn a certain type of pattern, that pattern will not be learned. This account implies an ingrained knowledge of phonetic factors that extends beyond the reaches of perception and production. This is the position taken in Stampe (1973), Flemming (1995, 2001), Boersma (1998), Steriade (2001), Hayes & Steriade (2004),

Hayes & White (2013), among others.

The second of the two theories, that of *channel bias*, states that there is a perception or production bias against unnatural patterns. Under this theory, both natural patterns and unnatural patterns are equally easy to learn, but unnatural patterns are less likely than natural patterns to survive across generations of speakers because they are more difficult to perceive or produce. A consequence of this hypothesis is that if the difference in perceptability or production were somehow reduced, or removed altogether, the difference in acquiring the two patterns would shrink or disappear. In other words, the natural pattern only has an advantage over its unnatural counterpart insofar as the perceptual or production advantage exists. This is the position taken in Ohala (1981, 1993), Blevins & Garrett (1998, 2004), Hale & Reiss (2000), Hyman (1976, 2001), Myers (2002), Yu (2003, 2004), Blevins (2004, 2006), among others.

### 1.1.1 Examples of naturalness effects in three phonological domains

This dissertation uses experimental methods to investigate naturalness effects within three phonological domains: weight-sensitive stress, utterance-final devoicing of obstruents, and coda sonority. The distinctions between the natural and unnatural patterns are provided below in (2).

- (2) Naturalness effects explored in this dissertation
- a. *Weight-sensitive stress*: Certain syllable types have the ability to attract stress away from whichever syllable would have been stressed by default. In one type of natural pattern in this domain, syllables with long rimes act as stress attractors (e.g., [pá.ta.ka] by default, but [pa.tán.ka] by stress attractor). In the unnatural pattern, only syllables with short rimes may act as stress attractors (e.g., [pán.tan.kan] by default, but [pan.tá.kan] by stress attractor) The latter is unattested.
  - b. *Utterance-final devoicing*: Voicing distinctions between utterances are neutralized in utterance-final position. In the natural pattern, all utterance-final obstruents are voiceless (e.g., /pas, paz/ → [pas]).<sup>1</sup> In the unnatural pat-

---

<sup>1</sup>This neutralization is often seen at the end of words, not just utterances. Myers & Padgett (2014)

tern, all utterance-final obstruents are voiced (e.g., /pas, paz/ → [paz]). Whether or not the latter is attested is disputed in the literature (Yu 2004; Blevins 2006; Kiparsky 2006), but regardless of the dispute, the unnatural pattern is at least typologically rare.

- c. *Coda sonority*: Languages that allow syllables to have codas often have restrictions on which types of sounds can occupy coda position. In one type of natural pattern in this domain, only sonorant consonants are allowed to occupy coda position ([pa, pan], \*[pat]). In the unnatural pattern, only obstruents are allowed to occupy coda position ([pa, pat], \*[pan]). Whether or not the latter is attested is disputed in the literature (Zec 1988; Blevins 2004, chapter 6.6; and others).

Both the channel bias and the universal learning bias theories explain the observation that the natural pattern has some sort of advantage over the unnatural pattern, but they differ as to the nature of that advantage.

Under channel bias, each natural pattern is associated with a phonetic precursor, or a set of phonetic precursors, that underlie how the pattern is tied to perception and/or production (Chapter 2). Natural patterns rely on the phonetic reality that some sounds or sound sequences are easier to perceive (or produce) than others. For example, certain syllables (like syllables with branching rimes, CVC and CVV) are more perceptually prominent than others (like open syllables, CV), apart from any prominence added from stress. In addition, stressed syllables have more phonological and phonetic prominence than unstressed syllables. Patterns in which exceptional stress is attracted to syllables with branching rimes encode a phonetic reality into the phonological grammar: the phonetically more prominent syllable also receives a prominence boost in the phonology. Increasing the prominence of a syllable that is *already* prominent facilitates perception of the pattern because the auditory system can identify the stressed syllable with considerable ease (Gordon 2002, 2005). Unnatural patterns, in turn, are based in phonetic *irregularities*. An example of this would be a stress pattern in which exceptional stress is only attracted to open, light (CV) syllables. This pattern is perceptually difficult be-

---

argue that, while utterance-final devoicing is an issue of naturalness, word-final devoicing lacks phonetic precursors and is instead applied through domain generalization.

cause the phonetically less prominent syllable receives phonological prominence, making it more difficult for the auditory system to distinguish between stressed and unstressed syllables. If the listener cannot accurately perceive where stress lies in the word, learning the pattern of which syllable attracts stress will be difficult or impossible.

The universal learning bias theory does not deny that these phonetic precursors exist, but they play a different role in how naturalness effects arise. Under the UG branch of this theory, phonetic tendencies are somehow encoded into the cognitive system as part of the principles of UG (Chapter 3). A consequence is that learners are predisposed to learn phonetically natural patterns and to struggle with phonetically unnatural patterns, sometimes to the point of absolute unlearnability. Extending naturalness effects to an Optimality Theoretical framework (Prince & Smolensky 1993/2004), each natural pattern is associated with a markedness constraint, which prefers natural candidates and disfavors unnatural candidates (Hayes & Steriade 2004). In the exceptional stress example, the markedness constraint WEIGHT-TO-STRESS prefers candidates in which syllables with branching rimes receive stress (e.g., [pa.tán.ka]) and disprefers candidates where branching rimes are unstressed (e.g., \*[pá.tan.ka]). When the markedness constraint is ranked above faithfulness constraints that would otherwise preserve an unnatural input (e.g., [pá.tan.ka] receiving default initial stress), the natural candidates will always be more harmonic than the unnatural candidates, resulting in the natural pattern surfacing in the grammar.

Table 1.1 summarizes the phonetic precursors and markedness constraints associated with each of the three patterns investigated in this dissertation.

## 1.2 An experimental framework

This dissertation uses artificial grammar experiments to support the hypothesis that naturalness effects are driven by channel bias alone. Specifically, I explore the perceptual side of the channel bias account, which suggests that patterns that are difficult to perceive should be learned less successfully than patterns that are easier to perceive. Although learning an artificial pattern in the lab is different from learning a natural language, this method can be used to test the ways in which perceptibility of the stimuli that make up a pattern affects how well the pattern is learned.

<b>Pattern</b>	<b>Channel bias: Misperception</b>	<b>Universal learning bias (UG branch): Markedness</b>
Exceptional stress	Syllables with lower perceptual energy are less perceptually prominent, less able to be perceived as exceptionally stressed (Gordon 2002, 2005)	WEIGHT-TO-STRESS
Final devoicing	Reduced vocal fold vibration at utterance endings leads to partial obstruent devoicing & misperception as voiceless (Jaeger 1978; Ohala 1983)	*VOICED-OBS] <sub>Word</sub>
Coda sonority	Reduced stop burst strength & intensity leads to perceptual deletion (e.g., Ohala & Kawasaki 1984; Bybee 2001; Wright 2004)	CODA CONDITION, coda markedness scale

Table 1.1: Examples of phonetically grounded patterns and their channel vs. UG bias explanations

The experimental methods used in this dissertation follow a structure that is commonly employed in naturalness research. Participants are first assigned to learn one (and only one) artificial pattern, and after some amount of training on this pattern, they perform a task that assesses how well they learned this pattern. Artificial grammar experiments can be used to determine whether unnatural patterns are unlearnable or underlearned compared to their natural equivalents (Wilson 2006; Hayes et al. 2009; Becker, Ketrez & Nevins 2011) or probe for the nature of any learning asymmetries between two patterns (Pycha, Nowak, Shin & Shosted 2003; Moreton 2008, 2012).

The overall picture of naturalness in the lab, however, is rather murky. Some experiments have been successful at recreating naturalness effects from typology in a laboratory setting (e.g., Carpenter 2010; Myers & Padgett 2014), while others claim to

have found no effect for naturalness (e.g., Pycha et al. 2003; Seidl & Buckley 2005). This has led to the question of whether or not capturing naturalness effects in the lab is a reliable endeavor (see Moreton & Pater 2012 for an overview).

I argue that the channel bias account provides an explanation for these inconsistencies (Chapter 4) If naturalness effects are driven entirely by perceptual (and/or production) differences, then properly replicating the naturalness effects seen “in the wild” entails properly replicating these phonetic differences in the lab. In order to reproduce naturalness effects within a given phonological domain, the researcher must know and reproduce whichever phonetic underpinnings drive the observed differences between the two patterns. Experiments that do not control for these differences run the risk of null results.

The three sets of experiments in this dissertation are discussed in Chapters 5–7. Chapter 5 presents a pair of experiments in the domain of weight-sensitive stress. In the Stress Learning Experiment, two separate stress patterns that are unattested in typology also fail to be learned in the lab. One of these patterns is unnatural, and the other is formally complex (but natural). By contrast, a pattern that is both natural and simple is learned successfully. The Stress Perception Experiment finds that perceptual issues can explain the lack of learning in the unnatural pattern, but not in the complex pattern. This experiment shows evidence that channel bias is at work in shaping naturalness effects and suggests that simplicity, instead, is an instance of universal learning bias.

The Final Devoicing Experiment in Chapter 6 demonstrates the role that stimulus clarity plays on naturalness effects. As before, two patterns were created for this experiment: a natural pattern (final devoicing) and its unnatural equivalent (final voicing). In addition, the condition of stimulus clarity is added, such that half the participants learned their assigned pattern in careful speech, and half learned their pattern in casual speech. This experiment found that speech clarity affected how the two unnatural groups performed, with the casual group performing the task less successfully than the careful group. In addition, this experiment saw evidence that perceptability of final voiced obstruents is correlated with the proportion of voicing during the closure. The less voicing during the closure of a final voiced obstruent, the more likely it is to be misperceived as voiceless. This correlation is trending in the same direction as sound

change and provides an explanation for the performance difference between the unnatural groups.

The Coda Sonority Experiment in Chapter 7 follows this structure and applies it to the phonological domain of coda well-formedness. Two artificial patterns were created for the purpose of this experiment. Both patterns allowed nasals and stops to occupy onset position, but only one type of coda was permissible in each: either nasals (in the natural pattern) or stops (in the unnatural pattern). As before, participants were assigned to learn one of these two patterns in either careful or casual speech. The results of this experiment demonstrated that the learnability of the two patterns differed only within the groups trained on casual speech. In casual speech, the natural group both perceived and learned their pattern more successfully than the unnatural group. In careful speech, both groups perceived and learned their pattern equally well. This is exactly in line with the predictions from the channel bias hypothesis.

The findings of this dissertation point to the crucial role that perception plays in determining how well a pattern is learned. The logic behind this is satisfyingly clear: if listeners cannot accurately perceive the stimuli that make up a phonological pattern, they will not be able to learn and reproduce it accurately either. While it is possible that universal learning biases are needed to explain other aspects of typology (for example, the propensity for phonological patterns to be formally simple; Gordon 2002; Moreton 2008; Moreton & Pater 2012; and others), I suggest that channel bias provides the whole story when it comes to naturalness. In this way, I adopt a view of phonology that is very similar to Moreton & Pater’s (2012) “structurally-biased phonology”, in which phonetic naturalness is associated with channel biases and formal structure is associated with learning biases.

The experiments run in this dissertation study pattern learning, and pattern learning only. Experimental designs that probe for a listener’s real-time judgments, such as wug testing (Berko 1958) and poverty of the stimulus designs (as seen in Wilson 2006; Finley & Badecker 2007; Becker et al. 2011; Bennett 2012), are outside the scope of this dissertation but could provide additional insight to the question of where naturalness effects of any sort come from.



## Chapter 2

### The channel bias approach

The experiments in this dissertation address naturalness effects within three different phonological patterns: weight-sensitive stress, utterance-final devoicing, and coda sonority (as summarized in (2)). This chapter addresses the channel bias theory explanation for these patterns by discussing the phonetic correlates associated with each one.

Under the channel bias theory, perceptually motivated naturalness effects arise over time through a series of misperceptions and productions based on those misperceptions (Ohala 1981, 1983, 1993). As these errors compound, the pattern that emerges from these phonetic correlates may become encoded as part of the grammar (Catford 1974; Hyman 1976; Blevins 2004; and others). This is the process of *phonologization*, which I discuss at the end of this chapter.

#### 2.1 A channel bias explanation

##### 2.1.1 Weight-sensitive stress

This dissertation adopts Gordon's (2002, 2005) account of weight-sensitive stress where the key phonetic cue is *perceptual energy*: a measure of perceived loudness over time (see also Lieberman 1960; Beckman 1986). Gordon (2005) finds that perceptual energy of a syllable's rime correlates with weight-sensitive stress in three different case studies of Pirahã, Banawá, and Arrernte, such that syllables with larger perceptual energy are heavier. In this subsection, I summarize the discussion of perceptual energy as it relates

to stress.

When the auditory system is exposed to a high-intensity stimulus, such as a vowel, a period of increased activity at the beginning of the stimulus is quickly followed by a sharp decrease in sensitivity. The auditory system is most sensitive to the stimulus within the first 0-40 ms. After this, the decline in sensitivity, called *adaptation*, continues throughout the duration of the sound (Delgutte 1982; Silverman 1995; Wright 2004). A period of silence or reduced intensity, such as a pause or a voiceless consonant, can then allow the auditory system to *recover* before hearing the next high-intensity sound. Figure 2.1, from Gordon (2005), is a schematic of how the silence provided by onset [d] in [da] gives rise to increase sensitivity at the beginning of the vowel [a].

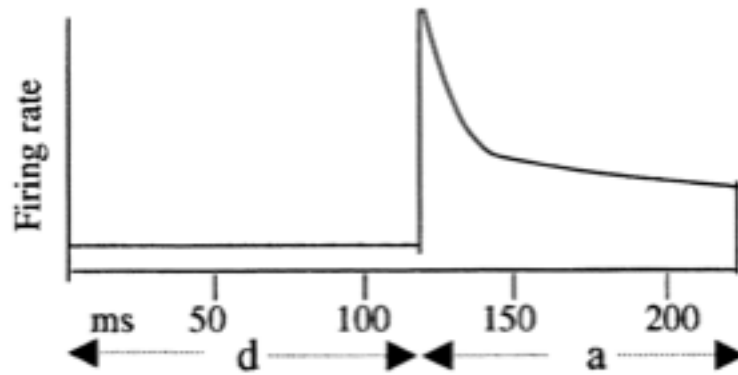


Figure 2.1: Auditory nerve response to [da] stimulus (from Gordon 2005)

The concepts of adaptation and recovery play into the calculation of perceptual energy. Perceptual energy is not a calculation of raw intensity over time. Instead, it integrates how a sound will be perceived as more or less intense depending on the intensity of its neighboring sounds. To calculate perceptual energy, the sound is broken down into a series of target windows of equal duration.<sup>1</sup> The raw intensity of each time window is compared to the raw intensity of the time window immediately preceding it. Sounds

---

<sup>1</sup>Averaging over more target windows of shorter duration should provide a more accurate measurement of perceptual energy, in much the same way as a higher sampling rate and smaller spectral slices provides more acoustic information. Gordon (2005) does not provide a suggestion for window size or window number, other than to say that each window should be of equal duration.

that follow a period of silence are perceptually louder (and receive a numerical boost to intensity), and sounds that follow a period of reduced sensitivity are perceptually quieter (and receive a numerical damper in intensity). Time windows at the beginning of the sound (e.g., the beginning of a vowel) should therefore be perceptually louder than time windows at the end of the sound. The adjusted intensities are then summed across the entire sound to produce a single measure of perceptual energy for that sound. Since perceptual energy is itself a summation, longer sounds will be perceptually louder than shorter sounds of the same intensity.

Following Gordon (2005), this dissertation takes perceptual energy to be the key phonetic correlate of weight-sensitive stress. Syllables with greater perceptual energy make more natural stress attractors than syllables with lower perceptual energy because the former are more perceptually prominent. To stress a syllable is to make that syllable prominent in both the phonetics and the phonology, and so stressing syllables that already carry a great deal of phonetic prominence helps to distinguish between stressed and unstressed syllables. Imagine a bisyllabic word, [ $\sigma_1$ . $\sigma_2$ ] where  $\sigma_1$  has a greater perceptual energy than  $\sigma_2$ , independent of stress (indicated by the underline). If  $\sigma_1$  is also in a position of stress, [ $\sigma_1$ . $\sigma_2$ ], its role as the most prominent syllable of the word is abundantly clear. If instead  $\sigma_2$  is stressed, the word will contain two prominent syllables:  $\sigma_1$ , which is phonetically prominent on its own, and  $\sigma_2$ , which receives phonetic prominence from the stressed position: [ $\sigma_1$ .' $\sigma_2$ ]. It should therefore be much easier for listeners to perceive which of the two syllables has phonological stress in the former than in the latter. In the same vein, stress patterns that offer larger differences in perceptual energy between stress attractors and non stress attractors are more natural than stress patterns with smaller perceptual energy differences (Gordon 2005).

The question that remains, then, is what the factors are that increase the perceptual energy of a syllable. Since the calculation of perceptual energy relies primarily on intensity, factors that increase intensity will also increase perceptual energy, and by extension, the naturalness of the stress-attracting syllable. Syllables with long rimes have more target windows that contribute to the final calculation than do syllables with shorter rimes. Given two syllables of equal intensity and different rime length, the syllable with the longer rime will have a greater perceptual energy than the syllable

with the short rime ( $[pa:]$ ,  $[pan]$ )  $>$   $[pa]$  and make a more natural stress attractor.

Intensity of the vowel is another factor that contributes to the syllable's perceptual energy. Low vowels have been shown to have greater intensity than high vowels (Lehiste 1970; Gordon 2005; Carpenter 2010). Given two syllables of equal rime length and different vowel quality, the syllable with the lower vowel will have a greater perceptual energy than the syllable with the higher vowel ( $[pa]$   $>$   $[pi]$ ). This effect was supported in Carpenter (2010), who found that a stress pattern in which low vowels attracted stress was learned more successfully than a stress pattern in which high vowels attracted stress.

In theory, stress patterns in which the stress attractor must have both a long rime *and* a low vowel may be more natural than stress patterns where the stress attractor has only one of these qualities (e.g.,  $[pan]$   $>$  others' more natural than 'CVC  $>$  others'). Gordon (2005), however, remarks that this stress pattern is unattested, and that there are no stress patterns in typology in which a single stress attractor is defined by two properties. The Stress Experiments in this dissertation (Chapter 5) support Gordon's (2005) claim that these stress patterns, although natural by a perceptual energy account, are too complex to be learned, resulting in their typological absence (see also Moreton 2008, 2012, for work on universal learning biases and intra-dimensional dependencies).

### 2.1.2 Final devoicing

To create a voiced sound, two components are necessary: the vocal folds must be configured in such a way so that they may vibrate, and there must be enough air flowing through the glottis to sustain the vibration. Both of these components are necessary, and if either one of them is removed, voicing is impossible (Jaeger 1978; Ohala 1983, 1989; Smith 1997). Towards the end of the utterance, speakers tend to spread their vocal folds in anticipation of the pause (Myers & Hansen 2007; Myers & Padgett 2014; and references therein). Doing so will result in partial or full devoicing of the final sound. Another property of the utterance is that subglottal pressure declines gradually over time, such that it is at its weakest point at the end of the utterance. Without sufficient air flow, voicing will not be sustained throughout the final consonant. Such phonetic devoicing is specific to obstruents, since devoiced sonorants tend to be perceptually deleted rather than perceived as voiceless (Ohala 1993; Myers & Hansen 2007).

### 2.1.3 Coda sonority

It has been observed that sonorants make better codas than do stops (Zec 1988; Clements 1990; Orgun 2001; Baertsch 2002; Baertsch & Davis 2003) – and I argue that the basis of this is rooted in channel bias. There are three types of errors that can be made when perceiving or producing codas: *deletions* (e.g., /pat.ka/ → [pa.ka]), *insertions* (e.g., /pat.ka/ → [pa.tə.ka]), and *substitution errors* (e.g., /pat.ka/ → [pap.ka], [pan.ka], etc.). While the reduced perceptual salience of stop codas leads to a variety of errors, I focus in particular on deletions.

Stops consist of three phonetic components: formant transitions at their offset and onset, a period of silence during the closure (or, in the case of truly voiced stops, voicing without formant structure), and the release. Stop releases, and bursts in particular, provide information about place (Malécot 1958; Winitz, Scheib & Reeds 1972; Ohala 1992a; Wright 2004; Jun 2004, and references therein) and voicing (Malécot 1958; Repp 1979). If the release is produced in a less perceptible way, it follows that the stop will be less perceptible as well.

It is generally assumed that consonants undergo weakening in final position (Benkí 2003; Krakow 1999, and references therein). For stops, greater changes in aerodynamics and pressure occur at their offset than their onset, leading to greater perceptual salience in CV sequences than in VC sequences (Ohala & Kawasaki 1984; Ohala 1992a; Wright 2004). Kochetov (2001) adds that coda stops are released less frequently than onset stops, and that the rate of release decreases in faster speech. These weakened productions can also be deleted outright. Greenberg, Carvey, Hitchcock & Chang (2003a) performed an analysis of the SWITCHBOARD corpus, comprised of spontaneous speech telephone dialogues from American English. The researchers hand-annotated a subset of the materials for segmental and prosodic material and investigated the role that stress accent level plays on productions of segments within the syllable. They found that deletion errors occurred frequently in coda position across all stress levels, but there was a greater percentage of deletions in unstressed syllables. In contrast, deletions in the onset and nucleus were very rarely encountered, even in unstressed syllables. Benkí's (2003) word recognition study found few *perceptual* deletion errors of onset consonants (CVC → VC), but a significant number of deletion errors of coda consonants (CVC → CV).

The increased perceptual salience of onsets (compared to codas) means that they can be heard and recognized more easily. This provides some phonetic suggestion for why the types of consonants that can be onsets are fairly unrestricted, cross-linguistically, while coda consonants are more restricted (Itô 1986; Beckman 1997; Zoll 1998; among others), and for why neutralization in coda position is a common phenomenon (Kiparsky 1995; Lombardi 1999; Benkí 2003; among others).

The decrease in perceptibility from onset to coda prompts an increase of deletion errors of stops in both perception and production. If the stop release is produced in such a way that it is perceptually weak – or if the stop is produced unreleased – the only cues to presence of the stop are formant transitions at the onset of the consonant and silence during the closure. Consider a cluster of heterorganic stops, the first of which is an unreleased coda ([pat̚.kɑ]). This cluster will have the formant transitions of [t] at its onset, but the release and formant transitions of [k] at its offset. Given how important the release is to the perception of a stop, it is conceivable that this cluster will occasionally be heard with a deletion error ([pa.kɑ]) (Ohala 1990, and references therein). The first consonant in such a cluster is generally the target for deletion, and rarely the second consonant (Wilson 2001; McCarthy 2006). Alternatively, as Bybee (2001) says, perceptually weak obstruent codas may “assimilate to the following obstruent, producing a geminate that degeminates.” In either case, lexical contrasts that rely on presence vs. absence of the coda will be less reliable with more perceptual deletion errors or more degemination in production.

The final piece of the puzzle is why deletion in the coda seems to disproportionately affect stops more than it affects sonorant consonants. Indeed, sonorant consonants are weakened in coda position as well. American English laterals in coda position have been noted to be produced with lower tongue tip positions and lower F2 when compared to onset laterals (Giles & Moll 1975; Browman & Goldstein 1995). Nasals in coda position involve a larger lowering movement of the velum compared to onset nasals (see Krakow 1999 for a discussion of the literature), leading to a larger effect of nasalization on the tautosyllabic vowel. Nasal codas can thus nasalize the preceding vowel and then delete (Ohala 1989; Bybee 2001; Benkí 2003). Given the overall weakening of consonants in coda position, the question of why some languages have developed a preference for

sonorant codas – as opposed to banning codas altogether – arises.

One possibility is the presence of audible material during the closure. Unlike stops, sonorants carry formant structure throughout the closure, signaling the presence of a phoneme distinct from a following onset. Thus, in a word like [pan.ka], the pervasive formant structure during the nasal may inform the listener of the coda’s presence better than in the case of [pat.ka]. Alternatively, if the coda is the target of anticipatory assimilation, a nasal coda will incur a change in place (/pan.ka/ → [paŋ.ka]), while a stop coda will become the first half of a geminate (/pat.ka/ → [pak.ka]) and delete upon degemination.

This may explain why stridents make particularly good codas, despite their obstruent status, and why languages like Italian, whose coda inventory is almost entirely restricted to sonorants (Krämer 2009), make exceptions for [s]. Stridents, unlike stops, have strong perceptual cues – aperiodic noise and a high frequency center of gravity – that persist throughout the coda and are less likely to be masked by low frequency background noise than are stop release bursts and non-sibilant fricatives (Wright 2004; Meyer, Dentel & Meunier 2013). It is reasonable to assume that deletion errors would be less frequent for strident codas than for other types of obstruent codas.

An alternative view from Blevins (2004) is that the apparent preference for sonorant codas is a result of many converging factors. She cites a sound change observed in both Gilbertese and Manam, wherein the only permissible codas allowed were nasals. This arose, she argues, from a gradual loss of post-nasal, word-final, voiceless high vowels [N<sub>i</sub>#], caused by nasalization blocking or reducing airflow during the vowel and rendering it difficult to perceive (Ohala 1983; Blevins 1997). In these cases, the presence of word-final nasal codas is not due to nasals themselves being “well-formed” or perceptually superior codas, but to their ability to perceptually dampen the following voiceless vowel. Blevins also notes a change in Fijian, Tawala, and some Eastern Bantu languages, in which a variant of unstressed *mu* is a singular *m* ([yámu, yám] ‘mosquito’, Fijian). She notes that this change is not limited to coda position (/mu-ma:/ > [mmá:]), but may explain why such languages allow [m] to arise as a coda and not other sonorants. Overall, she argues that there is no evidence of an innate preference for high-sonority codas, and that the typological distribution of coda types are a result of several converging

diachronic patterns.

## 2.2 Phonologization of phonetically natural patterns

A final step in the discussion is to address *phonologization* (coined by Hyman 1976, and also discussed in Catford 1974 and several works by Ohala, most notably 1981, 1993), a process of language change in which a phonetic feature, pattern, or tendency becomes part of the phonology. Through phonologization, gradient phonetic precursors become regularized as part of the grammar. Since this dissertation is primarily interested in the perceptual side of channel bias, I will focus this discussion on the role that perception – particularly, *misperception* – plays in phonologization.

Misperception-based sound change relies on the interaction between speaker and listener. It is widely agreed upon that speech is highly variable. As discussed in the previous sections of this chapter, variation arises when the vocal apparatus fails to produce its intended targets – from gestural overlap, different degrees of voicing, etc. Transmission from speaker to listener is also rife with variation. One of the listener’s roles is to interpret and identify words in the speech signal. When the listener perceives a speaker’s noisy, degraded, or accidental productions as intentional (Ohala 1993), it may lead them to categorize the components of the speech signal differently from how the speaker intended. When the listener then becomes the speaker, their productions will be affected by the tokens that they have heard and their own categorizations of those tokens (Wedel 2003, 2006). Systematic misperceptions of this kind can, over time, become part of the language’s phonology when multiple listeners interpret these phonetic regularities as phonological rules (Kiparsky 1995). Much of the work of misperception-based sound change is attributed to Ohala (1981; 1983; 1989; to name a few). This is also a key theme of the Evolutionary Phonology framework (Blevins 2004) and ties into the predictions of exemplar theory (e.g., Nosofsky 1986; Hintzman 1988; Johnson 1997; Pierrehumbert 2001).

For example, the phonetic precursors of final devoicing (described in Chapter 2.1.2) lead many productions of voiced obstruents like /z/ in utterance-final position to be partially or fully devoiced. This happens on a gradient scale: some tokens of voiced obstruents are more voiced than others, and some are less voiced than others. One pos-



sible way that the process of phonologization may unfold is in line with the predictions made in exemplar theory. In this account, listeners store vast amounts of detailed information about the speech signals they hear – indexical information about the speaker (Goldinger 1998), word and segment level categorical information (Vitevitch & Luce 1999), auditory properties, etc. (see Johnson 1997 for an extensive review). In our final devoicing example, then, listeners hear and store various tokens of final /z/, each with a different amount of voicing, and many of which are partially devoiced to some degree (Smith 1997). As this amasses, they come to accept a wider range of voicing fluctuations as viable productions of /z/. When the listener then goes to produce utterance-final /z/, their productions will be influenced by the content of their exemplar space – in this case, the body of tokens that have been stored within their memory and given the category label /z/. Since their exemplar space now contains partially devoiced tokens of /z/, they will, in turn, *produce* partially devoiced tokens of /z/. This process builds over time: as a new set of listeners hears and stores this greater body of partially devoiced tokens, they may start to devoice /z/ (and other final voiced obstruents) at a greater rate.

Phonologization may occur if, somewhere in the midst of this process, listeners start categorizing these partially devoiced productions of final /z/ instead as voiceless /s/. Models of exemplar theory state that categorization of an incoming token happens through comparison of that token to existing categories (Johnson 1997; Wedel 2006). The more similar that token is to a body of exemplars in a single category, the more likely it will be categorized as a member of that category. Listeners may begin to categorize partially devoiced tokens of /z/ as /s/, due to their increasing similarity to the /s/-tokens. Tokens that are ambiguous and cannot be stored with much certainty as either /s/ or /z/, meanwhile, may not be stored in memory at all, and should decrease in frequency (Wedel 2006). What results is sound change pressure coming from two areas: the phonetic tendency for final voiced obstruents to devoice, and the strategy on the part of the listener to only store tokens that they can categorize with some degree of certainty. In this scenario, the more /z/ becomes devoiced, the more likely it is to be stored as /s/, while the tokens that are still somewhat voiced may fail to be stored by way of them being too ambiguous.

The process, of course, is not as simple as this. Many languages, such as English, still maintain a phonological voicing contrast in final position. The likelihood of a merger between phonemes relies not only on its phonetic precursors but on the functional load of its contrast. King (1967) defines functional load as a measure of the “extent and degree of contrast” between segments and other linguistic units. Neutralization between two phonemes on which many lexical contrasts rely would result in a loss of those contrasts. A corpus study by Wedel, Kaplan & Jackson (2013) found statistical evidence that mergers are less likely to occur between two phonemes with a large number of minimal pairs. When two contrastive phonemes with high functional load do begin to merge, other properties that co-occur with the contrast might become amplified to preserve the salience of the contrast (Wedel 2006). For example, English has a contrast between voiced and voiceless obstruents, but in utterance-final position, voiced obstruents tend to be partially devoiced, making them sound more and more like their voiceless counterparts. Vowel length is not contrastive in English, but before voiced obstruents, vowels tend to be longer than before voiceless obstruents (Raphael 1972, and others). In Wedel’s (2006) model, the vowel length difference may have increased in response to voiced and voiceless obstruents sounding less perceptually distinct, thus taking on the burden of preserving the final voicing contrast.

Nevertheless, the speaker-listener tradeoff exemplifies why the naturalness of a phonological pattern is rooted in phonetics. From the production side, speaker variability is high, and the variability of patterns that are difficult to produce will, over time, gravitate towards a version that is simpler to produce (Garrett & Johnson 2012). From a perceptual side, listeners attempt to store exemplars of speech sounds and words that they encounter, and part of this storing process is categorization. Misperception of a sound may result in the listener’s categorization differing from the speaker’s categorization (Ohala 1981). All of this hinges on the understanding that, when the listener becomes the speaker, their productions will be affected by the content of their exemplar space. Fluctuations in the variability of a pattern then slowly gravitate towards tokens that are easier to produce by the speaker and easier to correctly categorize by the listener.

## 2.3 Summary

This chapter has provided an explanation of the channel bias account of naturalness, under which a phonetic tendency over time becomes incorporated as part of the grammar (e.g., Ohala 1981, 1993; Blevins 2004; Wedel 2006; Moreton 2008). This account relies on the interaction between speaker and listener and assumes that the speaker will produce a high amount of variability, which occasionally results in misperception on the part of the listener. A sound change can begin to occur when listeners allow these misperceptions to affect their mental representations of the language (as with the example from exemplar theory of shifting category labels from /z/ to /s/) and apply the phonetic tendency across a uniform set of domains, as if it were a phonological rule (Kiparsky 1995).

This argument has the benefits of being explanatorily simple, traceable through history, and testable in the lab. Little stipulation is required: the types of patterns that are suppressed in typology due to their phonetic structures are also the types that are unlikely to survive in sound change, or at least *less* likely to survive than a phonetically natural alternative. While the concepts of “exemplar spaces” and “mental representations” are, in some sense, theoretical, experimental research within exemplar theory can show how listeners’ productions are easily influenced by those of other speakers and the directions and influences of long-term sound change (Goldinger 1998; Vitevitch & Luce 1999; Wedel 2006). If this explanation is sufficiently powerful, Occam’s razor would suggest that an alternate explanation with more stipulation and less explicit evidence is inferior. Such, I argue, is the problem with the universal learning bias account of naturalness, presented in the next chapter.

## Chapter 3

# The universal learning bias approach

In arguing for a channel bias approach to naturalness, this dissertation also argues against an approach where unnatural patterns are at a cognitive, structural disadvantage. The universal learning bias account does not seem to explain any of the observations seen in typology that the channel bias account does not already explain, and it falls comparatively short in accounting for studies of pattern learning in the lab. Even so, a universal learning bias approach may still be needed to explain other observed asymmetries in phonology between patterns that are equally phonetically motivated.

In this section, I discuss how one type of universal learning bias, in which phonetic knowledge is encoded into Universal Grammar (UG), accounts for each of the three phonological domains investigated in this dissertation. I argue against the UG approach to naturalness here and shed light on the types of phonological issues where a universal learning bias approach – but not necessarily a UG-specific approach – seems necessary.

### 3.1 A closer look at cognitive biases

The term ‘cognitive bias’<sup>1</sup> is an umbrella term that encapsulates a variety of learning difficulties. These cognitive biases promote learnability of certain patterns while inhibiting learnability of others. An individual may struggle to learn certain patterns because of the structural makeup of the pattern itself, or because of particulars about that individual learner, as outlined in (1).

---

<sup>1</sup>Also called ‘analytic bias’, ‘inductive bias’, or ‘learning bias’.

(1) *Types of cognitive biases*

- a. **Universal learning bias:** The mental architecture that humans use to generalize patterns from large amounts of data biases them towards certain patterns and away from others (e.g., Pinker 1979; Mitchell 1982).
- b. **Individual learning bias:** Learners may be predisposed to learn certain patterns due to their individual experience (transfer effects from their native language, e.g. Grosjean 1989; Broselow, Chen & Wang 1998; age and brain development, e.g. Smolensky 1996; Hale & Reiss 1998; etc.).

This dissertation is concerned with the particulars of certain *patterns* that affect their learnability, as opposed to the particulars of different individuals that make them predisposed to learn certain patterns. As such, I investigate only the universal learning biases in (1-a). This category can be broken down further into two subcategories, as in (2) (a non-exhaustive list): biases that apply to all patterns, both linguistic and nonlinguistic, and biases that are unique to linguistic patterns (Moreton 2012; Moreton, Pater & Pertsova 2015).

(2) *Types of universal learning biases*

- a. **General:** The components that make up a pattern, as well as the relationship between those components, make certain patterns more difficult to learn than others.
- b. **Linguistic:** Linguistic patterns that do not conform, or actively disobey, the principles of Universal Grammar (UG) are more difficult to learn.

The central argument of this dissertation is that naturalness effects cannot be subsumed under either (2-a) or (2-b). Markedness as it relates to naturalness effects is not an issue of learnability, but of perception and production. Of course, I am not arguing that there is no cognitive element to pattern acquisition. Natural patterns, however they arise, must eventually be learned, and this in itself is a cognitive process. Phonologization, discussed in Chapter 2.2, involves making generalizations about phonetic regularities and applying them regularly to various domains. The question is why making such generalizations about unnatural patterns may be more challenging.

The universal learning bias account suggests that the learner is influenced by *a priori* biases against unnatural patterns, and that the learner will struggle to acquire these patterns across a variety of conditions. The channel bias account, instead, suggests that the problem lies with the stimuli themselves, not with the learner. Humans are equally willing to learn both natural and unnatural patterns if the stimuli are clear enough, but because unnatural stimuli push back against phonetic tendencies, they pose perceptual problems for the listener. Patterns that cannot be perceived properly also cannot be learned properly.

What can be said about asymmetries in typology that are not an issue of naturalness? Simplicity effects – or the tendency for formally simpler patterns to be at an advantage over complex patterns, both in typology and in the lab – are one example of this. I argue that simplicity effects observed in phonology are an example of (2-a), and not of (2-b). The cognitive factors that suppress complex linguistic patterns, like certain types of weight-sensitive stress patterns, are also at work in non-linguistic patterns (Shepard, Hovland & Jenkins 1961; Moreton et al. 2015). In other words, while simplicity effects *can* be seen in language, they are not *only* seen in language. Section 3.4 addresses this in greater detail. To conclude, I speculate as to whether (2-b)-type biases exist in phonology at all, or whether all phonological patterns driven by universal learning biases can be explained by larger pattern-learning mechanisms.

## 3.2 A universal learning bias explanation

Within the UG branch of universal learning bias, the difficulty of learning unnatural patterns is encoded into formal phonological systems through the notion of markedness. Certain phonological structures and patterns are more marked than others, leading to discrepancies in the typology.

Within an Optimality Theoretic (OT) framework, naturalness effects occur when a markedness constraint that prohibits unnatural forms ( $M_{Unnat}$ ) dominates a markedness constraint that prohibits natural forms ( $M_{Nat}$ ), or when  $M_{Nat}$  is absent from the grammar. Phonological patterns are associated with at least one markedness constraint, which is high-ranking and prefers natural candidates over unnatural candidates. In particular, the markedness constraint is ranked above any constraints that would otherwise

prefer an underlying marked segment or feature to surface in the output. The unnatural candidates thus fail to surface because the natural candidates are more harmonic.

### 3.2.1 Weight-sensitive stress

Under standard OT analyses of stress, quantity-sensitive languages assign prominence to syllables with branching rimes, which are linked to more than one mora. The markedness constraint **WEIGHT-TO-STRESS** prefers candidates in which bimoraic syllables are phonologically prominent.

- (3) **WEIGHT-TO-STRESS** (Prince & Smolensky 1993/2004, based on Prince 1990)  
 Heavy syllables are prominent in foot structure and on the grid.

Figure 3.1 displays one simplified example of a heavy syllable attracting stress. The alignment constraint **ALL-FT-L** prefers candidates in which all feet within a word are left-aligned with the left edge of the word. In words containing all light syllables, the ranking **ALL-FT-L**  $\gg$  **ALL-FT-R** would ensure left-aligned feet. Assuming other constraint rankings (not shown) prefer trochaic feet, this would result in a default initial stress system, e.g. [(pá.ta).ka]. In Figure 3.1, however, a non-initial heavy syllable disrupts this pattern. The winning candidate [pa.(tán.ka)] is preferred by high-ranking **WEIGHT-TO-STRESS**, because the bimoraic syllable [tan] is realized as the most prominent element of a foot. It receives one violation of **ALL-FT-L** because there is one syllable between the left edge of the word and the left edge of the only foot in the word. The losing candidate, \*[(pá.tan).ka], receives one violation of **WEIGHT-TO-STRESS** due to bimoraic [tan] surfacing as the weak member of a foot, and no violations of **ALL-FT-L**.

/pa.tan.ka/	<b>WEIGHT-TO-STRESS</b>	<b>ALL-FT-L</b>
☞ [pa.(tán.ka)]		*
*[(pá.tan).ka]	*!	

Figure 3.1: Stress attraction in LHL words

While ALL-FT-L is technically an alignment constraint and not a faithfulness constraint, the ranking depicted in Figure 3.1 produces the desired effect: a high-ranking markedness constraint determining the most harmonic candidate.

Naturalness effects in stress arise amongst quantity-sensitive languages because the effects of WEIGHT-TO-STRESS are seen in words containing heavy syllables. Light syllables, by way of being monomoraic, can surface as the prominent member of a foot through alignment constraints, but they cannot attract stress *away* from the default. Figure 3.2 depicts this, using the same ranking as Figure 3.1 but a different input (/HLH/, instead of /LHL/). The winning candidate [(pán.ta).kan] receives one violation of WEIGHT-TO-STRESS from unfooted [kan], while the losing candidate receives two: one from unfooted [pan] and one from weak [kan]. Simply put, this ranking leaves no opportunity for the peninitial syllable to attract stress if it is light.

/pan.ta.kan/	WEIGHT-TO-STRESS	ALL-FT-L	ALL-FT-R
☞ [(pán.ta).kan]	*		*
[pan.(tá.kan)]	**!	*	

Figure 3.2: No stress attraction in HLH words

While alignment constraints (at the foot- and word-level) control which syllable receives default stress, the markedness constraint WEIGHT-TO-STRESS controls the ability of certain syllables to attract stress away from the default. Since WEIGHT-TO-STRESS only works to align stress with bimoraic (or heavier) syllables, this encodes a built-in ability for heavy syllables to attract stress, which light (CV) syllables lack. The absence of an equivalent markedness constraint that prefers candidates containing stressed light syllables explains why this unnatural phonological pattern is absent in cross-linguistic typology.

### 3.2.2 Final devoicing

One way of capturing final devoicing in Optimality Theory is with high-ranking markedness constraint \*VOICED-OBS]<sub>Word</sub>, which prefers candidates without voiced obstruents



at the end of the word.<sup>2</sup> This markedness constraint outranks the faithfulness constraint  $\text{IDENT}_{IO}$ , which prefers candidates that have the same featural makeup as the underlying representation. Given an underlying representation with a word-final voiced obstruent, such as /paz/ in Figure 3.3, the ranking of Markedness  $\gg$  Faithfulness will prefer the natural candidate [pas], in which the final obstruent is realized as its voiceless counterpart [s], over the unnatural faithful candidate.<sup>3</sup> If the underlying representation already ends in a voiceless segment, such as /pas/ in Figure 3.4, the natural candidate is also the faithful candidate. It surfaces because it is preferred by the markedness constraint, but it is also preferred by the lower ranking faithfulness constraint.

/paz/	*VOICED-OBS] <sub>Word</sub>	IDENT <sub>IO</sub>
☞ [pas]		*
[paz]	*!	

Figure 3.3: Final devoicing tableau (UR with a final voiced segment)

/pas/	*VOICED-OBS] <sub>Word</sub>	IDENT <sub>IO</sub>
☞ [pas]		
[paz]	*!	*

Figure 3.4: Final devoicing tableau (UR with a final voiceless segment)

The tableaux in Figures 3.3–3.4 represent a language that has final devoicing in its phonological grammar. Languages that do not have phonological final devoicing would simply rank these two constraints in the opposite order,  $\text{IDENT}_{IO} \gg *VOICED-OBS]_{\text{Word}}$ . Note that neither ranking leads to a pattern of phonological final *voicing*, which would prefer that both /paz/ and /pas/ surface as [paz]. The difficulty of learning this pattern arises from the lack of the corresponding markedness constraint that

<sup>2</sup>Flack (2009) argues that markedness constraints that exist at the edge of a prosodic category also exist at the edges of other prosodic categories. In other words,  $*VOICED-OBS]_{\text{Word}}$  is part of a family of constraints that also includes  $*VOICED-OBS]_{\sigma}$  and  $*VOICED-OBS]_{\text{Utter}}$ .

<sup>3</sup>Another way to account for this pattern is with a markedness constraint that prohibits voiced obstruents ( $*VOICED-OBS$ ) or laryngeal features ( $*LAR$ ), which is outranked by a positional faithfulness constraint  $\text{IDENT}_{\text{Onset}}$  that protects identity in onsets (Lombardi 1999).

eliminates final voiceless segments, e.g. \*VOICELESS-OBS]<sub>Word</sub>. Either this markedness constraint does not exist within the available constraint set, or it is universally ranked below \*VOICED-OBS]<sub>Word</sub>. Under Optimality Theory, the unavailability of this constraint to decide the most harmonic candidate determines this pattern's absence from Universal Grammar.

### 3.2.3 Coda sonority

One explanation for why sonorant codas are less marked than obstruent codas within UG begins with the assumption that a sonority hierarchy exists. One very basic version of a sonority hierarchy is given in (4) (e.g., Zec 1988).

(4) *Sonority hierarchy*

Obstruents < Nasals < Liquids < Glides < Vowels

The hierarchy provided in (4) can be expanded to make finer distinctions between the segment types. The hierarchy in Clements (1990), for example, draws a distinction in the 'Obstruent' category between oral stops and fricatives (*oral stops* < *fricatives*). Dell & Elmedlaoui (1985) distinguishes between both voiced and voiceless stops and fricatives in the 'Obstruent' category (*voiceless stops* < *voiceless fricatives* < *voiced stops* < *voiced fricatives*), while Gouskova (2004) further expands the 'Liquids' category into *laterals* < *rhotics*. For the purposes of this discussion, however, the hierarchy in (4), which simply states that obstruents are less sonorous than all other segments, will suffice.

The second assumption is that the sonority hierarchy is encoded into phonology as part of UG. Clements (1990) argues this on the basis that, cross-linguistically, phonological patterns function very similarly in terms of sonority, but there is no clear phonetic correlate that encapsulates sonority in any uniform way (Ohala & Kawasaki 1984).

Given this, it has been well observed that sonority plays a large role in governing the well-formedness of syllables. Prosodic principles that rely on sonority act on the hierarchy in a set/subset relationship. Constraints placed on the well-formedness of syllables can target any node on the sonority hierarchy, as well as all higher nodes (Zec 1988). In the domain of coda sonority, if a given element on the sonority hierarchy can occupy the coda position in a language, then all elements of greater sonority must also

be allowed in coda position.

One way of cashing this out is through a markedness constraint that speaks directly to which types of segments can occupy coda position. Itô's (1986) CODA CONDITION (CODACOND), which restricts Place features in codas, is one example of this. Language-specific CODACOND constraints speak to coda markedness in individual languages (e.g., Arabic, McCarthy & Prince 1993, Ito & Mester 1994; Toba Batak, Crowhurst 2001). When CODACOND is ranked above faithfulness constraints like MAX and DEP, the resulting form either removes the illegal coda from the word or resyllabifies it as an onset. In Figure 3.5, the version of CODACOND from (5) is ranked above MAX. The faithful candidate [pat.ka] is dispreferred by high-ranking CODACOND and is ruled out. The illegal obstruent [t] is deleted, violating MAX to satisfy CODACOND.

- (5) CODACOND (hypothetical, for Figure 3.5)  
 Assign a violation mark for all obstruents that occupy coda position.

/patka/	CODACOND	MAX <sub>IO</sub>
پاتکا [pa.ka]		*
[pat.ka]	*!	

Figure 3.5: Coda sonority tableau (CODACOND ≫ MAX<sub>IO</sub>)

While this solution is adequate in some cases, Zec (1988) argues that CODACOND does not capture the effects of the hierarchy. Positing one CODACOND constraint limited to sonorants only, \*CODA<sub>[-son]</sub>, does not capture the generalization that coda well-formedness is hierarchical: it would be just as simple to posit its unattested opposite, \*CODA<sub>[+son]</sub>.

Hierarchies of this sort can, however, be encoded into a series of constraints (Prince & Smolensky 1993/2004; Gouskova 2004). Constraints of this sort lie in fixed order, but other constraints may be interspersed between them. Example (6) is one way to encapsulate the sonority hierarchy in (4) into a series of universally ranked constraints (following Orgun 2001 and similar to the M<sub>2</sub> hierarchy in Baertsch 2002).

- (6) *A constraint scale based on the sonority hierarchy in (4)*  
 \*OBSCODA  $\gg$  \*NASCODA  $\gg$  \*LIQCODA  $\gg$  \*GLIDECODA<sup>4</sup>

Given (6), we can envision a language in which obstruent codas are prohibited, but all higher sonority consonants are permitted in coda position. Figure 3.6 presents a hypothetical situation in which \*OBSCODA dominates MAX, which in turn dominates the rest of the constraints on the scale in (6). Just as in Figure 3.5, obstruent codas are deleted to satisfy high-ranking \*OBSCODA, violating MAX.

/patka/	*OBSCODA	MAX <sub>IO</sub>	*NASCODA	...	*GLIDECODA
☞ [pa.ka]		*			
[pat.ka]	*!				

Figure 3.6: Coda hierarchy tableau: Obstruent codas are deleted

Since MAX dominates the rest of the markedness constraints on the scale, however, sonorant codas will not be deleted. Figure 3.7 demonstrates how the underlying nasal in /panka/ is parsed as a coda, satisfying MAX while violating \*NASCODA. Consonants of higher sonority in this position would also be parsed as codas so as to satisfy MAX.

/panka/	*OBSCODA	MAX <sub>IO</sub>	*NASCODA	...	*GLIDECODA
☞ [pan.ka]			*		
[pa.ka]		*!			

Figure 3.7: Coda hierarchy tableau: Nasal codas are permitted

### 3.3 Interim summary

So far, this chapter has provided the explanation of a UG approach to naturalness, particularly within the framework of OT. Each unnatural pattern is associated with at

---

<sup>4</sup>One possible exception is the hypothetical constraint \*VOWELCODA. Since vowels are never parsed as codas, this constraint should either be universally undominated, or nonexistent (if the syllabification of vowels is subsumed under a different set of constraints). The status of \*GLIDECODA is dicey as well, given that glides are often parsed as the weaker element of a branching nucleus.

least one markedness constraint, which is satisfied by candidates that do not feature a particular unnatural, marked structure. When ranked above the relevant constraint, or constraints, that would otherwise allow marked structures in the input to be realized in the output, candidates that permit these marked structures are consistently less harmonic than candidates that prohibit them – in other words, a consistent suppression of phonetically unnatural forms. Natural forms, in turn, either have no markedness constraints that work to suppress them, or said constraints are universally ranked below the markedness constraints associated with the unnatural form.

Both the channel bias and the universal learning bias accounts comprise different ways of explaining the same phenomenon: the typological suppression of unnatural patterns. I am not convinced that the universal learning bias account adds anything to the investigation of typology that was not already accomplished by the channel bias account. The UG branch of the universal learning bias account assumes that there exists a constraint (or set of constraints) that suppresses all the same patterns that we already believe sound change to be suppressing. In this way, the universal learning bias account runs the risk of overproposing the amount of linguistic structure inside the minds of the learner. At this point, I have not argued that this is *not* the case. However, any linguist studying the effects of a supposed markedness constraint should be aware of where this constraint might come from. After observing some language tendency or universal – such as avoidance in some languages of final voiced obstruents – it may be tempting to posit that certain patterns are suppressed due to some mental architecture that all learners are meant to share, such as the violable constraint \*VOICED-OBS]<sub>Word</sub>. However, doing so may be amiss if the same phenomenon can be explained through examining the likely outputs of sound change.

An ideal explanation of any phenomenon should not only be explanatorily simple, but factually accurate: corroborated by existing evidence and able to accurately predict future outcomes. This, I will argue, is the other major problem with universal learning bias. Any universal learning bias account of naturalness effects – whether specific to UG or not – relies on some innate learning advantage of natural patterns over unnatural ones. Experimental research would suggest that we should see evidence of this advantage in studies of artificial pattern learning. The experiments in this dissertation do

see systematic problems in learning unnatural patterns, but only in the places where the channel bias account predicts them to be. Removing or lessening the perceptual difficulties of an unnatural pattern also removes or lessens any learning difficulties. In other words, an unnatural pattern that is easily perceived is also easily learned. This will be the main argument of Chapters 5–7.

### 3.4 The role of universal learning bias in phonology

Although this dissertation argues that universal learning biases do not suppress unnatural pattern learning, it does not assert that they play no role in *phonology*. On the contrary, they are necessary to explain the distribution of linguistic patterns that do not differ in terms of phonetic groundedness.

As mentioned earlier in this chapter, one example of this is simplicity effects – a learning and typological asymmetry between patterns that differ in terms of formal simplicity, but not in terms of phonetic groundedness. Simplicity effects govern a wide variety of patterns, both linguistic and nonlinguistic. The notion of simplicity that I refer to assumes, to begin, that the mind is prone to categorization. Humans tend to categorize items using a variety of dimensions, such as shape, color, and size. Patterns arise when generalizations can be drawn based upon those categories.

For example, consider the set of items in Figure 3.8. These items can immediately be categorized in two ways: shape (square or diamond) and color (black and white). Within these items, the black square and the black diamond are selected (`boxed`), and the white square and the white diamond are not selected. The pattern that arises indicates which of the four items should be selected, and draws upon the categorization of color. Patterns, then, rely on our ability first to categorize items in a set across various dimensions, and then to choose only the items that meet a certain description, such as ‘All shapes are black’.

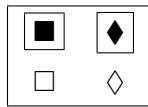


Figure 3.8: ‘All shapes are black’: A pattern with one predictable dimension (color)

The pattern in Figure 3.8 is considerably simple. Only one dimension of categorization – color – is needed to distinguish the items that are selected from the items that are not. In other words, the pattern ‘All shapes are black’ is *predictable by* one dimension.

Increasing the number of dimensions that a pattern is predictable by adds to its complexity. Experimental work has shown that participants are less successful at learning patterns that are predictable by two dimensions than patterns that are predictable by one dimension (e.g., Moreton 2008, 2012; Moreton et al. 2015; see Moreton & Pater 2012 for a discussion). This suggests an inverse relationship between number of dimensions and pattern learnability: as the number of dimensions necessary to define a pattern increases, the ease of learning the pattern decreases.

Another factor contributing to simplicity is the relationship between the dimensions that define a pattern. Within patterns that are predictable based on two dimensions, certain relationships between those two dimensions can contribute towards whether the pattern is learned more successfully. The interaction of number and relationship has led to a more nuanced, hierarchical definition of simplicity (Shepard et al. 1961; Pertsova 2012). In the simplicity hierarchy, patterns are arranged first by number (patterns with fewer defining dimensions are simpler than patterns with more defining dimensions), and then by the relationships between those dimensions. Figure 3.9 presents the hierarchical relationship of four types of patterns supported in Pertsova (2012). The simplest pattern, labeled AFF, is predictable based on a single dimension (color). The remaining three patterns are predictable based on two dimensions (shape and color).

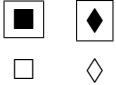
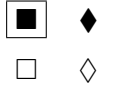
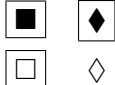
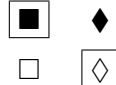
AFF	AND	OR	XOR
			
black	black AND square	black OR square	(black AND square) OR (diamond AND white)

Figure 3.9: Hierarchy of simplicity (from Pertsova 2012)

Previous experimental work has suggested that the simplicity hierarchy may govern

the learnability of phonological patterns. Evidence has been found not only for the learning advantage of simple (uni-dimensional) patterns over complex (multi-dimensional) patterns (e.g., Saffran & Thiessen 2003; Cristià & Seidl 2008), but of certain complex patterns over others (e.g., Pycha et al. 2003; Kuo 2009).

Crucially for this argument, the difference between the two patterns in question must be one of simplicity, not naturalness. Previous work on languages with weight-sensitive stress patterns suggests that universal learning biases play a role in determining which types of syllables may attract stress. In a typological survey, Gordon (2002; 2005) observes that patterns of weight-sensitive stress are predictable based on a single dimension, comparable to the AFF category in Figure 3.9. For example, many languages (e.g., Yana, Sapir & Swadesh 1960; Khalkha, Gordon 2002, and references therein) contain a stress pattern such that syllables with long rimes (CVC or CVV) are able to attract stress away from the syllable that is stressed by default. Other languages (e.g., Yimas, Foley 1991; Gujarati, Cardona 1965) contain a stress pattern such that syllables with non-high vowels (e.g., [a o]) are able to attract stress away from the default. Both of these observations are compatible with the concept of perceptual energy as the phonetic correlate of stress, described in Chapter 2.1.1. In both cases, the types of syllables that can attract exceptional stress are predictable by one dimension: rime length or vowel height.

Absent from Gordon's (2002) typological survey are patterns in which weight-sensitive stress is predictable by two dimensions. In other words, no such languages were found in which closed syllables with low vowels could attract stress (as in nonce word [pitánki]), but closed syllables with high vowels could not ([pítinkɑ]; \*[pitínkɑ]).

The difference between the attested pattern (closed syllables attract stress) and the unattested pattern (closed syllables with low vowels attract stress) cannot be boiled down to naturalness. If anything, the unattested pattern may somehow be *more* natural than the attested pattern. With both rime length and vowel height contributing to perceptual energy, a syllable like [tan] should have more perceptual energy than a syllable like [tin] and therefore should be a more natural stress-attracting syllable. Gordon (2005) attributes this asymmetry to a learning bias. The pattern that is predictable based on two dimensions is more difficult to learn than the pattern that is predictable



based on one dimension, in keeping with the simplicity hierarchies of Shepard et al. (1961) and Pertsova (2012). The pair of stress experiments discussed in Chapter 5 support Gordon's (2005) hypothesis and furthermore suggest that channel bias cannot account for the asymmetry between these two patterns. Together, both channel bias and universal learning bias shape the typology of weight-sensitive stress.

### 3.4.1 Do UG biases in phonology exist?

The simplicity biases that suppress certain complex linguistic patterns look very much like the biases that suppress the learning of complex non-linguistic patterns. In this dissertation, I address this only in the domain of stress, but previous experiments have found similar effects in other phonological domains (Saffran & Thiessen 2003; Cristià & Seidl 2008; Kuo 2009; see Moreton & Pater 2012 for an in-depth discussion of the literature). So far, I have argued that naturalness and simplicity, which have both been described as cognitive, linguistic markedness phenomena, both have other explanations. The former is not cognitive, and the latter is not purely linguistic.

The question that remains, then, is whether or not there are any cognitive structures that are used for phonology and phonology alone. Is phonology truly special, or is it simply an instance of larger pattern-learning structures applied to linguistic data? Likewise, are the *processes* used in generalizing over phonological patterns (phonologization, domain generalization, hyper- and hypo-correction, faithfulness) unique to linguistics, or can we see evidence of these processes active in non-linguistic patterns? This issue has formed much of the body of recent work by Moreton, Pater, and Pertsova (Moreton & Pertsova 2012; Moreton & Pater 2012; Moreton et al. 2015), who find evidence of similarities between phonological and non-linguistic pattern learning. Unless there is evidence that phonological pattern learning behaves unlike any other type of pattern learning, there may be little reason to propose that uniquely linguistic cognitive biases exist. It would, for example, be redundant to suppose that the cognitive system has one structure in place for language and another for color if learners exhibit similar strategies and similar problems for both types of data. Whether or not any phonological processes do fall under (2-b) is a pressing issue that I leave open to further investigation.

## Chapter 4

# Naturalness in the lab

In the pursuit of determining the underlying cause of naturalness effects, I take an experimental approach. Experimental research allows for controlled investigations of the circumstances under which patterns can, and cannot, be learned. The bulk of my argument for the channel bias approach to naturalness (and against the universal learning approach) comes from such experimental work. The channel bias approach is both supported by the experiments in this dissertation and capable of explaining some of the anomalies within existing research on naturalness in the lab. This chapter addresses these anomalies in detail.

Artificial grammar experiments in phonology allow us to test hypotheses regarding the underlying sources of naturalness and how effectively naturalness effects observed in natural language can be recreated in the lab (Pycha et al. 2003; Moreton 2008; Carpenter 2010, etc.; see Moreton & Pater 2012 for an overview). These types of experiments tend to be variants on the same, two-phase format. At least two artificial ‘languages’ are created for the purposes of the experiment, one based on an observed natural pattern and one based on its unnatural equivalent. Each participant is assigned to learn only one of these artificial languages. During the initial phase, participants are exposed to their assigned pattern, minimally by listening to the words of the pattern over headphones. In the second phase, participants are tested on their assigned pattern, usually by performing a task that requires them to sort a set of novel words into pattern-conforming or pattern-non-conforming. I will refer to these two phases as Exposure and Novel, respectively. Measures of accuracy and response time collected in the Novel phase

can provide telling information about how successfully each of the patterns was learned. Naturalness effects are said to be recreated in the lab if the participants learning the natural pattern perform significantly better on the Novel task (higher accuracy, lower response time, etc.) than participants learning the unnatural pattern.

Despite the number of experiments that have followed this format, whether or not naturalness effects can adequately be recreated in the lab remains unclear. Some experiments claim to have found evidence of such effects (Moreton 2008; Toro, Nespor, Mehler & Bonatti 2008; Carpenter 2010; Baer-Henney & van de Vijver 2012; Myers & Padgett 2014), while others claim to have found no such effect (Pycha et al. 2003; Wilson 2003; Seidl & Buckley 2005; Koo 2007; Kuo 2009). Possible conclusions from this mixture of results is that naturalness effects are not robust enough to be observed in the lab, or that this method of testing for them may not be effective (Moreton & Pater 2012).

I argue, instead, that prevailing assumptions about how to define and control for naturalness have muddled the overall picture of naturalness in the lab. If the driving force behind naturalness effects is, indeed, based in perception or production, then these effects should only appear in laboratory settings if the perceptual or production asymmetries that exist within typology are adequately recreated. In perception-based tasks where interact with the stimuli by listening to them, participants should have no problems learning an unnatural pattern whose stimuli are abundantly easy to perceive. With an increased understanding of what naturalness effects are and how to accurately tailor each experiment to recreate the specific pattern of interest, I expect that the overall results of naturalness experiments in the lab should become clearer.

This section outlines the problems involved in naturalness in the lab. First, I discuss the disagreements in how the term ‘naturalness’ is defined. I then take a closer look at some of the experiments probing for naturalness effects, outlining why some of them may not have been an adequate or appropriate test for naturalness and bringing attention to some of the more successful experiments. Finally, I provide the rationale for the experiments conducted for the purposes of this dissertation.

## 4.1 Defining naturalness

Amongst the definitions of naturalness provided in the literature, a common element is that naturalness must, in some way, be tied to phonetics. I have yet to see any definition of naturalness that does not reference phonetics in any way. This, it seems, is where the consensus ends. Large-scale differences between the definitions include whether a pattern can be ‘natural’ only because of phonetics, or whether adding another factor can make a pattern natural, such as phonetics plus typological frequency (Hayes et al. 2009; Carpenter 2010; Hayes & White 2013) or phonetics plus formal simplicity (Carpenter 2010). Additionally, some definitions make reference to the naturalness of *patterns* (Pycha et al. 2003; Seidl & Buckley 2005; Wilson 2006; Blevins 2006; Moreton 2008), while some refer to the naturalness of *phonological rules or constraints* (Hayes et al. 2009; Carpenter 2010; Hayes & White 2013). Amongst those papers that define naturalness only as a phonetic property of patterns, there are additional differences regarding the specific phonetic factors that contribute to a pattern’s naturalness.

To begin, consider the examples in (1)–(3). These definitions are alike in at least two respects: they speak to the naturalness of rules and constraints as opposed to patterns, and they define naturalness as a property of phonetics plus typological robustness (and, in the case of Carpenter 2010, in terms of formal simplicity as well).

- (1) “There are two ways to assess naturalness. First, a constraint should match typological data, particularly when it is related to the other constraints in the construction of a factorial typology...Second, a constraint can be asserted to be natural on phonetic grounds, when it can be shown to increase the ease of articulation or the saliency of contrasting forms in perception.” (Hayes et al. 2009)
- (2) “A natural rule is sometimes categorised as one that is formally simple (e.g. manipulates a single feature), as one that is phonetically or otherwise substantively grounded, or as one that is robustly attested cross-linguistically.” (Carpenter 2010)
- (3) “Phonological constraints are usually defended on two grounds: either typological or phonetic. The typological criterion can be expressed on the basis of

Greenbergian implicational universals. . . The phonetic criterion is that a constraint should be functionally effective, serving to form a phonological system in which words are easier to articulate or in which possible words are perceptually distinct from one another. . . We will refer to constraints that satisfy one or the other criterion as *natural*, and to other constraints as *unnatural*.” (Hayes & White 2013)

The choice to define naturalness in terms of patterns or in terms of constraints is deliberate. The former refers to an observational phenomenon that occurs in language, and the latter refers to the formalization of that pattern. I do not consider this distinction to be a problem in the naturalness literature, but it is worth noticing. One key difference between natural patterns and natural constraints is their connection to theory. To define naturalness as a property of rules or constraints is to assume a UG account of naturalness to some degree; in other words, it assumes an innate theory of markedness. Defining naturalness as a property of *patterns*, on the other hand, is theory neutral.

More problematic is the reliance on phonetics plus typological frequency. For one, including both aspects suggests that either both are necessary (a pattern is natural if it is *both* phonetically grounded *and* typologically frequent), or only one is (a pattern is natural if it is *either* phonetically grounded *or* typologically frequent). This is problematic because it assumes that both of these factors contribute equally to the naturalness of a pattern. Not all phonological patterns are rooted in phonetics, and so not all typologically frequent patterns are phonetically natural. For example, while utterance-final devoicing is a phonetically natural process, Myers & Padgett (2014) and Padgett (2014) argue that it is extended to word endings through domain generalization. During the process of phonologization, speakers, having misinterpreted the phonetic tendency to devoice as a phonological rule, apply that rule categorically at the word level instead of at the utterance level. The phonetic precursors for utterance-final devoicing do not apply to mid-utterance words, and so the authors argue that word-final devoicing is phonetically unmotivated and can be explained instead by cognitive means. Foot-based patterns provide another example of a frequent, yet phonetically neutral, phonological occurrence. An artificial grammar study by Bennett (2012) tested whether participants were sensitive to foot structure independently from stress. Participants were trained

on a phonotactic restriction within trisyllabic words wherein [i], and never [u], could appear after a stressed syllable: [(‘CV.Ci).CV], \*[(‘CV.Cu).CV]. This pattern could be interpreted as a restriction on post-tonic syllables (stress-dependent) or on the weak member of a foot (foot-dependent). When presented with five syllable words that bore only one stress, results found that participants generalized this restriction to the weak member of a covert foot: \*[(‘CV.Ci).(CV.Cu).CV]. These results provide evidence that speakers are predisposed to organize syllables into feet, even in languages with fixed stress.

The larger issue is that these definitions do not express the link between naturalness and frequency. As described in Chapter 2.2, natural patterns are more likely to survive across generations of speakers because of their perceptual and production ease. Patterns that are easier to perceive or produce are more likely to become phonologized as part of a language’s grammar, while phonetically difficult patterns are likely to morph into something less challenging. Typological frequency of natural patterns is thus a consequence of their phonetic groundedness, not a determiner of their naturalness. We may look to typological frequency to gain an idea of which patterns might be natural, but ultimately, the only defining factor of a pattern’s naturalness of its phonetic groundedness.

Now consider the definitions in (4)–(7). These definitions are alike in that naturalness is defined as a phenomenon of patterns, without reference to formalization of these patterns. In addition, the naturalness of a phonological pattern is defined only in terms of its phonetic grounding. I have added emphasis to highlight the specific discussions of the role of phonetics in each example.

- (4) “By ‘phonetically natural,’ we mean a pattern which could conceivably arise from listeners interpreting the **acoustic cues of speech** at face value – that is, interpreting them without reference to any grammar.” (Pycha et al. 2003)
- (5) “The crucial explanation in nasal assimilation and many other phonological patterns is that the direction of change is **phonetically grounded**. . . That is, the pattern arises due to **articulatory or auditory phonetic factors**.” (Seidl & Buckley 2005)

- (6) “In the field of generative phonology, which studies knowledge of linguistic sound systems, substance is now used in a broader sense to refer to any aspect of grammar that has its basis in the **physical properties of speech**. These properties include **articulatory inertias, aerodynamic pressures, and degrees of auditory salience and distinctiveness**.” (Wilson 2006)
- (7) “The most salient of all typological facts is that phonological patterns tend to be ‘phonetically natural’, in the sense that they resemble exaggerated or stylised expressions of **some phonetic fact**.” (Moreton 2008)

While these examples are alike in defining naturalness only in terms of phonetics, they differ in which aspects of phonetics they reference and how specific the references are. Within the first three examples, there are four broad categories that the authors use to define phonetic groundedness: acoustics, audition, articulation, and aerodynamics. None of these definitions make any explicit reference to perception. The definition in (4) refers only to acoustics, which I consider incomplete for two reasons. First, it excludes the role that production plays in shaping the naturalness of a phonological pattern. Second, listeners do not rely on the acoustics of a sound, but rather, on what they perceive. While perception of a sound is, in part, dependent on its acoustics, it is not a one-to-one relationship: listeners are more sensitive to certain acoustic distinctions and less sensitive to others. For example, whether or not a ten millisecond difference in a stop consonant’s voice onset time (VOT) affects identification of that stop depends on where in the continuum that ten millisecond difference lies. Stops with 20 ms VOT are more reliably identified by English speakers as voiced ([ba] or [da]), while stops with 30 ms VOT are more reliably identified as voiceless ([pa] or [ta]). However, the change from 0-10 ms VOT and from 40-50 ms VOT does not have as large an effect: the former is heard as voiced and the latter is heard as voiceless (Pisoni & Tash 1974; Ganong 1980). In addition, listeners have been shown to better perceive differences between tokens across some phonemic category than between tokens within the same phonemic category (Lieberman, Harris, Hoffman & Griffith 1957). If changes in acoustics cannot be perceived by the listener, they should not affect the naturalness of a pattern that relies on these acoustic changes.

The definition in (6) makes reference to articulation, aerodynamics, and audition of a sound. The terms “auditory salience and distinctiveness” more accurately reflect the connections between acoustic properties and the listener’s perception. A problem with this definition is its specificity: by listing only certain physical properties that are important for naturalness, the risk is that others will be left out. For example, the definition mentions “articulatory inertias”, but there are other properties of articulation that matter for naturalness as well. On the other hand, the definition in (7) is underspecified, making no reference to any categories of phonetic facts that may affect naturalness. I consider the mention of “articulatory or auditory phonetic factors” in example (5) to be nearly sufficient, yet lacking in its disregard of aerodynamics and perception.

I offer a potential solution to these problems: a definition of naturalness needs only to reference perception and production. Acoustics are wrapped up in perception, but only the differences in acoustics that matter to perception will affect the naturalness of a pattern. Production may refer to both articulation and aerodynamics, which are related but ultimately distinct. Both of these terms make indirect reference to the listener and the speaker, because what is ultimately important for naturalness is whether a pattern can easily be understood by the listener and easily produced by the speaker. With this in mind, this dissertation’s definition of naturalness, originally presented in the introduction, is repeated in (8). The term “natural” in this definition refers only to a phonological pattern and not to a rule or constraint.

- (8) A phonological pattern is natural if and only if it is phonetically grounded – that is, if and only if it is based in phonetic tendencies that facilitate perception and/or production.

I conjecture that naturalness is gradient, and it may be possible for one pattern to be “less natural” than another but not necessarily “unnatural” (Baer-Henney & van de Vijver 2012). Future research should work towards providing some measure of what it would mean for one pattern to be “more” or “less” natural than another.



## 4.2 Review of naturalness experiments

At minimum, a well-formed naturalness experiment must meet the following requirements laid out in (9):

- (9) *Requirements for well-formed naturalness experiments*
- a. The experiment must compare participants' performance on at least one phonetically grounded phonological pattern to that of its phonetically ungrounded counterpart.
  - b. The two patterns in question must differ only in their degree of phonetic groundedness.
  - c. The specific nature of the phonetic advantage that the natural pattern has over the unnatural pattern must be recreated in the lab.

I have not come across any naturalness experiments that do not at least attempt to meet requirement (9-a). However, simply comparing a so-called natural pattern to a so-called unnatural pattern is not enough to adequately recreate naturalness effects. Meeting requirement (9-b) ensures that any differences found between the two patterns really are representative of naturalness, and not of any other factor. A corollary of requirement (9-b), which I address in more detail in the following section, is that the two patterns in question must have the same degree of formal simplicity. Finally, meeting requirement (9-c) means having an understanding of the phonetic underpinnings of the pattern in question and how they give an advantage to the natural pattern and/or a disadvantage to the unnatural pattern, and controlling for those cues in the experiment. (This is similar, although not identical, to the recipe for eliciting hypo-correction in Ohala 1989.) This is crucial, because if naturalness is indeed a property of channel bias and not universal learning bias, then participants should be willing to learn natural and unnatural patterns, all things equal. Learning differences between the two should only be observed in the lab if the laboratory environment is set up to mimic, in some way, how these patterns behave in the wild. Chapter 4.2.2 addresses this. Finally, Chapter 4.3 provides some examples of experiments that meet all the requirements in (9).

### 4.2.1 Simplicity experiments are not naturalness experiments

In order to adequately test for naturalness effects in the lab, the natural pattern and its unnatural counterpart must be equally simple. By ‘simple’, I am referring to the metrics laid out in Chapter 3.4, which include, but is not limited to, number of and relationship between features. The patterns must be predictable by the same number of dimensions and, for multi-dimensional patterns, the relationship between these patterns must be the same (both *AND*, both *OR*, etc.). In phonological experiments, the “relationship” component may be extended to relationships between phonological tiers. A pattern that relies on an autosegmental relationship between two vowels – in other words, two segments on the vowel tier – may not be equally simple as one that relies on a relationship between a segment on the vowel tier and one on the consonantal tier (Moreton 2008; Moreton 2012).

A difference in formal simplicity may skew the results of the experiment, since the simpler pattern will likely be learned more successfully than the more complex pattern. If the unnatural pattern is more simple, it is unlikely that this pattern will yield a less successful performance. This may result in the unnatural pattern being learned just as well as, or even more successfully than, the natural pattern, masking any effect that naturalness would have otherwise had on performance. If the natural pattern is more simple, the results may yield something that looks like naturalness effects (more successful performance on the natural pattern), but for the wrong reason.

Wilson (2003) includes two experiments that test for naturalness in the domain of assimilation. In the first experiment, a pattern of nasal assimilation (suffix is [-na] if the final stem consonant is [+nasal], else [-la]) was tested against a so-called “random” pattern (suffix is [-na] if the final stem consonant is [+dorsal], else [-la]). The second experiment compared a pattern of nasal dissimilation (suffix is [-la] if the final stem consonant is [+nasal], else [-na]) to a different “random” pattern (suffix is [-la] if the final stem consonant is [+dorsal], else [-na]). The author remarks that there is no formal or substantive grounding for the random patterns. After being trained on their assigned pattern, participants completed a Novel phase task in which they reported whether or not they remembered hearing each Novel phase item in the Exposure phase. The Novel phase contained 20 old items and 60 new items; the pattern was considered “learned”

if the participants responded “yes” to significantly more new-grammatical items than to new-ungrammatical items. The results of the experiments showed no evidence that either of the random alternations had been learned, but that both the assimilation pattern and the dissimilation pattern had been successfully learned.

Is this result due to the lack of phonetic grounding in the random patterns, or to the degree of simplicity within the random patterns? Although, at a glance, these two patterns seem to be equally simple, they differ in terms of the relationship between the final consonant and the resulting suffix. In both the assimilation pattern and the dissimilation pattern, the final stem consonant and the suffix consonant are related in terms of one feature: if the final stem consonant is [ $\alpha$ nasal], the suffix consonant is also [ $\alpha$ nasal] (or [- $\alpha$ nasal] in the dissimilation pattern). Nasality is the only phonological feature that relates the two consonants, and as such, this is an example of an intra-dimensional relationship. In the “random” patterns, however, a second feature is needed to determine the nasality of the suffix: if the final stem consonant is [ $\alpha$ dorsal], the suffix consonant is [ $\alpha$ nasal] (or [- $\alpha$ nasal] in the second “random” pattern). This reference to dorsality makes the two “random” patterns inter-dimensional, and thus more formally complex than their intra-dimensional counterparts (Moreton 2012). The inability of participants to learn the “random” patterns can just as easily be explained by differences in simplicity as it can be by differences in naturalness. Without some way to tease these two explanations apart, this experiment cannot be considered an unambiguous test for naturalness effects.

A similar story is told in several other papers. Baer-Henney & van de Vijver (2012) studied the effects of both naturalness and locality on the learnability of an alternation pattern. In their artificial pattern in which the alternation was both natural and local, vowels in a  $C_1V_1C_2V_2$  word agreed in backness:  $V_2$  was [ $\alpha$ back] if and only if  $V_1$  was [ $\alpha$ back]. Their experiment included two unnatural variants: one in which the alternation was unnatural and local, and one where the alternation was both unnatural and non-local. In the former,  $V_2$  was [-back] *iff*  $V_1$  was lax. Some participants in this condition saw the reverse of this rule:  $V_2$  was [-back] *iff*  $V_1$  was [+tense]. In the latter,  $V_2$  was [-back] *iff*  $C_1$  was [+sonorant]. In the Novel phase, participants were asked to provide the correct suffix for a given word by speaking it into the microphone. The

natural and local pattern had the highest odds of a correct answer, followed by the unnatural/local pattern, followed by the unnatural/non-local pattern. The difference between the two unnatural patterns is certainly one of simplicity: the local pattern involves a relationship between segments on the same tier (the vowel tier), while the non-local pattern involves a relationship between segments on two different tiers (vowel and consonant). The difference between the two local patterns may be indicative of a naturalness effect, or they may be driven by formal simplicity: the natural pattern instantiates an intra-dimensional relationship that relies only on vowel backness, whereas the unnatural pattern is an inter-dimensional relationship that relies both on vowel backness and tenseness.

Hayes & White (2013) investigated the learnability of constraints based on patterns that all occur in English: some that were natural and some that were “accidentally true” – that is, encoded into the English lexicon but not phonetically based. The authors point out that all but one of their natural constraints refer to relationships between consonant sequences (same tier), and all but one of their “accidentally true” constraints involve consonant-vowel relationships (different tier). It cannot be determined whether the preference for the natural constraints came from their phonetic groundedness or their formal simplicity.

Similarly, Hayes et al. (2009) studied the degree to which phonetically natural and unnatural generalizations present in the lexicon could be internalized into the grammar. Their study of Hungarian vowel harmony within suffixes began with the observation that certain phonological environments have exceptionless, predictable harmony (e.g., right-most [+back] vowels take [+back] suffixes), while others demonstrate variability (e.g., words containing entirely [-back, -round] vowels may take [+back] or [-back] suffixes). Within these “zones of variation”, the experimenters identified statistical patterns in the lexicon that correlated with the frontness or backness of the suffix, some which were phonetically natural and some which were unnatural, and generated a set of weighted constraints that encapsulate these patterns (based on the results of a wug test). Their maximum entropy model of the wug test found that the unnatural constraints were underlearned when compared to a maximum entropy model based on the lexicon. However, the natural and unnatural constraints were not equally simple: as the authors

point out, the natural constraints referred only to backness harmony between stem and suffix, while the unnatural constraints referred to backness and height. It therefore cannot be guaranteed that the difference between the two sets of constraints is a facet of naturalness and not simplicity.

In order to provide a thorough test for naturalness effects, then, it is necessary that both the natural pattern and the unnatural pattern be equally simple. This involves adherence to at least all of the following requirements, if not more. First, for all naturalness tests, the two patterns must be predictable based on the same number of features.<sup>1</sup> There are further requirements placed on tests of dependency-based patterns. The two patterns must constitute the same type of relationship (i.e., both intra- or both inter-dimensional relationship). In addition, the two patterns must be equally complex in terms of the tiers that their predictable features are based on. A test in which both patterns refer to consonant-consonant sequences (same tier) satisfies this criterion, as does a test in which both patterns refer to consonant-vowel sequences (different tier). A test in which one pattern refers to consonant-consonant sequences and the other refers to consonant-vowel sequences does not satisfy this criterion. In short, any experiment in which the natural and unnatural pattern differ in formal simplicity cannot be considered a well-formed test for naturalness.

#### 4.2.2 Uncontrolled experiments do not capture naturalness effects

We now turn to naturalness experiments that meet the first two requirements of (9) – in that they compare performance on a natural pattern to an unnatural pattern, both alike in simplicity – but may fall short on the third. This is where the channel bias theory of naturalness makes its real contribution to experimental work. If it is true that naturalness effects, at their inception, are not ingrained into the grammar (until they have been phonologized), then we cannot expect experimental participants do perform differently on natural vs. unnatural patterns unless the circumstances that give the natural pattern an advantage in the wild are recreated in the lab. If the Novel phase includes a listening task, then providing participants with carefully-articulated stimuli

---

<sup>1</sup>Pertsova (2012), and others, discuss additional definitions of simplicity in patterns that make reference to more than one feature

with no background noise minimizes any perceptual differences between the patterns. Experiments that do not control for requirement (9-c) may still observe naturalness effects if channel biases arise in the lab by accident, but I contend that replicating the circumstances that capitalize on these phonetic inequalities may increase the likelihood that these effects will appear.

Many experiments that found no significant effect of naturalness lie in the domain of assimilation or harmony (Pycha et al. 2003; Wilson 2003; Koo 2007; Kuo 2009). In these experiments, patterns of assimilation are treated as “natural”, while equivalent patterns of dissimilation are treated as “unnatural”. In a review of naturalness experiments, many of which yielded null results, Moreton & Pater (2012) write that “results have been mixed at best, even within the same study” and that “the effect of phonetic substance is weaker [than the effect of formal simplicity] if it exists at all”. In this section, I first review some of the assimilation experiments and then speculate on why no effect was found.

Pycha et al. (2003) designed a series of patterns in which stem and suffix vowels could either agree in backness (“natural”, e.g. CiC-εk, CuC-λk) or disagree in backness (“unnatural”, e.g. CiC-λk, CuC-εk). They found that participants learning the vowel harmony pattern and participants learning the vowel disharmony pattern performed equally well at a task that asked them to determine whether or not novel forms fit their pattern. Upon finding a null result, (Pycha et al. 2003) write that “harmony and disharmony appear to be equivalently learnable patterns”.

Similarly, Wilson (2003, which I discussed in Chapter 4.2.1) found that both a pattern of nasal assimilation and a pattern of nasal dissimilation – where the suffix was either [-na] or [-la] depending on whether the final stem consonant was a nasal – were acquired during the Exposure phase. As these patterns were each statistically compared to an equivalent ‘random’ pattern and not to each other, it cannot be determined whether or not assimilation vs. dissimilation yielded a null result in this study.

Kuo (2009) studied place harmony within consonant-glide sequences: the sequences agreed in place in the harmony condition (e.g. [pwɔ tʃɔ]) and disagreed in place in the disharmony condition (e.g. [pʃɔ twɔ]). In a forced-choice task, where Mandarin-speaking participants selected which of two words, played auditorily, belonged to their pattern,

there was no significant difference between the two patterns. The author writes that the lack of naturalness effects in this experiment may be explained by the method of stimulus presentation, citing that participants were only played auditory stimuli, despite the role that articulation plays in the naturalness of consonant-glide sequences.

Koo (2007) is a variation upon this theme. Four patterns are considered: liquid harmony and disharmony (e.g. [r . . r] vs. [r . . l]) and backness harmony and disharmony (e.g. [i . . i] vs. [i . . u]). This dissertation allows for two major types of comparisons: harmony vs. disharmony and consonant vs. vowel relationships. These four patterns were tested using an auditory repetition task, in which participants were played both conforming and non-conforming words over the headphones and repeated them back into the microphone. This task found that the liquid harmony and disharmony patterns were both learned, but the vowel harmony and disharmony patterns were not learned. This is potentially problematic, since both backness harmony and disharmony were learned in Pycha et al. (2003), and suggests that auditory repetition may not be a suitable task for testing naturalness. In addition, the four conditions were not compared against each other, so there is no measure of whether or not harmony was learned more successfully than disharmony.

The two harmony conditions were tested a second time in Koo (2007) using a Yes-No task similar to that in Pycha et al. (2003). Both liquid harmony and vowel harmony were learned equally well in this task. Moreton & Pater (2012, and references within) claim that “phonological dependencies between non-adjacent vowels are thought to be more common than dependencies between non-adjacent consonants” and cite Koo (2007) as evidence against a naturalness difference between consonant vs. vowel dependencies.

Toro et al. (2008), on the other hand, found a conflicting result. In this experiment, vowel dependencies were learned, whereas consonant dependencies were not. The Exposure stimuli consisted of 12 CVCVCV nonce words. In the vowel-vowel pattern,  $V_1$  and  $V_3$  were identical, and in the consonant-consonant pattern,  $C_1$  and  $C_3$  were identical. The Exposure phase was set up a little differently from the previously discussed experiments, since the experimenters were partially interested in the information participants used to segment words. The Exposure words were played in a continuous stream, where the words were separated not by silence but by randomly assigned filler

syllables that did not adhere to any rule – thus, the vowel-vowel or consonant-consonant dependencies were one of the only cues to word segmentation in the Exposure phase.<sup>2</sup> One task of the Novel phase was a two-alternative forced choice generalization task, in which participants chose which of a pair of nonce words was more likely to belong to the pattern. This task included a new set of vowels in the vowel-vowel condition (and a new set of consonants in the consonant-consonant condition), such that participants not only had to generalize their pattern to novel words, but to novel segments. The results showed that participants in the vowel-vowel condition were able to generalize their pattern, whereas participants in the consonant-consonant condition were not.

While one interpretation of this collection of null or mixed results is that naturalness effects are capricious at best, I argue instead that it is indicative of the need for naturalness experiments to replicate more than just the patterns in question. In other words, in order to capture the advantage that assimilation has over dissimilation, the phonetic particulars of that advantage must also be included in the experiment. Kiparsky (1995) argues that assimilatory sound change is based in production, although his explanation takes a formalistic approach: a segment that is articulatorily weakened is de-linked from a particular feature, and then receives a new featural specification through feature spreading from other segments in its phonological environment. Vowel harmony, for instance, would be derived in this view first by reducing unstressed vowels and then through coarticulation from other vocalic gestures. Phonetic assimilations occur even more frequently in casual or fast speech, in which weakened articulations become even more masked by gestural overlap from the surrounding segments (Browman & Goldstein 1990; Ohala 1989; Keating 1996). From the perception side, these reduced gestures may be misperceived by the listener – for instance, if gestural overlap from the stressed vowel leads them to reinterpret the unstressed vowel as having the same features. In turn, the advantage over dissimilatory patterns is due, at least in part, to production ease. Assimilation is a process of reducing the number of articulatory gestures, whereas dissimilation creates gestures, or maximizes the differences between existing gestures. Continuing with the vowel harmony example, after unstressed vowels have been reduced, dissimilation works against the forces of coarticulation, while assimilation capitalizes on

---

<sup>2</sup>The consonants in the vowel-vowel pattern, and the vowels in the consonant-consonant pattern, were also designed to aid in segmentation.



them.

If the phonetic basis of assimilation is rooted partly in production, the null results could be a consequence of the experimental stimuli. Little is known about the specifics of how the stimuli are recorded. Many papers, including those discussed in this section, comment on the phonotactics and prosody of the pattern, age and dialect of the speaker, and the sampling rate of the recordings. The precise nature of how the speaker produced the stimuli is often unreported – whether they were using careful speech or not, how they were instructed to read the stimuli, whether they were instructed to aspirate stops or not reduce vowels, etc. Pycha et al. (2003) comment that their speaker used “normal prosody”, which is a step in the right direction but not very telling. If, for whatever reason, the phonetic correlates of the pattern did not make it into the recording of the stimuli, then the lack of a learning difference is unsurprising.

It may be possible to yield perception differences if the experimental stimuli take the production asymmetries into account – e.g., if the speaker produces the stimuli with heavy phonetic weakening or gestural overlap. My experiments in Chapters 6–7 adhere to this format. Such a design would constitute a mini sound change in the lab (in the sense of Ohala 1993) if listeners reanalyze the speaker’s degraded productions as phonetic correlates of a different category. For example, participants would hear a token of [pá.tə], in which gestural overlap from [ɹ] bled into the unstressed vowel, and reinterpret the word as [pá.tɹ]). If the stimuli of the two patterns are instead recorded in slow, careful speech with minimal phonetic weakening or gestural overlap and no mistakes in production, misperceptions are unlikely, and misproductions are impossible by the very nature of the task.<sup>3</sup>

Another potential solution to this problem is to present a production task in the Novel phase, instead of a perception task. This may provide a clearer picture of when

---

<sup>3</sup>I argue that Seidl & Buckley’s (2005) Experiment 1, while not an assimilation experiment, includes a task that may be ineffective in replicating naturalness effects. A headturn preference procedural study of 8- to 9-month-old infants found no performance difference between a natural pattern of intervocalic spirantization (only fricatives and affricates, but not stops, allowed inter-vocally) and an arbitrary pattern (only fricatives and affricates, but not stops, allowed word-initially). Spirantization, however, is rooted in production and correlates with faster speech rate (Soler & Romero 1999; DiCanio 2012). The task was perceptual in nature and did not control for speech rate, which may explain the null results.

and how gestural overlap (that may lead to assimilatory patterns) unfolds. A simple speaking task, however, may not do the trick. As mentioned, Koo's (2007) experiments included an auditory repetition task, which found no performance difference between consonant assimilation and dissimilation, or between vowel harmony and disharmony. Experiments of this nature have the benefit of replicating the production aspect but not the time course of naturalness. Sound change, of course, does not happen instantaneously, but gradually (Wedel 2003, 2006; Blevins 2006). A one-time production task may not be enough to capture how naturalness effects unwind across time.

A shadowing task, on the other hand, should prove more effective at this. In shadowing tasks, participants first produce and record a series of words that serve as their 'baseline' productions. They then listen to that same series of words, produced by a different speaker, and produce and record them again, either immediately after hearing each word or with a few seconds of delay. Their shadowed productions are compared to their baseline productions and the words they listened to to see how their shadowed productions were affected by listening to another speaker's voice (Goldinger 1998; Slowiaczek, McQueen, Soltano & Lynch 2000; Nye & Fowler 2003; and others). This method could be extended to replicate a mini sound change in the lab. Speaker A listens to a series of words, which have been recorded with some sort of degraded perceptual cue, and shadows them. Speaker B then listens to A's recordings and shadows them; Speaker C shadows B's words; and so on. Ideally, each speaker will introduce the degraded cue into their own shadowed speech in such a way that it becomes perceptually ambiguous to the next listener. This model of experimental task can be thought of as a small-scale, speeded version of how sound change occurs in the wild. Repeated shadowing tasks may be necessary to replicate naturalness effects in the lab when the pattern at hand is rooted in production.

I imagine that such shadowing tasks would also be beneficial to investigations of perceptually-based patterns. A task involving repeated iterations of listener/speaker interactions is certainly a better metaphor for how sound change functions in the wild than a task involving only one iteration. The experiments in this dissertation, however, suggest that one-time perception tasks are capable of replicating naturalness effects. Across all three experiments, the group of participants that demonstrated the most

difficulty with perception tasks also demonstrated difficulty with learning their assigned pattern. If the stimuli of the patterns in question are properly designed and controlled for, a one-time task can be successful at providing information about perception-based naturalness effects.

### 4.3 Experimental rationale

The picture of naturalness in the lab so far is wrought with inconsistencies. Papers disagree on how the term ‘naturalness’ should be defined, which leads to structural issues in designing experiments. Some papers’ supposed effect of naturalness is instead an effect of simplicity. Others fail to find an effect even when controlling for simplicity. Others still that meet both of the requirements in (9-a) and (9-b) do find a significant effect of naturalness. However, meeting just these two may not be enough. It is possible for experiments that do not explicitly intend to replicate the phonetic inconsistencies to still find an effect of naturalness. I expect that papers that adequately control for the phonetic correlates of a given pattern are even more likely to find an effect.

Using the experiments in this dissertation, I intend to create a guide for eradicating such problems in future studies of naturalness in the lab. All three experiments meet requirements (9-a) and (9-b). The Stress Experiments exemplify the importance of requirement (9-b) by demonstrating how, although unnaturalness and formal complexity can reduce learning in the lab, they do so for different reasons. The Final Devoicing and Coda Sonority Experiments demonstrate the role that requirement (9-c) plays: namely, that the learning difference between natural and unnatural patterns is significantly greater in fast, casual speech than in slow, hyperarticulated, careful speech. In other words, naturalness effects are greater when the Exposure stimuli replicate spontaneous speech found in the wild, with all its perceptual and articulatory degradation, but are smaller or nonexistent when the stimuli are exceedingly easy to perceive correctly.

The Stress Learning Experiment (one of the two experiments from this dissertation in the domain of stress) is based heavily on Carpenter (2010), which probed for, and subsequently found, an effect for naturalness in the domain of sonority-sensitive stress. In Carpenter’s experiment, participants were exposed to one of two artificial stress patterns. In both patterns, default stress fell on the initial syllable. In the natural

pattern, exceptional stress fell on the leftmost low vowel (either [æ] or [ɑ]), and in the unnatural pattern, exceptional stress fell on the leftmost high vowel (either [i] or [u]). The first pattern is natural from a perceptual energy viewpoint (Gordon 2005) because low vowels are more sonorous than high vowels (Kenstowicz 1994), which correlates with their higher intensity (Fant, Kruckenberg & Liljencrants 2000; Parker 2002; Gordon 2005).

Two versions of the experiment were run, one with American English-speaking participants and one with Québécois French-speaking participants. The patterns were presented in a series of alternative teaching and testing blocks, such that participants were continually tested on the words they learned in the Exposure phase. In the two Novel phases, participants were presented with pairs of words and were asked to select the word that conformed to their pattern.

Performance in the Novel phases on both the natural and unnatural patterns and in both sets of participants was significantly higher than chance. In both subject pools, accuracy on the Novel phases was higher for the natural pattern than for the unnatural pattern, although the French speakers performed overall less successfully than the English speakers. This paper shows evidence for a naturalness effect in a stress pattern that prevails even in a population that is thought to be “stress-deaf” (Carpenter 2010, and references therein).

The Stress Learning Experiment in this dissertation follows this design, but with some changes. The artificial patterns in my experiment focus in large part on rime length instead of vowel height. A third condition is added, which contains an artificial pattern that is phonetically natural but formally complex. These two patterns are compared side-by-side to a pattern that is both natural and simple. The Exposure phase is shorter than, but structurally similar to, that in Carpenter (2010). Finally, a Stress Perception Experiment is added to probe for the role of misperception in performance on the three patterns.

The Final Devoicing and Coda Experiments are not based on any previous works, but inspiration for the Final Devoicing Experiment comes from Myers & Padgett (2014). In their study, a pair of experiments demonstrated participants’ ability to learn a final devoicing pattern more successfully than a final voicing pattern. In the first experi-

ment, participants were assigned to learn one of two artificial patterns. Both [s] and [z] could appear word-initially, but in the pattern based on final devoicing, only [s] could be word-final ([pis], \*[puz]), and in the final voicing pattern, only [z] could be word-final ([puz], \*[pis]). After being exposed to their patterns, participants completed a Yes-No task where they decided whether novel utterances were pattern-conforming or pattern-non-conforming.<sup>4</sup> The group trained on the final devoicing pattern performed this task with higher accuracy than the group trained on the final voicing pattern. The second experiment tested participants' ability to learn alternations based on final devoicing. All participants were exposed to the same data, wherein the only obstruents that could occupy word-final position were voiceless. These were presented in both non-alternating ([git] ~ [giti]) and alternating ([git] ~ [gidi]) pairs, where the monosyllabic word represented a singular form and the bisyllabic word represented a plural. In the Novel phase, participants were played a novel word in plural form, followed by its singular counterpart, and were asked to decide whether or not this pair belonged to the language they had just learned. The results demonstrated an overall preference for final voiceless obstruents, even in situations where providing a final voiceless obstruent would create an alternating pair. Overall, these experiments demonstrate participants' ability to learn a natural pattern of final devoicing in the lab, both more successfully than its unnatural counterpart (Experiment 1) and in contexts where the unnatural form would give rise to non-alternation, which was hypothesized to be preferred (Experiment 2).

The Final Devoicing Experiment in this dissertation follows the basic format of comparing an artificial pattern of final devoicing to that of final voicing, but deviates from Myers & Padgett's (2014) Experiment 1 in most other ways. My experiment does not investigate domain generalization. All the stimuli are presented as single words, which had been recorded utterance-finally. In addition, my experiment adds stimulus clarity as a condition, with half of the participants learning their assigned pattern in hyperarticulated, careful speech, and the other half hearing fast, casual speech. The

---

<sup>4</sup>In addition to the utterance-final stimuli, participants also heard their Novel words in an utterance-medial position, which they had not yet been exposed to. This experiment therefore tested participants' ability to both learn their pattern in utterance-final position and to generalize their pattern to utterance-medial position. The latter is an instance of universal learning bias at work, as mentioned in Chapter §4.1.

idea is that the casual speech conditions more closely mimic one slice of how sound change functions in the wild, while the careful speech conditions demonstrate why using hyperarticulated stimuli in naturalness experiments so often yields null effects. The Coda Sonority Experiment uses this same format, applying it to a different phonological domain.

Ultimately, what we see is an emerging picture of channel bias at work in the lab. This account predicts that we should see naturalness effects in perceptually-driven patterns if and only if the unnatural stimuli are perceived less well than the natural stimuli. This explains why naturalness experiments based in assimilation do not tend to observe a difference between natural and unnatural patterns: the perception tasks that are so often used do not adequately capture that their asymmetry is based in production. Experiments that meet all three requirements in (9) but do not consciously aim to meet (9-c) can observe naturalness effects if the phonetic asymmetries are “accidentally” encoded into the stimuli – but if they are not, the experiment may yield null results. An additional task that checks for participants’ perception of their stimuli is therefore necessary to determine whether these asymmetries were accurately represented.

Explaining these observations from a universal learning bias point of view is more challenging. This account predicts that the advantage of the natural pattern should hold even in circumstances of high perceptibility. Thus, another point in favor of the channel bias account is its ability to explain existing problems between experimental papers on naturalness in a way that the universal learning bias account does not.

## Chapter 5

# Experiment: Stress learning and perception

This chapter contains the experimental methodology and results of two artificial grammar experiments studying naturalness in the domain of stress. The experiments test the weight-sensitive stress phenomena described in Chapter 2.1.1, wherein certain types of syllables can attract stress, while other types of syllables cannot. These experiments look at the stress-attracting ability of three types of syllables: closed (CVC), closed *and* containing a non-high vowel (shorthand CaC), and open (CV).

At least two constraining factors govern which types of syllables can be stress attractors. From a perceptual standpoint, syllables with greater perceptual energy make better stress attractors (Gordon 2005). Closed syllables have greater perceptual energy than open syllables due to the length of their rimes. For this reason, stress patterns in which syllables with branching rimes (but not syllables with simple rimes) attract stress are considered natural. The reverse, in which all monomoraic syllables (but no bimoraic syllables) attract stress, is considered unnatural. The experiments in this chapter show both a perceptual advantage and a learning advantage of natural pattern over its unnatural equivalent.

The same can be said for stress patterns based on vowel height. Low vowels have greater perceptual energy than high vowels due to their greater intensity. Patterns in which low vowels (but not high vowels) can attract stress are therefore natural, and patterns in which high vowels (but not low vowels) can attract stress are unnatural.

This experiment does not search for naturalness effects regarding vowel height, but this factor is used in the study of simplicity, which also appears to constrain stress patterns in typology.

Since both rime length and vowel height contribute to perceptual energy, closed syllables with low vowels (CaC) should make better stress attractors than closed syllables with high vowels (shorthand CiC). As Gordon (2002; 2005) notes, there are attested stress patterns in which all closed syllables (CVC) can attract stress (e.g. Yana, Sapir & Swadesh 1960; Manam, Chaski 1986, Lichtenberk 1983)<sup>1</sup>, and attested stress patterns in which all syllables with low vowels (Ca(C)) can attract stress (e.g. Yimas, Foley 1991). However, there is no attested pattern in which *only* closed syllables with low vowels attract stress (CaC, but not CiC or Ca). The unattested CaC pattern is phonetically natural, and perhaps even more natural than the attested patterns. If perception told the whole story, this typological asymmetry would be surprising. Gordon (2005) argues that the CaC pattern is ruled out by a simplicity bias. Because this pattern is predictable based on two features instead of one (rime length and vowel height), it is too complex for the learner. The experiments in this chapter show evidence in support of Gordon (2005).

This chapter contains two experiments on weight-sensitive stress. The first, named the Stress Learning Experiment, determines how well participants could learn each of the three patterns. The second, the Stress Perception Experiment, determines how well each of the stimuli of the three patterns could be perceived. The results of the two experiments together provide examples of both channel bias and universal learning bias in the lab. Channel bias rules out the CV pattern, while universal learning bias rules out the CaC pattern.

---

<sup>1</sup>As is noted in Gordon (2002) and Goedemans & van der Hulst (2013), the CVC-heavy distinction is typically extended to all syllables with branching rimes, CVX. I limit my discussion here to closed syllables to mirror the conditions in the Stress Learning Experiment, which did not contain a vowel length distinction.



## 5.1 Experiment 1: The Stress Learning Experiment

Three artificial stress patterns were designed for this experiment. Each followed the same basic pattern as to how stress was assigned, given in (1).

- (1) *Instructions for stress assignment*
- a. Stress the initial syllable, except:
  - b. If a **stress attracting syllable** is present in the word, stress the stress attracting syllable instead.

The assignment instructions follow a “stress the leftmost heavy syllable, else leftmost” pattern as seen in languages like Amele, Au, Lhasa Tibetan, Lushootseed, Murik, and Yana (list from Hayes 1995 and Walker 1995).

The three patterns differed, however, in what types of syllables could be considered **stress attractors**. In the ATTESTED pattern, all closed syllables (CVC) were able to attract stress. This pattern is exemplified in (2), which contains the three types of words designed for this experiment. Words in the ATTESTED pattern that contained all open (CV) syllables received initial stress by default. Words with an initial closed syllable also received initial stress because the stress attracting syllable occupied initial position. I refer to these words as having received initial stress “by stress attractor” (as opposed to “by default”). Finally, words with a closed syllable in peninitial position received peninitial stress, in compliance with the rule in (1-b). In trisyllabic words, the stress attracting syllable never appeared in final position; this was the case for all the patterns.

- (2) *Examples of stimuli from the ATTESTED pattern*      Stress attractor = CVC
- a. Default initial stress:                      kúpe      dógipa
  - b. Stress attractor in initial position:      tónta      térgati
  - c. Stress attractor in peninitial position:      kibóR      batíltu

The COMPLEX pattern is considered natural, from a perceptual standpoint, but differs from the ATTESTED pattern in terms of formal complexity. The stress attracting syllable is predictable by two features (rime length and vowel height), as opposed to one (only

rime length). To attract exceptional stress, syllables had to both be closed and contain a non-high vowel (CaC, CeC, and CoC). This means that words containing entirely closed syllables with high vowels (CiC and CuC), entirely open syllables with non-high vowels (Ca, Ce, and Co), or a mixture of the two received default stress on the initial syllable. CaC-type syllables could appear either initially or peninitially and received stress whenever present.

- (3) *Examples of stimuli from the COMPLEX pattern*      Stress attractor = CaC
- a. Default initial stress:                                      kápir      dóginpa
  - b. Stress attractor in initial position:                      **tónta**      **tár**getil
  - c. Stress attractor in peninitial position:                      kir**ból**      batéltu

Finally, the UNNATURAL pattern in (4) is, in a sense, the reverse of the ATTESTED pattern: the stress attracting syllable is open (CV), not closed. To demonstrate the stress attraction power of CV syllables, all other syllables in the UNNATURAL words needed to be closed (CVC). CV syllables could appear only in initial or peninitial position. Words that contain only closed syllables received default initial stress.

- (4) *Examples of stimuli from the UNNATURAL pattern*      Stress attractor = CV
- a. Default initial stress:                                      kúnper      dórginpan
  - b. Stress attractor in initial position:                      **tótan**      **té**gartil
  - c. Stress attractor in peninitial position:                      kir**bó**      baltítun

The experimental conditions in the Stress Learning Experiment are summarized in Table 5.1. The boxed syllables represent the stress attractors.

ATTESTED	COMPLEX	UNNATURAL
Closed > others	Closed and non-high > others	Open > others
n = 19	n = 23	n = 16
<div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px;">CaC</div> <div style="border: 1px solid black; padding: 2px;">CiC</div> </div>	<div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px;">CaC</div> <div>CiC</div> </div>	<div style="display: flex; justify-content: space-around;"> <div>CaC</div> <div>CiC</div> </div>
<div style="display: flex; justify-content: space-around;"> <div>Ca</div> <div>Ci</div> </div>	<div style="display: flex; justify-content: space-around;"> <div>Ca</div> <div>Ci</div> </div>	<div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px;">Ca</div> <div style="border: 1px solid black; padding: 2px;">Ci</div> </div>

Table 5.1: The three artificial stress patterns in the Stress Learning Experiment

## 5.1.1 Methodology

### 5.1.1.1 Participants

59 undergraduate students at the University of California, Santa Cruz, participated in this experiment. One participant was excluded from the analysis due to leaving the experiment early. The ages of the remaining participants ranged between 19 and 43 years old (mean = 21.19 years). Participants received either course credit or \$10 as compensation for their help.

51 of the 58 participants were coded by the experimenter to be native speakers of English. This information was culled from an intake form that was completed before the experiment began. Participants were asked to list all the languages they had experience with. For each language, they were asked to report the age at which they began learning it and rate, on a scale from 1–4 (4 = highest), their self-assessed competence in reading, writing, speaking, and understanding in that language. I considered participants to have native proficiency in some language if and only if: (a) they began learning that language at age 12 or earlier, and (b) they reported a combined average of at least 3.5 for the ‘speaking’ and ‘understanding’ categories. Since participants never had to interact with written forms of the stimuli, the ‘reading’ and ‘writing’ categories were excluded from this calculation of native proficiency. Under these same criteria, 13 of the 58 participants were coded as native speakers of Spanish, and all but one of the native Spanish speakers were also coded as native English speakers.

### 5.1.1.2 Stimuli

The three patterns contained a mixture of bisyllabic and trisyllabic nonce words. Stress could fall either initially or peninitially. Words with peninitial stress necessarily contained a stress attractor in peninitial position. Words with initial stress contained either no stress attractors or a stress attractor in initial position. This yielded six word types represented in the patterns, exemplified in Table 5.2. Every word contained exactly one stressed syllable (no secondary stress) and at most one stress attractor. The stress attractor never appeared word-finally in trisyllabic words.

The stimuli in this experiment were created using MBROLA (MBROLA Project Development Team 2010). MBROLA is a diphone-based speech synthesis program that

Stress placement	Attested	Complex	Unnatural
Initial (default)	páta túpeta	pátu túpirta	pántar túnpertal
Initial (attractor)	<b>bínko pálkode</b>	<b>bónko pálkode</b>	<b>bíkon pákolden</b>
Peninitial	godér batíltu	godér batéltu	gordé baltítun

Table 5.2: Examples of stimuli belonging to each pattern

allows users to create artificial stimuli without the use of recording equipment. For each phoneme, the user specifies its duration in milliseconds and zero or more pitch tuples, indicating a pitch value in Hz at a given percentage of the phoneme duration. This information is saved in a text file with the extension `.pho`, which MBROLA can then convert into an audio `.wav` file. As the experiment necessitated a large amount of MBROLA files, I wrote a Python script (<http://www.python.org>) that would automatically generate the appropriate `.pho` file for each given stimulus.

The phonemic inventory of all three patterns consisted of three nonhigh vowels [a e o], two high vowels [i u], six onset consonants [p t k b d g], and three coda consonants [n l r]. Codas were restricted to sonorants only to facilitate perception of rime duration. The stimuli were synthesized using MBROLA’s `de2` German male voice, chosen for its wider range of vowel height when compared to the other diphone databases. A German voice was chosen instead of an English voice for its monophthongal vowels (to counteract for unwanted effects of vowel length on stress perception).

Two phonetic correlates of stress were encoded into the stimuli: rime duration and pitch (Carpenter 2010; Bennett 2012). For duration, stressed syllables had longer rimes than their unstressed counterparts. The vowel of stressed syllables measured 200 ms; coda consonants inside stressed syllables measured an additional 100 ms. This resulted in stressed open syllables measuring at 200 ms and stressed closed syllables

measuring at 300 ms.<sup>2</sup> One consequence of this design is that, in the UNNATURAL pattern, syllables with default stress (e.g., [pán.tʌR]) were longer than stress attractor syllables (e.g., [bí.kon]). This was intended to discourage participants from hearing stress attractor syllables as containing a long vowel ('CV:, instead of the intended 'CV) – in other words, to discourage them from hearing a natural pattern in which long vowels attract stress. Unstressed open syllables had a duration of 100 ms and unstressed closed syllables had a duration of 166 ms (the coda consonant measured at 66 ms).<sup>3</sup> Stressed open syllables, therefore, were longer than both unstressed open and unstressed closed syllables. Most importantly, these measurements allowed duration to remain a cue for stress in UNNATURAL words containing a stress attractor (e.g. [bí.kon] → 200 ms [bí], 166 ms [kon]). All onset consonants had a duration of 95 ms (Bennett 2012), and every syllable contained a single onset consonant.

The pitch correlate was encoded by ensuring that every word had a high pitch accent that fell on the stressed syllable. At the beginning of the rime, the pitch of the stressed syllable was 120 Hz. Over the course of the rime, pitch rose to 140 Hz and then fell back to 120 Hz. The 140 Hz point fell at 50% of the rime's duration. Stressed open syllables reached 140 Hz at 100 ms after the beginning of the rime (195 ms overall), and stressed closed syllables reached 140 Hz at 150 ms after the beginning of the rime (245 ms overall).

All stressed syllables had identical pitch specifications regardless of where they appeared in the word. The pitch of unstressed syllables, however, was affected by their placement within the word in pitch. Pretonic syllables could only rise in pitch (e.g., [ba] in [ba.tʌl.tu]), and posttonic syllables could only fall in pitch (e.g., [gi] in [dó.gi.pɑ]). Under these parameters, all words with initial stress had falling pitch across the duration

---

<sup>2</sup>There seems to be some debate about the duration of vowels inside closed syllables. In a study of the SWITCHBOARD corpus, Greenberg, Carvey, Hitchcock & Chang (2003b) reports an average vowel duration of 172 ms inside CVC syllables. Hillenbrand, Getty, Clark & Wheeler (1995), on the other hand, measure an average of 225 ms for the five vowels used in this experiment inside /hVd/ syllables. Using a 200 ms vowel inside stressed syllables may serve as a happy medium between these two estimates but neglects to take into account more fine-grained effects of duration, such as duration based on vowel quality and based on the place and manner of the following consonant.

<sup>3</sup>Greenberg et al. (2003b) report similar measurements for codas in stressed and unstressed CVC syllables.

of the word, bisyllabic words with peninitial stress (e.g., [kir.**bó**]) had rising pitch, and trisyllabic words with peninitial stress had a pitch contour that rose pretonically and fell posttonically. In all trisyllabic words, the final syllable fell from 69 – 60 Hz, regardless of whether or not it immediately followed the stressed syllable.

Table 5.3 provides the numeric specifications of duration and pitch for each syllable type. Using the Python script, these specifications were entered into MBROLA, along with the phonemes of the word, to synthesize the stimuli.

Syllable	Parameter	Open (CV)	Closed (CVC)
Stressed	Rime duration	200 ms	300 ms (V = 200 ms)
	Pitch change	120 → 140 → 120 Hz over course of rime	120 → 140 → 120 Hz over course of rime
	Pitch peak	140 Hz	140 Hz
	Peak location	50% of rime = 100 ms	50% of rime = 150 ms
Unstressed	Rime duration	100 ms	166 ms (V = 100 ms)
Unstressed, initial	Pitch change	85 → 94 Hz over course of syllable	85 → 94 Hz over course of syllable
Unstressed, peninitial	Pitch change	92 → 69 Hz over course of syllable	92 → 69 Hz over course of syllable
Unstressed, final (third)	Pitch change	69 → 60 Hz over course of syllable	69 → 60 Hz over course of syllable

Table 5.3: Numeric specifications of duration and pitch for the Stress Learning Experiment stimuli

### 5.1.1.3 Procedure

The experiment consisted primarily of two phases. In the Exposure phase, participants were taught their assigned pattern by listening to short lists of words belonging to their pattern and being tested on them. This process repeated until all 32 Exposure words have been revealed. In the Novel phase, participants heard 64 words that were not presented in the Exposure phase and, for each word, determined whether or not it belonged to the pattern they were learning.

The experiment was written using the E-Prime software (Schneider, Eschman & Zuccolotto 2002) and was conducted in the phonetics lab at the University of California, Santa Cruz, supervised by the experimenter. Each participant sat at an individual computer in a soundproof booth and listened to the audio stimuli over AKG k271 headphones. A button box<sup>4</sup> was used to advance to the next screen and to respond to questions. Table 5.5 summarizes the procedure in the Stress Learning Experiment.

*Warm-up exercise.* Before the beginning of the Exposure phase, participants completed an AXB discrimination exercise that familiarized them with the task of listening for stress. In each trial, three trisyllabic nonce words were played in a row. All three words had exactly one stressed syllable, designed using the same parameters described in Chapter 5.1.1.2. The first word and the last word differed in which syllable carried the stress, and the middle word had stress on the same syllable as either A or B: e.g. [pó.bi.ke], [pó.bi.ke], [po.bí.ke]. Participants were instructed to press the ‘1’ button on the button box if the first word shared its stress pattern with the middle word, or the ‘5’ button if the last word shared its stress pattern with the middle word. The first 6 trials consisted of triplets that minimally differed in stress placement (e.g., [pó.bi.ke], [pó.bi.ke], [po.bí.ke]). The next 18 triplets also differed in their segments (e.g., [kí.da.tu], [pi.dú.bu], [ta.mí.bo]).

The triplets were equally balanced in terms of whether the correct answer was ‘1’ (first word) or ‘5’ (last word). They were also equally balanced in terms of where stress fell in the middle word (initial syllable: e.g., [pó.bi.ke]; peninitial syllable: e.g., [da.gú.si]; final syllable: e.g., [bi.go.fi]). Finally, the stress placement in the incorrect word was also varied throughout the warm-up (e.g., \*[po.bi.ké], \*[dá.gu.si], \*[bi.gó.fi]).

Initially, this warm-up exercise was designed as a method of excluding participants who may have more difficulty hearing stress. The intention, similar to Carpenter (2010), was to exclude from the analysis the results from participants who scored less than an average of 70% on the warm-up. However, adding or excluding the group that scored less than 70% did not affect the outcome of the experiment: the same statistical significances were found when the group was excluded and when the group was included. It is possible, then, that this exercise was not a meaningful test of stress perception.

---

<sup>4</sup>Serial Response Box from Psychology Software Tools, model #200A.

The final analysis (Chapter 5.1.2) presents the results from all participants, regardless of their performance in the warm-up exercise.

*Exposure phase.*<sup>5</sup> In this phase, participants were repeatedly played 32 nonce words that exemplified their assigned pattern and then tested on those words. At the beginning of this phase, participants were told that they would be learning a newly-discovered language of Papua New Guinea (a method that proved successful in Bennett 2012, to convince participants to perceive the synthesized stimuli as words of a real language).

Two sessions within the Exposure phase served to teach participants their pattern in incrementalized chunks. Before the Teaching session, participants were informed that they would hear a string of words in their assigned language and that they would be tested on the words they were learning. During each Teaching session, 4-6 of the 32 Exposure stimuli were played over the headphones. Each stimulus was accompanied with a generic photograph of an object, taken from the Bank of Standardized Stimuli (BOSS) corpus (Brodeur, Dionne-Dostie, Montreuil & Lepage 2010). Participants were told that each picture represented what its corresponding word (played over the headphones) meant. Stimuli were only ever presented auditorily and were never spelled out on the screen. The stimuli introduced in each Teaching session were played twice and in random order, yielding 8-12 trials in each session depending on the number of stimuli introduced.

A Testing session followed each Teaching session, in which participants were tested on the words they had just heard in a two-alternative forced choice (2AFC) task. In each trial of the Testing session, two stimuli were played in succession over the headphones (without an accompanying picture). One member of the pair was a word that had been presented in the previous Teaching session. The other member was segmentally identical but differed in stress placement (e.g., [tú.pe.ta], [tu.pé.ta]). Participants were instructed to choose which member of the pair belonged to the language they were learning and were given feedback after each pair. The feedback revealed whether their response was correct or incorrect, their response time for that question, and their overall accuracy within the Exposure phase up until that point.

A ‘Big Review’ and ‘Big Test’ session followed immediately after every two Teaching and Testing sessions. The Big Review played all the words presented in the previous

---

<sup>5</sup>The design of the Exposure phase is based off of Carpenter’s (2010) Experiment 1.



<b>Block</b>	<b>Description</b>
<i>Warm-up</i>	Heard triplets of nonce-words; either the first word or the last word shared its stress placement (initial, peninitial, or final) with the middle word. In an AXB task, asked to determine which word shared its stress pattern with the middle word.
<i>Exposure phase</i>	
Teaching 1	Heard 4 bisyllabic words with default initial stress, each played twice in random order with a corresponding image.
Testing 1	Tested on previous 4 bisyllabic words in a 2AFC task; played two words that minimally differed in stress placement and asked to determine which belonged to the language; given feedback.
Teaching 2	Heard 4 trisyllabic words with default initial stress, each played twice in random order with a corresponding image.
Testing 2	Tested on previous 4 trisyllabic words in a task identical to Testing 1.
Big Review 1	Heard previous 8 bi- and trisyllabic words with default initial stress, each played twice in random order with a corresponding image.
Big Test 1	Tested on previous 8 bi- and trisyllabic words in a task identical to Testing 1.
Teaching 3 & Testing 3	Heard 6 bisyllabic words (one default initial, two initial by stress attractor, three peninitial), then tested on those 6.
Teaching 4 & Testing 4	Heard 6 trisyllabic words (one default initial, two initial by stress attractor, three peninitial), then tested on those 6.
Big Review & Test 2	Reviewed previous 12 words and tested on them.
Teaching 5 & Testing 5	Heard 6 bi- and trisyllabic words (one default initial, one initial by stress attractor, four peninitial), then tested on those 6.
Teaching 6 & Testing 6	Heard 6 Heard 6 bi- and trisyllabic words (one default initial, one initial by stress attractor, four peninitial), then tested on those 6.
Big Review & Test 3	Reviewed previous 12 words and tested on them.
Final Big Review	Heard all 32 Exposure stimuli, each played once in random order with the corresponding image.
Final Big Test	Tested on all 32 Exposure stimuli in a task identical to Testing 1.
<i>Novel phase</i>	Played one novel nonce word at a time and asked to determine whether or not this word belonged to the language; 64 trials total; no feedback.

Table 5.5: Summary of the Stress Learning Experiment procedure

two Teaching sessions (10-12 new words). Again, each stimulus was played twice over the headphones in random order and accompanied by its original picture. The Big Test then tested participants on all 10-12 new words from the Big Review in the same 2AFC task described above.

The Exposure phase consisted of 6 Teaching and Testing sessions, with each Testing session following its corresponding Teaching session and with a Big Review and Big Test occurring after the second, fourth, and sixth Testing session. At the end of these six sessions, a Final Big Review replayed all 32 Exposure stimuli, again in random order and with their original picture. A Final Big Test tested participants on all 32 Exposure stimuli in the same 2AFC task they had completed throughout. The results from the Final Big Test are analyzed in the Results section (§5.1.2) and correspond to the ‘Exposure’ results.

It was crucial that participants learned the default stress pattern (initial stress in words with no stress attractors), the exception to the rule (stress falls on stress attracting syllables), and what type of syllable was the stress attractor in their pattern. To promote learning of the default pattern, the first two Teaching and Testing sessions presented only words with no stress attractors (bisyllabic words in the first Teaching session and trisyllabic words in the second). Not until the third Teaching session was it revealed that some words in the pattern have peninitial stress (in words where the stress attractor is peninitial). Participants needed to implicitly determine what type of syllable could trigger this exceptional stress and replicate their findings in the Testing sessions. To aid in learning this pattern, the last four Teaching and Training sessions presented a greater proportion of words with peninitial stress. Half of the stimuli introduced in the third and fourth Teaching sessions had peninitial stress, and two-thirds of the stimuli introduced in the fifth and sixth Teaching sessions had peninitial stress. Overall, the set of Exposure stimuli consisted of 12 words with default-initial stress, 6 words with initial stress by stress attractor, and 14 words with peninitial stress by stress attractor.

*Novel phase.* After the conclusion of the Final Big Test, the Novel phase began, in which participants were tested on 64 nonce words that had not been presented in the Exposure phase. Half of the novel words conformed to the pattern and half did not. In each trial, participants heard a single word over the headphones (not accompanied with

a picture) and were instructed to press the ‘1’ button if the word belonged to the pattern and the ‘5’ button if the word did not belong to the pattern. The instructions informed participants that none of these words had been presented in the previous phase, thereby assuring them that this was not a memorization task. 16 of the novel words contained default initial stress, 16 contained initial stress by stress attractor, and 32 contained peninitial stress by stress attractor. This set contained no trisyllabic words with final stress. The stimuli were presented in random order and no feedback was given.

#### 5.1.1.4 Hypotheses

This experiment design tests the hypotheses provided in (5) and (6). Performance in each pattern is measured in terms of the proportion of correct trials in both the Exposure phase (the Final Big Test only) and the Novel phase. More successful performance is reflected in a higher proportion of correct trials.

- (5) **Naturalness:** Phonetic naturalness affects how well a pattern may be acquired. Since the UNNATURAL pattern is not phonetically grounded, accuracy on this pattern should be lower than in the ATTESTED pattern, which is phonetically grounded.
- (6) **Simplicity:** Formal simplicity affects how well a pattern may be acquired. Since the COMPLEX pattern is predictable based on two features, accuracy on this pattern should be lower than in the ATTESTED pattern, which is predictable by one feature.

Note that this experiment does not distinguish between the channel bias and the universal learning bias accounts. The Stress Perception Experiment in Chapter 5.2 addresses this.

#### 5.1.2 Results

This section presents results from the 58 participants in the Stress Learning Experiment. 19 were trained on the ATTESTED pattern, 23 on the COMPLEX pattern, and 16 on the UNNATURAL pattern. The proportion of correct responses in each pattern and in each

of the phases is presented in Figure 5.1, while the  $d$  primes are presented in Figure 5.2. The results from the Exposure task correspond to the light gray bars, and the results from the Novel task correspond to the dark gray bars. Below this, Table 5.6 provides the numerical values for proportion of correct trials.

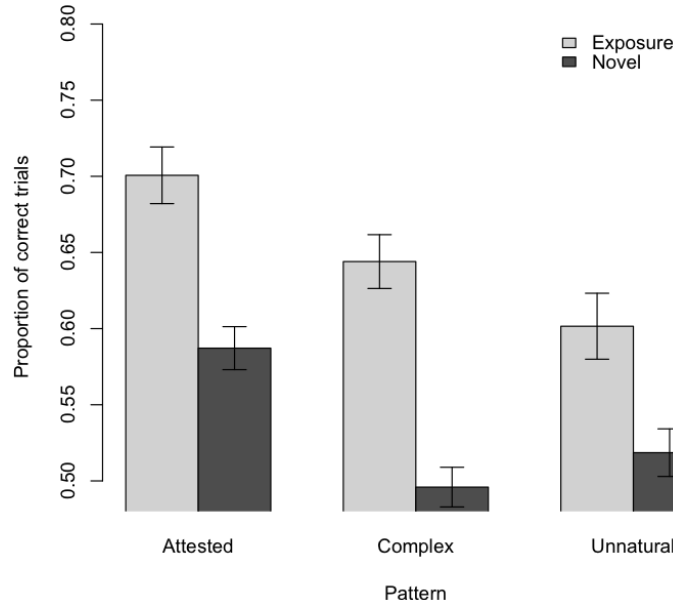


Figure 5.1: Proportion of correct trials by PATTERN and BLOCK ( $y$ -axis begins at 0.50)

Pattern		Accuracy		$d$ prime	
		Exposure	Novel	Exposure	Novel
Attested	n = 19	0.701	0.587	1.206	0.517
Complex	n = 23	0.644	0.496	0.794	-0.004
Unnatural	n = 16	0.602	0.519	0.603	0.121

Table 5.6: Average accuracy and  $d$  prime in each pattern

Each pattern had a higher proportion of correct trials in the Exposure phase than in the Novel phase. This is unsurprising, given that participants were explicitly and repeatedly trained on their Exposure words, whereas they heard the set of Novel words only once.

A more intriguing result is how the patterns compare in each of the phases. Figure

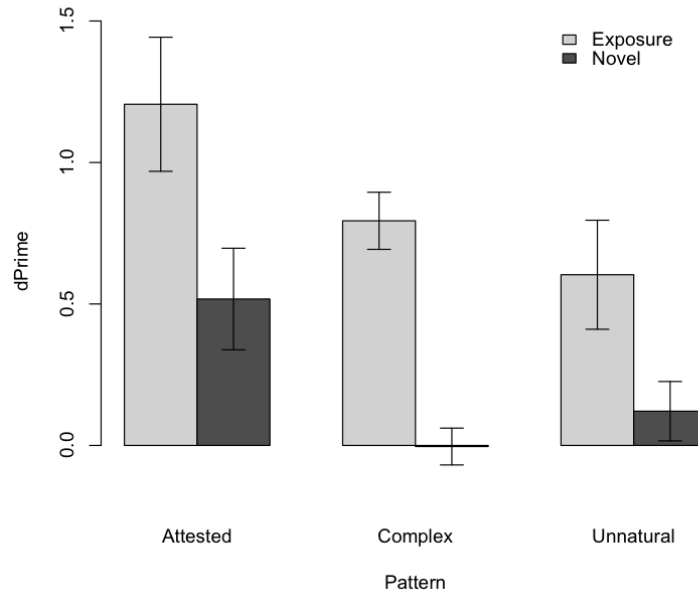


Figure 5.2: d primes by PATTERN and BLOCK

5.1 shows that, in the Exposure phase, the ATTESTED pattern had a higher proportion of correct trials than either the COMPLEX or the UNNATURAL patterns. The same trend holds in the Novel phase. Additionally, the proportion correct in the Novel phase for the COMPLEX and UNNATURAL patterns do not appear to be different from chance (0.50, or averaging half correct trials). Performance on the ATTESTED pattern appears higher than chance in the Novel phase.

Each subject’s individual average accuracy score in each pattern and phase was calculated. These scores were analyzed in R (Baayen 2009) with a linear regression, using the `lm()` function. The fixed effects were PATTERN and BLOCK. In Table 5.7, the ATTESTED pattern in the Exposure phase corresponds to the Intercept.

The linear regression shows a significant main effect of UNNATURAL ( $p < 0.01$ ). This means that, in the Exposure phase, the ATTESTED group had a higher average proportion of correct trials than the UNNATURAL group did. The main effect of COMPLEX fell just short of significance ( $p = 0.077$ ). The significant main effect of BLOCK confirms that the ATTESTED group performed better in the Exposure phase than they did in the Novel phase ( $p \approx 0$ ). There were no significant interaction effects, suggesting

that the decrease in proportion correct between the Exposure and Novel phases was essentially equal between the ATTESTED, COMPLEX, and UNNATURAL patterns.

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.701	0.024	29.795	≈0 ***
Complex	-0.057	0.032	-1.782	0.077
Unnatural	-0.099	0.035	-2.849	0.005 **
Novel	-0.113	0.033	-3.415	≈0 ***
Complex:Novel	-0.035	0.045	-0.770	0.443
Unnatural:Novel	0.030	0.049	0.620	0.537

Table 5.7: Linear regression (Intercept = ATTESTED:Exposure)

For both the COMPLEX and UNNATURAL groups, their average accuracy scores in the Novel phase were extremely close to chance. A series of one sample t-tests confirmed that neither of these groups performed significantly different from chance in the Novel phase (COMPLEX:  $\mu = 0.50$ ,  $t = -0.347$ ,  $df = 22$ ,  $p = 0.732$ ; UNNATURAL:  $\mu = 0.50$ ,  $t = 1.119$ ,  $df = 15$ ,  $p = 0.281$ ). The ATTESTED group, however, did perform the Novel task significantly differently from chance ( $\mu = 0.50$ ,  $t = 3.438$ ,  $df = 18$ ,  $p < 0.01$ ).

Finally, the results were analyzed to see if there was a significant correlation between how participants performed in the Exposure task and how they performed in the Novel task. The goal was to determine whether participants who were more successful at learning the explicit training words in the Exposure phase were also more successful at generalizing to new words in the Novel phase. A scatterplot of each participant's average proportion of correct trials is given in Figure 5.3, with performance in Novel as a function of performance in Exposure. The points are coded for the pattern that each participant was exposed to.

Visually, it appears that the correlation between the two phases is stronger in the ATTESTED pattern than in both the COMPLEX and UNNATURAL patterns. The ATTESTED group has a positive slope, such that performance in the Novel phase is positively correlated with performance in the Exposure phase. Recall that performance in the Novel phase was overall lower than performance in the Exposure phase, and so extremely high performance in Exposure did not guarantee similarly high performance

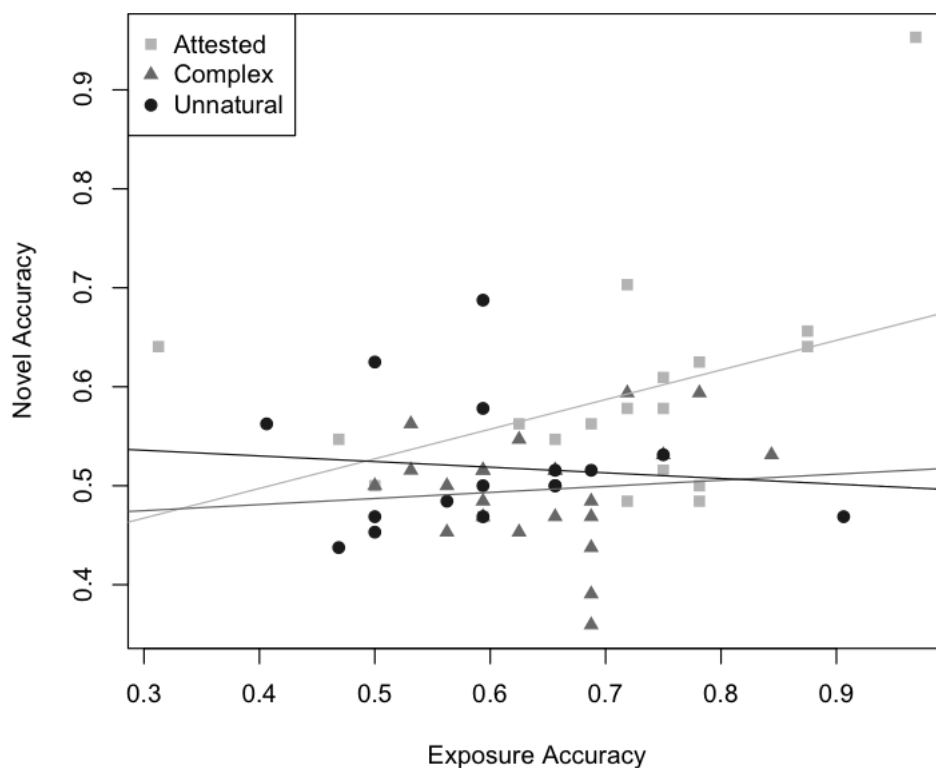


Figure 5.3: Effect of performance in Exposure on performance in Novel

in Novel (with the exception of a single outlier who performed almost perfectly in both phases). The COMPLEX group has a slightly positive correlation between Exposure and Novel, while the UNNATURAL group has a slightly negative correlation.

A Pearson’s product-moment correlation coefficient for the relationship between mean Exposure accuracy and mean Novel accuracy was calculated for each of the regression lines. The correlation for the ATTESTED pattern was numerically strong but did not reach significance ( $r = 0.416$ ,  $r^2 = 0.173$ ,  $p = 0.076$ ), likely due to a shortage of data points. In the two unattested patterns, the correlations are both weak and non-significant (COMPLEX:  $r = 0.092$ ,  $r^2 = 0.008$ ,  $p = 0.677$ ; UNNATURAL:  $r = -0.104$ ,  $r^2 = 0.010$ ,  $p = 0.700$ ).

### 5.1.3 Discussion

The results of this experiment show that the participants who were trained on the ATTESTED pattern performed more successfully at both of their tasks. In the Final Big

Test of the Exposure phase, in which participants chose which of two words conformed to their assigned pattern, the ATTESTED group performed significantly better than the UNNATURAL group. In the Novel phase, the ATTESTED group performed significantly above chance at applying their learned pattern to novel words. Neither of the COMPLEX and UNNATURAL groups performed above chance on this task.

Participants who were trained on the ATTESTED pattern showed some degree of pattern learning. In the same vein, it can be argued that the COMPLEX and UNNATURAL patterns were not learned at all. While these groups performed above chance on the words they were explicitly trained on, they could not reliably generalize their pattern to a set of new words. Essentially, they performed the Novel task no better than if they were guessing at random. It may be the case that the COMPLEX and UNNATURAL group had simply memorized some of the words heard in the Exposure phase but did not learn the pattern itself. The ATTESTED group may have memorized some of their Exposure words as well, but they were more successful at generalizing the overall pattern to words they had never heard before.

This experiment demonstrates an apparent naturalness effect in the domain of stress. The pattern that most closely resembles the canonical weight-to-stress pattern found in typology, in which stress is assigned to syllables that are perceptually prominent, was learned. The UNNATURAL pattern, which lacks phonetic correlates, was not learned. In addition, this experiment demonstrates an apparent simplicity effect in the domain of stress. The COMPLEX pattern is more formally complex than the ATTESTED pattern because it is predictable based on two features instead of one. While the simpler ATTESTED pattern was learned to some degree, the COMPLEX pattern was not learned.

From these results, we see that the naturalness of a pattern affects how well it can be learned. The Stress Learning Experiment does not, however, determine why the UNNATURAL pattern went unlearned. The Stress Perception Experiment was conducted to discover whether perception of the stimuli played any role in suppressing the learnability of the UNNATURAL pattern.

Simplicity also appears to be a factor, given that the poor performance on the COMPLEX pattern was not an issue of naturalness. Since the main focus of this dissertation is naturalness effects and not simplicity effects, I do not delve much further



into this investigation. The results of the Stress Perception Experiment (Chapter 5.2), however, do suggest that the performance difference cannot be explained by channel bias. This opens up the possibility of future, more in depth research into the particulars of how universal learning bias inhibits pattern learning.

## 5.2 Experiment 2: The Stress Perception Experiment

Although the Stress Learning Experiment showed that there were naturalness effects observed in the domain of stress, it said nothing about what the source of the naturalness effects could be. I entertain two possible scenarios for why the UNNATURAL group performed significantly worse than the ATTESTED group at their learning task: a foot-building scenario and a duration-based scenario.

The foot-building scenario assumes that the struggle with the UNNATURAL pattern is cognitive in nature. Participants attempt to complete their task by building feet, but struggle to do this for the UNNATURAL stimuli. A foot-based analysis of the ATTESTED stimuli is straightforward. Words containing all open syllables would receive an initial trochaic foot, resulting in default initial stress (7-a). Closed syllables disrupt this pattern by becoming the head of a foot. When the closed syllable is word-initial, it results in a left-aligned trochee, giving word types like (7-b) initial stress. In word types like (7-c), footing shifts rightward to stress the closed syllable at the expense of aligning the trochee with the left-edge of the word. Participants could have parsed the feet in (7-b) and (7-c) as either bimoraic or bisyllabic. The result is the same: a word containing a single binary foot with the stress attractor serving as its head.

- (7) Possible footing in the ATTESTED pattern
- a. Initial (default): (tú.pe).ta
  - b. Initial (by stress attractor): (pál).ko.de, (pál.ko).de
  - c. Peninitial: ba.(tíl).tu, ba.(tíl.tu)

Building feet in the UNNATURAL pattern is a more challenging endeavor. Words that receive default initial stress have two possible footings, as shown in (8-a). The left-edge trochee can either be quantity-insensitive, yielding a bisyllabic foot, or quantity-

sensitive, yielding a bimoraic (monosyllabic) foot. Both of these options result in initial stress. With words like [pá.kol.den] in (8-b), however, the parser is essentially forced to build a left-aligned, quantity-insensitive trochee. The duration of stressed open syllables is shorter than that of stressed closed syllables (such as in [tún.per.tal, which receives default stress), and so it is unlikely that participants would parse the open stress attractors as bimoraic. Taking this into consideration, a footing such as \*[pá).kol.den] is insufficient, but a quantity-insensitive footing such as [(pá.kol).den] is still viable. If we assume quantity-insensitivity, however, there is nothing to explain the stress shift that occurs in (8-c). The parser is required either to switch their footing type such that some words contain left-aligned iambs, [(bal.tí).tun], or to shift the alignment of the head foot on some words, [bal.(tí.tun)]. Either way, there is no rule that would support this stress shift, and so it would have to be lexically assigned and therefore more difficult to generalize over. The foot-building scenario therefore states that the UNNATURAL pattern is harder to learn than the ATTESTED pattern, explaining the difference in performance.

(8) Possible footing in the UNNATURAL pattern

- |                                   |                              |               |
|-----------------------------------|------------------------------|---------------|
| a. Initial (default):             | (tún.per).tal, (tún).per.tal |               |
| b. Initial (by stress attractor): | (pá.kol).den                 | *(pá).kol.den |
| c. Peninitial:                    | (bal.tí).tun, bal.(tí.tun)   | *bal.(tí).tun |

The duration-based scenario does not assume that either pattern is harder on the cognitive system, nor does it assume that participants rely on foot-building to complete their task. Instead, this scenario is perceptual in nature: the UNNATURAL pattern yielded lower performance because it was less clear to the listeners what the most perceptually prominent syllable was meant to be. In both patterns, stressed syllables were longer than unstressed syllables, and closed syllables were longer than open syllables (as described in Chapter 5.1). The stimuli were designed so that stressed open syllables were slightly longer than unstressed closed syllables (200 ms. vs. 166 ms.), a design which may have been crucial to performance in the UNNATURAL pattern.

Tables 5.8–5.10 exemplify how this plays out in the stimuli from the ATTESTED and UNNATURAL patterns. In the words that receive default stress, both languages have a duration difference of at least 100 ms between the stressed and the unstressed syllables.

This is shown in Table 5.8.

ATTESTED			UNNATURAL		
tú	pe	ta	tún	per	tal
CV	CV	CV	CVC	CVC	CVC
200 ms	100 ms	100 ms	300 ms	166 ms	166 ms
<i>Duration difference: 100 ms</i>			<i>Duration difference: 134 ms</i>		

Table 5.8: Syllable durations in default-stressed words

In the remainder of the words, the durational differences between stress attractors and non stress attractors increased in the ATTESTED pattern and decreased in the UNNATURAL pattern. The UNNATURAL stress attractor words all contained one stressed open syllable (the stress attractor) and at least one unstressed closed syllable. In these words, the duration difference between the syllables was quite small (34 ms). The ATTESTED stress attractor words, on the other hand, contained one stressed closed syllable (300 ms) and at least one unstressed open syllable (100 ms), resulting in a large durational difference between syllable types (200 ms). If participants used duration to determine stress placement, the 34 ms increase may not have been enough for them to determine definitively which syllable bore stress.

ATTESTED			UNNATURAL		
<b>pál</b>	ko	de	<b>pá</b>	kol	den
CVC	CV	CV	CV	CVC	CVC
300 ms	100 ms	100 ms	200 ms	166 ms	166 ms
<i>Duration difference: 200 ms</i>			<i>Duration difference: 34 ms</i>		

Table 5.9: Syllable durations in words with initial stress by stress attractor

The Stress Perception Experiment aims to differentiate between these two scenarios by determining whether participants were able to accurately perceive stress placement in the words designed for the Stress Learning Experiment. This design only tests for presence of channel bias and cannot explicitly rule out the foot-building scenario. However, presence of perceptual difficulty with the UNNATURAL stimuli suggests that, at

ATTESTED			UNNATURAL		
ba	<b>tíl</b>	tu	bal	<b>tí</b>	tun
CV	CVC	CV	CVC	CV	CVC
100 ms	300 ms	100 ms	166 ms	200 ms	166 ms
<i>Duration difference: 200 ms</i>			<i>Duration difference: 34 ms</i>		

Table 5.10: Syllable durations in words with peninitial stress

minimum, channel bias is active in shaping naturalness within this domain. If universal learning bias is also active, it appears from this experiment to be doing no additional work that channel bias does not already accomplish. A universal learning bias account is therefore stipulative, while a channel bias account is explanatorily satisfying.

## 5.2.1 Methodology

### 5.2.1.1 Participants

20 undergraduate students at the University of California, Santa Cruz, participated in this experiment. The ages of the participants ranged between 18 and 22 years (mean = 19.9 years). All of the participants were coded as native speakers of English using the same intake form described in Chapter 5.1. None of the participants from the Stress Learning Experiment participated in the Stress Perception Experiment; therefore, no participants in this experiment had encountered the data before. All participants received either course credit or \$10 as incentive.

### 5.2.1.2 Stimuli

The stimuli used in this experiment are exactly the same as the stimuli in the Stress Learning Experiment. No new stimuli were created and none were deleted. The main difference in stimulus presentation between the two experiments is that, in the Stress Learning Experiment, participants heard only the stimuli belonging to the pattern they were assigned to. In the Stress Perception Experiment, all participants heard all stimuli designed for the ATTESTED, COMPLEX, and UNNATURAL patterns. Each pattern had 96 words created for it: the 32 Exposure words, the 32 Novel words that belonged to

the pattern, and the 32 Novel words that did not belong. This should have resulted in every participant hearing 288 words.

However, occasionally, the same stimulus appeared in two of the patterns (e.g., [páɪ.ko.de] was an acceptable word of both the ATTESTED pattern and the COMPLEX pattern). These words were included only once in the Stress Perception Experiment, so that participants heard each stimulus only once. Repeated stimuli were coded in the analysis of belonging to more than one pattern (e.g., [páɪ.ko.de] contributes to the results of both the ATTESTED and COMPLEX patterns). After accounting for the 26 words that appeared in more than one pattern, each participant heard a total of 262 words.

### 5.2.1.3 Procedure

The Stress Perception Experiment consisted of two phases. The first phase was identical to the warm-up exercise presented in the Stress Learning Experiment. Again, this exercise was intended to exclude participants who performed below 70% accuracy on all the trials. Because this restriction was eliminated from the analysis of the Learning Experiment, it has also been eliminated from the analysis of the Perception Experiment.

In the second phase, participants listened to each of the 262 words created for the Stress Learning Experiment and indicated which syllable the word's stress fell on. There were six practice trials, followed immediately by the 262 test trials. Every trial began with a cross-hatch symbol appearing in the center of the screen for 1500 ms. Following this, a single stimulus played over the headphones and was spelled out in the center of screen, so that participants could hear the word and see it written simultaneously. Below the written word, the individual syllables of that word were also printed. The numbers 1, 2, and 3 appeared beneath each syllable, corresponding to the 1, 2, and 3 buttons on the button box. The prompt, "Which syllable does this word's stress fall on?", was written above the word. Participants were instructed to press the "1" button on the button box if the word had stress on the first syllable, the "2" button if the word had stress on the second syllable, and the "3" button if the word had stress on the third syllable. When the stimulus in question was bisyllabic, participants saw no corresponding syllable above the number 3 on the screen. After every 16 trials,

participants were allowed a self-timed break. Figure 5.4 provides two examples of the screen that participants saw in each trial.

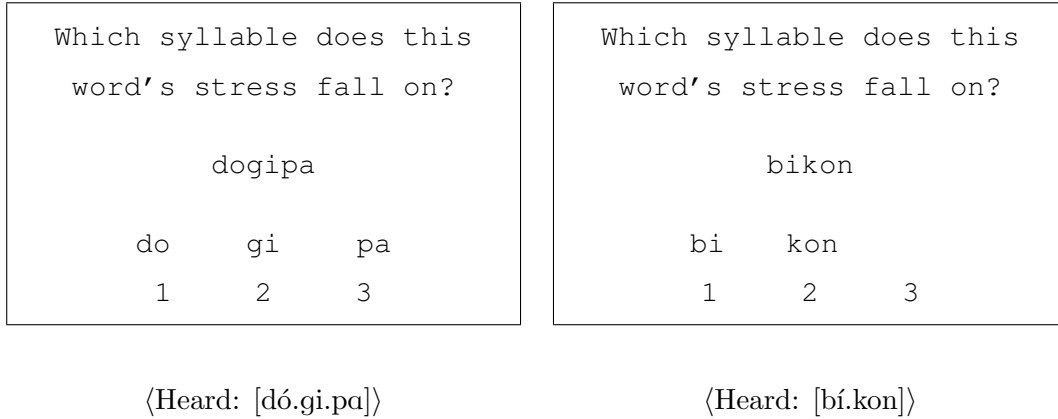


Figure 5.4: Example of two trials from the Stress Perception Experiment

#### 5.2.1.4 Hypothesis

The purpose of this experiment is to discover whether channel bias can explain the naturalness effects observed in the previous experiment. The Stress Learning Experiment found that the UNNATURAL stress pattern was learned less successfully than the ATTESTED pattern. I have proposed a foot-based analysis at the beginning of Chapter 5.2 as one possible type of universal learning bias against the UNNATURAL pattern. Under this analysis, perception plays little to no role: even if all participants perceived stress placement in their pattern extremely well, the UNNATURAL group is still predicted to perform worse on learning tasks than the ATTESTED group. In the Stress Perception Experiment, then, if these two groups perform equally well at perceiving their stimuli, this may indicate that a universal learning bias is at work in suppressing the learnability of the UNNATURAL pattern, found in the Stress Learning Experiment.

If, however, channel bias is active in some way in shaping the observed naturalness effects, we would expect that stress placement in the UNNATURAL stimuli should be misperceived more frequently than in the ATTESTED stimuli. This, in turn, would explain the poor performance of the UNNATURAL group in the Stress Learning Experiment: if stress placement in the UNNATURAL stimuli cannot be accurately perceived, then the

pattern of exceptional stress will not be accurately learned. The hypothesis that the Stress Perception Experiment tests is stated in (9).

- (9) The naturalness effects observed in the Stress Learning Experiment are driven by misperception of the UNNATURAL stimuli. It is predicted that stress placement in the stimuli that conform to the UNNATURAL pattern will be misperceived (and therefore misreported) on more trials than stimuli that conform to the ATTESTED pattern.

### 5.2.2 Results

This section presents results in which accuracy corresponds to proportion of correct trials. A portion of the trials have been removed from analysis. As a reminder, participants in the Stress Perception Experiment listened to every single stimulus created for the Stress Learning Experiment (both the Exposure stimuli and the Novel stimuli); however, half of the Novel stimuli were designed to *not* conform to the pattern. The analysis below includes only the pattern-conforming trials, as a way of determining how well each of the stimuli of the three patterns could be accurately perceived.

Figure 5.5 presents the proportion of trials in which participants correctly identified word-level stress in each pattern. Numerically, it appears that participants were more accurate on ATTESTED-conforming trials (mean = 0.870) than on UNNATURAL-conforming trials (mean = 0.796). Participants also responded more accurately to ATTESTED-conforming trials than the COMPLEX-conforming trials, although by a smaller margin (mean = 0.850).

These results were analyzed using a series of t-tests. Each subject's individual accuracy score on each of the three PATTERN types were calculated. This yielded three vectors of length 20, one vector for each pattern, corresponding to the 20 subjects of this experiment. A Welch's two sample t-test then compared the means of the ATTESTED scores to the means of the COMPLEX scores. The difference in these means did not reach significance ( $t = 0.39$ ,  $df = 37.608$ ,  $p = 0.700$ ). The difference between the ATTESTED and UNNATURAL identification scores also did not reach significance ( $t = 1.22$ ,  $df = 33.43$ ,  $p = 0.231$ ).

The t-tests show that the overall averages of correct perception trials did not sig-

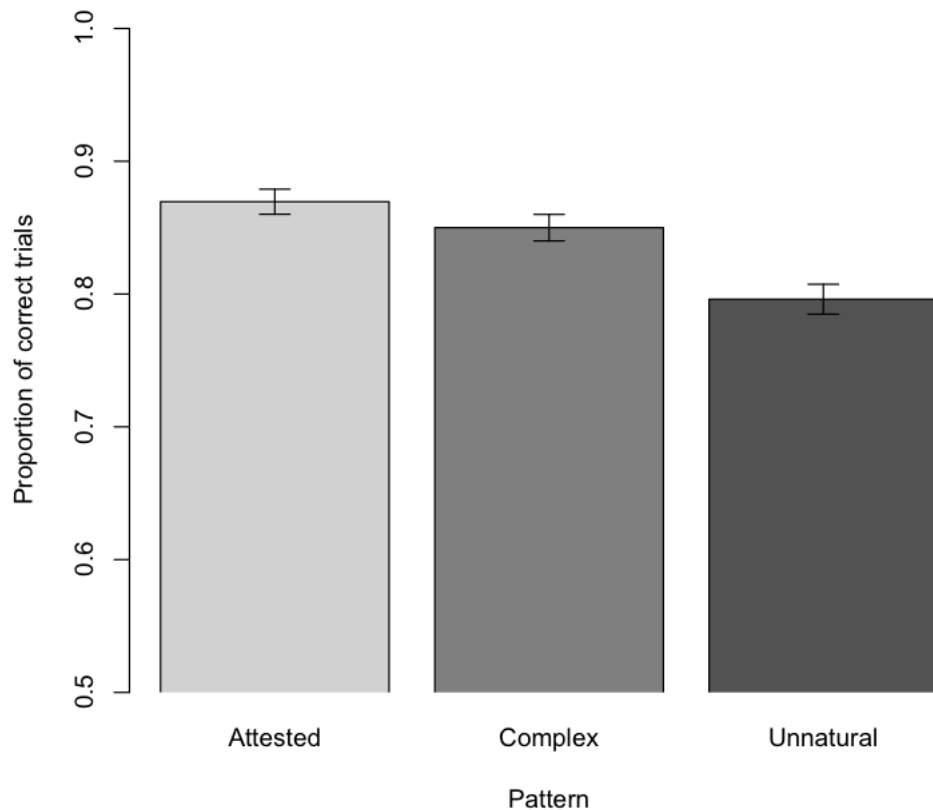


Figure 5.5: Proportion of correct trials in each pattern

nificantly differ. This experiment is less concerned with overall accuracy, however, and more concerned with accuracy at perceiving exceptional stress. It is possible that perception of initial stress was high across all three conditions, which would draw the averages up and mask any differences in perceiving exceptional or penultimate stress.

Two additional analyses were performed in an attempt to discern whether different word types were perceived differently. In Figure 5.6, the accuracy results are grouped by pattern and stress placement. Stress in each word fell either on the initial syllable – either by default stress assignment or by stress attractor – or on the peninitial syllable. Within the ATTESTED words, participants appear to be slightly more accurate in identifying initial stress than peninitial stress. Additionally, they appear to be slightly more accurate in identifying initial stress in the ATTESTED and words than in



the UNNATURAL words. The most noticeable difference, however, lies within the UNNATURAL-conforming words: participants responded much less accurately on trials in which the UNNATURAL word bore peninitial stress than on trials containing initially stressed UNNATURAL words. The COMPLEX pattern appears very similar to the ATTESTED pattern, with a small numerical decrease in each of the groups.

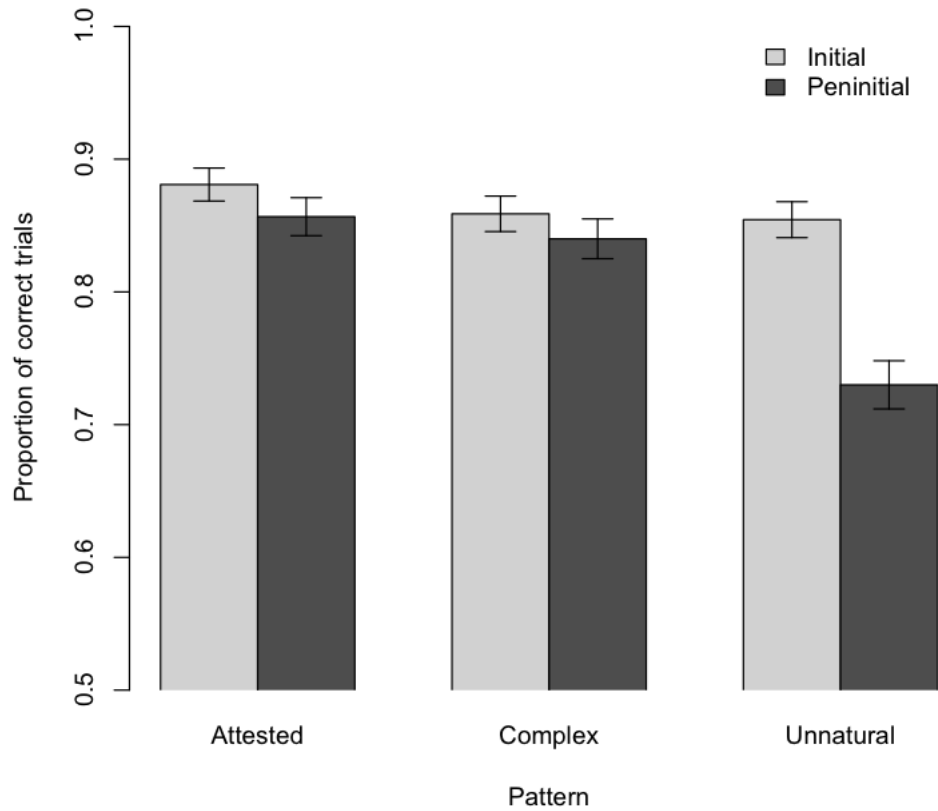


Figure 5.6: Accuracy in each pattern on initially stressed vs. peninitially stressed words

These differences were analyzed with another series of t-tests. To do this, I followed nearly the same procedure as with the overall accuracy scores, with one intervening step. Since one goal of this analysis is to determine whether perception of initial stress placement differed from perception of peninitial stress placement, this doubled the number of subject averages to calculate. For each subject, six different average scores were computed: their average on trials with initial stress and with peninitial stress in each of the three patterns. This is summarized in Table 5.11, using Subject 201 as an example.

Each of these averages were entered into six vectors, each of length 20, which corresponded to the 20 experimental subjects. From this, each subject's difference score was calculated: their average initial score minus their average peninitial score (for Subject 201, this would be -0.088 for ATTESTED, 0.067 for COMPLEX, and 0.137 for UNNATURAL). Compiling each of these difference scores yielded three vectors of length 20, corresponding to the 20 difference scores in each pattern.

Subject	Pattern	Placement	Accuracy
201	ATTESTED	Initial	0.912
201	COMPLEX	Initial	1.000
201	UNNATURAL	Initial	0.971
201	ATTESTED	Peninitial	1.000
201	COMPLEX	Peninitial	0.933
201	UNNATURAL	Peninitial	0.833
202	ATTESTED	Initial	1.000
...	...	...	...

Table 5.11: Sample table for calculating difference scores

A Welch's two sample t-test was calculated to determine whether distribution of difference scores in ATTESTED trials differed from the distribution of difference scores in COMPLEX trials. This difference did not reach significance ( $t = 0.13$ ,  $df = 38$ ,  $p = 0.894$ ). The difference between the distributions of ATTESTED and UNNATURAL trials, however, did reach significance ( $t = -2.44$ ,  $df = 37.935$ ,  $p < 0.05$ ).

A set of paired t-tests were then calculated to investigate the nature of this difference. These t-tests compared, for each pattern, the distribution of average scores on initially-stressed trials to that of peninitially-stressed trials. In the ATTESTED pattern, no significance was found ( $t = 0.85$ ,  $df = 19$ ,  $p = 0.404$ ). From this, we can conclude that participants perceived initial and peninitial stress equally well in the ATTESTED words. In the UNNATURAL pattern, the difference did reach significance ( $t = 4.20$ ,  $df = 19$ ,  $p \approx 0$ ). This confirms that participants perceived initial stress significantly better than peninitial stress in the UNNATURAL words.

In the pattern-conforming words in the Learning Experiment, the only way stress

could be peninitial was if the peninitial syllable was a stress attractor. The stress attractor itself could be either initial or peninitial. A final analysis determined whether the decrease in accuracy on the UNNATURAL trials could also be explained by stress falling on a stress attractor (independent of syllable position). Figure 5.7 shows a similar picture to Figure 5.6, in which performance on the UNNATURAL words decreases by a large margin in the trials where stress was assigned by stress attractor.

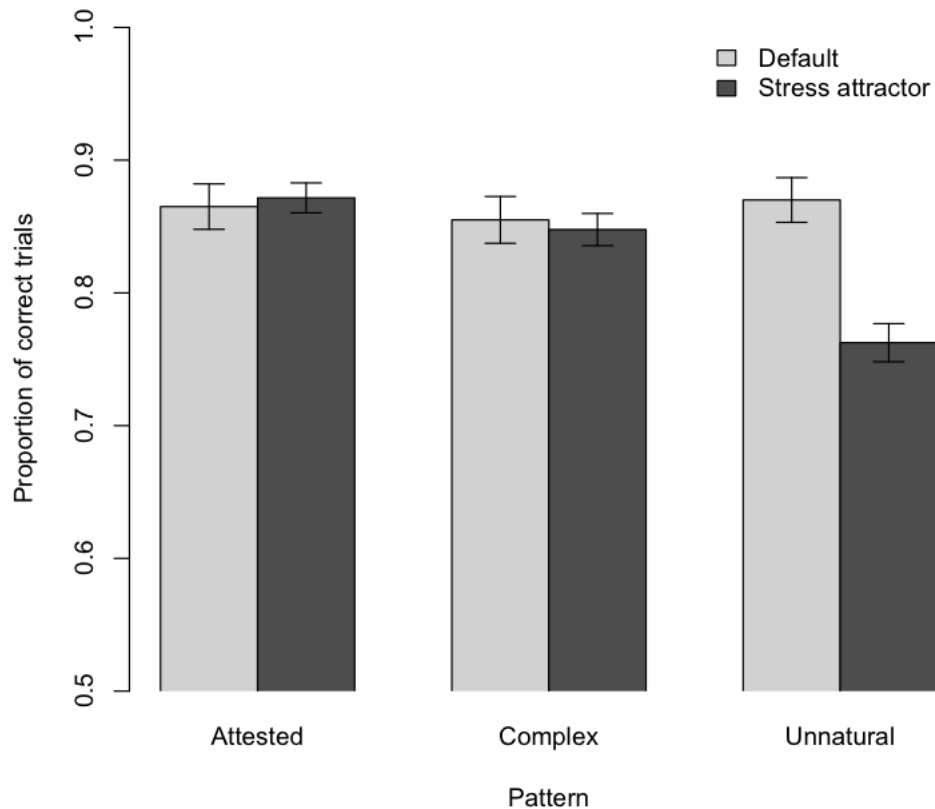


Figure 5.7: Accuracy in each pattern on default-stressed vs. stress attractor words

The procedure for analyzing perception of stress placement was repeated to analyze exceptional stress perception. The difference scores were calculated by subtracting each subject’s stress attractor average from their default average. The distributions of difference scores in ATTESTED vs. COMPLEX did not reach significance ( $t = -0.32$ ,  $df = 37.407$ ,  $p = 0.753$ ), but the distributions of difference scores in ATTESTED vs. UNNATURAL did ( $t = -2.87$ ,  $df = 37.85$ ,  $p < 0.01$ ). Again, this difference was investigated using

a set of paired t-tests, each of which compared the distribution of default averages to stress attractor averages in each pattern. In ATTESTED trials, no difference was found ( $t = -0.23$ ,  $df = 19$ ,  $p = 0.823$ ): participants did not perceive stress any differently whether it was assigned by default or by stress attractor. A difference was found, however, in the UNNATURAL trials ( $t = 3.95$ ,  $df = 19$ ,  $p \approx 0$ ). From this, we may conclude that participants perceived stress in the UNNATURAL pattern significantly better when it was assigned by default than by stress attractor.

### 5.2.3 Discussion

As predicted, stress placement in the UNNATURAL stimuli was identified less accurately than stress placement in the ATTESTED stimuli. This decrease in accuracy seems to come from the trials in which stress fell on a stress attracting syllable (e.g., [pá.kol.den], [bal.tí.tun]). Figure 5.7 showed that participants responded just as well to UNNATURAL words with default stress (e.g., [tún.per.tal]) as they did to the ATTESTED words but performed significantly worse on trials with exceptional stress.

A similar story is told in Figure 5.6, which compared performance on initially stressed words and peninitially stressed words. The initially stressed group includes all words that received default stress, as well as all words that received initial stress by stress attractor (e.g., [tún.per.tal], [pá.kol.den]). Grouping the initial-by-stress attractor trials in with the default-stress trials does not drag the accuracy down: these trials were still responded to with the same degree of accuracy as all the ATTESTED trials. Stress in the UNNATURAL:PENINITIAL trials, in which stress was assigned only by stress attractor, was identified significantly less accurately than stress in the UNNATURAL:INITIAL trials.

A final crucial observation is that this decrease in accuracy in either the Peninitial trials or the stress attractor trials was only observed in the UNNATURAL-conforming stimuli. In the the ATTESTED trials, participants were able to hear and report stress just as accurately on default-stress words (e.g., [tú.pe.ta]), on initial stress attractor words (e.g., [pál.ko.de]), and on peninitial stress attractor words (e.g., [ba.tíl.tu]). We observe that the pattern that was perceived less accurately also had lower performance in the Stress Learning Experiment. This experiment therefore shows evidence for channel

bias in the lab as a source of the naturalness effects. The UNNATURAL pattern was learned less successfully because it was perceived less successfully. Participants might have learned a generic pattern of initial-stress, but they were less able to learn which types of syllables could attract stress away from the initial because they were less able to perceive that these peninitial syllables were stressed at all.

These results suggest that the problems with the UNNATURAL pattern in the Stress Learning Experiment are perceptual in nature. It could be the case that the participants relied heavily on duration as a cue to stress, and the UNNATURAL stress attractor words did not have a large enough duration difference between stressed and unstressed syllables for them to make a reliable judgment of where the stress fell. It could also be the case that participants relied on information from both duration and pitch, since both were encoded into the stimuli as phonetic correlates of stress. All stressed syllables in all the stimuli had the highest pitch measurement within the word, and all contained a rise-fall pattern, measuring 120 Hz at the beginning and end of the rime and 140 Hz at 50% of the rime. In stressed closed syllables, which were always longer than stressed open syllables, the pitch rise towards that 50% mark was more gradual, taking 150 ms for pitch to rise from 120 Hz to 140 Hz. In the stressed open syllables, this change took place over 100 ms. It is possible that participants used information from both duration and pitch, such that it was easier to hear the pitch rise on the stressed closed syllables, when it was more gradual. Either description is concurrent with the participants' significant difficulty of identifying stress within the UNNATURAL stress attractor words.

Although my findings do not have clear consequences for the findings in Carpenter (2010), they may provide one explanation for why both experiments found a significant effect for naturalness. My experiment and Carpenter's studied two different types of weight sensitivity: rime length in mine and vowel duration in Carpenter's. Observing a perceptibility difference in one experiment does not necessitate that such a difference existed in the other. It is not unreasonable to think that Carpenter's stimuli may not have been equally perceptible, given that low vowels supposedly have greater perceptual energy than high vowels due to their increased energy (Lehiste 1970; Gordon 2005; Carpenter 2010).

Proponents of the universal learning bias account might argue that the two hy-

potheses are not mutually exclusive, and that both universal learning bias and channel bias are active in shaping naturalness. At this point, I am not ruling this out as a possibility. The question is whether or not we *need* to look to universal learning bias for an explanation. The results of the Stress Perception Experiment suggest quite a drastic perception difference between the ATTESTED and UNNATURAL patterns and provide a convincing argument that channel bias is at work. Both hypotheses would have the same effect on the Learning data (the suppression of the UNNATURAL pattern), but since we know channel bias to be active, this may be all that is needed to explain this phenomenon. This is parallel to the discussion of naturalness effects in typology: both accounts explain the same tendencies, but the channel bias account does so in a way that is less stipulative. By Occam’s razor – in typology and, perhaps, in this experiment – the channel bias account wins out (Hale & Reiss 2000).

#### 5.2.4 Universal learning bias shapes simplicity effects

As a final observation, the stimuli designed for the COMPLEX pattern were perceived just as well as the stimuli designed for the ATTESTED pattern. No differences were found in any of the t-tests: participants were equally able to identify the stressed syllable in both patterns whether it was initial or peninitial, or whether stress was assigned by default or stress attractor.

If participants were using duration as a cue to stress, these findings are not surprising. Because exceptional stress always fell on a closed syllable (C<sub>0</sub>C), the duration of that syllable was always 300 ms. The remaining, unstressed syllables could be either closed, measuring 166 ms, or open, measuring 100 ms. This resulted in a durational difference of at least 134 ms (similar to the default stressed words in the UNNATURAL pattern) and at most 200 ms (similar to the exceptionally stressed words in the ATTESTED pattern). These word types are summarized in Table 5.12. Participants showed no significant trouble identifying stress in the words of 134 ms duration difference between the syllables, and so they should not have trouble identifying stress in words with a greater durational difference.

The Stress Perception Experiment showed no evidence that the COMPLEX stimuli were misperceived any more than the ATTESTED stimuli were. This was intended in the

COMPLEX: Minimal difference			COMPLEX: Maximal difference		
pul	<b>báR</b>	dun	ba	<b>tél</b>	tu
CVC	CVC	CVC	CV	CVC	CV
166 ms	300 ms	166 ms	100 ms	300 ms	100 ms
<i>Duration difference: 134 ms</i>			<i>Duration difference: 200 ms</i>		

Table 5.12: Syllable durations in COMPLEX words with a stress attractor

design: durational differences were encoded into rime length, but not into vowel height. Even so, the Stress Learning Experiment found that the COMPLEX language was learned less successfully than the ATTESTED language. I contend even further that this language was not learned at all, given that participants could not apply the pattern to novel forms. Without a perceptual bias, we are left with the alternative that there is a cognitive bias against the COMPLEX pattern. This supports the predictions in Gordon (2005) based upon the simplicity bias literature (Shepard et al. 1961; Pertsova 2012; etc.). This may be an artifact of the featural makeup of the COMPLEX pattern. The stress attracting syllable was defined by two features (rime length and vowel height) instead of one (rime length). Increasing the number of features of the pattern is, apparently, enough to make this particular pattern too difficult for participants to learn.

#### 5.2.4.1 Potential limitations of the Complex pattern

This dissertation defines formal simplicity in terms of the number of features by which a pattern is predictable. It may be the case, however, that defining simplicity is not nearly as simple as it seems. This section addresses some of the potential limitations in this experiment surrounding the COMPLEX pattern.

The COMPLEX pattern in this dissertation’s experiment is predictable based on both syllable closedness and vowel height, and participants are expected to attend equally to both features. Native language effects from English may be a potential confound surrounding this task. If English stress does not depend on vowel height in any significant way, or if syllable closedness matters more than vowel height, then learning the COMPLEX stress pattern would involve attending to a feature that participants were

not used to (Kie Zuraw, p.c.). The poor performance on the COMPLEX pattern (as compared to the ATTESTED pattern) may then be due, in part, to participants' difficulty with learning a pattern different from one that exists in their ambient language. One way to determine the role of native language effects in this experiment would be to run it again, using a subject pool of participants who speak a fixed-stress or stressless language. Carpenter (2010) found that speakers of both English and Québécois French performed better on a natural pattern in which low vowels were stress attractors than on an equivalent unnatural pattern in which high vowels were stress attractors. The participants in her study were able to learn the pattern even though it was not part of their native language, and it is reasonable to assume that my participants have the same capability. Since my COMPLEX pattern was, in fact, not learned, one possible conclusion is that this pattern was suppressed by the forces of simplicity and not entirely by native language effects.

A related concern is that learners of the COMPLEX pattern needed to attend not only to multiple features, but also to multiple phonological tiers: one of the predictable features referenced the vowel, while the other referenced the coda consonant. This could have the effect of making the COMPLEX pattern more complex than intended. Future research that intends to replicate this study should work to avoid this problem. One potential solution is to design a COMPLEX pattern where both features refer to the vowel. A pattern in which stress attractors must contain vowels that are both low and long would yield exceptional stress on words like [pi.tá:ki], but default stress on all other words. Crucially, short low vowels [pí.ta.ki] and long high vowels [pí.ti:ki] would not have the ability to attract stress. This pattern would maintain the complexity of being predictable based on two features, but listeners would need to attend to only one tier. In addition, the high intensity and duration (and therefore high perceptual energy) of the stress attractor makes this pattern phonetically natural. I am unaware if any such pattern is attested in typology.

Finally, the two features of the COMPLEX pattern were chosen because they raise the perceptual energy of a rime. Part of Gordon's (2005) claim is that a weight-sensitive stress pattern may arise in a language if the perceptual energy of the stress attractors is higher than the perceptual energy of all types of syllables that are not stress attractors.



Participants in the COMPLEX group needed to learn that the optimal division of perceptual energy was “CaC > {CiC, Ca, Ci}” and not “{CaC, CiC} > {Ca, Ci}”. While closed syllables with a low vowel should have higher perceptual energy than closed syllables with a high vowel, it is unclear whether the difference between them is enough to warrant such a weight split. Thus, a potential confound might arise if the division participants were taught is not the optimal division based on perceptual energy (Donca Steriade, p.c.). This leads to a much larger issue that, to my knowledge, has not yet been solved (but has been addressed in Gordon 2002): how large does the perceptual energy division have to be for weight-sensitivity to arise?

#### 5.2.4.2 Is this phenomenon specific to language?

If the increased complexity of the COMPLEX pattern can be explained by feature counting alone, an additional question arises about the degree to which this is a property of *language*. While this phenomenon seems to arise in stress pattern learning, indeed an issue of language, it arises in non-linguistic pattern learning as well (Shepard et al. 1961; Moreton & Pertsova 2014; Moreton et al. 2015). The complexity that arises with increased features of a pattern – and, presumably, the relationships between those features – is not a language-autonomous phenomenon, but rather a larger cognitive difficulty that has applications in language.

Increasing the complexity of a pattern, however, does not automatically make it unlearnable. Complex patterns may instead be “less learnable”: still learned to some degree, but to a lesser degree than their simpler counterparts. Each simplicity effect warrants the question: how complex does a pattern have to be for it to be unlearnable? In the case of the Stress Learning Experiment, increasing the complexity by only one degree (feature counting from one to two features) seemed to be enough. These results mirror the typological survey from Gordon (2005), wherein bifatural exceptional stress patterns do not occur in natural language, despite their increased perceptual energy. Because of this floor effect, I also expect that increased complexity due to the relationship between features could not be observed in exceptional stress patterns. That is, an OR-type stress pattern such as “Stress closed syllables *or* non-high vowels” would be learned to the same degree as an equivalent AND-type pattern, despite the former being

a supposedly more complex relationship. Both patterns should be learned (or, rather, not learned) at an at-chance rate, and so any simplicity effects between the two would not be visible.

While featural complexity is not specific to language, perhaps the cutoff point between learnable vs. unlearnable may be different in language-related patterns than in non-linguistic patterns. In the same vein, it could be that this quick cutoff point is a property of phonological patterns, or prosodic patterns, or something of this type. Future research should delve into how complex patterns need to be to become unlearnable in various domains.

## Chapter 6

### Experiment: Final devoicing

The combined results of the experiments presented so far show evidence of channel bias underlying apparent naturalness effects observed in the lab. At least in the domain of stress, when a natural pattern is learned more successfully than an unnatural pattern, it can be explained by participants' decreased ability to perceive the unnatural stimuli.

The other side of the channel bias account is that, if the stimuli are made abundantly easy to perceive, there should be no observed naturalness effects in the lab. The Final Devoicing Experiment tests this. A natural pattern of final devoicing (hence DEVOICING) and an unnatural pattern of final voicing (hence VOICING) were recorded in two different ways: one in which the stimuli were spoken exceptionally clearly, and another in which the stimuli were spoken in a method resembling casual speech. This dissertation argues that the advantage of natural patterns over unnatural patterns depends entirely on the natural pattern's ability to facilitate perception or production, and not on an innate, cognitive bias towards phonetic naturalness. If that perceptual advantage is not recreated in the lab, the performance differences between the two patterns may shrink or disappear.

This experiment tests final devoicing as opposed to stress because the phonetic correlates of final devoicing, discussed in Chapter 2.1.2, are much easier to define and control. The stimuli were recorded so that voiced segments in word-final position would be fully or partially devoiced when the speaker produced them in a more casual register. In the more careful register, final voiced segments were either fully voiced, or a large proportion of the obstruent was voiced. In the more casual register, the proportion of

voicing during the obstruent was considerably shorter. In addition, the duration of the vowel preceding the final obstruent was longer in the careful register and shorter in the casual register. Participants learning the unnatural pattern in the casual register should therefore misperceive some, if not all, voiced segments as voiceless.

The goal of this experiment is to distinguish between channel bias and universal learning bias as potential sources of observed naturalness effects. To do this, we must first discuss what each of the two accounts predict. If naturalness effects are driven by universal learning bias, then the two VOICING groups should perform worse than the two DEVOICING groups in both registers. Figure 6.1 provides an extreme example of hypothetical results for this scenario, in which the two register groups perceive their pattern equally well (indicated by the null effect of voice clarity), but the VOICING pattern is learned significantly worse than the DEVOICING pattern.

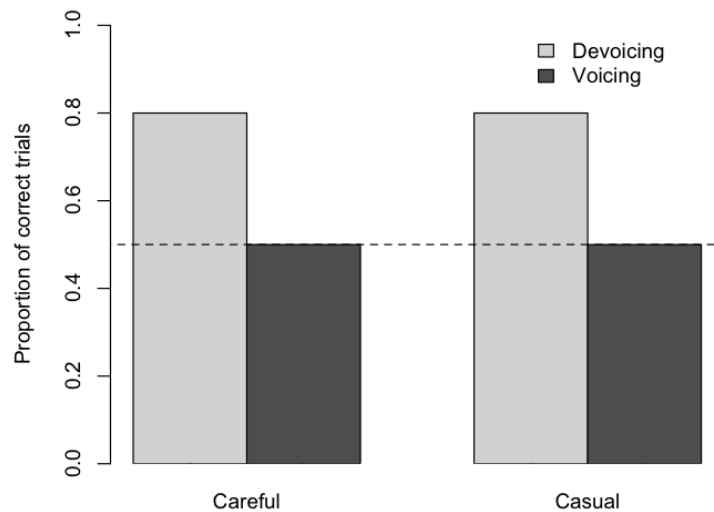


Figure 6.1: Hypothetical results: Only universal learning bias

If, on the other hand, channel bias is the sole explanation for naturalness effects, we would expect to see a difference between the DEVOICING and VOICING patterns only in the CASUAL register. When the stimuli are less clear, participants in the VOICING pattern should hear some of the final voiced obstruents as voiceless, due to partial or full devoicing produced by the speaker. As a result, they should perform worse on a task that requires them to accept only final voiced obstruents and reject final voiceless

obstruents, because they would have perceived some final voiceless obstruents during the Exposure phase. The DEVOICING group should not face this issue because nothing in the way these stimuli were recorded should lead participants to hear these stimuli as fully or partially voiced. In the CAREFUL register, on the other hand, final voicing is made abundantly clear. Participants should perceive only final voiced obstruents in their Exposure words and accept only final voiced obstruents in the Novel phase. These hypothetical results are provided in Figure 6.2.

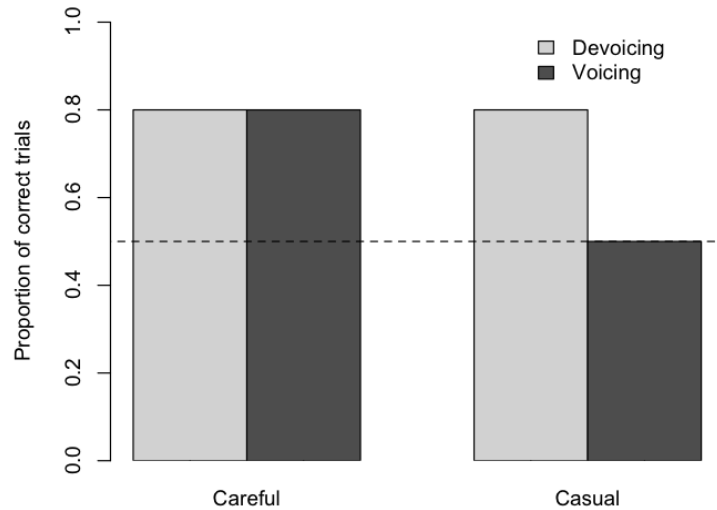


Figure 6.2: Hypothetical results: Only channel bias

The condition of interest, then, is the performance of the group trained on the VOICING pattern in the CAREFUL stimuli. Both hypothetical scenarios predict that the VOICING:CASUAL group should perform poorly, but for different reasons. The channel bias account predicts that learnability differences should only occur in places where perceptibility differences also occur. Since the stimuli of the two patterns in CAREFUL speech are expected to be equally perceivable, they are expected to be equally learnable as well. The universal learning bias account predicts instead that the VOICING pattern should be learned less successfully, regardless of voice clarity. Most notably, if the source of naturalness effects is indeed cognitive, the VOICING pattern should be underlearned even in CAREFUL speech.

The experiment was a between-subjects 2x2 factorial design. The two levels of the

PATTERN condition referred to the type of pattern that the participants were trained on. Half of the participants were trained on the DEVOICING pattern, and the other half were trained on the VOICING pattern during the Exposure phase. The CLARITY condition referred to the register in which the participants learned their pattern. Half of the participants were trained on their assigned pattern in careful speech, and the other half were trained on their assigned pattern in casual speech. The speaker recorded each stimulus in both registers. Thus, participants were assigned to one of four groups, each differing in the pattern heard during the Exposure phase and the register that pattern was heard in. These groups are summarized below in (1).

- (1) a. The DEVOICING pattern, heard in CAREFUL speech.
- b. The DEVOICING pattern, heard in CASUAL speech.
- c. The VOICING pattern, heard in CAREFUL speech.
- d. The VOICING pattern, heard in CASUAL speech.

## 6.1 Methodology

### 6.1.1 Participants

57 undergraduate students at the University of California, Santa Cruz, participated in this experiment. The ages of the participants ranged from 18 to 46 years old (mean = 21.09 years). Using the same intake form described in Chapter 5.1.1, all of the participants were coded as native speakers of English, and 11 were coded as native speakers of both English and Spanish. Participants received course credit as compensation for their help.

One participant was excluded from the final analysis due to a self-reported hearing disorder. Taking this into account, there were 14 participants in each of the four conditions.

## 6.1.2 Stimuli

### 6.1.2.1 Design

The stimuli in this experiment consisted entirely of nonce words. All of the words were monosyllabic and could be either closed (CVC) or open (CV). The consonant inventory contained 16 consonants: five voiceless obstruents [p f s t tʃ], five voiced obstruents [b v z d dʒ] and four sonorant consonants [m n l ɹ]. 6 vowels comprised the vowel inventory: [i eɪ ʌ u oʊ ə]. In all four conditions, the stops [p b t d] and the labiodental fricatives [f v] were restricted to onset position and never appeared word-finally. The sonorant consonants could be found in both word-initial (onset) and word-final (coda) position.

The two levels of the PATTERN condition differed as to what types of obstruents could appear word-finally. The DEVOICING pattern contained only voiceless obstruents [s tʃ] word-finally, while the VOICING pattern contained only voiced obstruents [z dʒ] word-finally. All four stridents also appeared word-initially in both patterns. Thus, in the Exposure phase, participants learned that both voiced and voiceless obstruents belonged to the pattern, but either only voiceless or only voiced obstruents could appear word-finally. Participants trained on the DEVOICING pattern heard words like [pas] and [dis], but not [paz] and [diz]. Participants trained on the VOICING pattern heard words like [paz] and [diz], but not [pas] and [dis]. The set of final consonants was limited to stridents to mirror the methodology in Myers & Padgett (2014), and because I had intended to measure sustained voicing during the closure, which would have proved more complicated with stops.

Table 6.1 presents the 36 words designed for the Exposure phase of the DEVOICING pattern. The 12 words in the “final voiceless obstruent” category represent the test words of interest; the remaining 24 are fillers, half of which are closed with a sonorant consonant (e.g., [nim]) and half of which are open (e.g., [vu]). The 12 test words were balanced across types of word-initial consonants. 4 of the test words contained a word-initial voiceless obstruent (e.g., [pas]), four contained a word-initial voiced obstruent (e.g., [dis]), and four contained a word-initial sonorant consonant (e.g., [mous]). Each of the six vowels in the inventory was represented twice within the test words, never preceded the same strident twice, and never followed the same initial consonant twice (e.g., [sutʃ] and [ɹus]).

Exposure words		Fillers			
Final voiceless obstruent		Final sonorant		Final vowel	
pas	‘poss’	tan	‘tahn’	sa	‘sah’
tertʃ	‘taych’	fʌl	‘fuhl’	tʃi	‘chee’
sutʃ	‘sooch’	boʊm	‘bome’	doʊ	‘dowe’
tʃʌs	‘chuss’	zɑɪ	‘zar’	zu	‘zoo’
dis	‘deese’	dʒoʊn	‘jown’	vu	‘voo’
vʌtʃ	‘votch’	dʌl	‘dule’	zeɪ	‘zay’
zʌtʃ	‘zutch’	voʊl	‘vole’	dʒoʊ	‘jowe’
dʒeɪs	‘jayce’	vʊm	‘voom’	dɑ	‘dah’
mouʃ	‘mowse’	zɑn	‘zahn’	boʊ	‘bowe’
ni:tʃ	‘neech’	bem	‘baim’	dʒeɪ	‘jey’
rus	‘roose’	nɪm	‘neem’	moʊ	‘mowe’
loʊtʃ	‘lowch’	lɑɪ	‘lar’	ɪ	‘ree’

Table 6.1: DEVOICING stimuli presented in the Exposure phase

The stimulus design controlled for a potential confound that could arise if participants had an overall preference for either voiced or voiceless obstruents, regardless of position (Myers & Padgett 2014). The number of voiced and voiceless obstruents that appeared across all 36 Exposure words was equal: all participants, regardless of their PATTERN assignment, heard 20 instances of voiceless obstruents and 20 instances of voiced obstruents.

To make this possible, the DEVOICING pattern contained a greater number of word-initial voiced obstruents than word-initial voiceless obstruents. 12 of the allotted 20 voiceless obstruents were necessarily word-final, leaving room for 8 word-initial voiceless obstruents: 4 in the “final voiceless obstruent” column, 2 in the “final sonorant” column, and 2 in the “final vowel” column. The 20 instances of word-initial voiced obstruents included 4 in the “final voiceless obstruent” column, 8 in the “final sonorant” column, and 8 in the “final vowel” column.

The VOICING stimuli presented in the Exposure phase, provided in Table 6.2, were nearly-identical to the stimuli for the DEVOICING pattern, save for two elements. First,



Exposure words		Fillers			
Final voiced obstruent		Final sonorant		Final vowel	
paz	‘pozz’	dɑn	‘dahn’	zɑ	‘zah’
teɪdʒ	‘taydʒe’	vɔl	‘vuhl’	dʒi	‘jee’
sudʒ	‘soodʒe’	poum	‘pome’	toʊ	‘towe’
tʃʌz	‘chuzz’	sɑɪ	‘sar’	su	‘soo’
diz	‘deeze’	tʃouɪn	‘chown’	fu	‘foo’
vɑdʒ	‘vodʒe’	tul	‘tule’	seɪ	‘sey’
zʌdʒ	‘zudʒe’	foʊl	‘fole’	tʃou	‘chowe’
dʒeɪz	‘jayze’	fum	‘foom’	tɑ	‘tah’
mouʒ	‘mowze’	sɑn	‘sahn’	pou	‘powe’
nɪdʒ	‘needʒe’	peɪm	‘paim’	tʃeɪ	‘chay’
ɹuz	‘rooze’	nɪm	‘neem’	moʊ	‘mowe’
loʊdʒ	‘lowdʒe’	lɑɪ	‘lar’	ɪ	‘ree’

Table 6.2: VOICING stimuli presented in the Exposure phase

the 12 test words contained only final voiced obstruents and no final voiceless obstruents (compare: [pas] in Table 6.1 vs. [paz] in Table 6.2). Second, this language contained more word-initial voiceless obstruents than word-initial voiced obstruents (compare: 8 voiced vs. 20 voiceless). To achieve this, the voicing feature of any given obstruent from the 24 fillers was reversed (compare: [tan] in Table 6.1 vs. [dɑn] in Table 6.2; [boum] in Table 6.1 vs. [poum] in Table 6.2). The only four words that were identical in both patterns were the only four that contained no obstruents: [nim, lɑɪ, moʊ, ɪ].

In the Novel phase, all four groups of participants heard the same exact set of 72 stimuli, provided in Table 6.3. This set used the same phonemic inventory as the words designed for the Exposure phase; however, none of the words played in the Novel phase were present in the Exposure set. 18 words contained a word-final voiceless obstruent; these words were coded as belonging to the DEVOICING pattern and not to the VOICING pattern. 18 words contained a word-final voiced obstruent, all of which belonged to the VOICING pattern but not to the DEVOICING pattern. In addition, there were 36 fillers (half with a word-final sonorant consonant and half with a word-final vowel). The fillers

were coded as belonging to both languages. Therefore, for each pattern, 54 of the Novel phase stimuli conformed to that pattern, and 18 words did not.

The stimuli presented in the Novel phase contained an equal number of voiceless obstruents, voiced obstruents, and sonorant consonants: 42 of each, with 24 in onset position and 18 in coda position. Each of the vowels in the phonemic inventory were represented six times in the set of test words: 3 times before a voiceless obstruent and 3 times before a voiced obstruent. Each vowel appeared at least once before each of the final obstruents, so that participants would never hear, for example, the same vowel only before [s] and never before [tʃ].

The stimuli were all phonotactically possible words of English, most of which were non-words. There were a few exceptions, where some of the fillers did have a real-world counterpart in English (e.g., [boʊ] ‘bow’, [dʒeɪ] ‘jay’). Whenever possible, these stimuli were presented to the speaker during the recording session using non-English spelling (e.g., ‘bowe’ for [boʊ], ‘jayze’ for [dʒeɪ]), to discourage recall of familiar words during production.

### 6.1.2.2 Recording

The two levels of the CLARITY condition differ as to whether participants were trained on their assigned pattern with stimuli recorded in careful speech or casual speech. Two sets of stimuli were recorded for each pattern, one in each register, yielding four sets of stimuli across the entire experiment. Note that the four groups of participants differed only in the pattern and register they heard in the Exposure phase. In the Novel phase, all participants heard the stimuli presented in Table 6.3 in the CAREFUL register.

The stimuli were recorded in the Phonetics Lab at UC Santa Cruz. The speaker was a 20-year-old male undergraduate at UC Santa Cruz and a monolingual speaker of California English. Out of a set of four undergraduates who completed the recording session, this speaker was chosen due to impressionistically performing the best at producing the intended targets of casual register: partial or full devoicing of voiced consonants, and failing to release final obstruents. The stimuli were recorded in Praat version 5.2.16 (Boersma & Weenink 2014) at a sampling rate of 44.1 kHz using an AKG HSC271 microphone headset and a Focusrite Scarlett 18i8 audio interface.

Novel words				Fillers			
Final voiceless		Final voiced		Final sonorant		Final vowel	
obstruent		obstruent					
pʌtʃ	‘putch’	pouɔdʒ	‘powdge’	pɑl	‘pahl’	peɪ	‘pey’
fouʃ	‘fowse’	fuz	‘fooze’	fun	‘foon’	feɪ	‘fay’
sɪtʃ	‘seech’	sadʒ	‘sodge’	fɑm	‘fahm’	tʃu	‘choo’
teɪʃ	‘tayse’	tʌz	‘tuhze’	sɑl	‘sahl’	soʊ	‘sowe’
tutʃ	‘tooch’	tɪdʒ	‘teedge’	teɪn	‘tayn’	teɪ	‘tay’
tʃɑs	‘choss’	tʃeɪz	‘chayze’	tʃɑm	‘chom’	tʃɑ	‘chah’
bɪs	‘beese’	beɪz	‘bayze’	bɑl	‘buhl’	bɑ	‘bah’
vʌs	‘vuss’	voʊz	‘vowze’	voʊm	‘vome’	veɪ	‘vay’
vutʃ	‘vooch’	vɪdʒ	‘veedge’	viː	‘veer’	vɑ	‘vah’
zeɪtʃ	‘zaych’	zadʒ	‘zodge’	ziː	‘zeer’	zi	‘zee’
doutʃ	‘dowch’	dudʒ	‘doodge’	doʊl	‘dole’	di	‘dee’
ʒɑs	‘joss’	dʒʌz	‘juhze’	dʒoʊl	‘jole’	dʒɑ	‘jah’
lɑtʃ	‘lotch’	leɪdʒ	‘laydge’	mɑl	‘mull’	meɪ	‘may’
lʌs	‘luss’	loʊz	‘lowze’	nɑː	‘nar’	nɑ	‘nah’
nutʃ	‘nooch’	nɑdʒ	‘nodge’	lɪm	‘leem’	li	‘lee’
neɪs	‘nayce’	nʌz	‘nuzz’	.ɪn	‘reen’	.ɪoʊ	‘rowe’
mitʃ	‘meech’	mudʒ	‘moodge’	mʌl	‘mool’	neɪ	‘nay’
.ɪoʊs	‘rowse’	.ɪz	‘reeze’	lɑn	‘lahn’	.ɪeɪ	‘rey’

Table 6.3: Stimuli presented in the Novel phase in all conditions

The stimuli were embedded inside carrier phrases, which were designed to help facilitate production of careful speech vs. casual speech. In examples (2) and (3), the nonce words *soof* [suf] and *leem* [lim] (boxed; from the Novel phase stimuli) are used to represent the stimulus that was extracted and incorporated into the experiment. The bolded words represent elements in phonologically contrastive focus.

- (2) *Carrier phrase for the CAREFUL stimuli*  
“No, I didn’t say **soove**, I said **soof**.”  
“No, I didn’t say **leen**, I said **leem**.”
- (3) *Carrier phrase for the CASUAL stimuli*  
“No, I’m not gonna talk about **that** soof, I’m gonna talk about **this** soof.”  
“No, I’m not gonna talk about **that** leem, I’m gonna talk about **this** leem.”

The stimulus was utterance-final in both carrier phrases. The CASUAL carrier phrase required the speaker to produce the target word twice; however, only the utterance-final token was incorporated into the experiment. This was designed to eliminate any effects that a following sound might have on the production of the final obstruents: e.g., sustained voicing, early devoicing, or durational effects.

The two carrier phrases mainly differed as to whether or not the stimulus word was focused. In the CAREFUL carrier phrase, the stimulus was not only focused, but the word against which it was contrasted differed in terms of the final consonant. For the test words (the non-fillers), the contrastive pairs differed in voicing of the final consonant. The idea behind this was to coerce the speaker into producing the voicing contrast especially clearly so that it would be easier for participants to determine whether the final consonant was voiced or voiceless. In the CASUAL carrier phrase, the stimulus is not focused, nor does it appear in a contrastive pair. Doing so should minimize the phonetic cues for voicing in the stimuli, thereby making the words in the VOICING pattern more difficult to hear in casual speech than in careful speech.

In addition to the carrier phrases, the speaker received explicit instruction on how he should produce the set of CAREFUL phrases and the set of CASUAL phrases. To produce the CAREFUL stimuli, he was instructed to speak as if he were addressing an elderly superior, attempting to be as polite as possible, keeping in mind to speak as clearly as possible so that his words may be understood. For the CASUAL stimuli, he was instructed to speak as if he were chatting with his best friend in his dorm room, in which there was no music playing, no one else around, and no particular need to speak clearly.

The speaker recorded a total of 280 sentences (carrier phrase + stimulus), 140 in

each voice style. During the recording session, the speaker was seated at a computer in a soundproof booth, and the experimenter stood behind him. The speaker was given a paper packet to read from, which contained the carrier phrases plus the written instructions for how to conduct the session. Before the session began, he was told that he would be reading aloud a series of sentences, each of which contained a non-English word. He was instructed to pronounce them as if they were English words and assume conventional methods of English spelling (a few practice words were given in the introduction to provide examples of how to pronounce the vowels and certain coda consonants).

The entire set of 140 CAREFUL sentences were recorded first, before any of the CASUAL sentences. They were presented to the speaker in a series of 14 lists, each of which contained 10 carrier phrases and one stimulus per carrier phrase. The lists were created by randomizing the 140-sentence set in Python, such that words from both patterns (DEVOICING and VOICING) and both experimental phases (Exposure and Novel) were intermixed across the lists. The carrier phrases were transcribed orthographically. After every ten-sentence list, the speaker was offered a self-timed break, although he did not choose to take any breaks longer than one minute.

Following completion of the CAREFUL stimuli, the speaker was encouraged to take a longer break, drink some water, and rest his voice. The experimenter then initiated a five-minute conversation with the speaker, in which the speaker was asked to recount the plot of his favorite movie (or book, or play, or TV show). The purpose of this mini-interview was to encourage the speaker to begin speaking in a more casual mode and train them out of speaking in his careful register. This session was not recorded.

Finally, the speaker recorded the remaining 140 CASUAL stimuli using the same methods as the recording of the CAREFUL stimuli. The entire recording session took exactly one hour, and the speaker was compensated \$10 for his participation.

Following recording, the target words were extracted from their carrier phrases using the “0 Chop” Praat script,<sup>1</sup> which extracts labeled intervals from the sound and TextGrid tiers and saves them as individual .wav files. The stimuli were then normalized using the “Equalizing Amplitude” Praat script.<sup>2</sup> This script takes, as its input, a series

---

<sup>1</sup>Written by Michael Wagner; edited by Erin Olson.

<sup>2</sup>Written by Shigeto Kawahara.

of .wav files inside a single directory, rescales the amplitude of each file to a specified dB value (in this experiment, 70 dB), and saves the series as a new set of .wav files with adjusted amplitudes. This is equivalent to using the command “Modify → Scale intensity...” in the Praat Objects window and has the apparent effects of setting all the .wav files to roughly the same intensity.

### 6.1.2.3 Acoustic Analysis

An acoustic analysis of the stimuli presented in the Exposure and Novel phases confirmed that the speaker was producing the intended cues to voicing described in this section. Final voiced obstruents were produced with more acoustic voicing in the CAREFUL speech stimuli than in the CASUAL speech stimuli, suggesting that the speaker was devoicing his obstruents to a greater degree in the CASUAL condition, as predicted.

Four different duration measurements were calculated in Praat for each stimulus: the durations of the word, of the final obstruent, of the voicing during the obstruent, and of the vowel preceding the obstruent. To determine the landmarks for where the strident began, I relied primarily on the waveform, placing the boundary at the point where the periodic property of the vowel began to change and where noise began to increase (Tine Mooshammer, p.c.). The end of formant structure in the spectrogram from the vowel was also used as a secondary measure. The offset of the strident was marked when the aperiodic noise in the waveform roughly reached equilibrium. The onset and offset of voicing during the strident was measured against the voicing bar in the spectrogram. The proportion of voicing during the obstruent was determined for each production by dividing the obstruent duration by the voicing duration, resulting in a number within the range  $0 \leq x \leq 1$ . Figures 6.3–6.4 provide sample spectrograms of stimuli with a final voiced obstruent. The stimulus in Figure 6.3 was determined to have a fully voiced final obstruent (followed by a short vocoid), while the final obstruent in Figure 6.4 is partially voiced at over 50% of the total duration.

The barplot in Figure 6.5 shows the average proportion of voicing during the final obstruent across both obstruent types (voiced and voiceless) and both CLARITY types (CAREFUL and CASUAL). The final voiced obstruents in the CAREFUL condition have overwhelmingly the highest proportion of voicing. In the CASUAL condition, this

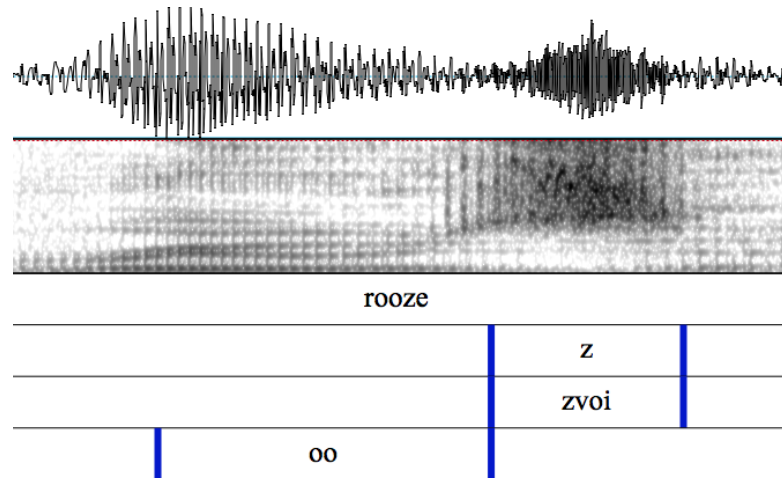


Figure 6.3: Spectrogram of ‘rooze’ [ɹuz] in CAREFUL speech (full voicing)

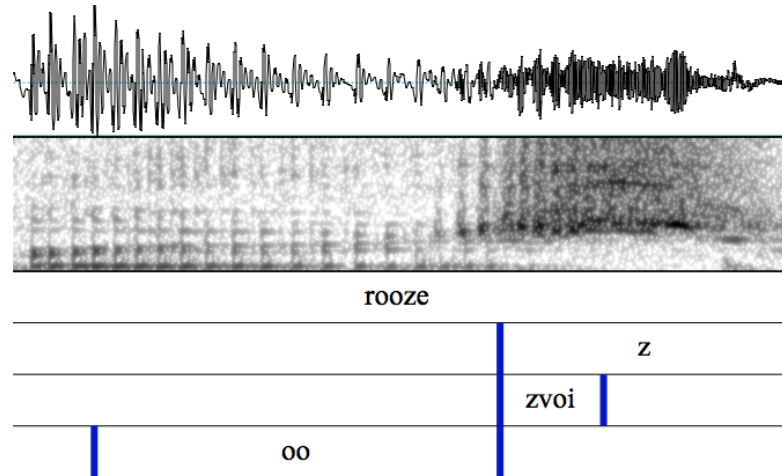


Figure 6.4: Spectrogram of ‘rooze’ [ɹuz] in CASUAL speech (under 50% voiced)

proportion has significantly decreased; a linear mixed effects model confirms that the effect of CLARITY is significant ( $t = -7.570$ ), as is the interaction of CLARITY and intended voicing of the final obstruent ( $t = 5.208$ ). The final voiceless sounds have a very small proportion of voicing during the obstruent, corresponding to a few milliseconds of voicing bleeding in from the vowel. While the final voiced obstruents in the CASUAL condition still have a higher proportion of voicing during the closure than the voiceless obstruents, it is predicted that this decrease in voicing proportion will lead participants

to perceive some of these sounds as their voiceless counterparts. If this occurs during the Exposure phase, it should affect their performance in the Novel phase, as they will not have correctly interpreted that their pattern allows only final voiced obstruents.

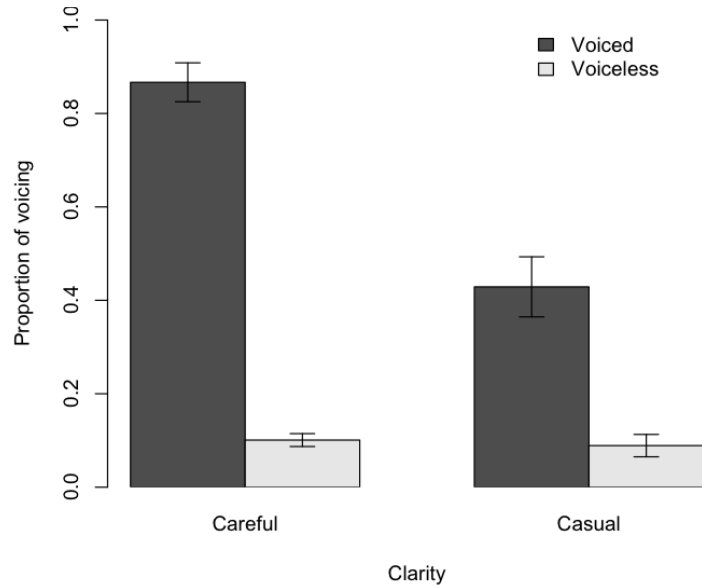


Figure 6.5: Proportion of voicing during the obstruent in each level of CLARITY

Since vowel duration is also a cue for final voicing of an obstruent, the duration of the vowel preceding the obstruent was measured as well. Figure 6.6 shows the average duration of the preceding vowel for each obstruent type and each CLARITY condition. On average, vowels are longest before voiced obstruents in the CAREFUL condition. A linear mixed effects model found that, compared to voiced obstruents in the CAREFUL condition, vowels are significantly shorter before voiceless obstruents in the CAREFUL condition ( $t = -6.511$ ) and before voiced obstruents in CASUAL ( $t = -4.599$ ). The interaction was also significant ( $t = 2.490$ ), signifying that the decrease in vowel duration before voiced obstruents from the CAREFUL to the CASUAL condition was greater than the same decrease before voiceless obstruents.

These two cues were analyzed to determine if there was any correlation between the proportion of voicing during the obstruent and the duration of the vowel before the obstruent. Two scatterplots are presented in Figures 6.7–6.8, where the  $x$ -axis corresponds to the voicing proportion and the  $y$ -axis corresponds to vowel duration. Figure



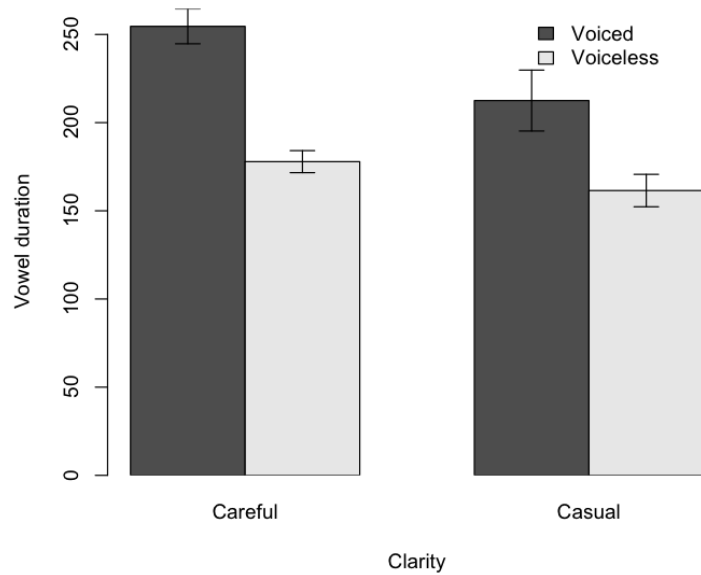


Figure 6.6: Vowel duration preceding each obstruent in each level of CLARITY

6.7 represents the tokens of the CAREFUL condition, while Figure 6.8 corresponds to the CASUAL condition. Note that there are fewer tokens within the CASUAL condition, because all of the Novel stimuli and half of the Exposure stimuli were presented in CAREFUL speech, while the CASUAL stimuli were only represented in half of the Exposure phases. In the CAREFUL stimuli, there is a significant positive correlation ( $r = 0.485$ ,  $r^2 = 0.235$ ,  $p \approx 0$ ), such that words with more voicing during the obstruent are also likely to have a longer vowel than words with less voicing. There is no such correlation in the CASUAL stimuli ( $r = -0.070$ ,  $r^2 = 0.005$ ,  $p = 0.745$ ).

What this means is that the acoustics of the words with final voiced obstruents are significantly different between the CAREFUL and CASUAL phases. In the CASUAL phases, these stimuli partially resemble the stimuli with final voiceless obstruents, due to their decreased voicing during the obstruent and their decreased vowel duration. Note, however, that the VOICING: CASUAL words are not identical to the DEVOICING words, but rather their cues lie somewhere in between the VOICING: CAREFUL stimuli and the DEVOICING stimuli in both CLARITY conditions. It is predicted that participants will perceive some of these words as ending in a final voiceless obstruent and some as ending in a final voiced obstruent.

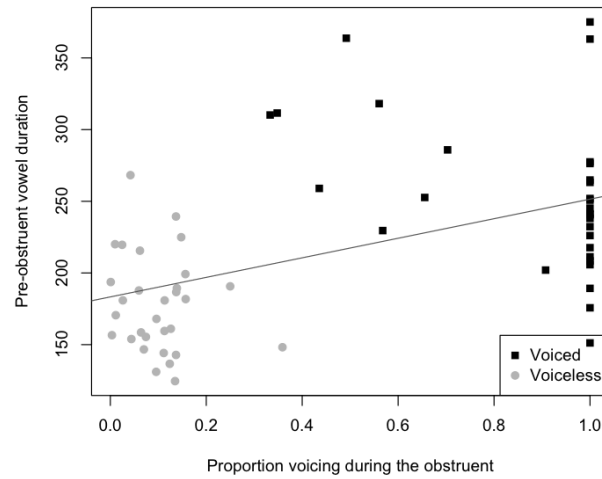


Figure 6.7: Correlation between proportion voicing and vowel duration in the CAREFUL stimuli

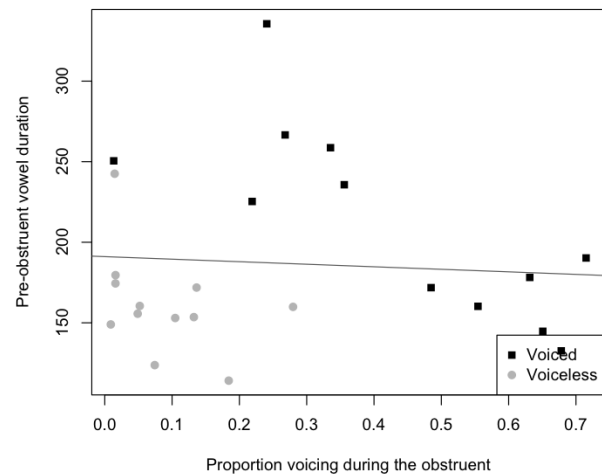


Figure 6.8: Correlation between proportion voicing and vowel duration in the CASUAL stimuli

Finally, obstruent duration depended both on CLARITY and on the voicing of the final obstruent. Obstruents were longer in CAREFUL speech than in CASUAL speech ( $t = -3.004$ ), and voiceless obstruents were longer than voiced obstruents ( $t = 8.593$ ). The interaction between voicing and CLARITY also reached significance ( $t = -2.865$ ), in that the difference in duration between voiced and voiceless obstruents was smaller in CASUAL speech than in CAREFUL. Based on obstruent duration alone, then, the

distinction between final voiceless and final voiced obstruents is smaller in CASUAL speech, and perhaps harder to perceive. Figure 6.9 displays a bargraph of these results.

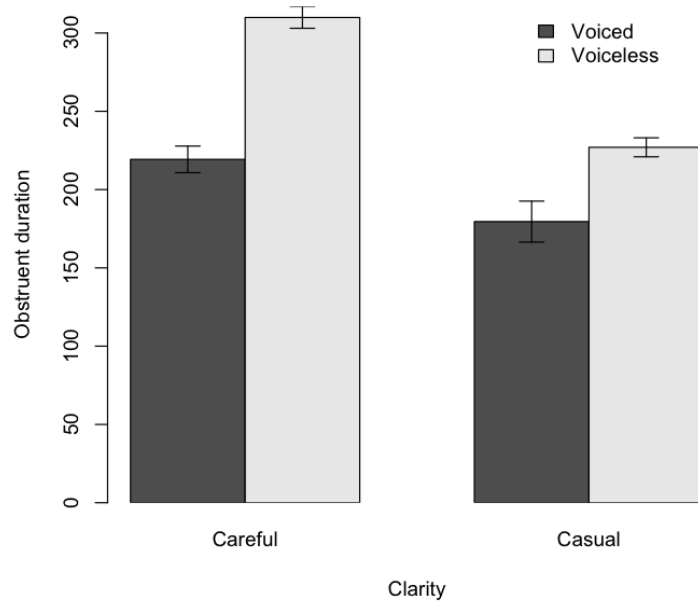


Figure 6.9: Obstruent duration in each level of CLARITY

### 6.1.3 Procedure

The experiment was conducted inside soundproof booths in the Phonetics Lab at UC Santa Cruz using the E-Prime software. Materials were presented over headphones. Each setup included a button box with labels numbered 1 through 5. Participants were instructed to press the button labeled ‘3’ to advance through the instruction screens and the ‘1’ and ‘5’ buttons to answer questions. Participants were assigned to one of the four conditions by the experimenter. The equipment used in this experiment is the same as the equipment described in Chapter 5.1.1.

The procedure consisted of three phases: an Exposure phase, a Novel phase, and an Identification phase. Before the Exposure phase began, participants were told that they were about to hear a series of words from a made-up language and would be asked questions about them.

*Exposure phase.* In the Exposure phase, participants heard all 36 words in the stimulus set designed for their pattern, each in isolation (i.e., not presented inside a carrier phrase). Participants were instructed to listen to each word carefully and then press the ‘3’ button to proceed to the next word. Each list of 36 words was presented a total of five times, in random order, with breaks in between each list, yielding a total of 180 trials in the Exposure phase. Participants were not asked to respond to any questions in the Exposure phase, nor were they instructed to speak the stimuli aloud.

*Novel phase.* Immediately following the Exposure phase, participants were presented with the set of 72 novel words, each of which either conformed to the given pattern or did not conform. Each stimulus was presented auditorily, in isolation, and one word at a time. For each trial, participants were instructed to press ‘1’ on the button box if the stimulus conformed to the pattern that they were learning and ‘5’ if it did not. Participants had to press the ‘3’ button to move forward to each trial. All participants across all conditions heard the exact same set of Novel stimuli, which the speaker recorded according to the CAREFUL speech instructions.

*Identification phase.* After completing the Novel phase, participants completed a task in which they identified the words they heard in the Exposure phase. In each trial, a single word from the Exposure phase of the assigned pattern and voice clarity would play through the headphones. Two possible spellings of that word were presented on the screen: one on the left side of the screen, accompanied by the number ‘1’, and one on the right side, accompanied by the number ‘5’. Participants were instructed to press the ‘1’ button if they thought the word on the left was the correct spelling of the word they heard, and to press ‘5’ if the word on the right was the correct spelling. The 12 test words differed only in the spellings of the final consonant, using the English orthography provided in Tables 6.1–6.2. For example, participants assigned to the DEVOICING pattern would hear [pas] through the headphones and see *poss* and *pozz* as the two choices.

The two-alternative forced choice method for this task was chosen in favor of having participants provide their own spellings to account for orthography-based confounds. English words spelled with final “se” correspond both to [s]-final words (e.g., *goose*, *mouse*, *chase*) and to [z]-final words (e.g., *choose*, *cheese*, *hose*). If a participant were

to hear the word [mouz] and provide the spelling *mouse*, it could not be determined whether that participant thought they heard [mouz] or [mous]. Giving participants with only two choices of spelling provides a better indication of whether they heard final voicing or final voicelessness. One caveat to this method, however, is that it hints to participants that the experiment is interested in the final voicing contrast. To help distract from this, most of the filler words followed a different pattern for how the spellings differed. For six of the filler words, the two spelling choices differed according to final consonant (e.g., the choices for [vum] were *voom* and *voon*). For the remaining 18 fillers, the spellings differed according to vowel (e.g., the choices for [tʃou] were *chowe* and *chah*).

#### 6.1.4 Hypothesis

If the channel bias account is correct, then performance differences between the natural and unnatural pattern should arise only when final obstruent voicing is harder to perceive, and should disappear when final voicing is easier to perceive. The hypothesis that the Final Devoicing Experiment tests is provided in (4).

- (4) Naturalness effects observed in the lab arise only when there is a difference of perceptibility between the natural and the unnatural stimuli. It is predicted that participants who learn the VOICING pattern in the CASUAL register will perform significantly worse than participants who learn the DEVOICING pattern in the CASUAL register, as well as the participants who learn the VOICING pattern in the CAREFUL register.

## 6.2 Results

### 6.2.1 Novel phase

As with the previous experiments in this dissertation, the findings are assessed based on average proportion of correct trials in each group. For the participants exposed to the DEVOICING pattern, a correct trial entails responding ‘1’ (does belong) to words with final voiceless obstruents and ‘5’ (does not belong) to words with final voiced

obstruents. For the VOICING pattern, the reverse is true: a correct trial means the participant either responded ‘1’ (does belong) to words with a final voiced obstruent or ‘5’ (does not belong) to words with a final voiceless obstruent. Fillers have been excluded from the analysis.

Figure 6.10 shows the proportion of correct trials in the Novel phase, broken down by the PATTERN and CLARITY conditions. The dashed line at  $y=0.5$  represents chance. Figure 6.11 shows the d prime results. All four of the groups performed numerically higher than chance. The two DEVOICING groups appear to have performed with roughly the same accuracy. For the VOICING pattern, however, the group who heard the CAREFUL stimuli during the Exposure phase performed remarkably high. The average accuracy and d primes in each group are provided in Table 6.4.

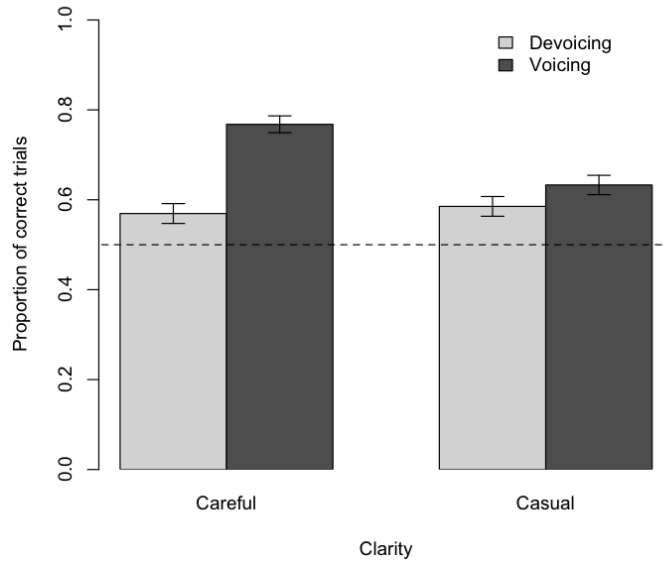


Figure 6.10: Performance in the Novel phase by PATTERN and CLARITY

Each subject’s individual average score on the Novel task was computed, and their scores were analyzed with a linear regression with fixed effects of PATTERN and CLARITY. The DEVOICING pattern in the CAREFUL register corresponds to the Intercept. The results of this test are provided in Table 6.5.

The main effect of CLARITY did not reach significance ( $p = 0.734$ ), so we may assume that the two DEVOICING groups performed about equally and that CLARITY did not affect their performance. The effect of PATTERN was, unsurprisingly, significant ( $p$

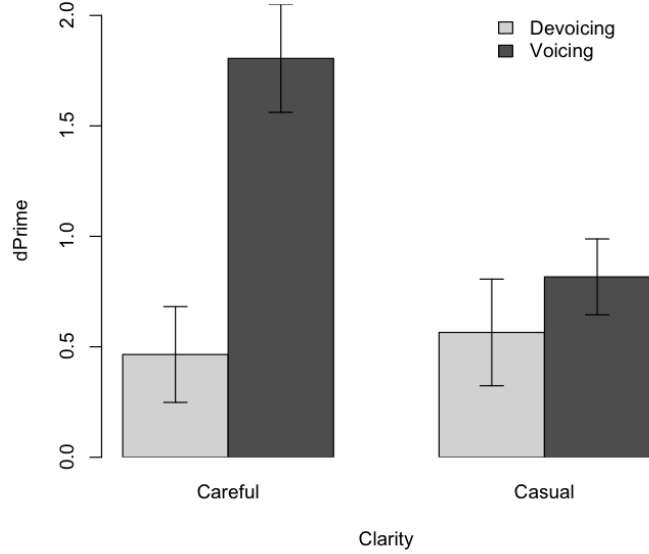


Figure 6.11: d prime in the Novel phase by PATTERN and CLARITY

Pattern	Clarity	Accuracy	d prime
DEVOICING	CAREFUL	0.569	0.465
VOICING	CAREFUL	0.768	1.805
DEVOICING	CASUAL	0.585	0.565
VOICING	CASUAL	0.633	0.816

Table 6.4: Average accuracy and d prime in the Novel phase

$\approx 0$ ). Out of the two groups trained on CAREFUL stimuli, the VOICING group performed significantly better on the Novel task than the DEVOICING group. Lastly, there was a significant interaction between PATTERN and CLARITY ( $p < 0.05$ ). A one sample t-test confirmed that the DEVOICING:CAREFUL group’s performance was marginally different from chance ( $t = 2.07$ ,  $df = 13$ ,  $p = 0.59$ ).

Two post hoc analyses of the simple effects investigated the effect of PATTERN in each of the two levels of CLARITY. Table 6.6 contains the output of the analysis of the CAREFUL groups, while Table 6.7 contains the analysis of the CASUAL groups. From these post hoc analyses, we can see that the reason for the significant interaction lies within the CAREFUL groups. The VOICING:CAREFUL group performed the Novel task

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.569	0.033	17.336	$\approx 0$ ***
Casual	0.016	0.047	0.342	0.734
Voicing	0.198	0.047	4.271	$\approx 0$ ***
Casual:Voicing	-0.151	0.066	-2.295	0.026 *

Table 6.5: Linear regression of Novel task (Intercept = DEVOICING:CAREFUL)

more successfully than the DEVOICING:CAREFUL group. In the CASUAL condition, the two groups did not perform significantly differently from each other.

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.569	0.033	17.307	$\approx 0$ ***
Voicing	0.198	0.047	4.264	0.0002 ***

Table 6.6: Post hoc: Effect of PATTERN in CAREFUL groups

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.585	0.033	17.849	$\approx 0$ ***
Voicing	0.048	0.046	1.027	0.314

Table 6.7: Post hoc: Effect of PATTERN in CASUAL groups

A striking finding of this experiment is the relatively poor performance of the DEVOICING groups. It was predicted that participants trained on the this pattern would perform successfully on the Novel task in both the CAREFUL and CASUAL conditions, leading the DEVOICING:CASUAL group to perform more successfully than the VOICING:CASUAL group and yielding a significant interaction (Figure 6.2). Instead, in the CASUAL condition, the two groups performed equally well, while in the CAREFUL condition, the DEVOICING group actually performed significantly *worse* than the VOICING group.

To investigate their performance, I investigated the nature of the errors that the DEVOICING groups were making. In the Novel phase, there were two ways that a



participant could respond inaccurately to a trial: by accepting words that did not belong to the pattern or by rejecting words that did belong to the pattern. The two DEVOICING groups should have a high acceptance rate of voiceless-final words and a low acceptance rate of voiced-final words. The two VOICING groups should perform in the reverse order: high acceptance of voiced-final words and low acceptance of voiceless-final words. Figure 6.12 provides a bargraph of each group’s acceptance rate of each word type. The bars correspond to each stimulus type: the “Voiced” bar, for example, represents how many of the trials containing [paz]-type words each group accepted as belonging to their assigned language.

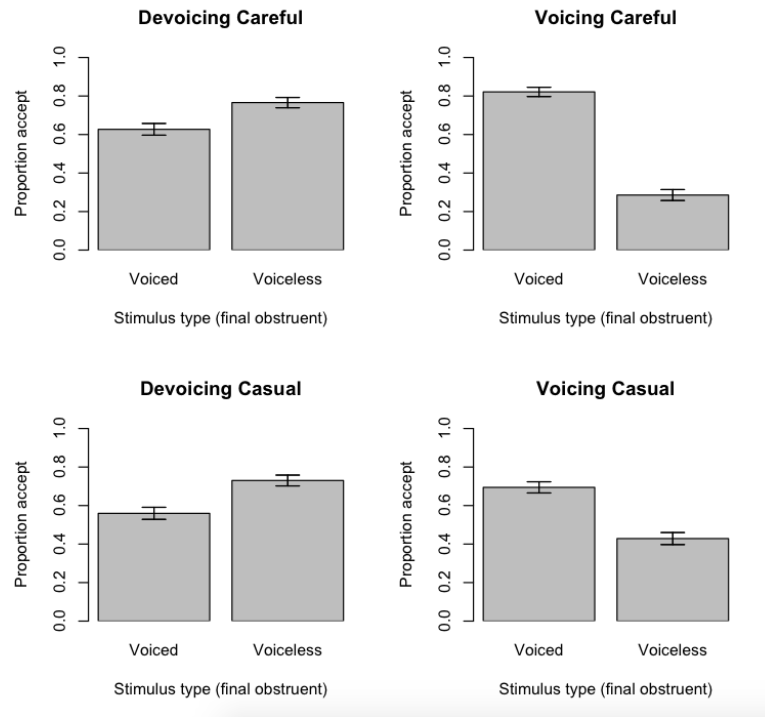


Figure 6.12: Proportion of accepted words by final obstruent type

The above figure shows that the VOICING:CAREFUL group accepted and rejected words much as they were supposed to, accepting an average of 82.14% [paz]-type trials and 28.57% [pas]-type trials. The other groups, however, seemed to struggle. While the two DEVOICING groups did accept [pas]-type words at a high rate (76.59% for CAREFUL; 73.02% for CASUAL), they also accepted a fair amount of [paz]-type trials (62.70% for

CAREFUL; 55.95% for CASUAL). Their willingness to incorrectly accept words ending in voiced obstruents is likely what drove down their performance in the Novel task.

A series of t-tests were run to test whether the DEVOICING and VOICING groups performed significantly differently in their acceptance rates. First, each subject's individual average acceptance rate of each trial type (Voiced and Voiceless) was calculated. From this, I calculated each subject's "difference" score: their average "Voiced" acceptance minus their average "Voiceless" acceptance score. This yielded four vectors of length 14, one for each experimental condition, corresponding to the 14 subjects (and thus 14 difference scores) in each condition. A Welch's two sample t-test compared the difference scores of the two CAREFUL groups and confirmed that the mean difference scores of the DEVOICING and VOICING groups were significantly different ( $t = -7.25$ ,  $df = 25.96$ ,  $p \approx 0$ ).

This test was followed up with a series of paired-sample t-tests, which investigated the nature of the interaction. A t-test that compared the DEVOICING:CAREFUL group's acceptance rate of [paz]-type and [pas]-type trials reached marginal significance ( $t = -2.07$ ,  $df = 13$ ,  $p = 0.059$ ). In other words, this group just barely accepted Voiceless words at a higher rate than Voiced words. Another paired-sample t-test repeated this process amongst the VOICING:CAREFUL group, which confirmed that [paz]-type words were accepted at a significantly higher rate than [pas]-type words ( $t = 8.31$ ,  $df = 13$ ,  $p \approx 0$ ).

This process was repeated for the two CASUAL groups. A Welch's two sample t-test determined that the DEVOICING:CASUAL and VOICING:CASUAL groups yielded significantly different difference scores ( $t = -4.71$ ,  $df = 23.47$ ,  $p \approx 0$ ). Within the DEVOICING:CASUAL group, the acceptance rate of [pas]-type and [paz]-type trials reached significance, with a greater (correct) acceptance of [pas]-type trials ( $t = -2.26$ ,  $df = 13$ ,  $p < 0.05$ ). The VOICING:CASUAL group also accepted more correct ([paz]-type) than incorrect ([pas]-type) trials, but this difference was significant by a smaller alpha ( $t = 4.95$ ,  $df = 13$ ,  $p \approx 0$ ).

In sum, all four experimental conditions accepted more correct trials than incorrect trials. The VOICING groups, however, seemed to perform better at this task, yielding a larger difference between their correct "accept" and incorrect "accept" trials than did

the DEVOICING groups. As mentioned, this likely explains the poor performance of the DEVOICING groups.

### 6.2.2 Identification phase

The results of the Identification (ID) phase suggest that participants were able to successfully perceive and report the voicing of the final obstruent. Figure 6.13 depicts the proportion of correct trials across all four groups in the Identification phase. As this image shows, all groups performed with high accuracy and well above chance. The linear regression in Table 6.8 confirms that the four groups performed significantly above chance and not significantly differently from each other.

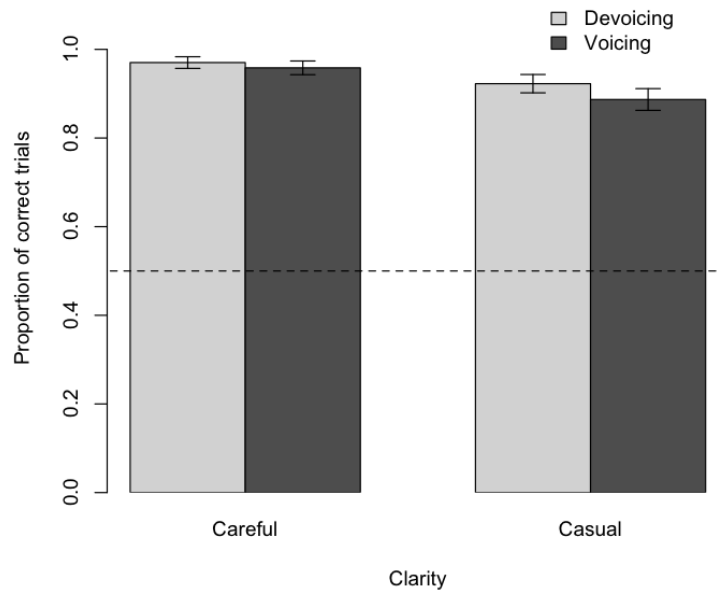


Figure 6.13: Performance in the ID phase by PATTERN and CLARITY

Although participants performed the ID task with high accuracy, there was an observable correlation between proportion of voicing during the obstruent and average performance on the ID task. The scatterplot in Figure 6.14 shows ID accuracy as a function of the proportion of voicing during the obstruent. This plot is restricted to the VOICING stimuli only, since proportion of voicing in the DEVOICING stimuli did not fluctuate very much. A Pearson's product-moment correlation coefficient was calculated

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.970	0.036	27.051	≈0 ***
Casual	-0.048	0.051	-0.939	0.352
Voicing	-0.012	0.051	-0.235	0.815
Casual:Voicing	-0.024	0.072	-0.332	0.741

Table 6.8: Linear regression of ID task (Intercept = DEVOICING:CAREFUL)

for the relationship between amount of obstruent voicing and accuracy on the ID task. The correlation is such that finally-voiced words with more voicing during the obstruent were more likely to be correctly identified as finally-voiced ( $r = 0.518$ ,  $r^2 = 0.268$ ,  $p < 0.01$ ). An additional test determined whether there was a correlation between vowel duration and Identification task accuracy; this correlation was not significant ( $r = -0.352$ ,  $r^2 = 0.124$ ,  $p = 0.092$ ).

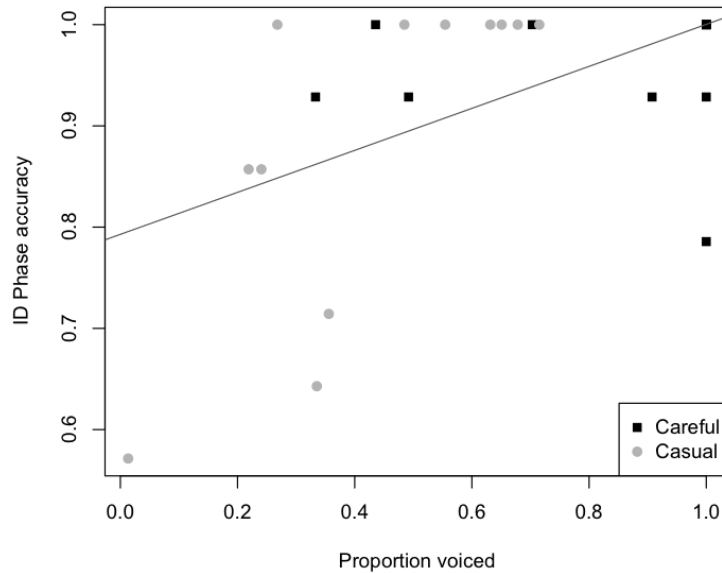


Figure 6.14: ID accuracy based on proportion of obstruent voicing (VOICING only)

### 6.3 Discussion

The goal of this experiment was to distinguish between the channel bias and universal learning bias accounts by manipulating the clarity of the stimuli that the participants were trained on. Evidence for channel bias would be seen if increasing the clarity of the unnatural stimuli increased performance, and if decreasing the clarity decreased performance. Under this hypothesis, performance on an unnatural pattern has a strong connection to clarity of the stimuli. Evidence for universal learning bias, on the other hand, would be seen if the group who learned the unnatural pattern performed worse than the group who learned the natural pattern, even when the stimuli were produced in CAREFUL speech. The latter scenario would suggest that the unnatural pattern is more difficult to learn for reasons unrelated to perception, since the Exposure stimuli for both groups should have (in theory) been exceptionally easy to perceive.

This experiment suggests that channel bias underlies the naturalness of final voicing and final devoicing patterns. This claim is supported by several pieces of evidence. In the Novel phase (Figure 6.10), we see that performance on the VOICING pattern is higher in the CAREFUL condition, in which obstruents are fully or mostly voiced, than in the CASUAL condition, in which obstruents are fully or mostly devoiced. This shows that clarity of the stimuli did have an effect on performance: participants trained on final obstruents with *more* voicing performed more accurately on the Novel task than participants trained on *less* obstruent voicing. In the ID phase (Figure 6.14), there was a positive correlation within the words belonging to the VOICING pattern between proportion of voicing during the obstruent and accuracy on the ID task. The correlation may have had categorical effects on the results of the ID phase, in that the VOICING:CASUAL group performed numerically less successfully than the other three groups, although this difference did not come out as significant (Figure 6.13).

These pieces of evidence suggest that channel bias was active but do not suggest on their own that universal learning bias was *inactive*. The evidence against the universal learning bias account comes from the two groups trained on the CAREFUL stimuli. The VOICING group performed much better than the DEVOICING group, a result that is problematic for the universal learning bias account but can be explained by the channel bias account. I speak to this distinction in the remainder of the discussion.

This experiment did not find evidence for naturalness effects in the lab, in the sense of the natural pattern being learned more successfully than the unnatural pattern. In the CAREFUL condition, the unnatural pattern was actually learned *more* successfully than the natural pattern, while in the CASUAL condition, there was no difference between the two patterns. What the experiment did find, however, is evidence for what I call channel bias in the direction of naturalness. That is, the relation between the speaker's production of the voiced stimuli and the participants' perception of the VOICING pattern is in line with our prediction. When final voiced obstruents are less voiced, participants are more likely to perceive them as voiceless. From a production-based stance, sound change over time should veer in the direction of final devoicing, since utterance-final voicing is more difficult to sustain than utterance-final voicelessness. This experiment supplies evidence that perception also favors sound change in the direction of final devoicing. If speakers do, in fact, partially devoice utterance-final voiced obstruents, listeners will store these productions in memory – and, when they become the speaker, they should produce partially devoiced final obstruents as well (see the discussion in Chapter 2.2). The more devoiced the productions of the voiced obstruents become, the more likely listeners will perceive these utterances as voiceless.

These results are problematic for a universal learning bias account of naturalness. This account predicted that participants would perform less accurately in the VOICING pattern than the DEVOICING pattern in both CLARITY conditions because, independent of any perceptual or production differences, this pattern is simply more difficult to learn. The high accuracy of the VOICING:CAREFUL group, and the lack of significance between the two CASUAL groups, could not be explained by this hypothesis. This experiment therefore lends support to the overall hypothesis of this dissertation: that naturalness effects can be explained by channel bias and not by universal learning bias.

Of course, the high performance of the VOICING groups and the low performance of the DEVOICING groups is surprising. The hypothetical channel bias scenario proposed in Figure 6.2 predicted that there would be no difference between the two groups trained on the CAREFUL stimuli, because the voicing or voicelessness of the final obstruents should have been easy to perceive. Between the groups trained on the CASUAL stimuli, the VOICING group was predicted to perform worse than the DEVOICING group, due

to hearing some of their Exposure words as finally voiceless instead of finally voiced. The results of this experiment are not in line with these predictions, because the two DEVOICING groups performed worse than expected. The prediction that the CAREFUL vs. CASUAL distinction would not be significant was borne out, but the two groups were predicted to perform higher overall than they actually did.

I had expected the DEVOICING groups to have no problem perceiving their pattern. The stimuli of the VOICING pattern were occasionally produced with some devoicing (and more so in the CASUAL stimuli than in the CAREFUL stimuli), leading to the prediction that some of these words would be misperceived as finally voiceless. Such misperception was not predicted to occur within the DEVOICING pattern, because the final voiceless obstruents should not have been produced as partially voiced. The acoustic analysis showed that indeed the speaker was not partially voicing the stimuli of the DEVOICING pattern. The slight amount of voicing that occurred was bled over from the vowel, did not last for very long during the obstruent, and did not significantly differ between the CAREFUL and CASUAL stimuli.

It is possible, however, that participants learning the DEVOICING pattern might have miscategorized the final obstruents due to influences from English. In the same vein as participants misperceiving final voiced obstruents with partial devoicing as devoiced, it is also possible that listeners might hear a final voiceless obstruent and perceive it as voiced. In English, voiced obstruents tend to be devoiced utterance-finally (Smith 1997). Keeping with the exemplar theory example from Chapter 2.2, the exemplar space of an English speaker should contain tokens of /z/ that are identical or nearly identical to tokens of /s/. Upon encountering a word like ‘poss’, listeners may store the final sound as an exemplar of /s/, but they also may reconstruct it as an exemplar of /z/ that had undergone devoicing. This is reminiscent of Ohala’s (1981) hypercorrection scenario.

If the DEVOICING groups were misperceiving in such a way, we would expect that their accuracy in the Novel task might be close to chance because they had not adequately learned that final voiced obstruents were not part of the pattern. As such, they should be accepting both types of words at an above-chance rate. This is exactly the pattern that we see. Figure 6.12 shows that participants in the DEVOICING groups

were accepting stimuli with final voiced obstruents as part of the pattern. The DEVOICING:CAREFUL group did so at a rate significantly greater than chance. This scenario would explain why the VOICING:CAREFUL group performed so much more successfully than the DEVOICING:CAREFUL group in the Novel phase. The VOICING stimuli were very clearly voiced and could not be misperceived as voiceless through hypercorrection, making it much easier for participants to perceive only finally voiced stimuli during Exposure and accept only finally voiced stimuli during the Novel phase. In short, the reason that the hypothetical results from Figure 6.2 were not observed in the experiment can be explained by the performance of the DEVOICING groups. The two VOICING groups performed as expected.

We did not see evidence in the ID phase that these participants had misperceived their stimuli. However, it is possible that the ID phase was too easy and may not have shown some of the more nuanced differences in perception between the groups. We might expect to see more differences between the groups if, for example, participants were asked to write out the word they heard as opposed to choosing one of two provided spellings. In addition, it is possible that participants' perception of the stimuli changed throughout the experiment, such that they may have perceived words like [pas] as 'pozz' at the beginning of the experiment and as 'poss' towards the end, after accruing more data. If this scenario is correct, it would be in the best interest of future researchers to design a more appropriate ID phase.

Another possible interpretation of these results is that participants' performance was driven largely by lexical biases from English. In a count from the WebCelex lexical database, Myers & Padgett (2014) found that English contains more [z]-final word forms than [s]-final word forms (22,023 ending in [z]; 12,770 ending in [s]).<sup>3</sup> This also held when the search was restricted to [s]- and [z]-final words immediately preceded by a vowel, which more closely resemble their experimental stimuli and my own (12,587 [z]; 4,385 [s]). If participants are using influence from the English lexicon as a strategy in completing the Novel task, then we would expect them to accept [z]-final words at a high rate, regardless of their Exposure training. Since this is what occurred, this interpretation remains a possibility.

---

<sup>3</sup><http://celex.mpi.nl>.



If my participants were heavily biased by the English lexicon, then it is curious why the participants in Myers & Padgett's (2014) first experiment were not. In both experiments, all participants were undergraduates at the University of California, Santa Cruz, who spoke English natively. It is reasonable to assume that these two sets of participants represent the same population. Any interpretation of the results of this dissertation's experiment should therefore not conflict with the results of Myers & Padgett's (2014) experiment. Under the hypercorrection interpretation, an explanation is needed for why my participants may have hypercorrected /pas/ to [paz] while theirs did not. Under the lexical bias interpretation, we would need to explain why the lexical bias towards [z] and away from [s] strongly influenced my participants and not theirs. It seems reasonable to assume that particulars about the two sets of stimuli may lead only one group to hypercorrect. For instance, Myers & Padgett (2014) presented sets of utterances to participants in their first experiment, while my experiment presented single words. I am unaware if this difference is enough to yield a difference in hypercorrection, but it is worth exploring. It seems less reasonable to assume, if all participants came from the same population, that a lexical bias would strongly affect performance in one set of experiments and weakly affect performance in the other.

The Final Devoicing Experiment in this dissertation provides evidence that channel bias is a viable explanation for naturalness in the domain of final devoicing. The results resemble typology: the more partial devoicing that occurs within final voiced obstruents, the more likely listeners will be to perceive them as voiceless. Under the assumption that speakers in the wild will veer towards devoicing final voiced obstruents in their productions, this experiment suggests that perception will also veer towards a pattern of final devoicing. This pattern hinges upon the observation that clarity of the production affects perception of the obstruent. As far as I am aware, the universal learning bias account could not explain why clarity of the stimuli only mattered for one of the patterns (VOICING) and not for the other, nor why the VOICING:CAREFUL group performed more successfully at rejecting non-conforming stimuli than the DEVOICING:CAREFUL group.

The results of the experiments in this dissertation have so far shown that channel bias is active in shaping naturalness in both weight-sensitive stress and in final devoicing. In addition, the universal learning bias account has been consistently more difficult to

defend in light of these results. For these reasons, I argue that we do not need to look to universal learning bias for answers about naturalness effects. A channel bias approach seems to suffice quite well on its own.

# Chapter 7

## Experiment: Coda sonority

### 7.1 Introduction

In Chapter 4, I proposed three requirements that a well-formed naturalness experiment must meet: that the experiment must compare a phonetically natural pattern to its unnatural counterpart, that the two must be equally formally simple, and that the phonetic circumstances that give an advantage to the natural pattern must be replicated in the lab. The third point is a significant contribution: while the channel bias account of naturalness states that stimulus clarity and presentation are crucial to yielding learnability difference, little focus is placed on controlling for these factors in the lab. The experimental design used in Chapter 6 – a between-subjects 2x2 factorial design in which participants learned one of two artificial languages in one of two registers – showed evidence that learnability of the unnatural pattern of final voicing was only suppressed in the *CASUAL* speech register, in which the key stimuli ended with an obstruent that had been heavily *devoiced*. The group trained on the unnatural pattern in the *CAREFUL* speech register, in contrast, were able to learn their pattern above chance and more successfully than the other three groups.

A strength of this design is that it lends itself well to naturalness effects in other phonological domains, so long as the natural pattern has a perceptual advantage over its unnatural counterpart. If the researcher has accurately grasped the nature of the perceptual advantage and adequately encoded it into the differences between the two registers, I hypothesize that a similar set of results will unfold, wherein the learnability

of the unnatural pattern is only suppressed in the CASUAL condition. Of course, there are other factors the researcher must consider as well, such as influences from participants' native phonetics (as seen in the previous experiment), phonology, or lexicon. Nevertheless, this design provides a solid basic format that builds in the opportunity to study channel bias in a laboratory setting.

The Coda Sonority Experiment is an effort to apply the design from the previous experiment to a new phonological domain. As discussed in Chapter 2, perception of stops is dependent, in part, on salience of the release burst (e.g., Winitz et al. 1972; Ohala 1992b; Wright 2004). In coda position, releases are weakened or absent (e.g., Ohala & Kawasaki 1984; Krakow 1999; Kochetov 2001), which may lead to more difficulty perceiving the presence of the coda. The two artificial patterns in this experiment allow syllables to be closed with a coda but place restrictions on which types of consonants can occupy coda position. The so-called SONORANT CODA pattern allows only nasal codas (e.g., [paŋ.ka]), while the so-called OBSTRUENT CODA pattern allows only stop codas (e.g., [pat.ka]). The two patterns were each recorded in two different registers, CAREFUL and CASUAL, using the same basic elicitation method as was used in the Final Devoicing Experiment. Notably, the speaker was instructed to place special focus on the coda consonant and to aspirate all stops while producing the CAREFUL register. In the CASUAL register, there was no such focus on the coda, the target word itself was also taken out of a focused position, and the speaker was instructed not to aspirate stop codas.

It was hypothesized that, without the presence of the stop burst, participants learning the OBSTRUENT CODA pattern in CASUAL speech would occasionally fail to hear the coda at all, thus misperceiving [CVC.CV] words as [CV.CV]. While the nasal codas were likely also phonetically weaker in the CASUAL condition than in CAREFUL, it was hypothesized that participants would not hear as many *deletion* errors in this condition (see Chapter 2.1.3 for a discussion). The group that learned the OBSTRUENT CODA pattern in CASUAL speech was predicted to have more trouble hearing whether or not a coda was present than any of the other three groups – and, by extension, have more trouble judging which types of codas were and were not permissible.

The four conditions with the number of subjects per condition are provided in (1).

- (1) a. The SONORANT CODA pattern, heard in CAREFUL speech. (n = 18)
- b. The SONORANT CODA pattern, heard in CASUAL speech. (n = 18)
- c. The OBSTRUENT CODA pattern, heard in CAREFUL speech. (n = 18)
- d. The OBSTRUENT CODA pattern, heard in CASUAL speech. (n = 17)

## 7.2 Methodology

### 7.2.1 Participants

71 undergraduate students at the University of California, Santa Cruz, participated in this experiment, with 17 participants in the OBSTRUENT:CASUAL condition and 18 participants in the other three conditions. The ages of the participants ranged from 18 to 26 years old (mean = 20.23 years). Using the same intake form described in Chapter 5.1.1, 60 of the participants were coded as native speakers of English, and 5 were coded as non-native speakers. Five participants did not report the age at which they began speaking English but reported an average combined ‘speaking’ and ‘understanding’ score of at least 3.5 out of 4, suggesting high, if not native, proficiency in English. One participant did not report English as a language that they spoke, most likely due to an error in interpreting the question on the intake form. Participants received either course credit or \$10 cash as compensation for their help.

### 7.2.2 Stimuli

#### 7.2.2.1 Design

The stimuli of this experiment consisted entirely of bisyllabic nonce words. In the experimental stimuli, the first syllable was closed with a single coda, and the second syllable was open (CVC.CV). Filler words contained only open syllables (CV.CV). All words were produced with stress on the initial syllable.

The phonemic inventory of the stimuli contained three stops [p t k], three nasals [m n ŋ], and four vowels [i u o ə]. The set of consonants were chosen so that the inventory of the artificial patterns would contain one obstruent and one sonorant consonant in each major place of articulation (labial, coronal, and dorsal).

The two levels of the PATTERN condition differed most notably in which consonants were allowed to occupy the coda position. In both the SONORANT CODA and the OBSTRUENT CODA patterns, all six consonants were permitted in onset position and could appear as the onset of either the first or the second syllable. The SONORANT CODA pattern, however, permitted only the nasals in coda position, thereby allowing words like [piŋ.ka] but not \*[pik.ta]. The OBSTRUENT CODA pattern permitted only the stops in coda position, allowing [pik.ta] but not \*[piŋ.ka]. Each coda appeared in four distinct words in the Exposure phase.

Additional phonotactic restrictions were placed on both of the artificial patterns. The onset of the second syllable was restricted to the stops [p t k]. Nasal codas always assimilated in place to the following onset ([piŋ.ka], \*[pim.ka]), so as to resemble the strong tendency in English. Stop codas never assimilated in place, as doing so would have created a geminate and may have interfered with participants' ability to hear a coda ([pik.ta], \*[pik.ka]). The onset of the first syllable never shared a place of articulation with any of the other consonants in the word. The velar nasal [ŋ] was never permitted in onset position; participants in the OBSTRUENT CODA condition therefore never heard this sound in the Exposure phase.<sup>1</sup> The vowels of the first and second syllable were never identical. The stimuli were balanced for vowel pair. Table 7.1 contains the entire set of stimuli presented to participants in the Exposure phase. The leftmost column contains the Exposure words from the SONORANT CODA pattern, while the second leftmost column contains the Exposure words from the OBSTRUENT CODA pattern.

---

<sup>1</sup>Originally, I was concerned that the OBSTRUENT CODA group would reject words containing the velar nasal in the Novel phase simply on the basis of not hearing this segment in the Exposure phase and not on the basis of having learned the pattern. An earlier version of this experiment therefore permitted [ŋ] in onset position in all conditions. This version also contained a blanket rule that prohibited all codas from place assimilating to the following onset, as a means of avoiding geminates and of keeping all experimental stimuli as uniform as possible. The results of this experiment, however, did not turn out as expected. Many participants across all conditions reported difficulty with the stimuli and incorrectly described the pattern they had learned on the debriefing questionnaire as simply "containing consonants and vowels". It is possible, then, that the decidedly un-English phonotactics of the original stimuli caused confusion in the Exposure phase and led them to complete the Test phase task using a different set of acceptance criteria than expected, or no coherent set. The experiment reported here addresses these concerns by designing the stimuli to resemble English phonotactics.

The same fillers were used in both patterns.

<b>Exposure words</b>	<b>Exposure words</b>	<b>Fillers</b>	
<b>Sonorant Coda</b>	<b>Obstruent Coda</b>		
kampi	kapti	patu	mapi
pinto	kitpo	pita	mipo
toŋku	poktu	poki	nopu
tumpa	kupta	puko	nupo
kanto	patko	tapo	matu
piŋku	piktu	tipo	mitu
nompa	nopta	toka	nota
munti	mutpi	tuka	nuti
maŋku	maktu	kapi	mako
nimpa	nipta	kipu	mika
monti	motki	kotu	noki
nuŋko	nukto	kuti	nuka

Table 7.1: Stimuli presented in the Exposure phase

72 nonce words were presented as stimuli in the Novel phase. This set consisted of 18 words containing nasal codas, 18 words containing stop codas, and 36 fillers (with no codas). Across the set of novel words, each of the six consonants appeared 30 times, for a total of 90 stops and 90 nasals. The set of Novel phase stimuli, provided in Table 7.2, was presented to all participants in all four conditions.

Stimuli for this experiment were intended to not resemble any English words; however, some participants transcribed certain stimuli as real words of English in the Identification phase (most commonly, “monkey” for [manki] or [muŋki], “me too” or “metoo” for [mitu], and “cup tea” or “cuptea” for [kupti]).

### 7.2.2.2 Recording

As in the Final Devoicing Experiment (Chapter 6), the stimuli described above were recorded under two different sets of instructions, yielding the two levels of the CLARITY

Novel words Sonorant codas only	Novel words Obstruent codas only	Fillers
kompu	koptu	pato mopu
ponta	kotpa	puti mupa
toŋki	pokti	pito mapu
timpa	kipta	poku napi
kintu	pitku	pika nipu
piŋko	pikto	puka nopi
kumpi	kupti	tapi mita
punta	kutpa	topa mota
tuŋka	pukto	topi muto
nampo	napto	taku natu
mantu	matku	tuko notu
naŋki	makti	tiku nuta
nimpo	nipto	kapo maki
montu	motpu	kipa miku
maŋki	nakti	kupi moki
nompu	noptu	kito nako
minto	mitko	kota niko
muŋki	nukti	kuto nuki

Table 7.2: Stimuli presented in the Novel phase in all conditions

condition. Half of the participants were exposed to their assigned pattern in what is referred to here as CAREFUL speech and the other half in what is referred to as CASUAL speech. All participants heard the Novel phase stimuli in CAREFUL speech, regardless of what they had heard during the Exposure phase.

The stimuli were produced by one 26-year-old male speaker of American English, a trained linguist. The 45-minute recording session took place in the Phonetics lab at UC Santa Cruz under the same recording conditions described in Chapter 6.1.2.2.

The differences between CAREFUL and CASUAL speech were elicited through the use of two carrier phrases and two sets of instruction. The two carrier phrases chosen



were identical to those used in the Final Devoicing Experiment. Examples of the carrier phrases are presented in (2)–(3). The nonce words that appear in boxes represent the target stimuli that were extracted and used as part of the experiment, and the bolded words are in phonologically contrastive focus. In the CAREFUL speech carrier phrases, obstruents in coda position were transcribed with a superscript <sup>h</sup> to indicate aspiration.

(2) *Carrier phrase for the CAREFUL stimuli*

No, I didn't say **kabti**, I said kap<sup>h</sup>ti.

No, I didn't say **kanpi**, I said kamp<sup>i</sup>.<sup>2</sup>

(3) *Carrier phrase for the CASUAL stimuli*

No, I'm not gonna talk about **that** kap<sup>ʔ</sup>ti, I'm gonna talk about **this** kap<sup>ʔ</sup>ti.

No, I'm not gonna talk about **that** kampi, I'm gonna talk about **this** kamp<sup>i</sup>.

In addition to the carrier phrases, the speaker was given the same instructions for producing each of the CAREFUL and CASUAL speech stimuli described in Chapter 6.1. The speaker was also explicitly instructed to aspirate obstruent codas when producing CAREFUL speech stimuli and to produce unreleased obstruent codas in the CASUAL speech stimuli.

The set of CAREFUL speech phrases was recorded during the first part of the session, after which the speaker was encouraged to leave the booth, drink some water, and rest his voice. Afterwards, the set of CAREFUL speech phrases was recorded. The speaker was compensated \$10 for his participation.

The target stimuli were then extracted and normalized using the same methods described in Chapter 6.1.

### 7.2.3 Procedure

The procedure of this experiment is roughly identical to the procedure of the Final Devoicing Experiment (Chapter 6.1.3), with the task in the Identification phase as the

---

<sup>2</sup>The original experiment ended the carrier phrases with 'today', so as to put the target words in an utterance-medial position. Some participants reported hearing a consonant at the ends of some words, likely influenced by the formant transitions from the final vowel into [t]. The newer experiment places the target words in an utterance-final position to avoid the problem of formant transitions.

one exception. As with the Final Devoicing Experiment, participants completed the Exposure and Novel phases before moving onto the Identification task. In the ID phase of this experiment, however, participants were asked to type the word they heard using the keyboard, instead of completing a two-alternative forced choice task. This was possible in this experiment because, regardless of the particular spelling conventions that each participant used, whether or not participants heard a coda in the first syllable should be relatively unambiguous. In addition, participants in the ID phase of this experiment were asked to identify the words they heard from both the Exposure and Novel phases.

#### **7.2.4 Hypothesis**

The hypothesis of this experiment is very similar in form to that of the Final Devoicing Experiment. Again, the premise is that naturalness effects lie in the CASUAL register; the differences between the two CAREFUL groups should be smaller (or non-existent).

- (4) Naturalness effects observed in the lab arise only when there is a difference of perceptibility between the natural and the unnatural stimuli. It is predicted that participants who learn the OBSTRUENT CODA pattern in the CASUAL register will perform significantly worse than participants who learn the SONORANT CODA pattern in the CASUAL register, as well as the participants who learn the OBSTRUENT CODA pattern in the CAREFUL register.

### **7.3 Results**

#### **7.3.1 Novel phase**

Performance in the Novel phase is determined by the average proportion of correct Novel trials in each of the four groups. For the SONORANT CODA groups, a correct trial entails responding ‘1’ (does belong) to words containing sonorant codas and ‘5’ (does not belong) to words containing obstruent codas. For the OBSTRUENT CODA groups, a correct trial entails pressing ‘1’ for words containing obstruent codas and ‘5’ for words containing sonorant codas. All the fillers were grammatical in both of the patterns and

have been excluded from the final analysis.

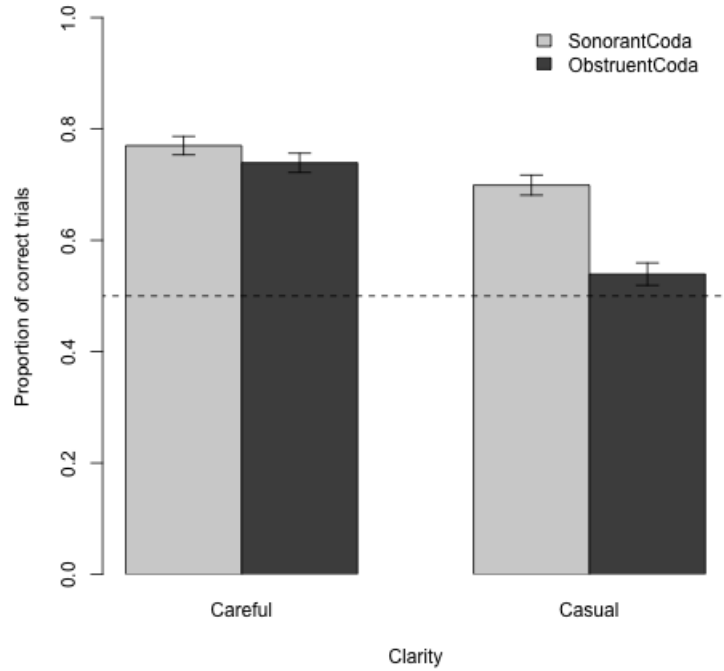


Figure 7.1: Performance in the Novel phase by PATTERN and CLARITY

Figure 7.1 shows the average performance in the Novel phase in each of the four groups (with the dashed line at  $y=0.5$  representing chance), while Figure 7.2 shows the average d prime in each group. One immediately clear finding is that the OBSTRUENT CODA: CASUAL group seems to be performing much less successfully than the other three groups. Both their average proportion of correct trials and their average d prime is numerically close to chance, while the other three groups appear to be performing above chance. The SONORANT CODA: CAREFUL group performed better than OBSTRUENT CODA: CAREFUL group and, by a slightly larger margin, than the SONORANT CODA: CASUAL group. Overall, though, the differences between these three groups are small, while the OBSTRUENT CODA: CASUAL group stands out in their low performance. Table 7.3 summarizes the average accuracy (proportion correct) and d prime for each of the groups.

Each subject's average performance on the Novel phase were fit to a linear regression model that contained two fixed effects, PATTERN and CLARITY. In the model printed

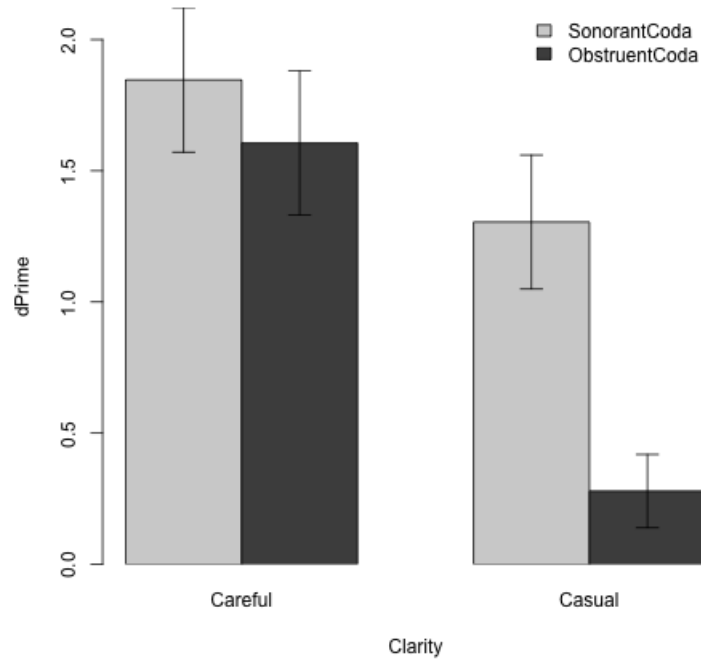


Figure 7.2: d prime in the Novel phase by PATTERN and CLARITY

Pattern	Clarity	Accuracy	d prime
SONORANT CODA	CAREFUL	0.770	1.846
OBSTRUENT CODA	CAREFUL	0.740	1.606
SONORANT CODA	CASUAL	0.699	1.305
OBSTRUENT CODA	CASUAL	0.540	0.279

Table 7.3: Average accuracy and d prime in the Novel phase

in Table 7.4, the group who learned the SONORANT CODA pattern in the CAREFUL register corresponds to the Intercept. The two main effects of PATTERN and CLARITY did not reach significance, but the interaction between PATTERN and CLARITY was marginally significant ( $p = 0.053$ ). A one sample t-test confirmed that the SONORANT CODA:CAREFUL group performed significantly above chance ( $\mu = 0.50$ ,  $t = 7.61$ ,  $df = 17$ ,  $p \approx 0$ ). An additional one-sample t-test found that the OBSTRUENT CODA:CASUAL group did not differ from chance ( $\mu = 0.50$ ,  $t = 1.88$ ,  $df = 16$ ,  $p = 0.08$ ).

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.770	0.032	23.726	≈0 ***
Casual	-0.071	0.046	-1.547	0.127
Obstruent	-0.031	0.046	-0.672	0.504
Casual:Obstruent	-0.129	0.065	-1.973	0.053 *

Table 7.4: Linear regression of Novel task (Intercept = SONORANT CODA:CAREFUL)

Two post hoc analyses of the simple effects were then completed, in which the effect of PATTERN was explored in each of the two levels of CLARITY. In Table 7.5, the model is restricted to the CAREFUL condition. Unsurprisingly, the main effect of PATTERN did not reach significance. From this, we may conclude that the two groups who learned their pattern in CAREFUL speech did not significantly differ from each other.

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.770	0.035	21.931	≈0 ***
Obstruent	-0.030	0.050	-0.622	0.538

Table 7.5: Post hoc: Effect of PATTERN in CAREFUL groups

Table 7.6 displays the effect of PATTERN in the CASUAL condition. Here, the main effect did reach significance: the SONORANT CODA:CASUAL group reached a significantly higher overall average than the OBSTRUENT CODA:CASUAL group ( $p < 0.001$ ).

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.699	0.029	23.72	≈0 ***
Obstruent	-0.160	0.042	-3.78	0.001 ***

Table 7.6: Post hoc: Effect of PATTERN in CASUAL groups

Together, these linear regression analyses depict exactly what was predicted: that the effect of PATTERN only mattered within the groups trained on CASUAL speech stimuli. Within this condition, the group who learned the natural SONORANT CODA pattern performed the Novel task more successfully than the OBSTRUENT CODA group.

In the CAREFUL condition, PATTERN had no effect: both groups were able to learn their pattern equally well.

### 7.3.2 Identification phase

Each trial in the Identification (ID) phase was coded as “correct” in only two situations: if one coda was reported in CVC.CV words, and if a coda was not reported in CV.CV words. This experiment was only concerned with perceiving the presence vs. absence of codas, and as such, deletion and insertion errors were recorded as “incorrect”. Identity errors, such as place of articulation errors (e.g., “kunpi” for [kumpi], “coctu” for [koptu]) and voicing errors (e.g., “mudpi” for [mutpi]), were noted and recorded as correct.<sup>3,4</sup> There were very few instances of a sonority identity error (e.g., “mogkey” for [maq̄ki], “monkey” for [motki]). These were excluded from analysis due to the importance of sonority in determining licit codas. No trial was ever excluded from analysis due to the identity of the vowels.

There were two types of insertion errors that indicated misperception of presence vs. absence of a coda. In one type, a vowel would be inserted after the coda, resulting in a word of shape CV.CV.CV (e.g. “motepoo” for [motpu]). This type of insertion error typically followed obstruent codas in the CAREFUL condition, wherein the hyperarticulated aspiration may have been heard as a heavily reduced vowel. In the other, a coda would be inserted in words of shape CV.CV (e.g., “monko” for [mako]).<sup>5</sup> Both of these types of errors were included in the analysis and marked as “incorrect”. Participants across all four conditions should have learned that the final syllable was always open,

---

<sup>3</sup>Due to English lacking an obvious letter for [ŋ], all instances of [ŋk] sequences transcribed “nk” were recorded as correct, with no identity error. All other identity errors were transcribed as such.

<sup>4</sup>Interestingly, a common reported identity error was one of dissimilation. All nasal codas present in the stimuli assimilated in place to the following onset, but many participants reported [mp] sequences as “np”. This may reflect a dissimilatory hypercorrection of the type discussed in Ohala (1981, 1993). It may also signify a preference for coronals, coronal codas, or [n] specifically (Hura, Lindblom & Diehl 1992 found similar results). Less common identity errors included [ŋk] sequences reported as “mk”, [mp] sequences reported as “ngk”.

<sup>5</sup>In rare cases, participants transcribed such words with a coda absent from the pattern’s phonemic inventory (e.g., “market” for [maki]). These trials were included in the analysis, so long as the coda was included as part of the first syllable and not the second.

and the first syllable could contain at most one coda consonant. Insertion errors that violated these rules (e.g., “cotop” for [kɔtɔ], “numpto” for [nukto]) were excluded from analysis. Exceptions to this were trials in which coda [k] was transcribed as “ck” (e.g., “puckto” for [pukto]), or in which coda [ŋ] was transcribed as “ng” (e.g., “tongki” for [tonki]). These are English spelling conventions, and so these trials were included and labeled “correct”.

Deletion errors overwhelmingly took only one form: deletion of the coda in the first syllable (e.g., “nupu” for [nompu]). These errors were crucial for the analysis and included as “incorrect”. Instances in which participants correctly transcribed the coda and second onset but failed to include the first onset (e.g., “aktu” for [maktu]), although technically a deletion error, were not coded as such because it was not a deletion error that affected perception of a coda; these trials were coded as “correct”. Trials in which the final vowel was deleted (e.g., “cap” for [kɔpɔ]) were excluded from analysis.

Rarely did participants transcribe a word with two simultaneous, adjacent consonants (e.g. “cuppa” for [kutpa]). It could not be determined whether these trials were heard with a singleton consonant or a geminate, and so they were excluded from analysis of the Identification phase. Trials in which no transcription was provided were also excluded from analysis.

The ID phase is intended to model how well participants perceived the pattern that they were originally taught in the Exposure phase. The analysis below is therefore limited to only those trials containing the Exposure words that each group heard in their assigned pattern and clarity level. The Novel words, which were all presented and then identified in CAREFUL speech, have been excluded. Fillers have also been excluded.

The results of the ID phase were analyzed similarly to the results of the Novel phase, in which accuracy is measured in proportion of correct trials. Table 7.7 summarizes performance on the ID phase in each of the four groups, including as well the average proportion of trials containing deletion errors and insertion errors. The percentages below the proportions of deletion and insertion errors indicates the percentage of total errors that were classified as either deletions or insertions. Figure 7.3 provides a visual representation of the groups’ performance in the ID phase, with the *y*-axis corresponding to proportion of correct trials.

Overall, performance was high, with three of the groups reaching performance close to 100% of correct trials. Again, however, the OBSTRUENT CODA:CASUAL group stands out. Although this group averaged at 70% correct trials, this average is still considerably lower than that of the other three groups. This means that, on an average of 30% of trials, the OBSTRUENT CODA:CASUAL group failed to report hearing a coda that had, in fact, been produced. There were two ways in which this could occur: deletion of the coda consonant (e.g., reporting “paka” for [pat̚kɑ]) or inserting a vowel after the coda consonant, resyllabifying it as an onset (e.g., reporting “pataka” for [pat̚kɑ]).<sup>6</sup> Across the board, however, insertion errors of this type were rare. Table 7.7 shows that 92.3% of the OBSTRUENT CODA:CASUAL group’s incorrect trials contained deletion errors, while only 7.7% contained insertion errors. This means that, overwhelmingly, failure to report a coda can be explained by failure to hear the presence of that segment.

Pattern	Clarity	Accuracy	Deletions	Insertions
SONORANT CODA	CAREFUL	0.995	0.005 (100% of all errors)	0.000 0%
OBSTRUENT CODA	CAREFUL	0.976	0.014 (60%)	0.009 (40%)
SONORANT CODA	CASUAL	0.956	0.037 (88.6%)	0.005 (11.4%)
OBSTRUENT CODA	CASUAL	0.716	0.262 (92.3%)	0.022 (7.7%)

Table 7.7: Average accuracy, deletion errors, and insertion errors in the ID phase

A linear regression based on each subject’s average ID performance was calculated, the results of which are printed in Table 7.8. The main effects of PATTERN and CLARITY did not reach significance, but the interaction did ( $p \approx 0$ ). A one-sample t-test confirms that the OBSTRUENT CODA:CASUAL group, which yielded the lowest average on the ID task, performed significantly better than chance ( $\mu = 0.50$ ,  $t = 5.02$ ,  $df = 16$ ,  $p \approx 0$ ).

Two post hoc analyses probed for the effects of PATTERN in each level of CLARITY.

<sup>6</sup>Since fillers have been excluded from analysis, this figure contains no instances of participants reporting a coda where none had been produced (e.g., reporting “patka” for [pɑkɑ]).



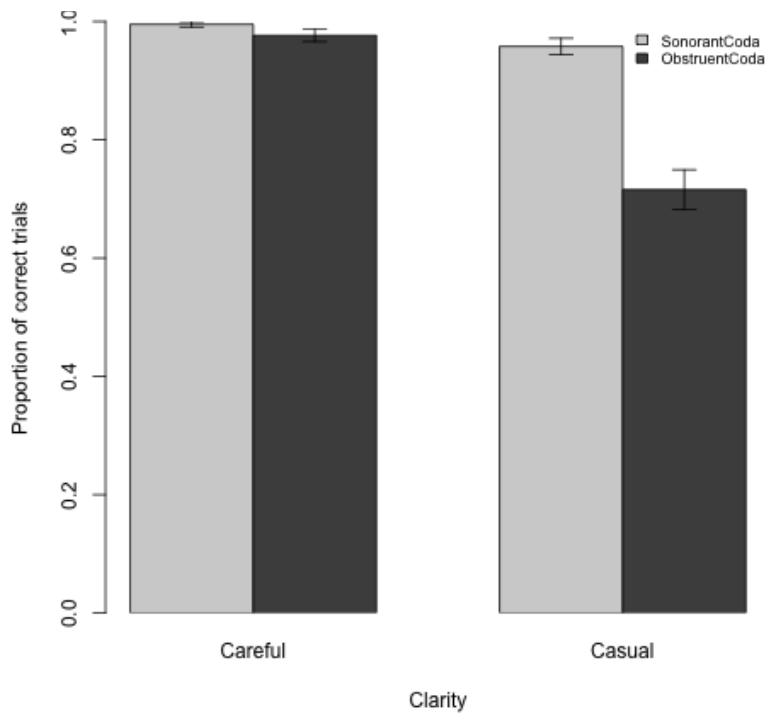


Figure 7.3: Performance in the ID phase by PATTERN and CLARITY

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.995	0.022	44.712	≈0 ***
Casual	-0.039	0.031	-1.243	0.218
Obstruent	-0.019	0.031	-0.588	0.558
Casual:Obstruent	-0.241	0.045	-5.372	≈0 ***

Table 7.8: Linear regression of ID task (Intercept = SONORANT CODA:CAREFUL)

Table 7.9 provides the results of the analysis based on the groups who learned their pattern in CAREFUL speech. The effect did not reach significance. In the analysis based on the groups trained on CASUAL speech, the effect of PATTERN did reach significance: the SONORANT CODA:CASUAL group performed this task better than the OBSTRUENT CODA:CASUAL group ( $p \approx 0$ ). This analysis is provided in Table 7.10. Again, these analyses combined show that PATTERN only has an effect in the CASUAL conditions.

Figure 7.4 depicts a bargraph of each group’s average proportion of trials containing

Effect	Estimate	St.E.	t value	$P$
(Intercept)	0.995	0.009	115.408	$\approx 0$ ***
Obstruent	-0.019	0.012	-1.518	0.138

Table 7.9: Post hoc: Effect of PATTERN in CAREFUL groups

Effect	Estimate	St.E.	t value	$P$
(Intercept)	0.956	0.031	31.36	$\approx 0$ ***
Obstruent	-0.260	0.044	-5.93	$\approx 0$ ***

Table 7.10: Post hoc: Effect of PATTERN in CASUAL groups

deletion errors (note that the  $y$ -axis is a scale from 0 to 0.5). Although the four groups overall performed the ID task quite well, with few errors of any kind, this figure shows that the OBSTRUENT CODA:CASUAL group produced more deletion errors than any of the other groups. An average of 26.2% trials that this group completed contained deletion errors, while the other three groups produced deletion errors in less than 4% of trials.

A linear regression computed from each subject's average proportion of deleted trials (based on SONORANT CODA:CAREFUL) confirms that these differences are significant. The model, printed in Table 7.11, found a significant interaction between PATTERN and CLARITY ( $p \approx 0$ ). The Intercept and the two main effects did not reach significance. A one-sample t-test showed that the OBSTRUENT CODA:CASUAL group yielded an average deletion error rate was significantly different from 0 ( $\mu = 0$ ,  $t = 7.75$ ,  $df = 16$ ,  $p \approx 0$ ).

Effect	Estimate	St.E.	t value	$P$
(Intercept)	0.005	0.021	0.223	0.824
Casual	0.035	0.029	1.176	0.244
Obstruent	0.009	0.029	0.316	0.753
Casual:Obstruent	0.234	0.042	5.603	$\approx 0$ ***

Table 7.11: Linear regression of deletions (Intercept = SONORANT CODA:CAREFUL)

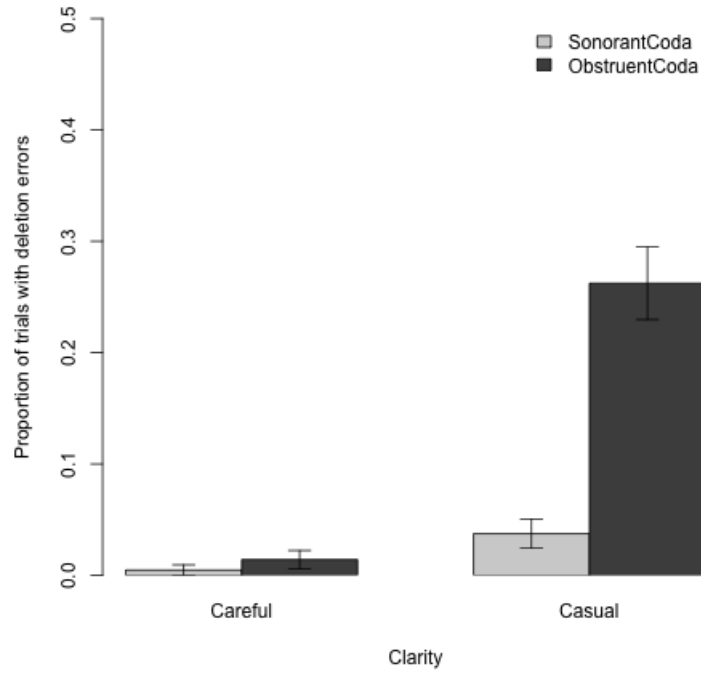


Figure 7.4: Proportion of ID trials containing deletion errors

The two post hoc analyses tell the same story as the analysis of overall ID performance. In the CAREFUL condition (Table 7.12), neither the Intercept nor PATTERN reached significance. In the CASUAL condition (Table 7.13), however, PATTERN did reach significance ( $p \approx 0$ ), confirming that the OBSTRUENT CODA: CASUAL group produced significantly more deletion errors than did the SONORANT CODA: CASUAL group.

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.005	0.006	0.741	0.464
Obstruent	0.009	0.009	1.047	0.302

Table 7.12: Post hoc: Effect of PATTERN in CAREFUL groups

### 7.3.3 Correlation between ID performance and Novel performance

The overall hypothesis of this experiment is that learnability of natural and unnatural patterns depends on perception, with the least perceptible pattern learned the least suc-

Effect	Estimate	St.E.	t value	<i>P</i>
(Intercept)	0.039	0.029	1.356	0.184
Obstruent	0.243	0.041	5.877	≈0 ***

Table 7.13: Post hoc: Effect of PATTERN in CASUAL groups

cessfully. To test this hypothesis, each individual subject’s average ID score (proportion of correct ID trials, limited only to the words heard in the Exposure phase) was plotted against their average Novel score. Figure 7.5 depicts this scatterplot, with each point corresponding to an individual subject across all conditions (perceived color differences are due to overlapping points). The points in this plot are pooled across all subjects in all conditions. Both deletion and insertion errors are included in the calculation of average ID score; excluding insertion errors, however, did not change the results in any meaningful way, since insertion errors were rare overall.

A Pearson’s product-moment correlation coefficient was calculated for the regression line. There was a positive correlation between performance on the ID phase and performance on the Novel phase ( $r = 0.444$ ,  $r^2 = 0.197$ ,  $p \approx 0$ ). Overall, increased ID performance correlated with increased Novel performance. The correlation itself was weak, however: many of the participants performed with 90-100% accuracy on the ID phase, but these participants performed as high as 100% and as low as 42% on the Novel phase.

19.7% of the variance in Novel performance could be explained by ID performance. Although this number may seem small, we have no *a priori* reason to believe it should be any higher. Perhaps only a small amount of misperception is needed to have large effects on pattern learnability. While we do not have enough information to quantify a minimum amount of misperception needed for decreases in pattern learning, what we can say right now is that this correlation is trending in the right direction. High ID scores do not necessarily mean high Novel scores, but *low* ID scores do seem to correspond to low Novel scores.

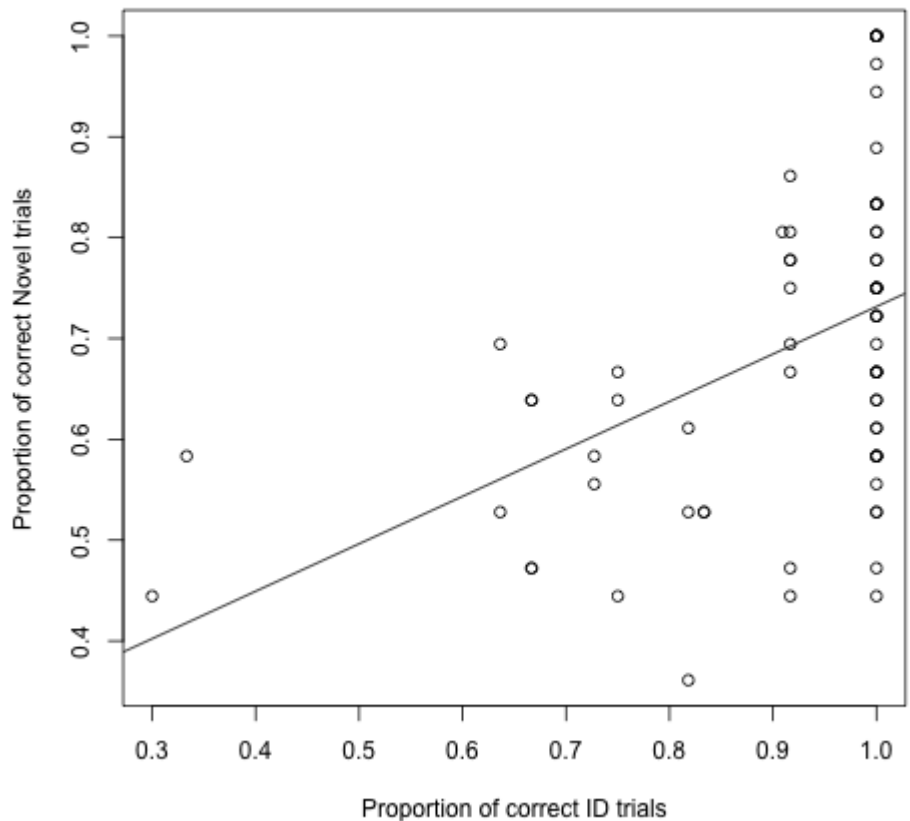


Figure 7.5: Correlation between subjects' ID performance and Novel performance

### 7.3.4 Individual subject variance

If learnability depended entirely on perceptibility (and nothing else), we would expect that, regardless of condition, participants who perceived their stimuli well would necessarily learn their pattern well, and vice versa. Of course, this is not what occurred: a subject's ability to learn a pattern depends in part, but not in full, on their perception of the pattern.

What is more surprising is that, at the between subject level, the relationship between ID performance and Novel performance is not consistent. As we have seen, when summarizing across all participants in each of the four groups, the condition that was perceived the least well (OBSTRUENT CODA:CASUAL) was also learned the least well. Within this group, however, some subjects exhibited relatively high perception

but low learning, while others who perceived less well learned more successfully.

Figures 7.6 and 7.7 contain bargraphs of the individual subject averages in the Novel phase (left) and in the ID phase. Figure 7.6 corresponds to the two CAREFUL groups, while Figure 7.7 corresponds to the two CASUAL groups; in both figures, the rows correspond to PATTERN and the columns correspond to phases of the experiment. The *x*-axis corresponds to individual subjects; the colors of the bars are meaningless except to aid in visual clarity.

Subjects in the two CAREFUL groups exhibited notably high ID averages. Of the 36 participants who learned their pattern in CAREFUL speech, only five performed below 100% accuracy, and each of these five participants scored over 80% accuracy. There was much more variability in the Novel phase, as tends to be the case with this phase of artificial grammar experiments. Both groups contained subjects who performed at ceiling (all trials were correct), and both contained subjects who performed at or around chance.

The picture of subject performance in the SONORANT CODA: CASUAL group does not look dissimilar from the other two groups discussed so far. Performance in the ID phase was remarkably high: 12 of the 18 subjects performed at 100% accuracy, and all but one performed higher than 80% accuracy. In the Novel phase, one subject performed at ceiling, with several more performing close to chance. It is the OBSTRUENT CODA: CASUAL group that stands out. Zero participants in this group performed at ceiling in either phase. Although there is considerable subject variation in all of the groups, this lack of perfect performance in this group may speak in favor of the overall hypothesis.

As mentioned, the hypothesis predicts that subjects who perform better in the ID phase will also perform better in the Novel phase (and, likewise, worse performance in ID should signal worse performance in Novel). This was expected to be the case regardless of condition, but did not uniformly hold true. For instance, Subject 328 received the highest ID score of all the OBSTRUENT CODA: CASUAL participants (91.67%). Their Novel score, however, was roughly at chance (47.22%), and was lower than many participants who performed relatively worse at the ID task. In sum, it appears that the subjects in the OBSTRUENT CODA: CASUAL group performed uniformly poorly, de-

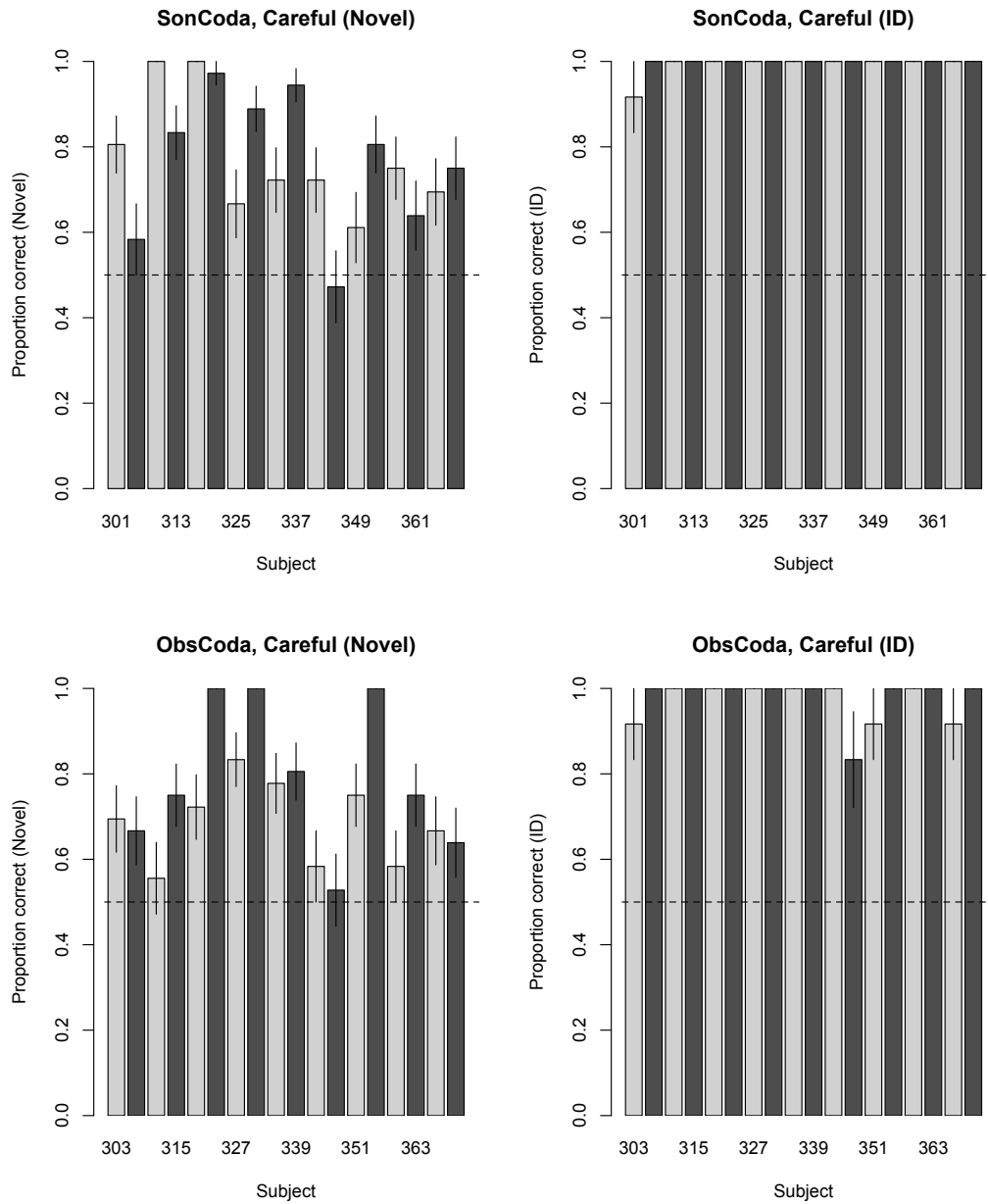


Figure 7.6: CAREFUL subject averages in the Novel phase (left) and the ID phase

spite how they performed in the ID phase. The overall group averages align with the hypothesis, but the individual subject averages do not.

The reason for this remains somewhat of a mystery. One possibility lies with the

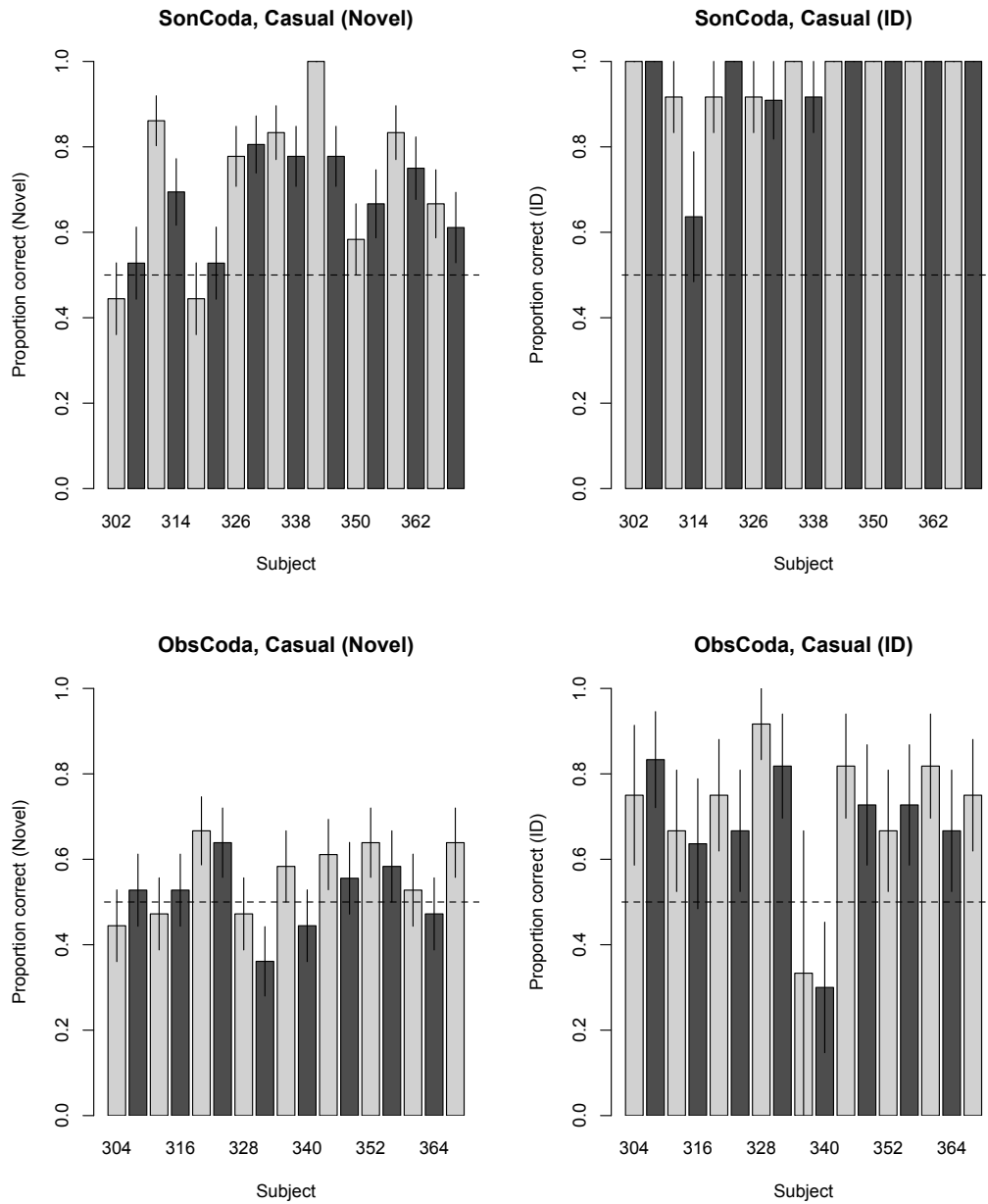


Figure 7.7: CASUAL subject averages in the Novel phase (left) and the ID phase

timecourse of the experiment. The ID phase was intended to give some measure of how well participants had perceived their Exposure stimuli during the Exposure phase, but it is possible that completing the Novel phase first informed how they perceived



their stimuli, such that they were able to perceive better in ID than in Exposure. For example, the OBSTRUENT CODA:CASUAL group spent the first ten minutes of their experiment listening to words like [kit<sup>h</sup>po], in which codas were not released. Before beginning the Novel phase, this group did not have a chance to hear any codas with hyperarticulated bursts, possibly leading them to believe that these words contained few or no codas at all. During the Novel phase, every participant heard the same set of words in CAREFUL speech, meaning that the OBSTRUENT CODA:CASUAL group was now almost unambiguously hearing codas in places they had not heard them before. This may have clued them in to the idea that this data set contains codas, and they need to start listening for them. Without having perceived codas in the Exposure phase, however, they were not certain whether words like [kompu] or words like [koptu] belonged to their language. The hyperarticulated stop codas in the Novel phase may have prompted them to attend to other cues for presence vs. absence of a stop (like formant transitions) during the ID phase. This group may have perceived the word [kit<sup>h</sup>po] as /kipo/ in the Exposure phase, but after completing the Novel phase, were able to perceive it as /kitpo/ in the ID phase. This scenario would explain why high performance in the ID phase did not correspond to high performance in the Novel phase.

This, of course, is only speculation, and there is no way to determine whether or not this scenario occurred given the design. Including a perception task during or directly after the Exposure phase could provide a better measure of how well participants perceived their stimuli *before* the Novel task, but it runs the risk of leading participants to attend to specific cues that might influence the way they perform the task. Another option is to run a followup perception experiment, as I did in the Stress Perception Experiment in Chapter 5, where participants are only asked to perceive words but not to learn a pattern. This option does not come with the risks of influencing how participants perform the Novel task, but it eliminates the possibility of directly comparing individual performance on Novel and ID, since the latter set of participants would not have completed the Novel task. Future researchers interested in using this experimental template should search for a way to measure how participants *actually* perceived their Exposure words during the Exposure phase without the task itself greatly affecting how they perform in the Novel task.

## 7.4 Discussion

This experiment aimed to determine whether the experimental design proposed and run in Chapter 6 could be extended to a different phonological domain. This endeavor was successful. The results of this experiment were in line with the predictions from the channel bias account. Learnability of the two patterns only differed in the CASUAL speech condition, where the ID phase also saw a difference in perception of the two patterns. In the CAREFUL speech condition, both patterns were perceived equally well, and both were learned equally well also.

In the Final Devoicing Experiment (Chapter 6), the results suggested that influence from English may have been active in shaping how participants performed. It may be the case that English influenced the results of the Coda Sonority Experiment, as well. If English contains more instances of [m n ŋ] in coda position than [p t k], this may lead participants to prefer stimuli containing [m n ŋ] codas, regardless of what they had heard in the Exposure phase. This would result in high performance of the two SONORANT CODA groups and low performance of the two OBSTRUENT CODA groups. This interpretation, however, could not explain why the difference in PATTERN only occurred in the CASUAL condition. If a lexical bias were active in this way, its effects should have held in both speech registers.

To investigate the possible effects of lexical bias, I calculated the token frequency of [m n ŋ p t k] codas in the WebCelex lexical database each time they appeared in a VC.CV sequence (resembling the items from this experiment).<sup>7</sup> This count yielded more [m] codas than [p] (3,450 [m]; 1,688 [p]), more [n] codas than [t] (16,915 [n]; 2,470 [t]), but more [k] codas than [ŋ] (4,694 [k]; 2,455 [ŋ]). If lexical bias is one of the primary factors determining how participants perform, then we may expect the two OBSTRUENT CODA groups to perform better than the two SONORANT CODA groups if we subset the Novel response scores to only those trials containing [ŋ] or [k] codas. Doing so raises the average scores of the OBSTRUENT CODA groups a little (75.9% correct for OBSTRUENT CODA:CAREFUL; 55.4% correct for OBSTRUENT CODA:CASUAL) and lowers the SONORANT CODA scores a little (73.1% correct for SONORANT CODA:CAREFUL; 67.1% correct for SONORANT CODA:CASUAL). These changes are enough to lose the

---

<sup>7</sup><http://celex.mpi.nl>.

significance of the interaction between PATTERN and CLARITY ( $p = 0.098$ ). Even so, the numerical trends are similar to the overall results: in the CASUAL condition, the SONORANT CODA group still trends towards a higher average of correct trials than the OBSTRUENT CODA group. A one sample t-test confirms that, even when restricting the trials to dorsals only, the OBSTRUENT CODA:CASUAL group did not perform significantly differently from chance ( $t = 1.26$ ,  $df = 16$ ,  $p = 0.226$ ). The greater frequency of [k] codas than [ŋ] codas may have had some impact on performance, but not enough to switch which pattern was learned more successfully.

This design continues to speak to why naturalness effects in typology fall out the way they do. Cross-linguistically, the overrepresentation of sonorant codas with respect to obstruent codas (and, in particular, stops) raises the question of whether this is due to an ingrained preference for sonorant codas. If this were so, however, we would expect this preference to show up in a variety of circumstances, even when both sonorant and obstruent codas can be perceived equally well. The performance of the two CAREFUL groups suggest that this is not the case. In most facets of everyday life, humans do not speak in the type of hyperarticulated, slow, citation speech that was elicited in the CAREFUL speech stimuli. Real human speech much more closely resembles the CASUAL condition, in which the natural and unnatural patterns yielded a difference in perceptibility *and* a difference in learnability. In other words, what may look like an innate learnability difference between the two is actually a perceptibility difference, and the asymmetry observed in typology holds if humans speak in such a way that continuously yields this perceptibility difference. Thus, the effects from typology are mirrored in the lab in the CASUAL speech condition, while the performance of the groups trained on CAREFUL speech provide further evidence that these observed naturalness effects are based in perception, not in universal learning bias.

## Chapter 8

### Conclusion

This dissertation makes strides toward unveiling the root causes of naturalness-based tendencies in both typology and in the lab. The overarching goal of this dissertation was to investigate in closer detail one such account: that perception and production are the driving forces – and the *only* driving forces – behind these effects. This issue was addressed through an experimental framework, in which a series of artificial grammar experiments probed for the connection between perceptibility of the stimuli that make up a pattern and how successfully that pattern was learned. Compared to an account in which unnatural patterns are more difficult to learn, I argue that the channel bias account is not only less stipulative but is also supported by the experiments in a way that the other is not.

One of the central arguments of this dissertation is that the channel bias account and studies of naturalness in the lab are in somewhat of a symbiotic relationship. On the one hand, experimental research of this nature provides evidence in favor of the channel bias account of naturalness. All three experiments observed that when the stimuli of the unnatural pattern were improperly perceived, the pattern as a whole was learned less successfully. Chapter 5 found that this relationship between perceptibility and learnability held for an unnatural pattern of weight-sensitive stress, but did not hold for a pattern that was formally complex (but phonetically natural). This pair of experiments served as evidence that perhaps naturalness is a different entity than simplicity, the latter of which is arguably a matter of universal learning bias. In Chapter 6, I proposed a 2x2 experimental design that is intended to serve as a template for any

perception-based naturalness domain, which takes voice clarity into account from the outset. The Final Devoicing Experiment found that the unnatural pattern was learned in both conditions, but was learned significantly more successfully in the CAREFUL speech condition, in which final obstruents were fully or mostly voiced throughout the closure. Voice clarity did not have an effect on learnability of the natural pattern. Chapter 7 observed that this experimental design also functioned in the domain of coda sonority: the unnatural pattern was not learned in the CASUAL condition, but was learned significantly above chance and no better or worse than the natural condition in CAREFUL speech.

Thus, across all three experiments, we see that perceptibility of stimuli matters, and it disproportionately affects how well unnatural vs. natural patterns can be learned. The learnability of unnatural patterns depends in a large way on phonetic factors, but the same was not observed for natural patterns, which tended to be learned without trouble in all conditions (the exception to this being the hypothesized native language effects discussed in Chapter 6). This aligns satisfyingly well with the channel bias account of naturalness effects in *typology*. Under this approach, the sole reason for the lack or scarcity of unnatural patterns in typology is that they are more difficult to perceive and/or produce than natural patterns, and are therefore more likely to morph into forms that align with phonetic tendencies.

Just as experimental research can provide support for a channel bias account of naturalness, so can the channel bias account provide an explanation for the pre-existing problems regarding naturalness in the lab. As detailed in Chapter 4, many researchers have attempted to recreate naturalness effects in the lab, but only some have been successful. From a channel bias perspective, this result is not surprising. Naturalness effects in the lab should only be observed if the perception and/or production differences between the two patterns have been adequately recreated. This does not mean that experiments that do *not* include, for instance, a CASUAL speech condition will always fail to find naturalness effects. The stress experiments in Chapter 5 did not include such a condition, and yet the natural pattern was still learned more successfully than the unnatural pattern. Even so, participants still found the stimuli of the unnatural pattern less easy to perceive, and so the relationship between perceptibility and learnability still

stands. An active effort to recreate the phonetic differences “from the wild” should only increase the likelihood that naturalness effects will be observed. This dissertation does not claim that every previous experiment that failed to find these effects did not adequately control for naturalness, but rather, that there needs to be a larger collective focus on how stimuli are designed and presented in order to do these effects proper justice.

In contrast, the universal learning bias neither holds up under experimental conditions nor explains the discrepancies between previous experiments. Chapter 5 saw the beginnings of this, where an Occam’s razor argument favored the channel bias account: since perception did prove to be a factor, any additional argument based abstractly in cognition is arguably stipulative. The bulk of this argument, however, came in Chapters 6 and 7, where the unnatural pattern was learned just as well as, or better than, the natural pattern in CAREFUL speech. This finding is crucial to the overall argument. While it is not surprising that participants would struggle to learn a pattern in cases of lower perceptibility, great success of learning a pattern in cases of high perceptibility is remarkable. This is problematic for the universal learning bias account, which posits an inherent learning bias toward natural patterns regardless of perceptibility of stimuli. If phonetic factors are not behind unnatural pattern learning, then it should not matter if the stimuli are remarkably clear (as in the CAREFUL speech conditions) or not adequately controlled for (as, perhaps, in previous naturalness experiments). The very findings that are problematic for the universal learning bias account are easily explained by channel bias.

Of course, this does not mean that a universal learning bias approach has no place in phonology. The stress experiments in Chapter 5, which looked both at effects of naturalness and simplicity, found that a complex stress pattern was not learned even though the stimuli of this pattern were perceived well. With no evidence of channel bias suppressing performance on this pattern, I argue that simplicity is probably an issue of universal learning bias: learners struggle more with formally complex patterns than simpler ones (e.g., Moreton 2008, 2012; Pertsova 2012; Moreton et al. 2015). These two approaches together very closely resemble Moreton & Pater’s (2012) “structurally biased phonology”, in which both approaches are needed to explain asymmetries in

phonological typology (channel bias for naturalness effects and learning bias for simplicity effects).

If, indeed, there are no learning biases against unnatural patterns, it is worth asking why some experiments do find that a natural pattern was learned more successfully than an unnatural pattern (e.g., Wilson 2006; Carpenter 2010; Myers & Padgett 2014). This, I believe, can also be explained by the channel bias account. If phonetic correlates are not controlled for in the lab, it is still possible that they may arise accidentally – that is, the two sets of stimuli may not be equally perceptible, but for accidental reasons as opposed to a conscious effort on the part of the researcher. In the absence of controlling for stimulus perceptibility, the channel bias approach to naturalness does not predict *no* learning differences: it predicts inconsistencies between experiments. Without including perceptibility data, however, it is impossible to know whether or not experiments yielded accidental channel bias in the lab.

Myers & Padgett (2014) addressed this by following up their first experiment with perceptibility data. A new set of participants (who had not taken part in the first experiment) listened to all the obstruent-final stimuli from the experiment and identified whether the final consonant was [s] or [z]. The results trended in the direction of more correct identification of [s], but this difference was not significant. Given how successfully the participants in my Final Devoicing Experiment identified [s]-final vs. [z]-final words, and how similar the procedures of these two perception studies were, one possible interpretation is that the perception test was too easy and did not adequately capture any channel biases that may have occurred in the lab.

Wilson (2006) studied asymmetries in velar palatalization using an artificial grammar experiment with a poverty of the stimulus design. Participants were trained on a pattern in which velars were fronted before front vowels. Half the participants were explicitly trained on stimulus pairs like [ki...tʃi], with palatalization before high front vowels, and [ka...ka], with no palatalization. These participants were crucially not exposed to any stimuli containing [e]. The other half of the participants were trained on [ke...tʃe] and [ka...ka], but were not exposed to any stimuli containing [i]. In the Novel phase, participants trained on [ke...tʃe] extended the pattern to [ki...tʃi], but participants trained on [ki...tʃi] did not extend velarization to other vowel contexts. These

results were argued to be in line with typological implicational laws and interpreted as supporting a substantively biased view to naturalness.

While this argument conflicts with the arguments made in this dissertation, they are not necessarily conflicting. The first experiment in Wilson (2006) was concerned with extending the pattern to novel environments, which this dissertation did not investigate. It could be the case that learning biases based on naturalness play a role in extending a pattern to novel environments, but they do not prevent learners from acquiring unnatural patterns. If markedness constraints exist within the mind of the learner, they need not be innate nor universal, and they are not necessary to explaining the typological facts about naturalness. In addition, as Moreton & Pater (2012) point out, the results from Wilson (2006) are not perfectly in line with typology. The [ke...tʃe] group did not distinguish between vowel quality in the Novel phase, extending their pattern not only to [ki...tʃi] (which they had not been trained on), but to [kɑ...tʃɑ] (which they had explicitly taught not to palatalize). This group also palatalized [gV...dʒV] more than [kV...tʃV] across all three vowels. This is problematic for a learning bias interpretation of their results, as the implicational laws the author cites state that [k]-palatalization implies [g]-palatalization, but not the other way around.

## 8.1 Implications

The findings of this dissertation have clear implications for future experimental work: when studying naturalness in the lab, the hypothesized phonetic correlates behind the patterns in question must be accurately represented. This includes, but is not limited to, presenting stimuli in registers other than careful, hyperarticulated speech and choosing production-based tasks for production-based patterns. I have provided three components that I argue are necessary for a sufficient naturalness experiment, as well as a template that achieves all three and that can, in theory, be modified to work for any naturalness domain. I hope that the accomplishments of this dissertation can serve as groundwork for future naturalness experiments.

In addition, this dissertation takes steps toward implications for a theory of markedness. The experiments saw no evidence that naturalness-based pattern learning is affected by ingrained learning biases against unnatural patterns. This, in turn, challenges



the notion of innate markedness constraints that explain naturalness typology. The typological tendencies that are often explained by markedness (as a cognitive principle) are explained just as well, if not better, by the forces of perception and production on sound change. This dissertation does not claim to be the final word on markedness, but if such ingrained biases against unnatural patterns do exist, it is suspicious that no evidence for them was seen here. Much of formal phonology treats markedness as an innate entity. If this turns out to be amiss, we may need to consider altering how we represent phonological patterns to align more closely with reality. A version of Optimality Theory that takes into account the components of sound change is one example of this (Boersma 1998).

In the pursuit of understanding naturalness, this dissertation has observed one type of phonological domain (natural vs. unnatural patterns whose asymmetries are rooted in perception) using one type of experimental framework (artificial grammar, in which participants are given the full pattern in the Exposure phase). Excluded from this are naturalness effects based in production: that is, where the two patterns differ in production ease but not necessarily in perception. Vowel harmony and disharmony may be one example of this type of pattern. I have suggested a shadowing task as one method of investigating patterns of harmony, and I expect that the desired effects would only occur if the speakers are instructed not to use careful speech, but I leave the particulars to future researchers.

Finally, while only one type of artificial grammar experiment was utilized for the purposes of this dissertation, other designs may weigh in on this question in important ways. One type that comes to mind is the poverty of the stimulus design, in which participants must generalize their pattern not only to novel words, but to novel phonological environments they had not previously been exposed to in the experiment (Berko 1958; Wilson 2006; Finley & Badecker 2007; Becker et al. 2011; Bennett 2012; Myers & Padgett 2014). In these experiments, generalizing to certain environments and not to others tends to be explained as a learning bias, as the participants show a preference towards certain forms that they were not explicitly trained on. It is possible that phonetic knowledge of some sort may not shape the type of pattern learning explored in this dissertation, but it may shape how we make decisions about whether patterns

can be extended to novel segments, environments, structures, etc. The driving forces of naturalness should continue to be researched so that we may gain a finer understanding of the roles and limits of each type of bias in phonology.

# Bibliography

- Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.
- Baayen, R. H. 2009. *languageR: Data sets and functions with “Analyzing Linguistic Data: A practical introduction to statistics”*.  
<http://CRAN.R-project.org/package=languageR>. R package version 0.955.
- Baer-Henney, Dinah & Ruben van de Vijver. 2012. On the role of substance, locality and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology* 3(2). 221–249.
- Baertsch, Karen. 2002. *An Optimality-theoretic approach to syllable structure: The Split Margin Hierarchy*. Bloomington: Indiana University dissertation.
- Baertsch, Karen & Stuart Davis. 2003. The split margin approach to syllable structure. *ZAS Papers in Linguistics* 32. 1–14.
- Becker, Michael, Nihan Ketrez & Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1). 84–125.
- Beckman, Jill. 1997. Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology* 14(1). 1–46.
- Beckman, Mary. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.
- Benkí, José R. 2003. Analysis of English nonsense syllable recognition in noise. *Phonetica* 60(2). 129–157.
- Bennett, Ryan. 2012. *Foot-conditioned phonotactics and prosodic constituency*: University of California, Santa Cruz Doctoral dissertation.
- Berko, Jean. 1958. The child’s learning of English morphology. *Word* 14(2-3). 150–177.
- Blevins, Juliette. 1997. Rules in Optimality Theory: Two case studies. In Iggy Roca (ed.), *Derivations and constraints in phonology*, 227–260. Oxford: Clarendon Press.

- Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.
- Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32(2). 117–166.
- Blevins, Juliette & Andrew Garrett. 1998. The origins of consonant-vowel metathesis. *Language* 508–556.
- Blevins, Juliette & Andrew Garrett. 2004. The evolution of metathesis. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically-Based Phonology*, 117–156. Cambridge: Cambridge University Press.
- Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.
- Boersma, Paul & David Weenink. 2014. *Praat: doing phonetics by computer*. <http://www.praat.org>. Computer program.
- Brodeur, Mathieu B., Emmanuelle Dionne-Dostie, Tina Montreuil & Martin Lepage. 2010. The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS one* 5(5). e10773.
- Broselow, Ellen, Su-I Chen & Chilin Wang. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20. 261–280.
- Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* 341–376.
- Browman, Catherine P. & Louis Goldstein. 1995. Gestural syllable position effects in American English. In Fredericka Bell-Berti & J. Lawrence Raphael (eds.), *Producing Speech: Contemporary Issues (for Katherine Safford Harris)*, 19–33. Woodbury, NY: American Institute of Physics Press.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Cardona, George. 1965. *Gujarati reference grammar*. Philadelphia: University of Philadelphia Press.
- Carpenter, Angela. 2010. A naturalness bias in learning stress. *Phonology* 27. 345–392.
- Catford, J.C. 1974. ‘Natural’ sound-changes: some questions of directionality in diachronic phonetics. In Anthony Bruck, Robert A. Fox & Michael W. La Galy (eds.), *Papers from the Parasession on Natural Phonology*, 21–29. Chicago: Chicago Linguistic Society.

- Chaski, Carole E. 1986. Linear and metrical analyses of Manam stress. *Oceanic Linguistics* 25(1/2). 167–209.
- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary Beckman (eds.), *Papers in Laboratory Phonology 1: Between the grammar and physics of speech*, 283–333. New York: Cambridge University Press.
- Cristià, Alejandrina & Amanda Seidl. 2008. Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4(3). 203–227.
- Crowhurst, Megan. 2001. Coda conditions and *um* infixation in Toba Batak. *Lingua* 111(8). 561–590.
- Delgutte, Bertrand. 1982. Some correlates of phonetic distinctions at the level of the auditory nerve. *The representation of speech in the peripheral auditory system* 131–149.
- Dell, Francois & Mohamed Elmedlaoui. 1985. Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7. 105–130.
- DiCanio, Christian T. 2012. The phonetics of fortis and lenis consonants in Itunyoso Trique. *International Journal of American Linguistics* 78(2). 239–272.
- Fant, Gunnar, Anita Kruckenberg & Johan Liljencrants. 2000. The source-filter frame of prominence. *Phonetica* 57(2-4). 113–127.
- Finley, Sara & William Badecker. 2007. Towards a substantively biased theory of learning. In *Annual Meeting of the Berkeley Linguistics Society*, vol. 33 1, 142–153.
- Flack, Kathryn. 2009. Constraints on onsets and codas of words and phrases. *Phonology* 26. 269–302.
- Flemming, Edward. 1995. *Auditory Representations in Phonology*. Los Angeles: UCLA dissertation.
- Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7–44.
- Foley, William A. 1991. *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1). 110.

- Garrett, Andrew & Keith Johnson. 2012. Phonetic bias in sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization*, 51–97. Oxford: Oxford University Press.
- Giles, Stephen B. & Kenneth L. Moll. 1975. Cinefluorographic study of selected allophones of English /i/. *Phonetica* 31(3-4). 206–227.
- Goedemans, Rob & Harry van der Hulst. 2013. Weight factors in weight-sensitive stress systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/16>.
- Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251.
- Gordon, Matthew. 2002. A phonetically driven account of syllable weight. *Language* 78(1). 51–80.
- Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23(3). 595–653.
- Gouskova, Maria. 2004. Relational hierarchies in Optimality Theory: The case of syllable contact. *Phonology* 21(2). 201–250.
- Greenberg, Steven, Hannah Carvey, Leah Hitchcock & Shuangyu Chang. 2003a. The phonetic patterning of spontaneous American English discourse. In *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, .
- Greenberg, Steven, Hannah Carvey, Leah Hitchcock & Shuangyu Chang. 2003b. Temporal properties of spontaneous speech – a syllable-centric perspective. *Journal of Phonetics* 31(3). 465–485.
- Grosjean, François. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36(1). 3–15.
- Hale, Mark & Charles Reiss. 1998. Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* 29. 656–683. Available on Rutgers Optimality Archive, <http://roa.rutgers.edu>.
- Hale, Mark & Charles Reiss. 2000. “Substance abuse” and “dysfunctionalism”: Current trends in phonology. *Linguistic Inquiry* 31. 157–169.
- Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: The University of Chicago Press.
- Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.). 2004. *Phonetically-Based Phonology*. Cambridge: Cambridge University Press.

- Hayes, Bruce, Péter Siptár, Kie Zuraw & Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85(4). 822–863.
- Hayes, Bruce & Donca Steriade. 2004. Introduction: The phonetic basis of phonological markedness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically-Based Phonology*, 1–32. Cambridge: Cambridge University Press.
- Hayes, Bruce & James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44(1). 45–75.
- Hillenbrand, James, Laura A. Getty, Michael J. Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97(5). 3099–3111.
- Hintzman, Douglas L. 1988. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review* 95(4). 528.
- Hume, Elizabeth & Keith Johnson (eds.). 2001. *The role of speech perception in phonology*. San Diego: Academic Press.
- Hura, Susan L., Björn Lindblom & Randy L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35(1-2). 59–72.
- Hyman, Larry. 1976. Phonologization. In Alphonse Juilland (ed.), *Linguistic studies presented to Joseph H. Greenberg*, 407–418. Saratoga, CA: Anna Libri.
- Hyman, Larry. 2001. The Limits of Phonetic Determinism in Phonology: \*NC Revisited. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, San Diego: Academic Press.
- Itô, Junko. 1986. *Syllable Theory in Prosodic Phonology*: University of Massachusetts, Amherst dissertation. Published 1988. Outstanding Dissertations in Linguistics series. New York: Garland.
- Ito, Junko & Armin Mester. 1994. Reflections on CodaCond and Alignment. In Jason Merchant, Jaye Padgett & Rachel Walker (eds.), *Phonology at Santa Cruz*, vol. 3, 27–46. Santa Cruz: Linguistics Research Center, UC Santa Cruz.
- Jaeger, J. J. 1978. Speech aerodynamics and phonological universals. In *Annual Meeting of the Berkeley Linguistics Society*, vol. 4, 312–329.
- Javkin, Hector R. 1977. *Phonetic universals and phonological change*: University of California, Berkeley Doctoral dissertation.
- Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. *Talker Variability in Speech Processing* 145–165.

- Jun, Jongho. 2004. Place assimilation. In Bruce Hayes, Robert Martin Kirchner & Donca Steriade (eds.), *Phonetically-Based Phonology*, 58–86. Cambridge: Cambridge University Press.
- Keating, Patricia. 1996. The phonology-phonetics interface. *UCLA Working Papers in Phonetics* 45–60.
- Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.
- King, Robert D. 1967. Functional load and sound change. *Language* 831–852.
- Kiparsky, Paul. 1995. The phonological basis of sound change. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 640–670. Cambridge, MA, and Oxford, UK: Blackwell.
- Kiparsky, Paul. 2006. The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32(2). 217–236.
- Kochetov, Alexei. 2001. *Production, perception, and emergent phonotactic patterns: a case of contrastive palatalization*. Toronto: University of Toronto dissertation.
- Koo, Hahn. 2007. *Change in the adult phonological processing system by learning non-adjacent phonotactic constraints from brief experience: An experimental and computational study*. Urbana, IL: University of Illinois dissertation.
- Krakov, Rena A. 1999. Physiological organization of syllables: a review. *Journal of Phonetics* 27(1). 23–54.
- Krämer, Martin. 2009. *The phonology of Italian*. Oxford: Oxford University Press.
- Kuo, Li-Jen. 2009. The role of natural class features in the acquisition of phonotactic regularities. *Journal of Psycholinguistic Research* 38(2). 129–150.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lieberman, Alvin M., Katherine Safford Harris, Howard S. Hoffman & Belver C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54(5). 358.
- Lichtenberk, Frantisek. 1983. *A Grammar of Manam*. Honolulu: University of Hawaii Press.
- Lieberman, Philip. 1960. Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America* 32(4). 451–454.
- Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala & Jeri J. Jaeger (eds.), *Experimental phonology*, 13–44. Orlando: Academic Press.



- Lindblom, Björn & Ian Maddieson. 1988. Phonetic universals in consonant systems. *Language, Speech, and Mind* 62–78.
- Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17(2). 267–302.
- Maddieson, Ian. 1996. Phonetic universals. *UCLA Working Papers in Phonetics* 160–178.
- Malécot, André. 1958. The role of releases in the identification of released final stops: a series of tape-cutting experiments. *Language* 370–380.
- MBROLA Project Development Team. 2010. *MBROLA speech synthesizer*. <http://tct.fpms.ac.be/synthesis/mbrola.html>. Computer program.
- McCarthy, John. 2006. Slouching towards optimality: Coda reduction in OT-CC. Ms. Available as ROA-878 on the Rutgers Optimality Archive, <http://roa.rutgers.edu>.
- McCarthy, John J. & Alan Prince. 1993. Generalized alignment. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology*, 79–153. Dordrecht: Kluwer.
- Meyer, Julien, Laure Dentel & Fanny Meunier. 2013. Speech recognition in natural background noise. *PloS one* 8(11). e79279.
- Mitchell, Tom M. 1982. Generalization as search. *Artificial intelligence* 18(2). 203–226.
- Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1). 83–127.
- Moreton, Elliott. 2009. Underphonologization and modularity bias. In Steve Parker (ed.), *Phonological argumentation: essays on evidence and motivation*, 79–101. London: Equinox Publishing Ltd.
- Moreton, Elliott. 2012. Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language* 67(1). 165–183.
- Moreton, Elliott & Joe Pater. 2012. Structure and substance in artificial-phonology learning. *Language and Linguistics Compass* 6(11). 686–701.
- Moreton, Elliott, Joe Pater & Katya Pertsova. 2015. Phonological concept learning. *Cognitive Science* 1–66.
- Moreton, Elliott & Katya Pertsova. 2012. Pastry phonotactics: Is phonological learning special? Talk given at NELS 43.
- Moreton, Elliott & Katya Pertsova. 2014. Pastry phonotactics: Is phonological learning special? In Hsin-Lun Huang, Ethan Poole & Amanda Rysling (eds.), *Proceedings of NELS 43*, Amherst: GLSA Publications, University of Massachusetts.

- Myers, Scott. 2002. Gaps in factorial typology: The case of voicing in consonant clusters. Ms.
- Myers, Scott & Benjamin B. Hansen. 2007. The origin of vowel length neutralization in final position: Evidence from Finnish speakers. *Natural Language & Linguistic Theory* 25(1). 157–193.
- Myers, Scott & Jaye Padgett. 2014. Domain generalization in artificial language learning. *Phonology* 31(3). 399–433.
- Nosofsky, Robert M. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115(1). 39.
- Nye, Patrick W. & Carol A. Fowler. 2003. Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics* 31(1). 63–79.
- Ohala, John J. 1981. The listener as a source of sound change. In C.S. Masek, R.A. Hendrick & M.F. Miller (eds.), *Papers from the Parasession on Language and Behavior*, 178–203. Chicago: Chicago Linguistic Society.
- Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer-Verlag.
- Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (eds.), *Language change: Contributions to the study of its causes*, 173–198. Berlin: Mouton de Gruyter.
- Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In John Kingston & Mary Beckman (eds.), *Papers in Laboratory Phonology*, vol. 1, 258–275. Cambridge: Cambridge University Press.
- Ohala, John J. 1992a. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Michael Ziolkowski, Manual Noske & Karen Deaton (eds.), *Papers from the parasession on the syllable*, vol. 26, 319–338. Chicago: Chicago Linguistic Society.
- Ohala, John J. 1992b. What’s cognitive, what’s not, in sound change. In G. Kellerman, M. D. Morrissey, G. Kellerman & M. D. Morrissey (eds.), *Diachrony within synchrony: Language history and cognition*, 309–355. Frankfurt: Peter Lang Verlag. Reprinted in *Lingua e Stile*, 1992, 27:321–362.
- Ohala, John J. 1993. The phonetics of sound change. In C. Jones (ed.), *Historical linguistics: Problems and perspectives*, 237–278. London: Longman Academic.
- Ohala, John J. & Haruko Kawasaki. 1984. Prosodic phonology and phonetics. *Phonology* 1. 113–127.

- Orgun, Cemil Orhan. 2001. English r-insertion in Optimality Theory. *Natural Language & Linguistic Theory* 19(4). 737–749.
- Padgett, Jaye. 2014. On the origins of the prosodic word in Russian. In Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O’Hara & Iris Chuoying Ouyang (eds.), *Proceedings of the 32nd Annual Meeting of the West Coast Conference on Formal Linguistics*, 41–53. Somerville, MA: Cascadilla Proceedings Project.
- Parker, Steve. 2002. *Quantifying the Sonority Hierarchy*. Amherst, MA: University of Massachusetts, Amherst dissertation.
- Pertsova, Katya. 2012. Linguistic categorization and complexity. In Lynne Cahill & Adam Albright (eds.), *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON2012)*, 72–81. Montréal, Canada: Association for Computational Linguistics.
- Pierrehumbert, Janet. 2001. Exemplar dynamics: word frequency, lenition, and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency effects and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.
- Pinker, Steven. 1979. Formal models of language learning. *Cognition* 7(3). 217–283.
- Pisoni, David B. & Jeffrey Tash. 1974. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* 15(2). 285–290.
- Prince, Alan. 1990. Quantitative consequences of rhythmic organization. In M. Ziolkowski, M. Noske & K. Deaton (eds.), *Parasession on the syllable in phonetics and phonology*, 355–398. Chicago: Chicago Linguistic Society.
- Prince, Alan & Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA, and Oxford, UK: Blackwell. Distributed in 1993 and as ROA-537 on the Rutgers Optimality Archive, <http://roa.rutgers.edu>.
- Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In M. Tsujimura & G. Garding (eds.), *Proceedings of the 22nd Annual Meeting of the West Coast Conference on Formal Linguistics*, 101–114. Somerville, MA: Cascadilla Press.
- Raphael, J. Lawrence. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America* 51. 1296–1303.
- Repp, Bruno H. 1979. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* 22(2). 173–189.
- Saffran, Jenny R. & Erik D. Thiessen. 2003. Pattern induction by infant language learners. *Developmental Psychology* 39(3). 484.

- Sapir, Edward & Morris Swadesh. 1960. *Yana dictionary*. University of California Press.
- Schneider, Walter, Amy Eschman & Anthony Zuccolotto. 2002. *E-Prime User's Guide*. Pittsburgh, PA. Computer program.
- Seidl, Amanda & Eugene Buckley. 2005. On the learning of arbitrary phonological rules. *Linguistic Learning and Development* 1(3-4). 289–316.
- Shepard, Roger N., Carl I. Hovland & Herbert M. Jenkins. 1961. Learning and memorization of classifications. *Psychological Monographs: General and Applied* 75(13). 1–42.
- Silverman, Daniel. 1995. *Phrasing and Recoverability*. Los Angeles: UCLA dissertation.
- Slowiacek, Louisa M., James M. McQueen, Emily G. Soltano & Michelle Lynch. 2000. Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language* 43(3). 530–560.
- Smith, Caroline L. 1997. The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics* 25(4). 471–500.
- Smolensky, Paul. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27. 720–731.
- Soler, Antonia & Joaquín Romero. 1999. The role of duration in stop lenition in Spanish. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 483–486.
- Stampe, David. 1973. *A Dissertation on Natural Phonology*. Chicago: University of Chicago dissertation. Published by Garland, New York, 1979.
- Steriade, Donca. 2001. Directional asymmetries in place assimilation: a perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, San Diego: Academic Press.
- Toro, Juan M., Marina Nespov, Jacques Mehler & Luca L. Bonatti. 2008. Finding words and rules in a speech stream functional differences between vowels and consonants. *Psychological Science* 19(2). 137–144.
- Vitevitch, Michael S. & Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40(3). 374–408.
- Walker, Rachel. 1995. Mongolian stress: Typological implications for nonfinality in unbounded systems. In Rachel Walker, Ove Lorentz & Haruo Kubozono (eds.), *Phonology at Santa Cruz*, vol. 4, 82–102. Santa Cruz: Linguistics Research Center, UC Santa Cruz.
- Wedel, Andrew. 2003. Self-organization and categorical behavior in phonology. In *Annual Meeting of the Berkeley Linguistics Society*, vol. 29 1, 611–622.

- Wedel, Andrew. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23(3). 247–274.
- Wedel, Andrew, Abby Kaplan & Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2). 179–186.
- Wilson, Colin. 2001. Consonant cluster neutralization and targeted constraints. *Phonology* 18. 147–197.
- Wilson, Colin. 2003. Experimental investigation of phonological naturalness. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, vol. 22, 533–546. Citeseer.
- Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30(5). 945–982.
- Winitz, Harris, Marlin E. Scheib & James A. Reeds. 1972. Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *The Journal of the Acoustical Society of America* 51(4B). 1309–1317.
- Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Martin Kirchner & Donca Steriade (eds.), *Phonetically-Based Phonology*, 34–57. Cambridge: Cambridge University Press.
- Yu, Alan. 2003. *The Morphology and Phonology of Infixation*: UC Berkeley dissertation.
- Yu, Alan. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80(1). 73–97.
- Zec, Draga. 1988. *Sonority Constraints on Prosodic Structure*: Stanford University dissertation.
- Zoll, Cheryl. 1998. Positional asymmetries and licensing. Ms. Available as ROA-282 on the Rutgers Optimality Archive, <http://roa.rutgers.edu>.
- Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language* 83(2). 277–316.