The Interaction of Speech Perception and Production in Laboratory Sound Change

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Bridget Joann Smith, B.A., M.A.

Graduate Program in Linguistics

The Ohio State University

2013

Committee:

Brian D. Joseph, Advisor
Cynthia G. Clopper, Advisor
Mary E. Beckman
Copyright by

Bridget Joann Smith

2013
Abstract

The Neogrammarians Principle, that sound change must be phonetically conditioned and exceptionless, has developed into a well-accepted heuristic that is the foundation for the comparative method. However, during the 20th century, researchers discovered a number of changes in sounds that had developed perfectly regular correspondences, but without the required phonetic conditioning environment, apparent sound-changes-in-progress that were not entirely regular, and confirmed that speech sounds are highly variable from utterance to utterance, and from talker to talker. A terminological dispute erupted over what counted as sound change, and was settled by the introduction of new terminology, such as the initiation or actuation and spread or diffusion of sound change. Because of the usefulness of the Neogrammarians principle to historical and comparative linguists, researchers who examine other aspects of sound change have largely avoided direct confrontation, allowing the gradual narrowing of the meaning of the term sound change to refer to something that is very small in scale in order to maintain its phonetic conditioning and exceptionlessness, even though the phenomena to which it first referred were very broad and sweeping changes.
While there is debate about just what may be considered sound change, the larger goals of all linguists studying sound change is to understand how sound change works, to isolate the causes and mechanisms of change, to better understand language, and hopefully one day be able to make predictions about sound change. In this dissertation, I outline a general model of phonetically gradual sound change, using experimental evidence to support its premises. This general model holds that while the basis for sound change is phonetic variation, the actuation period of sound change begins when the phonetic variation is associated with some other co-occurring factor, whether phonetic, phonological, physical, social, grammatical, syntactic, etc., or is generalized as being a characteristic of the sound category itself. In this way, all sound change is cognitive; although, the associations may range from below the level of consciousness (as in phonetic and some phonological associations) to conscious and purposeful adoption (as in socio-indexical markers), with an associated difference in the amount and type of cognitive effort required to adopt or abstain from a sound change.

Perceptual learning (e.g., Norris, McQueen and Cutler 2003) is a mechanism in which the perception of a sound category shifts to include a previously ambiguous range of exemplars. Shadowing techniques (e.g., Goldinger 1998) can induce a change in the production of speech sounds, as talkers adapt their speech to become more similar to the speech of another talker. I use perceptual learning and shadowing to recreate sound change in the laboratory, in
order to test the hypothesis that non-phonetic factors may play a key role in shaping sound change, including phonetically-conditioned sound change, and that regularity, as we have discovered with so many other linguistic concepts, is not truly a dichotomous characteristic, but rather a continuous one, relying on more than just phonetic universals.

Using this laboratory sound change paradigm, phonetically-conditioned sound change was replicated, with differences in perception and production based on, among other things, word familiarity and participant gender. These changes in perception were extended to novel talkers, words, and even phonologically different environments, showing that the paradigm was effective, but also suggesting that phonetic conditioning environments may not be as specific as we had thought. Part of the series of experiments attempted to condition a sound change with non-phonetic environments, which may be associated with differences in sound much in the way that phonetic environments are, with the result that talker and participant gender interact in much more complicated ways than simple association. I uncover evidence that the early stage of phonetically-conditioned sound change is not exceptionless, and is influenced by non-articulatory factors. “Universal” cognitive processes underlying perception are still bound by talker- and listener-specific factors (word familiarity and frequency are slightly different for each person, access to different speech partners, one’s own speech production feedback loop, age, sex, linguistic attitudes, etc.), which
are active from the beginning of a sound change, and may influence even the most pared down phonetically-conditioned sound change.
Dedication

Nancy Lee Smith (née Kienow).


Without you, I never would have started on this journey.
Acknowledgments

This is the part where I say that I love graduate school so much that I never want it to end, otherwise why would I have picked a topic that is so antagonistic to my advisor and advocate? Brian, you have never been anything less than kind to me, but usually so much more: encouraging, challenging, thought-provoking, and patient. Thank you for basically double-dog-daring me to write this thesis, and for everything. Thank you Cynthia for being the voice of reason that I so sorely needed. I have learned so many practical lessons from you, which is amazing because I also spent a lot of time laughing in your office. Thank you Mary for engaging my curiosity and encouraging me to chase down the answers to empirical questions. Thank you to the many former and current grad students in the department who make me feel like the dumb kid. I respect you all, and would be glad to work alongside any of you, no matter how dumb I feel. Thank you all for challenging me, and not letting me ever, ever, feel complacent about anything. Thank you to the participants in the many research and reading groups I’ve attended over the years, especially Changelings, Phonies, and the Psycholinguists lab groups for brainstorming,
research design, presentation practice, revision ideas, intellectual engagement, and of course, snack ;)

Thank you to the many interns who have helped my research move forward and made me a better mentor: Amy Roberts, Elizabeth Marasco, Christopher Finley, Allison Collins, Maxwell Boecker, Amber Torelli, Corey Campos, Stephen Birk, Theresa Lucius, Michael Rauh. Thanks especially to Amber, Theresa, and Chris, for being exceptional interns, and fun and interesting people that I’m glad I got to know. Thank you to everyone who sat in the stuffy old soundbooth to make stimuli for one or many of my experiments, some of you many, many times, for some very long periods of time, you know who you are 😌. Extra special thanks to Kodi and Rory, for your boss comments on this thesis, which were an unexpected gift.

Thank you to my three favorite undergraduate professors who took a special interest in me from the beginning: Karl Obrath, William Lasher, and Dan Goodman. Without your help, encouragement, high expectations (and sometimes a good kick in the pants), I wouldn’t have made it to grad school. I’m sorry I haven’t been in touch very often, but please know that each of you has played a vital role in my life.

Thank you to the Fulbright Association for selecting me out of the many qualified applicants to spend a year doing research in Germany. What I learned there not only changed my Weltanschauung, but really was the start of the questions that led me to write this thesis. Thank you to the Ohio State University for the Presidential Fellowship that allowed me to devote the entire last year to this monumental undertaking. Thank you to
the Linguistics Department, for supporting me the rest of my time here, giving me practical experience as well as a living, and helping me to get to conferences to share my research with others.

Thank you to my family, for encouraging me to go back to school, for your love, and your prayers. Thank you especially to my son, William, who has forced me to learn how to prioritize and manage my time more effectively. Thank you for teaching me patience and compassion, and that I don’t have to solve every problem. Thank you for being yourself. You have always been an interesting person, but you get more interesting every day, and I appreciate the insights you share with me, about physics and philosophy and the world around us. Thank you for helping me keep that sense of wonder and curiosity alive. Thank you for following me around on this crazy journey. I know I’ve been in school for most of your life, and haven’t given you the kind of attention you deserve. But hopefully the life I’ve given you has been rewarding in other ways. As you prepare to start out on your own adventure soon, I wish you the very best of luck, and these words of wisdom:

sá einn veit
er viða ratar
ok hefr fjöloð um farit
hverju geði
stýrir gumna hverr
sá er vitandi er vits
Vita

2005.............................................................. B.S. (honors) Linguistics, University of Cincinnati

2005.............................................................. B.S. (honors) English Literature, University of Cincinnati

2008.............................................................. M.A. Linguistics, Ohio State University

2007 to 2011 ....................................................... Graduate Teaching Associate, Department of Linguistics, The Ohio State University

Publications


Fields of Study

Major Field: Linguistics
Table of Contents

Abstract ................................................................................................................. ii
Dedication ........................................................................................................... vi
Acknowledgments .............................................................................................. vii
Vita ....................................................................................................................... x
Publications ........................................................................................................ x
Fields of Study ..................................................................................................... xi
Table of Contents ............................................................................................... xii
List of Tables ....................................................................................................... xvii
List of Figures ..................................................................................................... xviii
Chapter 1: About Sound Change ......................................................................... 1
  1.1 History of Sound Change ............................................................................. 1
    1.1.1 The Uniformitarian Principle: ............................................................. 2
    1.1.2 The Neogrammariian Principle ........................................................... 2
1.1.3 The Neogrammian Controversy ......................................................... 7
1.2 A Model of Sound Change (s/) ............................................................... 13
1.3 Plasticity in Perception and Production .................................................. 19
1.4 Experimental Investigation of Sound Change ............................................ 21
  1.4.1 The Innovative Variant ........................................................................ 22
  1.4.2 Experiment 1 – Methodology for Laboratory Sound Change ............... 25
  1.4.3 Experiment 2 – Gender-Conditioned Sound Change ............................ 29
  1.4.4 Experiments 3 And 4 – Flexibility of Phonological Representations ....... 31
1.5 Summary ........................................................................................................ 34

Chapter 2: Laboratory Sound Change ............................................................... 36
2.1 Introduction .................................................................................................... 36
  2.1.1 Kinds of Sound Change ....................................................................... 37
  2.1.2 Shadowing, Imitation, Accommodation, and Convergence ............... 41
  2.1.3 Perceptual Adjustment and Perceptual Learning ............................... 44
  2.1.4 The Perception/Production Link ........................................................ 45
  2.1.5 Source of variation ............................................................................. 51
2.2 Methodology .................................................................................................. 54
  2.2.1 Materials .............................................................................................. 54
  2.2.2 Procedure ............................................................................................ 57
2.2.3 Participants ................................................................. 60
2.2.4 Familiarity and Frequency Analysis .................................. 62
2.2.5 Acoustic Analysis Methods .............................................. 63
2.3 Results ............................................................................. 68
  2.3.1 Production Measurements – training and lexical decision stimuli .... 68
  2.3.2 Participants’ production results – baseline .............................. 69
  2.3.3 Participants’ production results – effect of training ...................... 75
  2.3.4 Individual Differences ..................................................... 87
  2.3.5 Perception Results .......................................................... 100
2.4 General Discussion ................................................................ 110

Chapter 3: Non-linguistic factors in laboratory sound change ................ 117
3.1 Introduction ....................................................................... 117
  3.1.1 Phonology as probabilistic abstraction from richly encoded exemplars 119
  3.1.2 An exemplar account of sound change ..................................... 122
3.2 Methods ............................................................................ 128
  3.2.1 Participants ..................................................................... 128
  3.2.2 Materials ...................................................................... 129
  3.2.3 Procedure ..................................................................... 131
3.3 Eye Tracking Results ............................................................ 135
List of Tables

Table 2.1: Average Euclidean distance of training stimuli from group average .......... 69
Table 2.2: Distribution of participants’ pronunciation variants for /tr/ and /str/ .......... 71
Table 2.3: Distribution of participants’ pronunciation variants for /tw/ ................. 71
Table 2.4: Comparison of plain [tw] training stimuli and participants’ baseline /tw/ productions .............................................................................................................. 72
Table 2.5: Baseline measurements for participants’ /tr/ and /ʧ/ productions .......... 74
Table 2.6: Post-release duration (VOT) of /t/ in tw-words .................................. 84
Table 2.7: Acceptance rates of trained and untrained words ................................ 101
Table 2.8: Measurements of endpoint stimuli from ID task ............................... 105
Table 2.9: Acceptance rate of untrained words .................................................. 108
Table 3.1: Percentage endorsement of /tw/ pronunciation ................................. 137
Table 3.2: ANOVA 0-400 ms after onset of stimulus ........................................ 143
Table 3.3: ANOVA 400-1200 ms after onset of stimulus .................................... 144
Table 4.1: Accuracy rates for /f/, /b/, and /v/ stimuli ........................................ 181
Table 4.2: Lexical decision accuracy rates ......................................................... 190
List of Figures

Figure 2.1: Diagram showing training procedure .......................................................... 59
Figure 2.2: Center of gravity of stimuli ......................................................................... 65
Figure 2.3: Spectral slope of stimuli ............................................................................. 65
Figure 2.4: Normalized rise time of stimuli ................................................................. 66
Figure 2.5: Center of gravity by participants before training ....................................... 73
Figure 2.6: Spectral slope by participants before training ........................................... 73
Figure 2.7: Normalized rise time by participants before training .................................. 73
Figure 2.8: Distance between \textit{chw}- training stimuli and the control groups’ /tw/ productions ............................................................................................................ 76
Figure 2.9: Distance between \textit{tsw}- training stimuli and control condition participants’ /tw/ productions ............................................................................................................ 77
Figure 2.10: Distance between \textit{chw}- training stimuli and \textit{chw}- condition participants’ /tw/ productions ............................................................................................................ 79
Figure 2.11: Distance between \textit{tsw}- training stimuli and \textit{tsw}- condition participants’ /tw/ productions ............................................................................................................ 81
Figure 2.12: Non-affricators’ Euclidean distance from training stimuli ...................... 90
Figure 2.13: Early affricators’ Euclidean distance from training stimuli .................... 91
Figure 2.14: Convergers’ Euclidean distance from training stimuli ........................................... 93
Figure 2.15: Non-convergers’ Euclidean distance from training stimuli .......................... 94
Figure 2.16: Conditional inference tree showing final convergence scores .................. 96
Figure 2.17: ID task results for the male talker ................................................................. 104
Figure 2.18: ID task results for the female talker ............................................................... 104
Figure 3.1: An example of four word and non-word choices participants saw on the computer screen during eye tracking trials. ................................................................. 135
Figure 3.2: Looks to tw- target relative to looks to ch- competitor, for chw- stimulus, female talker .................................................................................................................. 140
Figure 3.3: Looks to tw- target relative to looks to ch- competitor, for chw- stimulus, male talker .................................................................................................................. 141
Figure 3.4: Looks to tw- target relative to looks to sw- competitor, for tsw- stimulus, female talker ................................................................................................................ 141
Figure 3.5: Looks to tw- target relative to looks to sw- competitor, for tsw- stimulus, male talker .................................................................................................................. 142
Figure 3.6: Looks to tw- target, relative to sw- competitor, male talker, single training conditions ......................................................................................................................... 146
Figure 3.7: Looks to tw- target, relative to sw- competitor, female talker, single training conditions ....................................................................................................................... 147
Figure 3.8: Looks to tw- target, relative to ch- competitor, male talker, single training conditions ......................................................................................................................... 147
Figure 3.9: Looks to tw- target, relative to ch- competitor, female talker, single training conditions ....................................................................................................................... 148
Figure 3.10: Looks to tw- target, relative to ts- competitor, male talker, single training conditions by participant gender ...................................................................................... 149
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.11</td>
<td>Looks to <em>tw</em>-target, relative to <em>ts</em>-competitor, female talker, single training conditions by participant gender</td>
<td>149</td>
</tr>
<tr>
<td>3.12</td>
<td>Looks to <em>tw</em>-target, relative to <em>ch</em>-competitor, female talker, single training conditions by participant gender</td>
<td>150</td>
</tr>
<tr>
<td>3.13</td>
<td>Looks to <em>tw</em>-target, relative to <em>ch</em>-competitor, male talker, single training conditions by participant gender</td>
<td>150</td>
</tr>
<tr>
<td>3.14</td>
<td>Looks to <em>tw</em>-target relative to looks to <em>ch</em>-competitor, for <em>chw</em>-stimulus, female talker, split-training</td>
<td>152</td>
</tr>
<tr>
<td>3.15</td>
<td>Looks to <em>tw</em>-target relative to looks to <em>sw</em>-competitor, for <em>tsw</em>-stimulus, male talker, split-training</td>
<td>152</td>
</tr>
<tr>
<td>3.16</td>
<td>Looks to <em>tw</em>-target relative to looks to <em>ch</em>-competitor, for <em>chw</em>-stimulus, male talker, split-training</td>
<td>152</td>
</tr>
<tr>
<td>3.17</td>
<td>Looks to <em>tw</em>-target relative to looks to <em>sw</em>-competitor, for <em>tsw</em>-stimulus, female talker, split-training</td>
<td>153</td>
</tr>
<tr>
<td>3.18</td>
<td>Looks to <em>tw</em>-target relative to looks to <em>ch</em>-competitor, for <em>chw</em>-stimulus, female talker, split-training by participant gender</td>
<td>155</td>
</tr>
<tr>
<td>3.19</td>
<td>Looks to <em>tw</em>-target relative to looks to <em>ch</em>-competitor, for <em>chw</em>-stimulus, male talker, split-training by participant gender</td>
<td>155</td>
</tr>
<tr>
<td>3.20</td>
<td>Looks to <em>tw</em>-target relative to looks to <em>sw</em>-competitor, for <em>tsw</em>-stimulus, female talker, split-training by participant gender</td>
<td>156</td>
</tr>
<tr>
<td>3.21</td>
<td>Looks to <em>tw</em>-target relative to looks to <em>sw</em>-competitor, for <em>tsw</em>-stimulus, male talker, split-training by participant gender</td>
<td>156</td>
</tr>
<tr>
<td>3.22</td>
<td>Results of identification task for male talker, female listeners only</td>
<td>160</td>
</tr>
<tr>
<td>3.23</td>
<td>Results of identification task for male talker, male listeners only</td>
<td>160</td>
</tr>
<tr>
<td>3.24</td>
<td>Results of identification task for the female talker, female listeners</td>
<td>161</td>
</tr>
</tbody>
</table>
Figure 3.25: Results of identification task for the female talker, male listeners .......... 161

Figure 4.1: Selective adaptation is a kind of perceptual sharpening that results from accumulation of exemplars in a very tightly focused perceptual space...................... 173

Figure 4.2: Perceptual adjustment is a kind of shift in the perceptual space that results from the accumulation of exemplars in a different distribution from that normally experienced. ........................................................................................................ 173

Figure 4.3: Perceptual broadening is a broadening of the perceptual space resulting from the widely dispersed accumulation of recent exemplars................................. 174

Figure 4.4: An example of four word and non-word choices participants saw on the computer screen during eye tracking trials. ................................................................. 176

Figure 4.5: Distribution of /f/, /v/, and /b/ test stimuli ......................................... 178

Figure 4.6: Distribution of /v/ training stimuli ..................................................... 179

Figure 4.7: Trained and untrained participants log-odds looks to /b/ targets vs. /v/ competitors .................................................................................................................. 183

Figure 4.8: Trained and untrained participants log-odds looks to /f/ targets vs. /v/ competitors .................................................................................................................. 183

Figure 4.9: Trained and untrained participants log-odds looks to /v/ targets vs. /b/ competitors.................................................................................................................. 184

Figure 4.10: Trained and untrained participants log-odds looks to /v/ targets vs. /f/ competitors.................................................................................................................. 184

Figure 4.11: Identification results for male talker, Appalachian group................. 192

Figure 4.12: Identification results for female talker, Appalachian group............... 192

Figure 4.13: Appalachian trained participants convergence................................ 196

Figure 5.1: Partly cloudy with a chance of sound change .................................... 208
Chapter 1: About Sound Change

1.1 History of Sound Change

Scientific thought does not exist in a vacuum. Every scientific theory is a synthesis of the theories of the past combined with new evidence and new ways of thinking. Parts of old theories that seem useful are incorporated into new theories, while the parts that don’t work are discarded or adapted. Such has been the history of theories of sound change. The goal of this thesis is to outline a general theory of sound change that is flexible enough to account for large-scale sound changes across a community, as well as small-scale changes in an individual’s phonology, and to provide supporting evidence from experiments designed to elicit sound change in the laboratory.

As a point of clarification, throughout this thesis, when I use the word “sound change,” it means sound change sensu lato, that is, in the broad sense. Whenever I mean “Neogrammarian sound change”, with the qualification that it follows the Neogrammarian principles of phonetic conditioning and exceptionlessness, I will specify “Neogrammarian sound change.” Because my main theoretical point is that sound change operates in the same way whether it be big or small, regular or seemingly irregular,
ancient or modern, I feel justified in issuing this disclaimer. Even if I do not succeed in pleading my case, this practice will avoid the excessive use of scare quotes.

1.1.1 *The Uniformitarian Principle:*

The 19th century was a time of great advances in the sciences. In biology, evolution and natural selection gained the spotlight (e.g., Darwin 1859). In geology, the concept of uniformitarianism propelled that field into modern times (e.g., Lyell 1830). In the realm of philology, soon to become the modern field of linguistics, these same ideas appealed to philologists studying the connections between ancient and modern languages. Just as the geologists were looking, for example, at a deep and wide canyon, with the newfound realization that time, water, and wind must have acted on the stone to carve it, in the same way that time, water, and wind erode the modern landscape, philologists looked at the results of centuries of language change, and realized that similar natural forces must have been acting on the language over time to sculpt it into the linguistic landscape of the present. This is the essence of the *uniformitarian principle*, that in order to understand the natural processes of the past, one can look to the natural processes of the present, on the assumption that the same processes are in effect today that operated in the past (e.g., Lyell 1830). In fact, because we cannot directly study processes in the past, but only the outcomes of those processes and the processes that occur in the present, we must assume continuity between the two.

1.1.2 *The Neogrammarian Principle*
During the 19th century, two competing models of language change had some very strong supporters. The wave model described the (often uneven) spread of language change across words in geographical and social space, as successive waves of change spread across a language, yielding different changes (or changes differing in degree) in different lexical items. For example, Schuchardt (1885) explains:

\[
\text{Die Veränderung eines Lautes, sein Fortschreiten in einer bestimmten Richtung, wobei natürlich von der nothwendigen Wirkung rein physiologischer Veränderungen abgesehen wird, besteht au seiner Summe der allerkleinsten Verschiebungen, ist also von der Zahl seiner Wiederholungen abhängig. [...] Sehr selten gebrauchte Wörter bleiben zurück, sehr häufig gebrauchte eilen voran}
\]
The change of a sound, its progress in a particular direction, of course, excluding the essential effect of purely physiological changes, is composed of a sum of the tiniest sound changes, and is therefore dependent upon the frequency of repetition...Very seldom used words hang back, while very frequently used words progress quickly (in the sound change) [24-5, my translation].

The family tree model (or genetic model) described an unbroken chain of changes being passed down through generations of language users, with the assumption of general homogeneity across the lexicon and within a language community. Hockett (1965) credits Sir William Jones with achieving the first breakthrough regarding the genetic model, with his 1786 presentation on the apparent relatedness of Greek, Latin, Sanskrit, and the Germanic languages, but it wasn’t until the second half of the 19th century before the hypothesis began to be fleshed out. A group of philologists at this time, who became known as the Junggrammatiker (or in English, the Neogrammarians), codified their thoughts on the matter (Osthoff and Brugmann 1878), and engaged in heated arguments against the contemporaneous position which embraced the wave model, and proposed
their own heuristic. The Neogrammarian principle, as it has come to be known, can be summed up by two propositions: 1) There are three primary methods by which sounds in a language may change over time. These are analogy, borrowing, and sound change. 2) Of these, only sound change can be counted on to operate in a regular, exceptionless manner (e.g., Osthoff and Brugmann 1878). Applying the uniformitarian principle, they deduced that because borrowing and analogy could be witnessed in the synchronic language, that it must also have been available as a mechanism for language change in the past. By excluding those items whose lineage could reasonably be assigned to borrowing or analogy, a clearer picture of the genetic descent of a language could be formed.

The distinction between the three types of change in sounds is an important one because each has a different effect on the resulting phonological structure of a language. Borrowing of words from another language variety may introduce an entirely new sound into a language, such as the introduction of /ʒ/ into English from French. An introduced sound may create confusion about the relationship between two languages or the shape of the reconstructed proto-language, if it is not recognized as an import. Borrowing may also introduce new contrasts and new phonotactic generalizations into a language. For example, Middle English saw a number of words beginning with voiced fricatives borrowed from French (e.g., Lass 1994: 58-9), where previously word-initial fricatives had been voiceless. These voiced fricatives could create confusion about the voicing patterns of fricatives in earlier stages of English.
Analogy can act in many ways to change the sounds of a language. The easiest to explain is four-part analogy, in which morphological forms are leveled to a base analog form. For example, the past tense form *dived* gave way to an analogical form *dove*, by way of analogy to the past tense of *drive*, as in: *drive: drove :: dive: (dived -> dove)*. The workings of analogy on language can be much more complex than simple four-part analogy, which can make it tricky to identify across diachronic stages of a language, and difficult (though not impossible) to formalize. The most difficult aspect of using analogy in comparative linguistics is that guessing the process by which language users invented new forms is often purely speculative, and not without controversy.\(^1\)

All analogical reasoning, linguistic or general (that is, not restricted to language use) is described by cognitive scientists as being composed of a series of processes:

*In a typical reasoning scenario, one or more relevant analogs must be accessed. A familiar analog must be mapped to the target analog to identify systematic correspondences between the two, thereby aligning the corresponding parts of each analog. The resulting mapping allows analogical inferences to be made about the target analog, thus creating new knowledge to fill gaps in understanding. These inferences need to be evaluated and possibly adapted to fit the requirements of the target. Finally, in the aftermath of analogical reasoning, learning can result in the generation of new categories and schemas, the addition of new instances to memory, and new understandings of old instances and schemas that allow them to be better accessed in the future.* (Holyoak, Gentner, and Kokinov 2001: 9-10).

---

\(^1\) For example, it is widely believed that the colloquial form “warsh” for *wash* was a hypercorrection based on analogy to epenthetical r-insertion in otherwise r-less dialects based on the prestige of the /r/ sound (Hock and Joseph 1996: 187). It is my alternative analysis that the /r/ was analogically inserted based on the lack of /ɔ/ forms followed by a voiceless fricative, while /ɔ/ regularly appeared before {w,l,r}. But the origin of this change is essentially unknowable until we develop time-travel.
This description of analogy could apply to the processes that first or second language learners use to form generalizations about the forms in the language they are learning, as well as processes that well-established language users might fall back on in case of infrequently encountered forms. In the Neogrammarian spirit, I prefer to make use of general principles before resorting to special rules and exceptions, and view most linguistic processes as being explainable by ordinary cognitive processes, rather than a special innate language ability. The Neogrammarians used similar reasoning, when they explained irregular changes in terms of speakers’ use of analogical mappings to create new linguistic forms, based on the idea that contemporary language users created similarly analogical forms; therefore, language users in the past must also have done so (Schleicher 1860: 60). Their reasons for isolating cognitive processes as a separate mechanism for language change was because cognitive processes were unpredictable and often acted sporadically on the language, while purely physiological processes were hypothesized to be less capricious in their progress (Brugmann 1885: 15-16). Therefore, genetic comparison should only be undertaken on word forms that had no possible suspicion of analogical development. As such, this principle was generally viewed as axiomatic. For some, however, it was a theoretical backlash against the rival position that sound change also had a capricious element, in that a sound could evolve in multiple directions, even in the same phonetic environment, within the same language community (see Brugmann’s (1885) portrayal of Curtius’s objections, esp. pp 47-8).
Any of the components of language can change over time, such as the morphology, the ways words can be combined in a sentence, the meanings of words, and the sound shape of words. In comparing languages to look for relatedness or a pattern of changes from one time period to another, the most noticeable units of language are the words. Two related languages may have similar words, with some differences in pronunciation. The differences may be slight, as in the difference between American and British English, or between Norwegian and Swedish. Or they may be too advanced for speakers of the two languages to understand each other at all, such as between English and Danish. Or they may be so advanced that most words do not seem resemble each other at all, such as between English and Russian. But English and Russian are related just as surely as English and Danish are related. And it is some form of the comparative method (in which corresponding forms (words, morphology, syntax, etc.) are compared across languages) that was used to prove the relatedness of the languages in the Indo-European language family, including all the languages mentioned above. By reducing the pool of potential correspondences to those clearly unaffected by analogy or borrowing, the mass of data becomes somewhat less chaotic to sort through, and there are fewer cases in which a sound might show multiple possible trajectories for the same conditioning environment.

1.1.3 The Neogrammarian Controversy

The 20th century saw explosive growth in the field of linguistics, and a partitioning of the field into multiple subfields, dealing with the different components of language, such as syntax, semantics, morphology, phonetics, and phonology, or different aspects of
linguistic inquiry, such as psycholinguistics, sociolinguistics, and historical linguistics.

Many linguistic frameworks have been proposed that attempt to account for as much linguistic data as possible (e.g., Bloomfield 1933, Chomsky and Halle 1968, Greenberg 1963, Lieb 1993, Prince and Smolensky 1997, Bybee and Hopper 2001, Lyons 2006). Most of the 20th century boom has focused on synchronic analyses of language; however, with each linguistic framework that was developed, historical linguists who came of age, so to speak, in each tradition tended to phrase language change to fit in that overarching framework. One of the most enduring remnants from the early 20th century that remains standard in modern historical linguistics, is the formulation of a sound change in a form similar to a structuralist phonological rule (x -> y / z: sound x became y in an environment z), with the main difference being that synchronic alternations were denoted by a shaftless arrow. During the 20th century, the controversy that started between the Neogrammarians and their contemporaries did not die out, but rather took on new life, as new data pointed to possible flaws in the tenets of the Neogrammarian principle.

As phonetic analysis has become more widespread, and ever greater amounts of data are gathered, the greatest problem for regularity is the lack of invariance in the speech signal. No two utterances are exactly the same, and different talkers may even have different articulatory strategies for producing sounds that are perceived as equivalent: for example, American English /r/ may be bunched or retroflex (e.g., Mielke, Baker, and Archangeli 2007). With the lack of perfect regularity in the synchronic language, how can we expect the outcome of a sound change to be perfectly regular? Even without the extremely fine-grained tools now at our disposal, this question was raised by Schuchart
and Curtius and other non-Neogrammarians in the 19th century, but was never satisfactorily answered. Additionally, modern technology allows us to collect thousands of data points from thousands of individuals (as in Labov et al 2006), and even when we allow for continuous phonetic variation, the closest we come to exceptionlessness is statistical regularity of a stable sound category. Variationist studies (such as Labov 1966 and the boom that followed) showed that even in an individual, multiple forms may be in competition, with different expression based on pragmatic and social usage, and even the most predictable distribution was not exceptionless, which, again, reflected the earlier position of Curtius (1885) and others.

While the Neogrammarians had narrowed down sound change to include only those forms not suspect by analogy and borrowing, their successors in the 20th century were compelled to narrow the definition even further, to explain other possible mechanisms of change in terms of analogy and borrowing, and to reinterpret new facts about language within the Neogrammarian framework. For example, the lack of invariance in the speech signal was postulated as the source of sound change (e.g., Ohala 1974), and the idea of stable variation became acceptable. While Wang (1968) and Chen and Wang (1975) ruffled quite a few feathers with their evidence of *lexical diffusion*, as the sometimes regular, but not necessarily phonetically conditioned or immediately exceptionless implementation of sound change (which was one of the cornerstones of dialect geography), neo-Neogrammarians reiterated that lexical diffusion was most likely a case of dialect borrowing between changed and unchanged dialects, or postulated that it was a case of lexical analogy (Kiparsky 2003). Labov (1981) declared that sound change
proceeds slowly and phonetically gradually, while lexical diffusion is a different creature, proceeding in phonemic jumps, one word at a time, which suggests analogy or borrowing.

This 20th century controversy of “what counts as sound change?” eventually settled into a terminological truce by the end of the century, with researchers in different fields answering the question based on their own findings and theoretical inclinations. Ohala (1973) sidestepped defining sound change by focusing on “mini-sound changes” in the individual as the moment of initiation of sound change, which he asserted is carried out by the next generation of language learners in the transmission stage, which is echoed by Janda and Joseph (2003), who narrowed down what counts as sound change as the “Big Bang” moment of change when only physiological phonetic conditioning factors apply. The aftermath (transmission or diffusion) was not guaranteed to be regular or phonetically conditioned. Those who advocate a view of sound change as an adaptive enhancement (e.g., Lindblom et al 1995, Mufwene 2001, Blevins 2004) tend to name variation as the origin of sound change, and label actuation as the adoption of an innovative variant by the speech community. Yu (2010), for example, follows the convention of other sociolinguists (e.g. Labov 2001, Eckert 1999), holding that actuation is the point at which pronunciation variants (the result of the initiation) acquire social meaning and begin to spread along social lines. All have wondered whether a change in perception is enough to constitute sound change in the individual, or if a change in production is required (for example, see the discussion in the Introduction in Solé and Recasens 2012). The fact that this question is under debate indicates just how small-scale
we have come to imagine the process of sound change, in comparison to the widespread phonemic changes under investigation in the 19th century.

We must keep in mind that the Neogrammarians were limited by the kind and quantity of data that they could find. The texts that they used in reconstruction were sometimes written by only a handful of individuals, and therefore suffered from a lack of the kind of variability that one would expect in a larger sample of language users.\(^2\) Texts are also limited in the degree of phonetic detail they can render, having to use an orthography that necessarily converts continuously varying sounds into categorically different letters. Additionally, the time between synchronic samples that they analyzed was often many hundreds of years. In that amount of time, many irregularities would have the opportunity to be leveled by language users. 20th century linguists struggled to take modern data into account while holding onto the basic principles of the Neogrammarians. It seems to me that researchers, working on what they believed was sound change, were looking at different aspects of the same object, like the blind men and the elephant (e.g., Saxe 1868), and while trying to make progress in their fields (historical linguistics, sociolinguistics, articulatory and acoustic phonetics, phonology, psycholinguistics), did not want to alienate researchers in the other fields, especially

\(^2\) For example, what we know about the Gothic language is based primarily on the writings of one man, Wulfila, the bishop who translated the Greek Bible into Gothic. The Romance family, for example, is much better documented, but still greatly limited in that only a very small subset of language users who were educated enough to be able to write are represented. It is true that there is variability in these samples, but frequently, the variability is explained away as dialect mixing or dialect borrowing, as it is often (though not always) viewed as a hindrance to reconstruction, rather than a source of additional information. To be fair, it is incredibly difficult to attempt a reconstruction on greatly variable data, and being able to narrow down the potential candidate forms using some principled method is, I think, the true purpose of the Neogrammian principle.
historical linguists, who had the most to lose. This struggle might explain why no one argued very strongly about the use of terminology to narrow the scope of sound change, or the origin of sound change, to some event that fulfilled the basic requirements for the comparative method. This need for regularity in reconstruction and lack of regularity in synchronic data have left us with this terminological, and I would argue, theoretic and methodological split, in which each field looks at one aspect of sound change according to the interests and specialties of the researchers.

The benefits gained by limiting sound change to those events conditioned by phonetic factors alone, was believed to be threefold, answering why, how, and when a change would be regular: the lack of cognitive interaction allowed physics to explain why the change took place; physical principles therefore guaranteed regularity, which would last for as long as the conditions held (Janda and Joseph 2003, my interpretation). But, I will argue that one cannot separate cognition from any language usage, and definitely not from the processes that lead to sound change. A phonetic conditioning environment may be, in principle, the same environment as coarticulation occurs in, and because coarticulation is not only unavoidable, but generally does NOT lead to sound change, it is not a guarantee of regularity or an explanation of why sounds change. Indeed, many regular sound changes that are whole-scale unconditioned sound shifts are not directly attributable to/cannot be explained by coarticulation. And while coarticulation itself is attributable in some degree to universal constraints of motor limitations and aerodynamics, the perceptual compensation required to abstract out the intended target from coarticulation can also be quite language-specific (e.g., Beddor, Harnsberger, and
Lindemann 2002; Beddor and Krakow, 1999; Kawasaki, 1986; Samuel and Pitt 2003; Sonderegger and Yu 2010), suggesting that the realization of coarticulation is also, at least in part, language-specific.

While researchers who examine various aspects of sound change all have the same goal, to discover how sound change works, each researcher has his own ideas about how best to proceed, and his own theoretical biases to contend with. It is my position that the Neogrammian principle has the potential for grave misapplication. It is one thing to remove questionable lexical items for the sake of a reconstruction (see footnote 2); although, less scrupulous researchers could hide inconvenient exceptions under the guise of possible analogy or borrowing, while doing two comparisons might yield an interesting alternative hypothesis. It is another matter entirely to make an a priori claim about the nature of sound change and then refuse to admit evidence that doesn’t fit that assumption. I believe that while the Neogrammian principle may be a valid heuristic for comparative research, it is not a valid heuristic for undertaking a modern theory of sound change.

1.2 A Model of Sound Change (sl)

By sl, I mean sensu lato, that is in the broad sense. This model attempts to account for all the variously proposed mechanisms for phonetically gradual sound change under the heading of sound change. I will variously refer to this model as “unified” in the sense proposed by Hockett (1948), when he reasoned, “Newtonian mechanics, involving Euclidean geometry and Aristotelian logic, works out well for bodies of medium size moving at fairly slow velocities relative to the observer. It proves embarrassingly inaccurate for high velocities and for very small-scale phenomena. The problem therefore arises of constructing mechanics which will account well for what is observed at high velocities and on a small scale, but which will contain within itself, as a special case applicable within certain limits, the earlier Newtonian mechanics. The new mechanics will then constitute a more general field than the old, in precisely our sense” (560-1). And as a model, it is inevitably “wrong” but hopefully useful.
The purpose of this thesis is not to disprove or displace the Neogrammian principle as a heuristic for the comparative method, but rather to outline a new general model of sound change, to develop a methodology for studying how sound change works in the individual, and to gain evidence for how the early stage of sound change proceeds in the individual, and in principle, how it spreads across the individuals who make up a speech community. Over hundreds or thousands of years, one sound change in a million potential sound changes may develop into as wide-spread a phenomenon as Grimm’s Law, as almost an accident of history, just as one in a million streams over billions of years may turn into a river capable of carving a Grand Canyon. But the principles that explain soil erosion from the smallest rivulet are the same principles that apply to the formation of the Grand Canyon, and the principles underlying sound change in the individual are the same principles that should apply to wide-spread, language-splitting, Neogrammian-style sound change. This thesis provides evidence for a model of sound change that should apply to sound changes great and small, based on general cognitive principles, that will hopefully contribute to a conversation in which we can discuss the elephant as more than just the sum of its parts.

This general model of sound change begins with some basic assumptions concerning language more generally:

1. General cognitive principles apply to speech production and processing. For example, probabilistic learning and pattern matching do not require any special language module because they are mechanisms employed in other aspects of
cognition. (As clearly articulated in Bybee and MacClelland 2005: “The experience that users have with language shapes cognitive representations, which are built up through the application of general principles of human cognition to linguistic input” (382).)

2. Sound categories are abstract generalizations across exemplars of speech, which contain detailed phonetic information as well as additional linguistic and extra-linguistic information (as in e.g., Lindblom, MacNeilage, and Studdert-Kennedy 1984).

3. An individual’s speech perception and production are dynamic, changing not just over the course of a lifetime, but also continuously updating to meet the input of any given situation (as in, e.g., Bowie 2002, Goldinger 1998, and McQueen, Cutler, and Norris 2003).

Therefore, the flexible and continuously updated representations of sounds in a language are affected by the changing distributions in the language and the rich traces stored with each of the incoming exemplars. I posit that Neogrammarian sound change is one outcome of a long-term process of sound change, but that all sound change occurs due to the association (or lack of association) of phonetic variation with some other co-occurring factor, including phonetic or phonological environment, lexical generalization, talker-specific or general characteristics, register, word class, etc., which takes place in the individual.

My general model of sound change holds that:
1) Sound change occurs when the observed variation in a sound is no longer normally distributed around the prior mean values for that sound, which requires that the representation change in order to reflect the actual distribution of that sound (theorized by e.g. Hockett 1965, and empirically demonstrated by e.g. Nosofsky 1988).

2) Variation in a sound may come to be associated with one or more co-occurring factors, such as the phonetic or phonological environment (phonetic conditioning), lexical class, pragmatic or social usage, or talker characteristics.

3) All sound change is cognitive. While phonetic drift may be the least cognitively driven form of sound change, it too is a result of changes of the mental representations of sounds, in that the basic programming of speech gestures is tied to the mental representations of sound categories, and the abstraction of one sound category from multiple realizations of that sound. Coarticulation requires planning of consecutive speech sounds, and striking a balance between being informative as to the upcoming sounds, and possibly conveying some social and/or pragmatic information, and assimilating neighboring sounds and reducing to the point of unintelligibility. Learning to associate a pronunciation variant, whatever its original cause, with a co-occurring factor, such as a phonetic or phonological environment, certain words or classes of words, grammatical or pragmatic usage, a talker characteristic, or because it’s Tuesday, requires an even greater cognitive investment than a wholesale shift caused by phonetic drift, or the association of sound variation
with other sounds. But this is the driving force behind sound change. Coarticulation does not develop into a sound change unless there is some association that keeps the variant in use after the original conditioning factors become irrelevant.

As general linguistic theories of “how language works” have progressed through different phases, so, too have theories of language change, from rules and features to constraints and processes. Most recently gaining traction are theories of language change that are informed by various probabilistic, usage-based models and information-theoretical accounts (e.g., Johnson 1997, Bybee and Hopper 2001, Harrington 2011, etc.); although, a probabilistic view of sound change was prefigured over half a century ago (Hockett 1958; 1965), but was not well received by many scholars until much more recently. The theory guiding this thesis is also a usage-based model. A multidimensional model, that encodes talker and contextual information as well as acoustic and other linguistic information with each exemplar of incoming speech, could account for Neogrammarians sound change as well as other kinds of sound change. I take the position in this thesis that each sound category is an abstract generalization in a multidimensional acoustic (and/or articulatory) perceptual space that emerges from a group (or multiple groups) of exemplars that share acoustic (and/or articulatory) characteristics, and are linked by associations to the words and phonetic environments in which they are found, associations to talkers and talker characteristics, and associations to contexts and utterances. In this framework, it makes sense that the common associations among
groupings of sounds would provide the (analogical) associations necessary not just for phonetically conditioned sound change, but all kinds of sound change, including those shaped by shared extra-linguistic and other non-phonetic information. Differences in sound categories are a result of different accumulations of exemplars with associations including linguistic information, talker characteristics, situational or other relevant information. These abstract categories are flexible and dynamic, as they are updated by accumulations of exemplars sorted by linguistic information, talker characteristics, situations, or other relevant information, in order to more efficiently handle the incoming signal. This way of thinking about a phonological system as a process of pattern matching and sorting of the speech signal cuts through many layers of added complexity that are unnecessary to a theory of sound change. For example, Ohala (1974) grants that while the pronunciation of adults may change, he maintains that it is not until the next generation of language learners reanalyzes the adult production as belonging to a different category that sound change actually takes place. However, we do not need to wait for a new generation of language learners for change to take place if we postulate that our language faculties are regularly updated based on the input they receive, which matches up with recent findings that both perception and production are extremely dynamic and sensitive to changes in input, and can adapt quickly to changes in the probabilistic distributions of incoming exemplars, such as occurs in perceptual learning, selective adaptation, and gestural drift (respectively, e.g., Norris et al. 2003; Sawusch and Jusczyk. 1981; Sancier and Fowler 1997). With this in mind, I turn to the work being
done on the interface of psycholinguistics and phonetics for the inspiration for my experimental methodology.

1.3 Plasticity in Perception and Production

Much recent work has been done exploring selective adaptation (e.g., Sawusch and Jusczyk 1981), perceptual adjustment (e.g., Dupoux and Green 1996), perceptual learning (e.g., Norris, McQueen and Cutler 2003), gestural drift (e.g., Sancier and Fowler 1997), and phonetic convergence (e.g., Goldinger 1998). These phenomena are implicitly related to sound change, and may demonstrate some of the underlying mechanisms of sound change. Perceptual adjustment and perceptual learning are two methods of changing the boundaries of sound categories. Perceptual adjustment occurs when contextual cues induce a change in perception, such as when the vowels in a carrier sentence cause a systemic shift in perception of the target vowel (Ladefoged and Broadbent 1957), or when speech rate causes a shift in perception of VOT (Miller and Volaitis 1998), or when perception of a male or female voice changes the perception of whether an ambiguous sound is /s/ or /ʃ/ (Mann and Repp 1980; Johnson 1981; Strand and Johnson 1996). The premise of perceptual learning is based on the so-called “Ganong effect” (after Ganong 1980), whereby an ambiguous sound in a word is more likely to be perceived as an exemplar of the sound that is normally found in that word. Listeners can be trained to shift their perceptual boundaries based on the ambiguous sounds they perceive as being members of the sound category that is normally found in the word (e.g., Samuel 2001, Miller 2001, Norris, McQueen and Cutler 2003). This perceptual shifting
can be talker-specific, dialect-specific, or contextual (e.g., Maye, Aslin and Tanenhaus 2008; Kraljic, Brennan and Samuel 2008), and may generalize across a class of sounds sharing some common feature or characteristic (Kraljic and Samuel 2006, but cf. Kraljic, Samuel and Brennan 2008). Perceptual adjustment is generally short term, and talker-specific, while perceptual learning may be longer term in its effects, especially if multiple talkers are involved (Greenspan, Nusbaum and Pisoni 1988; Lively, Logan and Pisoni 1993; Lively et al 1994). An exemplar model of sound change would suggest that the effects of short-term training involved in perceptual learning would be overridden by later accumulations of differently pronounced tokens, but that if surrounded by speech that reinforces the perceptual shift, it may become long-term (Kraljic and Samuel 2005). Perceptual learning is a technique that closely mimics sound change, and can be used to investigate some of the finer nuances of sound change, especially in language processing by the listener.

Research on accommodation and convergence can give us insight into how production changes proceed. Convergence is a phenomenon in which listeners or collocutioners tend to imitate sub-phonemic aspects of the productions of conversational partners or the voices of trainers in shadowing tasks, including aspects such as pitch and amplitude (Gregory, Webster, and Huang 1993), increased VOT (e.g., Shockley, Sabadini, & Fowler 2004), or vowel height (Babel 2009). The imitation effect is stronger for less familiar words, but also for words repeated more often by the model talkers during training (Goldinger, 1998). The imitation effect is at its strongest during direct shadowing (in which listeners repeat words after hearing them spoken aloud), but may
persist six days or more after hearing model talkers (Goldinger and Azuma, 2004). Imitation is also stronger for the words that were heard during training or shadowing, but may also extend to novel words, and specific sub-phonemic qualities may be extended to new sounds, such as an increase of VOT in /p/ generalizing to /k/ (Nielsen, 2011). Some aspects of convergence mirror likely patterns of sound change, especially regarding gender and familiarity effects (in which women are more likely to be leaders of sound change and familiar words may be affected earlier than unfamiliar words).

1.4 Experimental Investigation of Sound Change

The experimental sections of this dissertation are designed to test the key assumptions of the Neogrammarian principle, as articulated in the “Big Bang” theory of sound change (Janda and Joseph 2003), and to provide evidence for the principles of a general model of sound change, as detailed in section 1.2, including the hypothesis that phonetically conditioned sound change is not subject to phonetic conditioning factors alone, but rather that all types of gradual sound change (that is, opposed to more drastic changes in sound, such as alternate pronunciations, in which one sound in a word is replaced with another, categorically different sound (such as the English and French-like pronunciations of the first vowel in the noun, envelope as [ɛnvələp] versus [ənvələp]), or borrowing, in which an entirely new word is adopted) operate in a similar fashion, regardless of whether the conditioning factors are phonetic, phonological, social, grammatical, or anything else.

A positive outcome of the experiments will also set up an effective methodology for studying sound change in the laboratory, in which changes in production and perception
can both be measured at a very early stage of sound change in the individual. Whether changes in production and perception in the individual, as a result of exposure to variation that is skewed in one direction or another, count as sound change, will depend entirely on one’s definition of sound change. But I argue that a much slower, more gradual shift in variation would result in similar outcomes in perception and production, with longer-lasting effects; although, this is also an empirical question. I also hope to provide enough evidence to support the efficacy of this methodology, as well as my theoretical claims.

1.4.1 The Innovative Variant

The experimental section uses existing variation in pronunciation as a springboard to replicate sound change in the laboratory. In order to create a plausible sound change, there had to be existing variation in a phonetically conditioned environment, that had not yet been assigned any social or linguistic significance. The variation needed to be natural, but not necessarily explanatory, in that multiple measurably different pronunciations were possible outcomes of a sound change. The sound in American English that best fit the requirements was the stop+approximant /tw/ cluster in American English. Approximants are known to increase the degree and length of aspiration in preceding stops because the degree of closure in approximants is greater than in vowels, just as aspiration is increased more preceding high vowels than low or mid vowels (e.g., Sievers 1901, Ohala 1983, Guion 1996, 1998, Ohala and Solé 2008). And the narrower constriction also allows the noisy frication of the burst to increase, changing the character
of the early aspiration to a noisier, more turbulent sound, which can lead to the development of affricates. The lip-rounding that accompanies /w/ may spread to the preceding stop, which, by lengthening the front cavity, may create the percept of a retracted /t/. The narrower constriction may allow the aspiration to become outright frication, in which the resulting sound may be similar to an post-alveolar affricate, similar to what may occur with the /tr/ cluster. This variant will be denoted by *chw*-, which stands in for the phonetic notation [tʃw] or [tʃw]. Alternatively, the character of the aspiration/frication could be fronted, as the place of articulation transitions from the alveolar to bilabial place of articulation, developing into a more anterior sounding frication, yielding something more similar to a dental affricate. This variant will be denoted by *tsw*- , which stands for [tʃw] or [tʃw]. Both variants are attested in sound changes in multiple languages, and as such, are “natural” outcomes of the coarticulation of /t+w/.

In American English, many alveolar stops became post-alveolar affricates, [tʃ] and [dʒ], before /j/, as in *congratulations* and *fortune* or *soldier* and *cordial*, respectively. Most American English speakers also palatalize and affricate /t/ preceding /ɾ/, as in *train* and *trick* (e.g., Read 1971; Olive, Greenwood, and Coleman 1993). Some speakers have carried this change to the voiced alveolar stop (e.g., Read 1971), as in *drain* or *drink*. The alveolar fricative /s/ continues to undergo retraction preceding /j/ and /ɾ/, sometimes as a

---

4 The uneven, continuing application of /s/ -> /ʃ/ before palatals is probably an example of phonetically-conditioned lexical diffusion, in which the spread is gradual and phonetically conditioned, with some words
complete phonological alternation resulting from past sound changes, and sometimes as a partially realized /ʃ/ (Zsiga 1995). Longer-distance assimilation of /s/ and /t/ affects /str/ clusters for some speakers, resulting in a retracted [ʃtr]–like cluster (Shapiro 1995). These examples suggest that affrication of /t/ before approximants is a process that is not unfamiliar to American English speakers, and that retraction may also be familiar. In /t/ before /w/, multiple kinds of affricates are possible, including [tsw], [tfw], and [tfw], all arising “naturally” from the coarticulation of /t+w/. The first two variants were used throughout the experimental sections as innovative variants, and targets of sound change. Both the fronted and retracted affricate may be “explained” by the coarticulation of /t+w/, but it is, in this way, not actually explanatory as a causative factor. /tw/ may remain stable, a cluster composed of a voiceless alveolar stop + labio-velar approximant, or it may develop into a retracted affricated cluster, or a fronted affricated cluster, among other possible variants. The phonetic environment does not specify that sound change must occur, or in what direction it should go. And even though it is true that some phonological (or phonetic) contexts are less stable, or more prone to sound changes of a certain type, we find sounds in those configurations remaining that way for long periods of time without undergoing sound change. If an explanation for sound change is sought, it should be more specific, such as “why now?”

having completely changed, and others showing phonemic alternation (such as in the word grocery) or more continuous phonetic variation.
The possible range of variation in itself suggests that the conditioning environment of /t+w/ is not an explanation of how or why sound change would occur in this environment. The possibility of a front and a retracted variant introduces at least three non-phonetic factors into the equation. One factor is analogical, in that the analogous changes in /t/ before {j,r} yields a retracted affricate. This parallel would make a retracted affricate [ʧw] more “familiar” seeming. The systemic factor is the lack of /ts/ as an independent or widely used sound, which would also suggest a greater likelihood of adoption of the retracted variant [ʧw]. Another factor is physical and social, in that men are generally expected to use retracted variants of the alveolar fricative, while women are expected to produce more anterior versions, due to social enhancement of physiological differences (e.g., Strand 1999). It then follows that men might be expected to use a more retracted affricate for an alveolar sound than women. This gender expectation may interfere with the direction of sound change, conditioning a split in affrication of /tw/, as men might be disinclined to adopt a front affricate.

1.4.2 Experiment 1 – Methodology for Laboratory Sound Change

As mentioned in section 1.3, changes in perception of a pronunciation variant can be elicited by exposing listeners to familiar words containing that pronunciation variant. Using the two /tw/ variants described in 1.4.1, an experimental design was constructed in which participants were exposed to familiar and unfamiliar words containing none, one, or both of the pronunciation variants. After hearing each word, participants were required to say the word out loud, with the expectation that most participants would adapt their
pronunciation to be closer to the training stimuli, thus creating a change in production as well. In the first experiment, a lexical decision task was used to determine participants’ acceptance rates of words containing the /tw/ variants, and was designed to find out if listeners would generalize their perceptual changes to include /tw/ words that they had not heard during training. Post-test productions were compared to productions from the beginning of the experiment to find out if changes persisted after shadowing, and if changes were generalized to new words not spoken during shadowing. An identification task was used to measure whether participants generalized perceptual changes in /tw/ to a similar but new phonetic environment, /tu/.

In addition to testing the efficacy of the methodology for studying sound change in the laboratory, the following questions were addressed by the research: Are there differences in adoption of a gradual sound change across familiar and unfamiliar words (as suggested by the results in Goldinger 1998, and postulated as a factor in sound change as early as Schuchardt 1885)? Does the sound change generalize to novel words in production, as expected by the Neogrammarian Principle, or is it learned one word at a time, as in the case of lexical diffusion as dialect borrowing? Does the perceptual component of sound change likewise generalize to novel words and novel talkers (as in Kraljic and Samuel 2005)? Is the sound change generalized beyond the original conditioning environment (also as in Kraljic and Samuel 2005), suggesting general involvement of the abstract sound category rather than an immediate association with a specific environment? Are there effects of language specific constraints, such as the
existence of neighboring phoneme categories (as in the difference in results between voiced and voiceless stops in Neilson 2007)? This would suggest greater involvement of cognition in the early adoption of sound change. Are there social constraints, such as the role of gender, on the adoption of different pronunciation variants (as would be suggested by the results of Strand 1999, etc.)? Does a shift in perception result in a shift in production (as would be predicted by gestural theories such as Liberman et al. 1967 or Fowler 1986, but not by observed results in near-mergers as in Warren, Hay, and Thomas 2007)? What is the role of the individual’s baseline production on adoption of a sound change? That is, if language users already display a shift in production, do they continue to shift or reverse course when encountering similarly shifted speech? And if they have a different variant, are they more or less likely to adopt the new variant?

It was expected that participants would exhibit production changes during the shadowing section of the first experiment, which would demonstrate convergence, which could be the first step to a long-term change in production. It was also expected that the retracted variant, \(chw\)-, would elicit stronger convergence than the front variant \(tsw\)-, as an analog to the pronunciation of /t/ before /r/ and /j/. It was expected that the difference in convergence rates between these two variants would be stronger for men, who would be expected to have a more retracted pronunciation, and if /t/ is analogous to /s/, could be seen as less masculine if they had a more front pronunciation. If these differences in convergence developed, it would be evidence that analogy and social factors play a role
in short term production changes, and by extension, probably also in longer-term sound changes.

It was also expected that these production changes that began during shadowing would persist through the end of the experiment, and that gradual production changes would also occur in /tw/ words that were not heard during shadowing. The extension of pronunciation changes would provide evidence of phonetic conditioning, in which coarticulation of /tw/ leads to an innovative variant, which occurs everywhere the conditioning environment applies. The persistence of changes after shadowing would validate this method of studying sound change in the laboratory as an effective, if not realistic, method of delivering a concentrated dose of variation in order to observe the effect on production and perception. The persistence of production changes would also provide evidence for a theory of sound change that holds that sound change is a result of accumulating variation changing the shape of phonological categories.

It was also expected that participants would adjust their perceptual mappings to make room for the incoming pronunciation variants, as revealed by endorsement of /tw/ variants as exemplars of /tw/ during the lexical decision task. This would provide evidence of perceptual adjustment. It was expected that participants would generalize the pronunciation variants to new words and to new talkers not heard during training, providing evidence of perceptual learning, a deeper, longer-lasting form of perceptual adjustment. It was considered a possibility, but not expected that participants would generalize the affrication of /t/ to a new environment, /tu/, because they did not hear any
evidence of this change during training, and during the lexical decision task, they heard two normal tokens of /tu/, plus tokens of normal /t/ preceding other vowels, which provided evidence that /t/ remained unaffected in other environments. If they did make the generalization of affrication from /tw/ to /tu/, it could indicate that even during phonetically conditioned sound change, analogical generalizations were at work, and that analogy could overcome negative evidence.

1.4.3 Experiment 2 – Gender-Conditioned Sound Change

Experiment 2, which is covered in chapter 3, was designed to address the question of whether non-linguistic factors may condition a sound change, in this case, a sound split of /tw/ into a front and back variant, based on the gender of the talker. Experiment 2 was actually composed of two separate experiments with the exact same training procedure and training conditions, but different testing procedures. The materials and procedure used in training were drawn from the first experiment, with the main difference being the training conditions. In one experiment, participants completed the same kind of lexical decision and identification tasks as in the first experiment, and in the other, participants completed an eye tracking identification task. Training conditions included a group that heard the retracted affricated variant (chw-), and a group that heard only the front affricated variant (tsw-), plus the addition of two conditions that presented the pronunciation variants in a completely gender-predictable way. The “traditional” group heard men pronounce a retracted affricated /tw/ (chw-), and women pronounce a front affricated /tw/ (tsw-). It is referred to as traditional because the relationship between
men’s and women’s productions is an exaggerated version of the traditional men=retracted and women=front relationship as is found for /s/. The “non-traditional” group heard the reverse relationship: men with a front affricate (tsw-) and women with a retracted affricate (chw-). The eye tracking groups (all four conditions) were seated in front of an eye tracker for the testing phase. Mouse click identification of words and non-words containing /tw/, with sw- and ch- competitors showed participants’ acceptance rate of the /tw/ variants. Eye tracking data, in which the gaze location was recorded every 17-20 ms, showed the time to decide which orthographic form was the intended target, and showed degree of confusion by how often participants looked to the target versus competitor. The other groups (all four conditions) performed lexical decision and identification tasks as in experiment 1.

The second set of experiments were designed to address whether non-linguistic factors could condition a sound change based on gender. The question, ‘do participants learn to associate the gender of the talker with a specific pronunciation variant, and does the association generalize to new male and female talkers?’ addresses the ability of language users to form associations between pronunciation variants and non-phonetic factors, forming generalizations that apply to an entire group of talkers rather than idiolectal usage. I also wondered if there would be differences in performance between groups depending on whether the distribution of variation was an exaggeration of the customary relationship between men’s and women’s productions (men chw-, women tsw-) versus if the variation presented was the reverse of the traditional relationship (men tsw-, women chw-)? And are there differences in perceptual adaptation based on the gender
of the listener as well as the genders of the talkers? The interference of pre-existing gender-based distributions in the sound category, and an interaction with listener gender could point to an even greater role for social factors in shaping the outcome of sound change, and give us insight into individual differences in how sounds are associated with non-phonetic factors.

It was expected that the front and retracted participants would perform similarly to those in the first experiment, with greater endorsement of the variant they were trained on, as the result of a perceptual shift. It was expected that the perceptual shift would be extended to novel words and talkers, as in the first experiment. This result would lend further weight to the appropriateness of the methodology for studying sound change in the laboratory, by replicating the first results. If there was an asymmetry in the adoption of these variants, it was expected to be due to the fact that chw- is a closer analog to processes of /t/ affrication that have historically occurred and are presently active in English. It was expected that the traditional and non-traditional groups would experience shifting of their perceptual categories for men and women, depending on the variant they heard during training, providing evidence that sound+gender associations can be learned in the same way that sound+phonetic conditioning environments can be learned. As in the first experiment, it was expected that the front variant would be more difficult for men to incorporate into their repertory, which would indicate a social bias in sound change.

1.4.4 Experiments 3 And 4 – Flexibility of Phonological Representations
The third experimental chapter provides evidence from two side investigations of the main two experiments that can be used to explore how changes in the organization of the phonological system relate to sound change. The first experiment in this chapter involves filler trials from the experiment in chapter 3. Training in which listeners hear many different tokens of word-initial /v/, while with a normal range of variation, leads to broadening of a sound category along more than one acoustic dimension. Perceptual broadening was elicited from most of the participants in all the experiments to some degree, as the range of variation heard was greater than that usually experienced over such a short time period. Chapter 4 outlines how perceptual broadening is related to perceptual adjustment and selective adaptation. The second experiment in chapter 4 involves a different condition of the experiment in chapter 2, in which participants believed that some or all of the trainers might be speaking an Appalachian dialect of English. As such, when they were introduced to the chw- variant through training, they processed it differently than participants in the other conditions, who believed that the trainers spoke educated Standard English. It was believed that participants in the Appalachian condition might fail to converge with the trainers due to the stigma attached to the Appalachian dialect; however, they converged during training, but did not continue to affricate after training, unlike the other participants. Also, due to their expectation of wild variation, these participants experienced perceptual broadening, in which they endorsed both of the variants, and to a greater degree than those trained on a single variant.
The purpose of chapter 4 is to outline some of the ways that the phonological system adapts to changes in the language as experienced by the language user. These changes in how sound categories are mapped onto the perceptual space allow flexibility in adapting to speakers known and unknown, but they can also lead to sound change over time. Two of the types of changes discussed in this chapter are fairly well-known: perceptual adaptation, which is a shift in the perceptual space of a sound; and selective adjustment, which is a narrowing of the perceptual space of a sound. A third type of change is introduced: perceptual broadening, which is a broadening of the perceptual space in response to broad variation, or the expectation of variation. While selective adjustment, which involves repeated exposure to a very narrow band of variation in a sound, seems unlikely to be encountered very often in nature, perceptual adjustment seems a likely configuration at the early stages of a sound shift, when the variation in a sound tends toward a certain, changed direction. Perceptual broadening, while discovered by happenstance as a result of experimental setup, seems likely to be the configuration at the early stages of a merger or sound split. In case of a merger, two categories may broaden, resulting in wide, shallow categories and overlapping boundaries, which allow a great range of variation. In case of a sound split, there may be great variation in one sound, so that any kind of association with the variants produced at one end of the sound category may be enough to allow them to break off and form their own category, with those associations. This type of perceptual broadening was found in chapter 3, as participants hadn’t yet assigned meaning to the variation. Perceptual broadening is also a very likely method of attempting to understand a talker who speaks an unfamiliar variety. Any
category that exhibits variation beyond that normally encountered gets a boost, so that a maximal number of possible words can be identified. This necessarily requires overlap of multiple categories and, while an inefficient use of the perceptual space, may be preferable to misunderstanding one’s interlocutor.

All three of these configurations can be easily accounted for in a probabilistic, usage-based model, which takes recency into account in configuring the current phonological settings. A similar and much more specific account has been detailed by Clayards et al. (2008), which showed that listeners do take into account the distribution of variation that they encounter in a short period of time. Certainly, it would make sense for long-term variation to have at least as great an effect on re-tuning the perceptual system.

1.5 Summary

As mentioned in the introductory section, there are theoretical disagreements about what sound change is, and how it is connected to Neogrammarian sound change. It is anticipated that the results of the following experimental chapters will provide support for a model of sound change sl, that is based on general cognitive principles, and follows these three principles: 1) that sound change is a result of a change in the balance of observed variation in a sound, 2) that variation in a sound may come to be associated with one or more co-occurring factors, which depend on what associations are salient to the language user, and 3) that all sound change is cognitively based.

It is hoped that by breaking the truce that was established among the different fields involved in the study of sound change, it will not plunge us back into civil war, but rather
engender greater cooperation, and more experimental work on the part of historical linguists, and greater care in doing one’s historical linguistics “homework” on the part of the experimental linguists who want to study sound change.
Chapter 2: Laboratory Sound Change

2.1 Introduction

In the early days of linguistic study, sound change was a phenomenon that could only be inferred from textual analysis of written language across different time periods or comparative analysis across related languages. As technology has advanced and theories have evolved, we have developed new ways of investigating sound change, using recordings and phonetic analysis to look at synchronic and even diachronic variation within and across groups of speakers, and using perceptual measures to see how such variation may contribute to sound change. However, many have believed that the precise moment in which variation turns into sound change could not be witnessed. I would argue that there is no such moment, but rather a continuous process in which variation accumulates, altering the nature of the mapping from sound category to acoustic-perceptual space, and that even small shifts in perception and/or production could rightly be called sound change (as in Ohala’s (1974) “mini-sound change”). Just as a child does not suddenly become an adult at one moment (regardless of the arbitrary age denoted by
legal standards), variation does not suddenly become sound change. Rather, sound change emerges from the variation, over time. While it is likely that succeeding generations of language learners help a sound change to advance, there must be some reason for them to acquire a different phonological system than that which their parents (and even older siblings) acquired. And that reason is that the sound change was already present in their predecessors’ language usage, in slowly shifting categories. While adults phonological systems may change more slowly than children’s, we now know that individuals’ perception and production can change over the lifetime, and are not as inflexible as once imagined. Long-term gradual (or rather continuous) sound change must occur in the grammars of individuals rather than being transmitted from one generation to the next, or there would be no change. Thus, the key to studying sound change is to examine it within the individual.

2.1.1 Kinds of Sound Change

The Neogrammarians took the contemporarily radical position that the expected outcome of diachronic sound change should be complete regularity, and that, instead of inventing a new sound change to explain every exceptional outcome, two other processes, borrowing and analogy, should be used to explain most irregularities. Borrowing can result in irregularities when a word with a phonological form that is different from the original native form (e.g., kirk versus church) is borrowed into the language from a different dialect or language. Analogy creates irregular sound correspondences when a word with a phonological form that is different from the original native form is created
by association with similar forms (e.g., frozen was formed by analogy to the present tense form freeze, but would have been froren if it had developed without analogical processes). The Neogrammarians also postulated that regular sound change must proceed solely on a physiological basis, conditioned only by a phonetic environment that causes or explains the change, as opposed to the cognitive processes involved in analogy (Paul 1880: 43-4). On the basis of this physiological explanation, sound change was thus expected be completely exceptionless, with the change occurring in every word in which the conditioning environment applied (or in the case of unconditioned shift, every instance of the sound without regard to words or phonetic environments), for every member of a speech community, unless resulting in dialect split.

In order to examine sound change at the earliest stages, there must be ambient variation that has not yet reached a point in which it is consciously associated with a conditioning or indexical factor, because after this time many theorists consider the actuation or inception of the sound change to be over, and it may be considered the aftermath of sound change (Janda and Joseph 2003), called by others a sound change in progress (e.g., Labov, Yaeger and Steiner 1972), in which an innovation is diffused, or passed along from talker to talker, or word to word. If a pronunciation variant is

\[5\] In fact, Georg Curtius was scorned by his former student, Karl Brugmann (1885: p 47-8) for not believing this principle: “er hielte einerseits für möglich, dass eine phonetische Änderung nur einen Theil der Wortformen, in denen der Laut unter denselben Bedingungen vorliegt, ergreife und den anderen unberührt lasse…anderseits erschien ihm möglich, dass eine einheitliche Sprachgenossenschaft derselben Laut in gleichartigen Formen, und sogar in denselben Worte in verschiedenen Richtungen abändere” [On the one hand, he believes it is possible for a phonetic change to encompass only one part of the lexical items in which the sound appears, and yet, under the same conditions, to leave another set of words untouched…on the other hand, he believes it is possible that within a unified speech community, the same sound could change in different directions, within the same forms, and even within the same word.] (my translation).
consciously associated with a group of talkers or any other conditioning factor, it may be perceived very differently (especially in terms of social meaning) from the pronunciation that is not associated with that factor. Indexical pronunciations can be categorical, such as *tomayto* versus *tomahto*, or continuous, such as the American Northern Cities raised /æ/. Categorical pronunciation variants are phonemic, in that one phoneme replaces another, while continuous pronunciation variants may be consciously noticeable, but their adoption does not necessarily constitute a phonemic change, especially in the case of shifts (as opposed to mergers or splits).

Labov (1981) argued that categorical changes are likely to be spread by *lexical diffusion*, that is phoneme replacement word-by-word, but that sub-phonemic changes proceed phonetically gradually in all conditioned environments, as in regular sound change, which may eventually result in a phonemic change, over a long period of time. In order to maintain the Neogrammarian principle, it has been claimed that lexical diffusion may be viewed as a type of *analogical change*, as a variant pronunciation is generalized from one word to another (Kiparsky, 1995). Diffusion of a sound change (sound change-in-progress) may be viewed as a type of *borrowing*, in which a variant pronunciation is passed from speaker to speaker. The definition of regular sound change to include only those changes which are phonetically conditioned (or widespread unconditioned drift) and exceptionless is problematic, in that exceptionlessness seems to imply an eventual categorical change (or else how would we discover its existence?), while phonetically-conditioned processes are gradual or continuous by nature. During the course of a gradual
phonetic change, irregularities may surface, which is why the dichotomy between sound change actuation and diffusion was adopted. However, I take the position that regular, phonetically-conditioned sound change is the same creature as the diffusion of sub-phonemic phonetically-conditioned sound change, and that exceptionlessness does not apply during the actuation or diffusion of sound change, but is rather a concept that must be applied well after the sound change has reached completion. I also argue that extension of the conditioning environment to new phonetic or phonological conditions operates in the same way that sound change proceeds through all relevant lexical items, and that both of these are analogical processes, in that they proceed by the interaction of different levels of abstraction from the distributions of accumulated exemplars. These are all caused by the same underlying mechanism as perceptual learning, which is the accumulation of unevenly distributed phonetic variation, and the learned association of that variation with another linguistic or non-linguistic object, such as a phonetic or phonological environment, a word class, or a talker characteristic. Using a usage-based, exemplar-type model, I theorize that sub-lexical abstractions (i.e., sound categories) are generalizations that arise from the grouping together of similar acoustic or articulatory patterns that occur across the lexicon. Such associations are inherently cognitive processes; therefore, I argue that sound change is not a purely physical process, rather that it is highly analogical. This study attempts to provide evidence for these hypotheses.

In a model that assumes that memories of sounds are tagged with other linguistic and extra-linguistic information, it is reasonable to assume that these other possible
associations are also accessible, as is the association of a phonetic conditioning environment. The mind will automatically search for patterns in order to create efficiency in processing. Thus, the best predictor of variation may not always be a phonetic conditioning environment, but whichever co-occurring factor or combination of factors is most salient.

Experiment 1 makes use of perceptual adjustment and shadowing to elicit changes in perception and production. It was expected that sub-phonemic changes in perception and production would not be limited only to the words presented during training, as might occur in the case of lexical diffusion, but that the changes would be generalized to novel words that were not heard with the pronunciation variant, as in regular, phonetically-conditioned sound change. This finding would provide evidence that the early stages of sound change could be recreated in a laboratory. Furthermore, it was hypothesized that perceptual changes might be generalized to a new conditioning environment, suggesting phonological analogy, and it was expected that gender and word frequency would also play a role in the outcome of perception and production changes. If phonetically gradual sound change could be found to contain such “irregularities” in its actuation phase, this would provide a solid argument against the actuation-diffusion dichotomy, and in favor of a cognitively-driven, usage-based theory of sound change.

2.1.2 Shadowing, Imitation, Accommodation, and Convergence

Research on accommodation and imitation can give us insight into how sound change progresses. Imitation or convergence is a phenomenon that we can observe by using a
shadowing task, in which listeners repeat words after hearing them spoken aloud. In conversational interaction, the terminology is slightly different, as are the results. *Accommodation* is the term for the act of changing one’s speech to be more like that of an interlocutor. *Convergence* is the resulting closeness of speech; although, the term is also used to describe imitative aspects of speech during shadowing (which may not be considered a properly “conversational” setting).

It has been found that listeners tend to imitate sub-phonemic aspects of the production they’ve heard during shadowing, such as increased VOT (e.g., Shockley, Sabadini, & Fowler 2004) or vowel height (Babel 2009). The imitation effect is stronger for less familiar words, but also for words repeated more often by the model talkers during training (Goldinger, 1998). The imitation effect is at its strongest during direct shadowing, but pronunciation changes may persist six days or more after hearing model talkers (Goldinger and Azuma, 2004). Pronunciation changes are stronger for the words that were heard during training or shadowing, but may also extend to novel words, and specific sub-phonemic qualities may be extended to new sounds, such as an increase of VOT in /p/ generalizing to /k/ (Nielsen, 2011). These aspects of convergence mirror likely patterns of sound change.

Differences in degree or likelihood of convergence can also reveal how sound change progresses in the individual. The degree of convergence between talkers in a conversational setting has been found to vary by sex and role, with women more likely to accommodate if they are giving directions, and men more likely to accommodate if they
are receiving directions (Pardo, 2006). Namy et al. 2002 showed that in a shadowing task, women were more likely to converge to the model than men. Convergence may also be mediated by the participant’s attitude toward the model talker (Babel, 2009), and the initial degree of linguistic difference between the talkers, with more accommodation observed when the linguistic distance is greater to start with (Kim, Horton, and Bradlow, 2011).

The traditional method for measuring similarities between models’ and shadowers’ speech is an AXB task, in which a word from the model is played, flanked by two words from the shadower, and the listener must determine which of the first and last stimuli sounds more like the model. This method has the benefit of taking into account multiple cues, but this can also be seen as a limitation, because it is unknown just what the shadowers are imitating. Because the AXB task does not inform us about what aspects of speech may be imitated, some researchers have used acoustic measurements to get at which sub-phonemic qualities are imitated in these experiments, such as pitch (f0) and intensity (Gregory et al. 1993.), lengthened VOT of voiceless stops (Shockley et al. 2004), vowel formants (Babel, 2009), formant differences between light and dark /l/ (Honorof, Weihering, and Fowler, 2011), and those which do not show imitation effects, such as reduced VOT in voiceless stops (Nielsen, 2011). Because the present study aimed to look at the change in a specific sound, it was determined that spectral and durational measurements were more appropriate to the purpose of discovering whether participants exhibited the expected change in pronunciation, and how strong this change was.
2.1.3 *Perceptual Adjustment and Perceptual Learning*

Production changes are not the only sign of sound change, and in fact, may be the end result of perceptual changes. It is therefore important to consider speech processing and the mental organization of variation in pronunciation. *Perceptual adjustment* is a general cognitive process whereby any perceptual process may be re-tuned to take into account new expectations regarding perceptual input (e.g., Gibson and Gibson, 1955). Perceptual adjustment for speech sounds may occur when a listener hears familiar words that contain a variant pronunciation of one of the segments, and top down processes allow that sound to be categorized as an instance of the normally occurring phoneme, thus altering the perception of that phonemic category, possibly by shifting the balance of accumulated exemplars in the direction of the variant pronunciation. The resulting change in perception extends to new words with the pronunciation variant, or even more extreme versions of the variant (Norris, McQueen, and Cutler, 2003). The effect may be limited to one talker (Eisner and McQueen 2005), or may become generalized to new talkers and new related phonemes, such as VOT from /d-t/ to /b-p/ (Kraljic and Samuel, 2006), and can be applied to talker characteristics such as dialect (Clopper and Pisoni, 2004) or foreign accent (Bradlow and Bent, 2008). *Perceptual learning* is the term used when describing these more general effects, which also turn out to be fairly long-lasting, even without further input (Kraljic and Samuel, 2005). Perceptual adjustment has been robustly demonstrated for VOT, but also occurs for place of articulation of fricatives (/s-ʃ/) (Norris, McQueen, and Cutler, 2003), /s-ʃ/ (Kraljic and Samuel, 2005), and vowels (Maye, Aslin, and Tanenhaus, 2008).
Perceptual learning essentially demonstrates how the perceptual system is re-tuned when variation in a sound is consistently shifted in one direction, rather than being distributed evenly around the prototypical pronunciation. Variation is important: repeated exposure to the same stimulus or the same talker does not create generalized perceptual learning, but only a specific perceptual adaptation to the one stimulus (as in selective adaptation) or one speaker (e.g., Eisner and McQueen, 2005). It is generally accepted that variation is the source of sound change (e.g., Ohala, 1974; Harrington, 2012). How variation proceeds to become full-fledged sound change has not been documented in any existing sound change, especially given the theoretical split between sound change actuation and diffusion; however, it seems reasonable to assume that the mechanism underlying perceptual learning is the same as that engaged in the perceptual aspects of gradual, phonetically-conditioned sound change, as the incoming variation gradually alters the existing distribution, and the perceptual system compensates for the change.

2.1.4 The Perception/Production Link

In order to develop a model of sound change, we must have some understanding of how perceptual and production changes fit together. A change in the mental representation of a sound, does not, in itself, constitute a sound change (c.f. Ohala’s (1974) concept of a “mini sound change”), without a corresponding change in production. That speech perception and processing must somehow be linked to speech production is clear. The nature of this link, and at what level it occurs is much less clear. During language acquisition, we must certainly base the output in speech production on the input
from perception, else why would speakers in a community sound more like each other
than they do outside speakers? However, when adults move into a new speech
community, they sometimes adapt their speech to the norms of the new community;
sometimes they do not; and sometimes, only parts of their speech are affected (e.g.,
Bowie 2000). In some communities undergoing merger of two previously distinct sounds,
speakers may be able to perceive a distinction even though they cannot produce it (e.g.,
Warren, Hay, and Thomas 2007). Additionally, some speakers may be able to produce a
distinction without discerning it, which may be called a near merger (e.g., Labov, Karen,

Motor/articulatory theories of speech perception are attractive in their ability to
explain imitation, convergence, and some aspects of sound change (e.g., Motor Theory:
Liberman et al., 1967, Articulatory Phonology: Browman & Goldstein, 1992). Some
claim that speech is perceived as constellations of gestures, that the acoustic pattern of a
sound activates the gestures required for producing it, and that the gestural configuration
is the source of the representation of linguistic sounds (e.g., Direct Realism, Fowler
1986). These theories are supported by the speed and near automaticity of imitative
speech while shadowing (e.g., Fowler, Sabadini, and Weihsing 2003), and the below-
consciousness effect of gestural drift (Sancier and Fowler 1997); however, they do not
explain the existence of near-mergers, or the ability of merged speakers to perceive
merged sounds distinctly. Many other theories make use of abstract representations to
indirectly link gestures and acoustic patterns, as do most theories that make use of
distinctive features (e.g., Chomsky and Halle, 1968). A model that allows abstract
representations would enable two phonemes to connect to one set of articulatory gestures, as when a contrast is perceived but not produced. And one phoneme could have two or more sets of gestural specifications, if the grammar called for it, but it is unclear how a purely abstract representation would allow non-allophonic variation, specified at the lexical level with two different sets of gestures for one abstract representation, as would have to be the case in a near-merger.

Each of this type of model has its own additional shortcomings, as reviewed by Pierrehumbert (2002). An exemplar model has the advantage of storing whole word memory traces, accounting for specific lexical effects, as in a near-merger. Kirchner (2012) presents several arguments for an exemplar-based account of sound change, which addresses the issues of incremental sound change, frequency sensitivity of lexical items, recency effects in imitation, individual and socio-phonetic variation, lexical diffusion, and generalization across the lexicon. Kirchner also suggests that “phonological units such as segments are simply local patterns obtaining over speech signals, involving relatively stable correlations between auditory cues and articulatory gestures. Such units, like all phonological patterns, should emerge bottom-up from comparison over the exemplars, rather than being treated as primitives” (341). Other exemplar accounts allow greater degrees of abstraction, which would be necessary to in order to associate one’s own speech gestures with memory traces of others’ productions, as in Fowler et al (2003). Hypothetically, speech gestures from one’s own productions that most closely match the acoustic pattern associated with the category of incoming exemplars could be activated during speech processing, accounting for speed and imitation while shadowing.
Memory traces of words would also be tagged with talker-specific information (such as socio-indexical information), which would allow for specific knowledge about other talkers’ productions to guide categorical perception of the incoming acoustic pattern. This function could account for perceptual discrimination without contrast in production, as in Warren et al (2007).

In studies that have attempted to combine perceptual learning with changes in production, the results have been mixed. Kraljic, Brennan, and Samuel (2008) measured participants’ productions of words containing /str-/ clusters before and after completing a task that was designed to elicit perceptual learning of an ambiguous s~ʃ sound, but observed no significant change in production of /str-. Perceptual learning took place through a speeded lexical decision task, in which ambiguous segments were presented in the context of words, and participants had to quickly choose if what they heard was a word or a non-word; therefore, no shadowing or imitation of the model talkers was elicited before the post-training production task. The hypothesis was that perceptual learning on its own might also effect a change in production, as would be expected by a gestural account, such as Browman & Goldstein (1992), or by any account that requires parity of representation across perception and production, such as the mechanistic convergence discussed in Pickering and Garrod 2004. Kraljic et al (2008) explained the failure of participants to alter their /str/ pronunciation as evidence against motor theories of speech processing.
Convergence is not always an automatic consequence of conversational interaction, much less passive listening. We may hear any number of dialects and idiolects but not adopt pronunciation variants that we hear. For example, if a native English speaker listens to a Japanese learner of English who does not distinguish /l/ and /r/, the English speaker will need to adjust his perception in order to understand the Japanese accented speaker, but will not merge his own productions of /l/ and /r/, especially if he is only listening to a recording, as is generally the case in a perceptual learning experiment.

Perceptual adjustment is not unique to speech; rather it is a way of fine-tuning any kind of perception to be more effective and efficient. Perceptual learning is a long-studied mechanism in the psychology literature, in which learning may be defined as “an increase in the specificity of an identifying response or […] a decrease of the class of items that will elicit the response” or any of the following: “progressive change in acuity, variability, and accuracy of perception” (Gibson and Gibson, 1955, pp. 37, 39). The mechanism behind convergence is also not limited to speech, but also occurs with body movements, gestures, facial expressions, laughter, pausing, even walking or foot-tapping (Chartrand et al., 2005). Convergence necessarily requires both perception and production, but the perceptual system does not necessarily have to be re-tuned in order to repeat a sequence of sounds. And it is unclear whether perceptual learning requires anything more than the perceptual system to be involved.

Even in primary language acquisition, babies do not passively listen to speech sounds, but rather also produce them, first through babbling, and later through words, phrases,
and sentences. Unsurprisingly, children born with disorders of the physical speech production system, such as cleft palate or esophageal fistula, are reported to have delayed language acquisition, both in production and perception, even when corrective surgery is performed relatively quickly (Kuehn and Moller, 2000). Changes in production that are triggered by adjusted auditory feedback require multiple trials to level out at the new target production or to return to the original state after feedback reverts to normal (e.g., Munhall et al., 2009), which suggests that production changes do not occur concomitant with changes in perception, but rather occur after the two systems have taken some time to negotiate a new target that is different from the practiced one. Thus it is not unexpected that simply hearing a novel pronunciation would not lead to convergence, even if it has been correctly identified as belonging to a particular sound category. In order to study sound change in the laboratory, it is important to be able to elicit changes in both perception and production, but we are still in the early stages of discovering just how the two systems are connected.

Most language use occurs in interactive situations involving both listening and speaking; therefore, it would be odd to expect sound change to develop in the absence of either listening or speaking. Harrington (2012) suggests “that imitation is a consequence of cooperative interaction between two individuals and that sound change may be a derivative of such imitation” (330). Under an exemplar-theoretic approach, I hypothesize that using shadowing techniques will not only enhance perceptual learning, but also allow generalization of production changes to extend to novel words and to persist post-
training. The reason for this is simple: if a listener hears a word with a variant pronunciation once, then one exemplar is created. If the same listener then also says the word, with some degree of imitation, a second exemplar is created, along with an associated configuration of gestures, and if this is repeated for a number of words, it should strengthen the neural connection between the sub-lexical acoustic pattern and gestural configuration. Once a sufficient accumulation of exemplars containing acoustic and gestural information is established, I hypothesize that the innovative variant should be able to be generalized to new words containing the target sound. For these reasons, the current experiment was designed to elicit both perception and production of variant pronunciations during the training phase, in order to see if sound change, entailing changes in both perception and production, could be created – and thus studied – in a laboratory setting.

2.1.5 Source of variation

For this experiment, existing variation in pronunciation of the stop+ approximant /tw/ cluster in American English was used as a basis for pushing underlying phonetic variation to the point of sound change. Approximants have long been known to increase the degree and length of aspiration in preceding stops, which sometimes can lead to the development of affricates (Sievers, 1901). In American English, an alveolar stop may become an alveo-palatal affricate, [ʧ], before /j/, as in congratulations, mature, and infatuated. Most, if not all, American English speakers also palatalize and affricate /t/ preceding the rhotic approximant /ɾ/, as in truck, trap, etc., as evidenced by children’s
spellings *chrac, chrap*, etc. (Treiman, 1985), and seen in the productions in this study. These examples suggest that affrication is a process that continues to be active in English. In /tw/, the lip-rounding that accompanies /w/ may spread to the preceding stop, which, by lengthening the cavity in front of the constriction, may create the percept of a retracted /t/. If the constriction is narrowed, increased frication may develop, and the resulting sound may be similar to an alveo-palatal affricate [tf]. However, the /t/ could also be produced with a dental place of articulation and less rounding, developing more anterior frication, yielding something more like [ts], which may be found in some English dialects, such as London Cockney, before high front vowels (Ladefoged 2001, 145-6). This front variant has also been the result of sound change of /t/ preceding a high back vowel in Japanese (Itō and Mester, 1995), and in German, which developed [ts] from /t/ during the High German Consonant Shift, including in /tw/ clusters. After the /w/ changed to /v/ during Middle High German (Schmitt, 1970), forms such as modern German Zwilling /tsvɪlɪŋ/, zwei /tsvaɪ/, etc., remain. In Greek, *tw changed into /s/, presumably through an intermediate [tsw] form (PIE *twe * (you ACC) -> Gk. σε [se]). These examples demonstrate that both a front and a retracted variant could “naturally” arise from the coarticulation of /t+w/. These two variants were used in separate training conditions as the target of sound change. A third control group heard plain alveolar [tw].

Another kind of affricate that could develop from /t/ before /w/ is the heterorganic [tf], which is attested in Northern Sotho (Johnson 2003, p. 144) and some other Bantu languages, in the same environment (that is, preceding a high back vowel) that
conditioned /t/ -> [ʦ] in some languages and t -> [ʧ] in other languages (Guthrie, 1967). Heterorganic affricates develop frication at a secondary constriction, rather than the primary place of articulation; in this case, the lip-rounding allows turbulence to be produced at a constriction of the upper teeth and lower lip. An intermediate production also exists, and is much more difficult to tell apart from plain [tw]: an affricate in which the frication originated from the lips, but without labio-dental contact, [twʰ], which could be considered a less affricated version of [tfw] or perhaps a slightly noisier version of [tw]. During the design phase of the experiment, the [tfw] and [twʰ] variants were not considered, but they did show up in some participants’ pronunciation. For the purposes of the experiment, only the affricates resulting from frication at the primary constriction ([ʧ] and [ʦ]) are used as training stimuli.

In addition to setting up a methodology for studying sound change, the following questions were addressed by the research: Are there differences in adoption of a gradual sound change across familiar and unfamiliar words (as suggested by the results in Goldinger 1998, and postulated as a factor in sound change as early as Schuchardt 1885)? Does the sound change generalize to novel words in production, as expected by the Neogrammarians Principle, or is it learned one word at a time, as in the case of lexical diffusion as dialect borrowing? Does the perceptual component of sound change likewise generalize to novel words and novel talkers (as in Kraljic and Samuel 2005)? Is the sound change generalized beyond the original conditioning environment (also as in Kraljic and Samuel 2005), suggesting general involvement of the abstract sound category rather than
an immediate association with a specific environment? Are there effects of language specific constraints, such as the existence of neighboring phoneme categories (as in the difference in results between voiced and voiceless stops in Neilsen 2007)? This would suggest greater involvement of cognition in the early adoption of sound change. Are there social constraints, such as the role of gender, on the adoption of different pronunciation variants (as would be suggested by the results of Strand 1999, etc.)? Does a shift in perception result in a shift in production (as would be predicted by gestural theories such as Liberman et al. 1967 or Fowler 1986, but not by observed results in near-mergers as in Warren, Hay, and Thomas 2007)? What is the role of the individual’s baseline production on adoption of a sound change? That is, if language users already display a shift in production, do they continue to shift or reverse course when encountering similarly shifted speech? And if they have a different variant, are they more or less likely to adopt the new variant?

2.2 Methodology

The experiment was a sequence of several different tasks, including a production task that was done before, during, and after the training (to test the degree of convergence in production), a shadowing task that constituted the training phase, a lexical decision task and a final identification task (to test whether there was perceptual learning).

2.2.1 Materials

The words used in training/shadowing and words and non-words for the post-training lexical decision task were all naturally produced stimuli, elicited from Linguistics
students from the Ohio State University. Talkers first read over a printout of all the words and non-words, and explanations of the target variant pronunciations, and prior to recording, each talker was encouraged to practice the variant pronunciations, with varying amounts of training and rehearsal required for each talker. Variant pronunciations were written out using <chw> for the retracted variant (e.g., chwin), and <tsw> for the front variant (e.g., tswin), while the unaffricated plain [tw] was represented with normal orthography (e.g., twin). The words were presented one at a time on a computer screen, and talkers were recorded saying the words with a slightly falling intonation, in a sound booth, using a desktop microphone to record directly into digital format using Audacity on a PC running Windows XP. All of the words were segmented in Praat (Boersma and Weeninck 2005) and normalized to 70dB and downsamplied from 44,100 Hz to 22,050 Hz so that they would be compatible with early versions of E-Prime. 8 talkers (4 men, 4 women) were used as model talkers for the perceptual learning/shadowing/training, and an additional 4 talkers (2 men, 2 women) were used for the lexical decision task.

Because there are only a handful of very familiar /tw-/ initial words in the English language, the initial difficulty was in finding a way to elicit and test perceptual learning. The traditional methods for inducing perceptual learning include lexical decision tasks (e.g., Norris et al. 2003) and passive listening (Maye et al. 2008), in which is it certain that listeners are familiar with the words they hear. Then testing whether perceptual learning has taken place may use lexical decision or identification tasks on other familiar but novel words. With only a handful of familiar /tw/ words to split up, there were not
enough to ensure sufficient training and comprise a robust sample for post-training measurements. So a wider variety of high and low familiarity words had to be used. Suitable words beginning with the $tw$- cluster, and their definitions, were sought out in the Oxford English Dictionary. 14 archaic, obscure, or otherwise unfamiliar $tw$, $en$, and $vi$- words were chosen to be presented during training, as well as 14 each of more familiar $tw$, $en$, and $vi$- words, for a total of 28 each $tw$, $en$, and $vi$- words. Furthermore, 12 somewhat familiar $tw$, $en$, and $vi$- words were used for the post-training lexical decision task, in addition to the trained words, for a total of 40 each $tw$, $en$, and $vi$- words. In order to construct a plausible experiment format, it was decided that the words should be presented in the guise of a vocabulary learning task, in which participants could be exposed to familiar words with a variant pronunciation, while also learning new words, which also contained the variant pronunciation. In addition to $tw$-words, the words beginning with $en$, and $vi$- were introduced as fillers so that participants would not immediately guess the purpose of the experiment.

As much as was possible within these limits, the words were distributed evenly among trained and untrained words with regards to familiarity, pronunciation variant ($chw$- and $tsw$-), number of syllables, and following vowel/rhotic/palatal. An equal number of $tw$, $en$, and $vi$- non-words for the lexical decision task were created, in part, by replacing phonemes or diphones in training words, to make the task challenging. Some of the non-words were created by putting together consonants and vowels in phonotactically legal combinations. All non-words began with $tw$, $en$, or $v$. 
Stimuli for the identification task were created by selecting one male and one female talker from among those whose pronunciations were used in the lexical decision task. For both talkers, one token each of *two, chew, and tsu* were carefully selected, matching closely in affricate/stop duration, vowel formants and duration, and intonation. Two continuua for each talker were created by blending an intensity and duration normalized stop and affricate (/t/+/ts/ and /t/ + /ʧ/), in 16.67% steps, which were re-spliced onto the vowel from the word *two*. The end result was a male continuum and a female continuum, each ranging from *chew to two to tsu.*

2.2.2 Procedure

The first task was a lexical familiarity rating task, which included a production component. Participants were presented visually with a subset of the familiar and unfamiliar training words (14 each *en-, vi-, and tw-* words in a range of familiarities) plus a subset of untrained words (7 each *en-, vi-, and tw-* words, plus words containing initial *tV-, tr-, str-, ts-* (*tsu*) and *ch-* (*chew*), as well as medial *tw-, tr-, and str-* for eventual comparison), one word at a time on a computer screen, and were instructed to say each word out loud and then rate how familiar they were with that word by clicking the box next to the appropriate rating, as follows:

5. very familiar – I know this word and use it
4. somewhat familiar – I know this word, but I may or may not use it myself
3. neither familiar nor unfamiliar – I may know this word, but do not use it
2. somewhat unfamiliar – I may have heard this word before, but have never used it
1. very unfamiliar – I have never heard or used this word before
The productions were recorded using a headset microphone, which participants wore during the entire experiment, at a 44,100 Hz sampling rate using Audacity. All experimental stimuli were presented and mouse click responses were recorded using E-prime 1.1.

The second task was the perceptual training. Participants were randomly assigned to one of three different conditions. A control group heard normally pronounced /tw/. The front affricate training group heard /tw/ words pronounced with the front affricated [ʦw] variant. For ease of reference this group will be hereafter referred to as the tsw- group. The retracted affricate group heard words pronounced with the retracted affricated [ʧw] variant. For ease of reference, this group will be hereafter referred to as the chw- group. Participants were told that they would learn words from three different sections of the dictionary, including en-, vi-, and tw-. They were also told that words they already knew from each section would also be presented, to aid in learning by associating new words with the old. Training was divided into alphabetical blocks, with en- words presented first, followed by vi- and then tw- words. Participants were instructed that they would see the word on the screen and hear it pronounced, at which time, they should say the word out loud themselves. Next they would see a slide with the word written on top, which they heard pronounced another time, followed by a definition of the word or, upon encountering the word for a second time, a sentence using the word in context, both of which should be read silently. Then they would again see the word, hear the word, and say the word, before progressing onto the next word. Definition trials preceded sentence
trials. Multiple choice quiz questions were asked every 5-10 words, in which a definition was provided, and they had to choose from one of 4 words, to make sure they were paying attention and learning the words. Each word was pronounced by a model talker three times in a definition trial, and 3 times by talker of another gender for the sentence trial, for a total of 6 times, by 2 different talkers. The words were to be pronounced out loud by the participant twice on each trial, for a total of 4 times each. The sequence of events on a given trial is illustrated by the diagram in Figure 2.1.

Figure 2.1: Diagram showing training procedure, in which a participant 1) sees, hears, and then says the word out loud, 2) sees and hears the word and then reads a sentence or definition silently, and 3) sees, hears, and then says the word out loud again.

The third task was a lexical decision task, in which words and non-words containing en-, vi-, tV-, and both variant pronunciations of tw- were presented binaurally over
headphones. Participants indicated whether the stimulus they heard was a word or non-word by pressing the appropriately labeled button on a 5-button serial response box.

The fourth task was a post-training production task. All of the tw-, en-, and vi- words from training plus the rest of the tw- words from the lexical decision task, plus the tV-, tr-, str-, ts- (tsu) and ch- (chew), words from the pre-training production list, were displayed on the screen, one at a time for 2 seconds. Participants were instructed to say each word out loud as quickly and naturally as possible.

For the final section, participants completed an identification (ID) task, in which they had to decide whether an ambiguous stimulus sounded most like the word chew, two, or tsu. The ID task was designed to measure changes to the perceptual category for /t/ before /u/, to see if the perceptual learning of the /tw/ pronunciation variant could be generalized to a new conditioning environment.

The entire experiment, including filling out paperwork, lasted anywhere from 65 minutes up to a maximum of two hours. Almost all participants were finished within 90 minutes, but those who took longer than 105 minutes were excluded for falling asleep.

2.2.3 Participants

Participants in this experiment received participation credit in partial fulfillment of a course requirement for introductory linguistics classes. The experimental protocol and

---

6 The sound booth was very warm and the experiment was very long. It is possible that others nodded off, but were not noticed because they did not exceed the normal time allotment.
materials were reviewed and approved by the Ohio State University IRB, in compliance with human subjects research standards.

A total of 109 students participated in the sequence of five tasks that included the familiarity rating and vocabulary training tasks. In addition, 17 students participated in just the final identification task, to provide a baseline control set of responses because the identification task could serve as an additional training of sorts, by exposing control group participants to both /tw/ variants. 30 participants were excluded from the results as non-native English speakers, 2 were excluded for speech and/or hearing disorders, and 6 were excluded for experimental errors or for falling asleep during the experiment. The remaining 72 participants included 23 men and 49 women between the ages of 18 and 50 (mean=21.3, sd=4.3), divided into 21 participants in the \textit{tsw-} condition (13 female, 8 male), 31 participants in the \textit{chw-} condition (24 female, 7 male), and 20 participants in the control group (12 female, 8 male). All of these 72 participants were included in the analysis of the lexical decision and identification tasks.

A subset of these participants was quasi-randomly selected for acoustic analysis by drawing from a list containing an identifying number, sex, and training condition. 7 participants of each gender from each of the variant training condition groups plus 5 each from the control group were selected, for a total of 19 men and 19 women.
2.2.4 Familiarity and Frequency Analysis

Familiarity rather than frequency was used in the analysis of this experiment because most of the obscure *tw*-words obtained a frequency of 0 in a search of several spoken and written corpora (Brown Corpus (Kučera and Francis 1967); Open American National Corpus (Ide 2008); Buckeye Corpus (Pitt et al. 2007); Google (using a simple Google search to find how many web pages the searched-for word occurred on)), and the range of frequency of the remaining words had an extremely large spread. Thus, frequency would not be a robust or reliable measure for these words.

Familiarity ratings from the first section of the experiment were averaged across participants to gauge the familiarity of these items. Words that were not rated in the first task were given familiarity ratings based on the Hoosier Mental Lexicon by converting from a 7-point to a 5-point scale, using the formula in (1), which allowed a very close match between words rated in the present experiment and the converted HML equivalents (F(1,48)=258, p<0.05, slope=0.92, r²=0.84).

\[
(1) \quad x = 0.68 \times \text{HML} + 0.2
\]

It was decided that the best way to examine familiarity and frequency differences of the words was to simply split the words into familiar and unfamiliar words based on whether the word scored lower than 2.5 on the 5-point scale. It has been found that familiarity is correlated with spoken word frequency, but that the relationship is more
stable for higher frequency words (Tanaka-Ishii and Terada 2011), so it is not unreasonable to split the words into only two levels of familiarity.

2.2.5 Acoustic Analysis Methods

Training and lexical decision stimuli were measured and used as comparison for participants’ /tw/ words, to chart the amount and direction of imitation of the pronunciation variant, and whether the effects would persist in post-training productions, and if subjects would extend the pronunciation variant to untrained words. All segmentation and measurements were performed using Praat (Boersma and Weeninck 2005). The following landmarks were marked: burst – the initial burst or in the absence of a burst, the onset of the /tw/ sound as indicated by increased perturbation of the waveform; VOT – the onset of voicing, as indicated by periodicity in the waveform; w-onset – the onset of the following approximant, as indicated by the presence of at least three formants in the spectrogram; v-end – the end of the vowel (or diphthong or triphthong) following the approximant; and end – the end of the word.

Participants’ recordings were downsampled to 22,050 Hz so that their productions could be directly compared with productions they heard during training. A number of measurements were taken to see what could best distinguish trainers’ productions across the different pronunciation conditions. Three measurements were chosen that best discriminated among training variants and plain [tw]. Center of gravity, also known as
centroid frequency, which is the weighted average frequency of the spectrum (Figure 2.2), was calculated over five overlapping 20 ms windows centered from 20%-30% of the duration of the stop/affricate. The 25% point was used as a landmark because affricates tend to show greater differences between place of articulation in the frication portion than at the burst (Recasens and Espinosa 2007). The greater the constriction, the less the lower-frequency resonances of the open vocal tract are noticeable in the sound wave.

Spectral slope (Figure 2.3), takes the amplitude of one frequency band subtracted from another band or the entire spectrum. In this case, the slope was calculated for a high range of 3600-6700 Hz for men and 4500-7400 Hz for women, minus the corresponding low range 0-3600 Hz or 0-4500 Hz, also over five 20 ms windows centered from 20-30% of the duration of the stop/affricate. The reason these limits were chosen is that in the trainers’ affricated productions, bands of stronger intensity could be found starting around 3600 Hz for men and 4500 Hz for women.

It is unclear whether the frequency differences are a result of anatomical or socially acquired differences, or, more likely, some combination of the two. The front affricated *tsw*- variant has greater amplitude in this range plus another strong band of amplitude in higher frequencies, while the retracted affricated *chw*- variant has increased amplitude in this mid-range band, and very little amplitude in higher frequencies, which means the difference between slope and centroid should help to distinguish the two affricated variants. But also, the stronger the low-frequency band, the lower (more negative) the slope will be, because the token is less affricated. Normalized rise time (Figure 2.4) is the
Figure 2.2: Center of gravity measurement taken from five 20 ms windows around the 25% point of stop/affricate produced by trainers.

Figure 2.3: Spectral slope as calculated by subtracting low frequency amplitude from mid to high frequency amplitude for training and lexical decision stimuli.
time from the beginning of the burst to the point of highest amplitude in the stop/affricate, divided by the duration of the stop/affricate (from burst to VOT). Generally, stops have a very short rise time as the constriction is quickly released, while affricates take longer to reach peak intensity (Stevens 1998, 117-8).

Figure 2.4: Normalized rise time of training and lexical decision stimuli: rise time (time from onset until peak intensity) divided by the duration of the stop/affricate (time from burst to VOT).
In order to ascertain whether the participants exhibited any change in pronunciation during shadowing and if they maintained it after training, a measurement of the acoustic distance between each /tw/ token and the average of the training stimuli was calculated as follows: Each participant’s measurements for rise time, slope, and center of gravity were normalized against the appropriate training condition tokens, using the z-score method, which calculated the difference of each token from the trainers’ mean, divided by the standard deviation of the trainers’ values, as shown in equation (2). This step essentially captured the individual distance of the rise time, slope, and centroid, from the averages of the training stimuli, resulting in 3 z-scored values for each word.

\[
(2) \quad z = \frac{x - \bar{x}}{sd}
\]

Then a Euclidean distance function was used to combine the three measurements, by squaring each z-score and adding them, then taking the square root, as in equation (3).

\[
(3) \quad \sqrt{slope^2 + rise^2 + centroid^2}
\]

The resulting numbers are a normalized measure of acoustic distance between each instance of /tw/ uttered by each participant and the mean value of the trainers’ variant
/tw/ pronunciations. The larger the number, the greater the distance between shadowers’ and trainers’ productions.

2.3 Results

2.3.1 Production Measurements – training and lexical decision stimuli

Results of the acoustic analysis of the trainers’ and lexical decision talkers’ productions were combined for statistical analysis. Center of gravity measurements (Figure 2.2) revealed no difference between sexes, or stimulus type (training vs lexical decision), but only a main effect of pronunciation variant $F(2,16)=38.6$, $p<0.05$, with highest values for $tsw$- and $chw$, and lowest values for plain [tw]. Spectral slope measurements (Figure 2.3) showed no difference between sexes or stimulus type, but a main effect of variant: $F(2,16)=72.8$, $p<0.05$, with three ranges for each $tsw$-, $chw$-, and plain [tw]. Normalized rise time (Figure 2.4) showed a main effect of condition ($F(2,16)=64.1$, $p<0.05$) and sex ($F(1,8)=5.5$, $p<0.05$). $tsw$- and $chw$- affricated variants had the longest rise time, while plain [tw] had generally shorter rise times. T-tests showed differences between men’s and women’s normalized rise time for $chw$- stimuli ($t(9)=4.5$, $p<0.05$), and a marginal difference in rise time between men and women’s [tw] productions ($t(9)=2.04$, $p=0.074$). It is also pertinent to note at this point that some of the trainers’ productions of plain /tw/ show indications of affrication in all three measures, probably as a result of their extensive training with the variants before recording, although it is unclear where the frication is located (at the primary or secondary place of articulation - that is, the alveolar ridge, palate, or lips). Average Euclidean distances from
each stimulus to the group mean of the training stimuli, according to pronunciation variant, is listed in Table 2.1, so that participants’ /tw/ productions can be compared to the relevant trainers’ productions. The closer the participants’ productions to the trainers’, the more convergence is demonstrated.

Table 2.1: Average Euclidean distance of individual training stimuli from the group average. This measure illustrates how much dispersion existed in the training stimuli and gives a reference point for location of trainers’ productions.

<table>
<thead>
<tr>
<th></th>
<th>Mean Euclidean distance from trainers’ average</th>
</tr>
</thead>
<tbody>
<tr>
<td>chw- training stimuli</td>
<td>1.56</td>
</tr>
<tr>
<td>tsw- training stimuli</td>
<td>1.57</td>
</tr>
<tr>
<td>[tw] stimuli distance from chw- stimuli</td>
<td>3.75</td>
</tr>
<tr>
<td>[tw] stimuli distance from tsw- stimuli</td>
<td>4.56</td>
</tr>
</tbody>
</table>

2.3.2 Participants’ production results – baseline

As a baseline for reference, participants’ initial pronunciations of three consonant clusters (/tr/, /tw/, and /str/) was judged by listening to the pre-training recordings. A summary is outlined in Tables 2.2 and 2.3.

All participants regularly used the retracted affricate [ʧr] for /tr/, with three participants using an aspirated stop in a small minority of words. These participants, though not only these participants, exhibited other indications of hyper-articulation, such as medial or final aspirated [tʰ]. The other 35 participants did use an affricated /tr/ in
every case (10 word-initial instances plus 4 medial instances, with a wide range of familiarity).

Participants who frequently (i.e., more than a third of the time) used an affricated variant of /tw/ in their baseline productions will be referred to hereafter as early affricators, or simply affricators. Six participants frequently used the retracted [ʧw] variant for /tw/, while two male participants frequently used the front affricated [tsw]. These two front affricators showed much less affrication than the tsw- training stimuli, however, while the [ʧw] affricators’ productions were more similar to the trainers’ fully affricated productions. Four participants used some combination of [tsw], [ʧw], [tfw] and [twh], and another generally favored [tfw]. Another participant generally used plain [tw] but for one specific word, tweak, used [ʧw], which he continued to do post-training. This participant is not counted as an affricator in Table 2.3. The distribution of the affricated pronunciation variants by word appears to be influenced by a complex interaction of following vowel, following rhotic or other palatal sound, familiarity, and lexical neighborhood, but some unknown factors also seem to be at work. Broadly generalizing, /i/ conditioned the greatest affrication among front [tsw] affricators, while /e/ and /a/ conditioned more affrication among retracted [ʧw] affricators. None of the /tw/ affricators used any affricated variant exclusively, but rather also used plain or heavily aspirated [tw] for one or more words.
Additionally, eleven participants showed evidence of some /s/ retraction in /str/ clusters; only four of these were also /tw/ affricators. A case study on each of these individuals is planned for a later date.

Neither /str/ retraction nor /tw/ affrication had reached the point of a complete change across the entire vocabulary of sampled words. It seems fair to say that /tr/ affrication is a sound change nearly completed, /str/ retraction is in some state of variability, though it is arguable whether it is an ongoing change-in-progress (e.g., Durian 2007), and /tw/ affrication is sporadic enough to say that it is at a point of stable (or possibly unstable) underlying variation, but perhaps not yet a sound-change-in-progress.

Table 2.2: Distribution of pronunciation variants for /tr/ and /str/ from participants’ baseline recordings, separated by training condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>/tr/ realization</th>
<th>/str/ realization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[ʧʃr]</td>
<td>[tr]/[ʧʃr]</td>
</tr>
<tr>
<td>Front tsw-</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Retracted chw-</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Control tw-</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.3: Distribution of pronunciation variants for /tw/ from participants’ baseline recordings, separated by training condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>/tw/ realization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[tw]</td>
</tr>
<tr>
<td>Front tsw-</td>
<td>10</td>
</tr>
<tr>
<td>Retracted chw-</td>
<td>9</td>
</tr>
<tr>
<td>Control tw-</td>
<td>6</td>
</tr>
</tbody>
</table>
Measurements of participants’ baseline pronunciations of /tw/ are shown in Figures 2.5-2.7. Male participants had longer rise time for baseline /tw/ productions than female participants (t(27)=2.2, p<0.05), which is similar to the pattern in the training stimuli. There is no known physiological reason why men would exhibit longer rise time in /tw/.

Additionally, while not statistically significant, the trend appears that plain [tw] stimuli produced by trainers of both genders, were already slightly affricated, relative to participants’ baseline productions, as can be seen by comparing figures 2.5-2.7 to Figures 2.2-2.4. Measurements for plain [tw] stimuli and participants’ baseline productions, summarized in Table 2.4, indicate that the rehearsal of the affricated /tw/ variants impacted trainers’ productions of plain [tw] stimuli to some degree.

Table 2.4: Comparison of plain [tw] training stimuli and participants’ baseline /tw/ productions. Training stimuli show slightly more affrication than participants baseline productions.

<table>
<thead>
<tr>
<th></th>
<th>Rise time</th>
<th>Slope</th>
<th>Center of gravity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain [tw] stimuli</td>
<td>0.113</td>
<td>-6.43 dB</td>
<td>2667 Hz</td>
</tr>
<tr>
<td>Participants’ baseline /tw/ productions</td>
<td>0.109</td>
<td>-8.17 dB</td>
<td>2041 Hz</td>
</tr>
</tbody>
</table>
Figure 2.5: Center of gravity measurement taken from five 20 ms windows around the 25% point of stop/affricate produced by participants before training.

Figure 2.6: Spectral slope as calculated by subtracting low frequency amplitude from mid to high frequency amplitude for participants’ baseline /tw/ productions.

Figure 2.7: Normalized rise time of participants’ baseline productions: rise time (time from onset until peak intensity) divided by the duration of the stop/affricate (time from burst to VOT).
Rise time, center of gravity, and spectral slope were also measured for /tr/ and /ʧ / (in the word *chew*) as displayed in Table 2.5, to illustrate the characteristics of the phonemic alveopalatal affricate and the affricated variant of /tr/. Men exhibited lower centroid values than women for both /ʧ / (t(33)= 6.5, p<0.05) and /tr/ (t(35)=4.6, p<0.05). These differences can possibly be explained by the fact that men have larger lips (e.g. Ferrario, et al. 2009, Vorperian et al. 2009), which, when rounded, could extend the cavity in front of the constriction, resulting in a lower centroid frequency; however, it is also possible that men do have more retracted productions for both sounds, as they do for /s/, as discussed in, e.g., Strand 1999. Rise time between men and women was not significantly different for /ʧ /, but men exhibited significantly longer rise time than women in their /tr/ productions (t(34)=2.3, p<0.05), which paralleled their /tw/ productions. Spectral slope for both /tr/ and /ʧ / was not significantly different between men and women. By comparing these measurements with the measurements for *chw*- shown in Figures 2.2-2.4, the similarity of these sounds is evident: longer rise time, higher center of gravity, and greater spectral slope are all indicative of affrication.

<table>
<thead>
<tr>
<th></th>
<th>Rise time</th>
<th>Center of gravity</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tr/ men</td>
<td>0.379*</td>
<td>2939 Hz*</td>
<td>-2.42 dB</td>
</tr>
<tr>
<td>/tr/ women</td>
<td>0.296*</td>
<td>3644 Hz*</td>
<td>0.12 dB</td>
</tr>
<tr>
<td>/ch/ men</td>
<td>0.480</td>
<td>3462 Hz*</td>
<td>3.27 dB</td>
</tr>
<tr>
<td>/ch/ women</td>
<td>0.554</td>
<td>4289 Hz*</td>
<td>5.91 dB</td>
</tr>
</tbody>
</table>
2.3.3 Participants’ production results – effect of training

2.3.3.1 Control group

Figures 2.8 and 2.9 show the Euclidean distance measures, as explained in equations (2) and (3), for the control group participants, relative to the chw- training stimuli (Figure 2.8) and the tsw- training stimuli (Figure 2.9). Both chw- and tsw- comparisons are made so that participants in both of those training conditions can be compared to the control group, to demonstrate that the changes observed as a result of training are not an artifact of the experimental procedure, but rather only occur in certain training conditions. The results are further divided by the time in the experiment the word was produced, whether or not the word was heard during shadowing, and whether the word was familiar or completely unfamiliar. The unfamiliar words are those 14 tw- words described in the stimuli section that most participants had never encountered before the experiment, such as twayblade and twissel. Words that were not heard and pronounced during training are referred to as untrained or novel words. The y-axis, showing Euclidean distance of participants’ productions from the training stimuli, is held constant across Figures 2.8-2.11 for easier comparison across training conditions.
Figure 2.8: Distance between \textit{chw-} training stimuli and the control groups’ /tw/ productions before, during, after training, separated by novelty and familiarity. The gray dashed line shows the average baseline measurement, to facilitate comparison of shadowing and post-training productions to the baseline measurement. The black dashed line indicates the Euclidean distance of the control condition [tw] stimuli from the \textit{chw-} condition stimuli, and indicates the target that participants in the control condition heard during shadowing.
Participants in the control condition do appear to affricate more during shadowing than during baseline, as could be expected due to the trainers’ mild affrication as seen in Figures 2.2-2.4. Control group participants also appear to retain some of the affrication after training. Paired t-tests, however, do not show a significant difference between participants’ average Euclidean distance before and during shadowing, or before and after shadowing (relative to either chw- stimuli (Figure 2.8) or ts\\w- stimuli (Figure 2.9)). Differences when split by familiarity and novelty of stimulus are also not significant. If
the control group participants were measured relative to the control group training stimuli, these differences might well be significant, but the purpose of the control group is to illustrate that the changes in the speech of the two variant trained groups is not just an artifact of the procedure.

2.3.3.2 Retracted affricated (\textit{chw}-) training condition

As can be seen in a comparison of Figures 2.8 and 2.10, control group participants and \textit{chw}- trained participants started out with a similar average baseline distance from the \textit{chw}- training stimuli. However the average degree of convergence for the \textit{chw}- group (Figure 2.10) is significant, while the control group’s is not. The average of \textit{chw}-condition participants’ productions were significantly closer to the trainers during shadowing than baseline ($t$(13)=4.04, $p$<0.05, mean difference=0.83) and were still significantly closer to trainers after training than at baseline ($t$(13)=2.3, $p$<0.05, mean difference=0.42), demonstrating convergence during shadowing, and persistence of pronunciation changes after training. Words that were not heard and said during shadowing were significantly closer to trainers after training than at baseline ($t$(13)=2.99, $p$<0.05, mean difference=0.62), indicating generalization of pronunciation changes to novel words; however, trained words, on average, were only marginally closer to trainers after training relative to baseline ($t$(13)=1.86, $p$=0.085, mean difference=0.33). Splitting up trained words into familiar and unfamiliar words reveals a significant difference in baseline pronunciations, with familiar words beginning closer to trainers than unfamiliar words ($t$(13)=2.5, $p$<0.05, mean difference=0.22). Both familiar trained words
(t(13)=3.25, p<0.05, mean difference=0.72) and unfamiliar trained words (t(13)=4.08, p<0.05, mean difference=0.93) were significantly closer to trainers during shadowing than baseline, illustrating convergence during shadowing. While there is no significant difference between familiar and unfamiliar words during shadowing, the difference in their rate of change between baseline and shadowing is 0.21.

Figure 2.10: Distance between chw- training stimuli and chw- condition participants’ /tw/ productions before, during, and after training, separated by whether the word was heard and repeated during training, and familiarity of the words. The horizontal gray dashed line shows the average baseline measurement, to facilitate comparison of shadowing and post-training productions to this original baseline measurement. The average Euclidean distance of the chw- training stimuli is 1.56.
2.3.3.3 Front affricated (*tsw*) training condition

Figures 2.9 and 2.11 display similar average baseline measurements for control group and *tsw*-condition participants, but the *tsw*-group (Figure 2.11) demonstrates a significant change in average /tw/ production between baseline and shadowing \(t(13)=3.2, p<0.05, \text{mean difference}=0.66\), showing convergence. The average change between baseline and post-training is not significant (mean difference=0.10), but indicates only a slight degree of persistence of pronunciation changes. Trained words, i.e., words that were heard and repeated during shadowing show significant convergence during shadowing \(t(13)=2.96, p<0.05, \text{mean difference}=0.65\), but are only 0.10 sd closer to trainers after training than at baseline (compare to control condition 0.13). Words that were not heard and said during shadowing are also not significantly closer to trainers after training than at baseline (mean difference=0.11, compare to control condition 0.04). When split by familiarity, unfamiliar trained words are closer to the training stimuli during shadowing than familiar words \(t(13)=2.12, p=0.05, \text{mean difference}=0.20\), though after training, familiar words are 0.16 sd closer to trainers than baseline, while the unfamiliar words show only a 0.08 sd difference.
Figure 2.11: Distance between tsw- training stimuli and tsw- condition participants’ /tw/ productions before, during, and after training, separated by whether the word was heard and repeated during training, and familiarity of the words. The horizontal gray dashed line shows the average baseline measurement, to facilitate comparison of shadowing and post-training productions to this original baseline measurement. The average Euclidean distance of the tsw- training stimuli is 1.57.

2.3.3.4 Discussion of production results by training condition

Both tsw- and chw- trained groups show convergence during shadowing, but only the chw- trained group shows significant persistence of pronunciation changes after training. Unfamiliar words show greater pronunciation changes during shadowing for both groups, which is consistent with an exemplar model’s predictions, in which those words with few or no memory traces are more flexible. However, generalization to novel words does occur, especially for the chw- group, which shows more post-training affrication of
untrained words than words that were heard during shadowing. This finding indicates some level of sub-lexical abstraction.

It is not unexpected that participants may generalize the new pronunciation variant to novel words, but what is unexpected is that the trained words do not continue to show greater effects of training than untrained words. According to an episodic model (e.g., Goldinger, 1998), we would expect the words that were shadowed to show the strongest influence of training. The fact that the words pronounced by the chw- trained group do not fit this pattern invites closer inspection. Additionally, according to an episodic account, words with fewer traces, such as lower frequency or unfamiliar words, should show the greatest effect of convergence, but the average familiarity rating for the familiar trained words is 4.05 with a standard deviation of 0.74 (out of 5, with 5 being very familiar, and 1 being completely unfamiliar), and 1.2 (sd=0.37) for the unfamiliar trained words, while the average familiarity rating of the untrained words is 4.66 (sd=0.46). However, the unfamiliar trained words and familiar untrained words both showed greater affrication after training than the familiar trained words, with the untrained words showing the greatest amount of change from baseline.

In general, affricates are characterized by having a longer duration of higher intensity noise than aspirated stops (that is, “a more prolonged period of frication after the [stop] release” Ladefoged and Maddieson 1996, p.90), which requires the articulators to retain a tighter constriction with greater turbulence in the air stream for a longer time. Reduction in duration and intensity generally correspond to unstressed or reduced forms (Gay
Reduction in an affricate might be realized as a faster release of the constriction, which would yield shorter rise time, reduced amplitude in higher frequencies, and more generally, reduced duration. The fact that the untrained words show greater affrication than the trained words may be due to a repetition effect creating reduction in the trained words, especially in the already familiar trained words, which reduces the degree of affrication, and creates the appearance of less persistence of pronunciation changes after training.

This hypothesis is upheld by looking at the post-release duration, indicated by VOT, before and after training (Table 2.6). Untrained familiar words (which have the highest familiarity: 4.66) have the shortest VOT in the baseline recordings, while trained unfamiliar words (mean familiarity=1.2) have the longest VOT. While familiarity does not directly equate to frequency, they are highly correlated in very familiar words, and it is known that the unfamiliar trained words have a near zero frequency count. Thus, the baseline VOT may reflect reduction effects due to familiarity. The VOT in both familiar and unfamiliar trained words is significantly shorter after several repetitions by participants in all conditions (t(51)=2.4, p<0.05), suggesting reductive processes due to repetition. This result parallels the shorter duration due to repetition seen by Fowler and Housum (1987). The untrained words do not show a significant reduction in VOT, which suggests that the lack of repetition is linked to a lack of further reduction. Thus, the effects of affrication gained during training may be lessened due to reduction. It is unclear why the tsw- trained group does not show a similar pattern, but they do pattern
more like the control group, so it may be related to the general lack of persistence of pronunciation changes after training.

Table 2.6: Post-release duration (VOT) of /t/ in tw- words, before and after training.

<table>
<thead>
<tr>
<th>Training Condition</th>
<th>Duration before (ms)</th>
<th>Duration after (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained unfamiliar words</td>
<td>122</td>
<td>101</td>
</tr>
<tr>
<td>Trained familiar words</td>
<td>111</td>
<td>97</td>
</tr>
<tr>
<td>Untrained familiar words</td>
<td>102</td>
<td>101</td>
</tr>
</tbody>
</table>

Goldinger (1998) and others have reported repetition effects that actually increase the similarity between participants and model talkers (up to a certain point), because as participants become more familiar with the pronunciation, they improve their ability to reproduce sub-phonemic aspects of what they have heard. But high frequency words seem to be less prone to convergence, and Goldinger (2000) and Goldinger and Azuma (2004) show an interaction between frequency and repetition in which shadowers’ productions of higher frequency words are judged less like the model’s when heard a dozen times before shadowing, but lower frequency words continue to increase in similarity with repetition. While the offered explanation for this frequency effect is that higher frequency words have a higher concentration of memory traces, episodes, or
exemplars, and are less likely to get dragged about by additional tokens, one might consider that the effect is compounded by the participants’ tendency to reduce high frequency and repeated words. This effect results from the participants’ hearing the target word multiple times, even though the number of times they actually produce it does not change, somewhat like the selective adaptation effects found for VOT, for example, in Cooper (1974). Hearing a word mentioned in conversation leads to reduction in the second mention by the listener-turned-talker (Bard, et al. 2000), and production of a word multiple times is also linked to a sharp decrease in duration for each of the first few repetitions, and thereafter steadily decreasing duration with increased mention (Aylett and Turk 2004), as is also seen in the difference in durations before and after repetition (Table 2.6) in this study.

Additionally, Babel (2012) showed a probable repetition effect of this sort in her Figure 4, which shows that the more times a participant repeated a word, the less imitation it reflected, with the exception of words containing the vowel /æ/, which became more raised with greater imitation. However, when vowels are reduced, not only are they shorter in duration and intensity, but they are also centralized (Lindblom 1963, i.a.). Notably, when the vowel /æ/ is reduced, it is raised, as seen in Lindblom (1963), but as also can be observed in common pronunciations of highly frequent words and phrases, such as can [ken] or [kən], thank you [θəŋkju], and and [ən]. In all likelihood, the repetition effect seen in Babel 2012 is due to processes of reduction, which are known to occur with repetition (see Fowler and Housum 1987, i.a.). Hyper-articulated /æ/ is also
raised (Lindblom 1963), however, so it is difficult to untangle the two processes if one only looks at formant values. Thus, one could view all of the results in Babel 2012 as evidence for an interaction between spoken repetition and convergence, in which the reduction resulting from repetition mitigates the imitation effect. All of these facts support the interpretation that reductive processes occurring on a lexical level interact with the affrication learned during training in the current project to decrease the persistence of convergence in the post training productions.

Another result that merits discussion is the lack of persistence in pronunciation changes in the front affricated *tsw*-trained group (although note the (not significant) change in familiar trained words). One clue to this result lies in anecdotal information from participants’ debriefing sessions. More of the front (*tsw*) training condition participants noticed that the *tw*-words were pronounced with a variant, and many were unhappy with the productions of the trainers, reporting that the /tw/ sounded more like /sw/, or that “they weren’t saying it right”, or “it sounded like an ‘s’ but I didn’t say it that way – I said it the right way.” It may also be supposed, based on the initial pronunciations of participants, that the retracted variant is more common. Only 6 out of 38 participants ever used the front variant in the baseline recording (and only 2 used it as the preferred variant), while 9 participants used the retracted variant (with 4 using it most often as their preferred variant). Additional evidence is anecdotal, but over the course of this study, several helpful lab members have sent the author links to youtube videos containing media personalities, such as Rachel Maddow and Michael Savage, who use the retracted variant, while only one youtube example of use of the front variant has been
passed along\(^7\). It is also notable that most trainers had a much easier time learning to pronounce the retracted variant than the front variant. Due to the common use of the retracted affricate created by /t+r/ and /t+j/, the analogy might make the retracted *tchw-* pronunciation feel ‘more familiar’.

2.3.4 *Individual Differences*

2.3.4.1 Early affricators and non-affricators

As mentioned earlier, some participants, who are being labeled as *early affricators*, displayed affrication of /tw/ before they were exposed to the training variants. Because the affricators were distributed unevenly across training conditions, they added a great deal of noise to the analysis, being different from each other and from the other participants in their groups, especially in their baseline pronunciations. Besides having different kinds of affrication (front, retracted, etc.) that were not evenly matched across conditions, each one had a different distribution of which words received which /tw/ pronunciation. It has been shown that convergence may be increased when the original distance between model and participant is greater (Kim, Horton, and Bradlow, 2011), which points to a probable difference in response for affricators and non-affricators, with non-affricators likely to show greater convergence to the affricated training stimuli. Therefore, the affricators and non-affricators were subjected to separate analyses to see

---

\(^7\) Link to one Rachel Maddow example: [http://www.youtube.com/watch?v=XS3rR5N5vAA&t=49s]. Link to one Michael Savage example: [http://youtu.be/D4rq53Ztvbg?t=4m18s]. Link to single [tsw] example, brought on by the 2012 twinkie crisis: [http://youtu.be/G1IJSKJr3Mk?t=1m2s]. Another twinkie example, this one with *chw-*: [http://youtu.be/QMhxH7WWWOQ?t=3s].

87
what the effects of training were on participants based on whether they already showed
signs of early adoption of an affricated /tw/ pronunciation variant. The retracted (chw-)
training condition had 5 affricators (3 m, 2 f), and 9 non-affricators. The front (tsw-)
training condition had 4 affricators (all male), and 10 non-affricators. The control group
(plain [tw]) had 4 affricators (1 m, 3 f), and 6 non-affricators.

Figure 2.12 shows the results for the non-affricators split by training condition,
illustrating an increase in convergence, relative to full group analyses, both during
shadowing and in its persistence post-training in both variant training conditions, and to a
lesser degree in the control group (who heard slightly affricated plain /tw/). All of the
non-affricators’ baseline productions were further from the training stimuli than
affricators’ baseline productions (control: t(7.7)=4.1, p<0.05; tsw-: t(4.4)=4.2; p<0.05;
chw-: t(7.8)=7.2, p<0.05), as can be seen by comparing Figure 12 (non-affricators) to
Figure 2.13 (affricators). During shadowing, pronunciation differences between
affricators and non-affricators were insignificant for the tsw- group and the control group,
as the non-affricators developed a more affricated pronunciation that was similar to the
affricators’ normal pronunciation. The chw- trained participants maintained a difference
between affricators’ and non-affricators’ productions (t(12)=3.5, p<0.05), because some
of the chw- condition affricators increased the degree of affrication to sound even more
like trainers, and the non-affricators did not reach that advanced degree of affrication.
After training, both the control group and the chw- group affricators’ productions were
again closer than the non-affricators’ productions to the training stimuli (control:
t(5.7)=3.9, p<0.05; **chw**: t(8.8)=3.7, p<0.05), but **tsw**- condition non-affricators were not significantly different from the affricators after training, as some of the affricators in this condition continued to increase the distance between their own productions and the affricated training stimuli. Affricators in the control condition actually decreased the amount of affrication during shadowing, as they tried to emulate the (only slightly affricated) plain [tw] stimuli. Affricators in the **tsw**- condition also changed their productions to sound less like trainers both during and after shadowing, but **chw**- condition affricators displayed an insignificant degree of convergence.

Both the front and retracted training groups of non-affricators exhibited significant convergence during shadowing (**tsw**: mean=0.98, t(9)=8.4, p<0.05; **chw**: mean=1.12, t(8)=4.4, p<0.05), and persistence after training (**tsw**: mean=0.41, t(9)=2.7, p<0.05; **chw**: mean=0.57, t(8)=2.4, p<0.05). The degree of convergence that occurred during shadowing was not statistically greater than the degree of convergence after training, which indicates that the persistence of convergence was quite robust. For the most part, the effect of convergence during shadowing and after training was much more robust for participants who did not initially affricate /tw/. This analysis shows that a significant lasting change in production occurred as a result of the shadowing task, for the participants who did not already use a variant /tw/ pronunciation.
Figure 2.12: Non-affricators’ Euclidean distance from training stimuli before training, during shadowing, and after training. The lighter gray solid lines show the control group’s productions, relative to the *tsw*-stimuli (top) and *chw*-stimuli (bottom). Convergence is indicated by a reduction in distance from the training stimuli. Compare to Figures 2.10 and 2.11 to see relative improvement in convergence.
Figure 2.13: Early affricators’ Euclidean distance from training stimuli before training, during shadowing, and after training. The lighter gray solid lines show the control group’s productions, relative to the \texttt{tsw}-stimuli (top) and \texttt{chw}-stimuli (bottom). Compare to Figures 2.10 and 2.11 to see differences from training groups.

### 2.3.4.2 Convergers and non-convergers

Although the group results indicate a significant trend of convergence, not all of the participants actually exhibited convergence on an individual level. Many of the affricators did not change their pronunciations to be more like the trainers, and two of them even changed to be less like the trainers, which explains some of the improvement in degree and persistence of convergence in the analysis of non-affricators only. Some affricators did show convergence, however, and some non-affricators did not show convergence. While some basic background information was collected from each
participant, the sample size was not large enough to point to any particular factor or set of factors that could always predict convergence or non-convergence.

All participants in the front and retracted training conditions were divided into convergers (Figure 2.14) and non-convergers (Figure 2.15) based on the difference in individual Euclidean distance scores between baseline and shadowing. If a two-sample t-test revealed a significant reduction in distance between the first and second measures, the participant was marked as a converger. Those who did not have a significant p-value, or whose productions moved further away from the trainers’ during shadowing were marked as non-convergers. 4 out of 14 participants in the front condition (3 male, 1 female), and 5 out of 14 in the retracted condition (3 male, 2 female) were marked as non-convergers. The control group participants were not marked for individual convergence because they did not hear either of the two variant pronunciations during training. 4 out of 5 non-convergers in the chw- condition were affricators, and 2 out of 4 non-convergers in the tsw- condition were affricators.
Figure 2.14: Convergers’ Euclidean distance from training stimuli before training, during shadowing, and after training. Convergence is indicated by a reduction in distance from the training stimuli. Compare to Figures 2.10 and 2.11 to see relative improvement in degree of convergence.
Non-convergers retracted (chw-) and front (tsw-) training
temporal location in experiment

Figure 2.15: Non-convergers’ Euclidean distance from training stimuli before, during, and after training. The *chw*- condition non-convergers move slightly toward training stimuli, but the *tsw*- condition non-convergers show divergence from training stimuli.

Essentially, convergers in both variant training groups converged during shadowing, and to some extent, these pronunciation changes persisted after training even though there was some movement back toward the baseline pronunciation after shadowing, which differed by participant. As can be seen in Figure 2.14, *tsw*- condition convergers showed less overall change between baseline and post-training productions, but the cause of this difference may be an artifact of the measurement scale. Non-convergers (Figure 2.15) in the *chw*- group showed no significant change over the course of the experiment, except for one male, who adjusted his pronunciations only slightly toward the training stimuli during shadowing (mean difference=0.25, t(39)=1.7, p=0.09), and then doubled
the difference after training, for a significant difference between baseline and post-training pronunciations (mean difference = 0.55, t(55) = 3.02, p < 0.05). One *tsw-* condition non-converger diverged significantly from trainers’ pronunciations during shadowing (t(23) = 7.7, p < 0.05) while the other three non-convergers in the *tsw-* condition showed essentially just a lack of difference from baseline during shadowing. After training, that diverger, plus two more non-convergers (all 3 male) each showed significantly increased divergence from the training stimuli compared to baseline, though that divergence is not significant for the whole group of non-convergers according to a paired t-test, due both to the small sample size and the fact that the one female non-converger did not show divergence, but rather a lack of significant convergence.

In order to find out the effectiveness of training on the persistence of pronunciation changes according to various configurations of factors, a final convergence score, as in example (4), was calculated by subtracting each individual post-training Euclidean distance measure (E3<sub>i</sub>) from that participant’s average baseline Euclidean distance measure (E1).

\[(4) \text{ Final convergence score } = E1 - E3_i\]

Because the number of factors was great and the levels unbalanced, traditional ANOVA could not be used to explore the effects and interactions of word familiarity and novelty, participants’ baseline /tw/ pronunciation, sex, and training condition. A
conditional inference tree (Hothorn, Hornik, and Zeileis, 2006) illustrates the interaction of affrication status, sex, and training condition (Figure 2.16). Word familiarity and novelty did not emerge as significant. Control group participants’ productions were omitted in order to simplify the analysis, and to further reduce the large number of observations (1,546) and resulting variance, the convergence scores were averaged for each participant for each level of word familiarity and novelty (end n=309). The conditional inference tree uses a two-sample linear statistic (T), as described by Hothorn et al., 2006, which numerically describes goodness of the split as the discrepancy between two samples, given the calculated conditional expectation and covariance of all possible permutations of the response values.

Figure 2.16: Conditional inference tree showing final convergence scores split by baseline affrication status, training condition, and sex.
The baseline pronunciation of /tw/ emerged as the first grouping factor (T=67.03, p<0.001), with non-affricators and front (tsw) affricators on the left branch (number of participants = 21), with generally more affrication in the post-training productions. The right branch (number of participants=7) contains retracted (chw) affricators and those who used a mix of /tw/ variants, showing no change from baseline or even less affrication in post-training, as indicated by smaller values and negative numbers. Following the leftmost branch, the non-affricators and front affricators are split by the factor of sex (T=13.95, p<0.001), with females showing generally greater post-training affrication, though there are no females who used the front affricate, so this grouping represents only non-affricators. Both men and women are split by training condition, with a crossed pattern: women showed greater post-training affrication in the chw- trained group than in the tsw- trained group (T=12.99, p<0.001), but men showed greater post-training affrication in the tsw- trained group than in the chw- trained group (T=9.47, p=0.006). The chw- trained group contained the two tsw- affricators, plus two non-affricators, so the difference between training conditions in this case is actually be a result of some reduction of Euclidean distance measure as the front affricators adopted a more retracted pronunciation. Recall that the front variant has generally higher slope and centroid than the retracted affricate. If a participant began by using the front variant, but switched to the retracted affricate, his Euclidean distance score would decrease slightly for the retracted productions.
Following the retracted and mixed affricators (number of participants=7) on the upper right branch in Figure 2.16, condition emerged as a significant factor (T=39.45, p<0.001), with the *tsw*- training condition eliciting the greatest divergence, shown by a great reduction in affrication post-training. Only males were represented in this group (n=3) because there were no female affricators in the *tsw*- training condition. In the *chw*-condition, the participants who already used the *chw*- variant extensively showed little change, while those participants who used a mix of /tw/ variants showed an increase in post-training affrication, as they more consistently used the retracted variant (T=10.04, p<0.001).

To summarize, non-affricators showed the greatest degree of post-training convergence, with women converging more than men, and the *chw*- training eliciting greater convergence than the *tsw*- training. Little or no change was seen in the affricators’ productions where the training variant overlapped with their previous productions, but the three male affricators (*chw*- and mixed) in the *tsw*- condition appear to have rejected the training variant, and greatly reduced the amount of affrication post-training, while two mixed affricators in the *chw*- condition increased their use of the training variant.

2.3.4.3 Discussion of production results by sex, affrication, and convergence status

Initial affricator status determined how close participants’ baseline productions were to the training stimuli, and also gave a reasonable idea of how much convergence could be expected. A smaller initial distance from the training stimuli resulted in a generally smaller change in production during shadowing and after training, which is similar to the
result found by Kim, Horton, and Bradlow (2011), even though the current experiment is not conversational, and the convergence under study is sub-phonemic. There were other reasons for non-convergence. For example, 6 out of 9 of the non-convergers were male, while only 3 were female, and statistical tests indicate much greater degrees of convergence for women than men. These results suggest a possible trend of greater non-convergence in males, in agreement with results obtained in Namy et al. 2002. Additionally, it is likely that there are differences between the chw- and tsw- conditions that can be explained by the pronunciation variants themselves. For example, the non-convergers in the chw- condition either showed a complete lack of change, or else a very slight degree of change toward training stimuli between baseline and post-training productions, with the exception of one male participant, whose productions were significantly further from the training stimuli after training than at baseline. This suggests a general trend of convergence for this group, which is why, in the initial analysis including all participants, both shadowing and post-training productions are significantly closer to the training stimuli than baseline productions for the chw- trained group. On the other hand, the non-convergers in the tsw- condition tended toward divergence away from the training stimuli, and 5 out of 7 of the male participants (including the 3 male non-convergers) in the tsw- condition had productions that were further from the training stimuli after training than at baseline (compared to only 2 out of 7 males in the chw- condition). This pattern suggests that the tsw- stimuli did not encourage lasting convergence among male participants, most especially those who were using an affricated variant before training.
One reason for this effect of gender bias against certain pronunciation variants may be that men generally adopt a more retracted pronunciation of /s/, while women usually adopt a more front production (e.g., Strand 1999), and this may hold true for other alveolar sounds, such as /t/. Thus, it may be more “natural” (in conforming with social norms) for men to dis-prefer the front *tsw-* variant, especially if they have already developed a more retracted affricate. For this reason, a follow-up experiment explores the interaction of gender and /tw/ pronunciation variants in the following chapter.

Separating participants based on convergence or non-convergence reveals that 19 out of 28 variant-trained participants exhibited robust effects of convergence during shadowing and after training. The majority of those who did not exhibit significant convergence had already adopted one or more affricated pronunciation variants, and some of these, especially in the *chw-* condition, were already close enough to the training stimuli that any convergence would be so slight as to be insignificant. In the *tsw-* condition, however, the male affricators were more likely to exhibit divergence. On the whole, over two-thirds of participants developed lasting pronunciation changes, and thus the methodology can be considered generally successful.

2.3.5 *Perception Results*

2.3.5.1 Lexical Decision

The lexical decision task revealed whether participants had undergone perceptual learning of the pronunciation variants, by their acceptance (or endorsement) rate of words containing those pronunciation variants. Acceptance rates of /tw/ words for all 72
participants (Table 2.7) were tested using a repeated measures ANOVA, with training condition, whether the words were heard during training, and pronunciation variant as factors.

A marginal effect of /tw/ variant (F(1,69)=3.94, p=0.0593) indicates that the retracted chw- variant was more acceptable than tsw- to all participants, regardless of condition. This finding may be due to the fact that chw- is a more “familiar” variant in the ambient language, as discussed in the context of the production results. A significant interaction between training condition and /tw/ variant (F(2,69)=5.65, p<0.05) indicates that perceptual learning occurred for the retracted chw- and tsw- trained groups, because the chw- and tsw- trained participants were more accepting of the variant they were trained on than those who were not trained on that variant. An interaction between stimulus novelty (whether the words were heard during training) and pronunciation variant did not reach significance, but the tsw- variant was generally less acceptable in untrained words for the two groups that were not exposed to it during training.

Table 2.7: Acceptance rates of trained and untrained words containing training variants, separated by training condition and variant. en- and vi- words are included for comparison.

<table>
<thead>
<tr>
<th>Training condition</th>
<th>Novel (untrained) words</th>
<th>Trained words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tsw- variant</td>
<td>chw- variant</td>
</tr>
<tr>
<td>Front (tsw-) trained</td>
<td>89.4%</td>
<td>84.1%</td>
</tr>
<tr>
<td>Retracted (chw-) trained</td>
<td>79.1%</td>
<td>88.7%</td>
</tr>
<tr>
<td>Control group</td>
<td>78.9%</td>
<td>84.2%</td>
</tr>
</tbody>
</table>
2.3.5.2 Identification task

The identification task examined the overall effect of /tw/ training variants on the perception of a related phonetic environment: /tu/. Participants heard stimuli that ranged along a continuum from chew to two to tsu, and their mouse click responses indicated which word they thought they heard. The dependent measure is proportion of /t/ (two) responses for each stimulus. All 72 participants were included in a repeated measures ANOVA of /t/ responses, with talker gender, stimulus, and training condition as factors. The responses for the male talker are illustrated in Figure 2.17, and the female talker in Figure 2.18. A main effect of stimulus (F(12, 828)=556, p<0.05) indicates different points along the continuum had different numbers of /t/ responses. An interaction of talker and stimulus (F(12, 828)=69, p<0.05) illustrates that the responses are different for the male versus female continuum. More of the retracted affricated stimuli are accepted as instances of /t/ for the male than for the female talker, and more of the front affricated stimuli are accepted as instances of /t/ for the female than for the male talker.

Although the anchor stimuli were selected to be typical sounding and categorically the same between the male and female talker, and the combination of sounds proceeded in exactly the same increments, the effect of gender on category boundaries once again illustrates a non-biological basis for pronunciation and perceptual differences between men’s and women’s speech. A post-hoc acoustic analysis of the sounds (Table 2.8) reveals very little difference between the male and female productions of /ts/, or between
the male and female productions of /tʃ/. The /t/ stimuli, however, show the male pronunciation to be more affricated, with longer rise time and greater slope, which corresponds to the greater affrication of /tw/ found for male participants and trainers. If spectral characteristics alone were responsible for /t/ judgments, we would expect to find the boundary drawn closer to the /t/ rather than closer to the /tʃ/, especially because of this already increased affrication. The greater number of retracted affricated stimuli judged to sound like /t/ for the male talker occurs despite this original difference in /t/, pointing to a strong effect of gender.

These results compare to research on gender effects on English /s/ and /ʃ/ (e.g., Strand and Johnson 1996; Strand 1999), and indicate that the gendered pattern extends beyond sibilants to include /t/. The reason why the two~tsu boundary differs between the male and female talker is because participants were already generally willing to accept some front affricates as instances of /t/ for the female talker, as demonstrated by the control group. Thus, there is also more room for this boundary to shift for the male talker. Relatedly, the chew~two boundary is less flexible because the minimal pairs chew and two are already in direct competition for that phonological space. /tʃw/, by contrast, encounters no competition, so incorporating the new variant into one’s phonological repertory is unproblematic, as regards phonological space. To fully extend the sound change into the /tu/ environment, however, a language user would be confronted by the choice of either allowing the possible creation of homophones or shifting the existing chew category back to make room for the incoming variant.
Figure 2.17: ID task results for the male talker, by training condition. The y-axis shows proportion of results labeled as beginning with /t/. The x-axis shows the percentage of each variant included at that stage of the continuum. The male elicits more /t/ responses with more retracted affricated pronunciation than with front affricated pronunciation.

Figure 2.18: ID task results for the female talker, by training condition. The y-axis shows percentage of results labeled as beginning with /t/. The x-axis shows the percentage of each variant included at that stage of the continuum. The female elicits more /t/ responses with more front affricated pronunciation than with retracted affricated pronunciation.
An interaction between stimulus and condition (F(24, 828)=2.9, p<0.05) indicates that participants in the different training conditions responded differently to the stimuli, having different thresholds for identification as /t/. More front variants were acceptable as /t/ for the tsw- group, while the chw- group accepted fewer front variants, which provides evidence for generalization from the /tw/ conditioning environment to a new, but related one: /tu/. The main difference between /w/ and /u/ is that of manner: /w/ is an approximant, while /u/ is a vowel. For this reason, /w/ should have a narrower constriction than /u/, which makes affrication more likely as the air stream increases speed as it travels through the narrower constriction. Generalization of a change that is conditioned by an approximant to a change that may also be conditioned by a similarly articulated vowel could be considered analogical.

Differences in responses to different stimuli can be seen in the graphs as the difference in curve shape to the left (chew~two) and to the right (two~tsu) for both the male and female talker. Because /ts/ is not a phoneme in English, the boundary between /t/ and /ts/ is much more gradual than the categorical boundary between /t/ and /ʧ/.

Table 2.8: Measurements of endpoint stimuli from ID task

<table>
<thead>
<tr>
<th>Word</th>
<th>Gender</th>
<th>Rise (ms/ms)</th>
<th>Centroid (Hz)</th>
<th>Slope (dB)</th>
<th>VOT (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>chew</td>
<td>male</td>
<td>0.538</td>
<td>3863</td>
<td>2.25</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>0.609</td>
<td>3451</td>
<td>1.86</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>0.606</td>
<td>6707</td>
<td>12.26</td>
<td>125</td>
</tr>
<tr>
<td>tsu</td>
<td>female</td>
<td>0.439</td>
<td>6895</td>
<td>14.80</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>0.264</td>
<td>5636</td>
<td>11.22</td>
<td>65</td>
</tr>
<tr>
<td>two</td>
<td>female</td>
<td>0.042</td>
<td>4821</td>
<td>4.76</td>
<td>65</td>
</tr>
</tbody>
</table>
the *chew-two* boundary, it only takes two steps to go from almost no /t/ responses to about 85% /t/ responses, but following the *two–tsu* boundary, it takes 4 steps to cover the same amount of perceptual distance. Additionally, it is the *two–tsu* continuum where the greatest training differences occur. For the female talker in Figure 2.18, participants in the retracted condition perceived fewer front tokens as instances of /t/ than the control group and front-trained group. Similarly, for the male talker in Figure 2.17, the *chw*-trained participants’ /t–ts/ boundary is well behind the front and control groups. The *tsw*-trained participants show a very slight difference for the endpoint *tsu* stimulus for the female talker, and a robust shift for the male talker, with more front variants being acceptable as instances of /t/. The differences between the training groups suggest a change in the perceptual acoustic space for /tu/ as an extension of the perceptual changes in /tw/. The fact that fewer front affricated stimuli were accepted as /t/ by the *chw*-trained group than by the control group suggests two possible explanations, both of which are upheld by the lexical decision results. One explanation is that the control group participants experienced some degree of perceptual learning for the *tsw*- pronunciation variant during the lexical decision task, which generalized to /tu/. The other explanation is that the *chw*-group participants became less willing to accept the front variant because of training, and instead of generally broadening the range of variant pronunciations acceptable as /tw/ or /tu/, as may be the case for the *tsw*- and control groups, experienced a shift in the whole /tw/ (and /tu/) category. It is likely that the interaction of these two effects was responsible for the differences between the *chw*- and control groups’ *two–tsu* boundary.
2.3.5.3 Comparison of perception and production results

By analyzing the perception results for the subset of participants whose productions were also analyzed, we can more directly investigate the interaction of speech production and perception during the actuation of sound change. Participants in the front and retracted training conditions were coded as convergers and non-convergers, as described in the production section, and their average acceptance rates of /tw/ words from the lexical decision task were subjected to a repeated measures ANOVA, with training condition, convergence status, /tw/ variant, and stimulus novelty (whether they heard the word during training) as factors. There was a main effect of convergence status (F(1,24)=7.7, p<0.05), showing convergers to have generally higher acceptance rates of words containing both /tw/ variants. A significant interaction of training condition and pronunciation variant (F(1,24)=14.4, p<0.05), showed that convergers and non-convergers alike were more likely to endorse words beginning with the variant they were trained on, which provides evidence for overall perceptual learning in both training conditions among convergers and non-convergers. A marginal interaction between convergence status * /tw/ variant * training condition * novelty of stimulus (F(1,24)=3.5, p=0.0756) motivated separate t-tests for trained and untrained words for convergers and non-convergers in each condition, by /tw/ variant. These tests showed that non-convergers performed worse than convergers especially on untrained words containing the variant they were trained on (t(25)=5.96, p<0.05). This result indicates that not only did the non-convergers perform worse than convergers at incorporating pronunciation variation, but also that they found it especially difficult in words that they had not heard
with a pronunciation variant during training. Thus, non-convergers did not generalize perceptual learning of the trained pronunciation variant to new words. Table 2.9 shows the convergers’ and non-convergers’ unequal performance in the lexical decision task on novel words that were not presented during training, with their acceptance rates of *vi-* and *en-* words for comparison. The entire control group’s results from Table 2.7 are reprised in Table 2.9, also for comparison.

Table 2.9: Acceptance rate of familiar but untrained words containing training variant, split by training condition, variant, and whether they displayed significant movement toward trainers during shadowing.

<table>
<thead>
<tr>
<th></th>
<th>Convergers</th>
<th></th>
<th>Non-convergers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>tsw-</em> variant</td>
<td><em>chw-</em> variant</td>
<td>vi- and <em>en-</em> words</td>
</tr>
<tr>
<td>Front trained (tsw-)</td>
<td>94.0%</td>
<td>84.3%</td>
<td>97.0%</td>
</tr>
<tr>
<td>Retracted (chw-) trained</td>
<td>82.2%</td>
<td>98.4%</td>
<td>99.3%</td>
</tr>
<tr>
<td>Control group</td>
<td>All control group participants combined</td>
<td>78.9%</td>
<td>84.8%</td>
</tr>
</tbody>
</table>

In summary, both front and retracted trained participants were more likely to accept words containing the variant they were trained on (as opposed to thinking the word contained an *s* or *ch*, which would make it a non-word, like *swiddle* or *chuídle*, rather than *twiddle*), which suggests some degree of perceptual learning across the board. The control group shows an average acceptance rate of 78.9% for untrained *tsw-* words, and 84.8% for untrained *chw-* words, even without training, which brings into stark contrast
the underwhelming performance of the non-convergers. The front and retracted trained convergers perform better than non-convergers on both variants in both trained and novel words, containing both the variant they were trained on and the untrained variant, though the difference is greatest for the novel tw- words. The relatively high and stable acceptance rates for all participants of novel vi- and en- words (from 94.9-99.3%) suggests that the differences in responses to the /tw/ variants is not coincidental with a lack of attention or some other general inability to distinguish words from non-words. The fact that convergers and non-convergers perform similarly except for on words containing a /tw/ variant, especially novel words, suggests that convergence requires a deeper level of processing, that convergers incorporate the new pronunciation variant into their mental representation for that phonemic category, and are thus better able to generalize to new words. Non-convergers may engage in a more superficial type of pattern recognition in words they were trained on, but without association to a phoneme category or phonological environment, this does not generalize as well to new words. Because of the low acceptance rates by non-convergers for both /tw/ variants, there may be some additional factors about the non-convergers that makes them less accepting of phonetic variation, at least in some circumstances. It is possible that they made certain judgments about the talkers in the experiment that influenced their perception.

As noted earlier, some non-convergers exhibited non-significant changes in pronunciation, while some were complete divergers. It was hypothesized that the degree of convergence or divergence might be predictive of the degree of perceptual learning.
The difference between Euclidean distance scores at baseline and during shadowing were calculated for each participant in the front and retracted training conditions, and subjected to linear regression, with acceptance rate of the trained variant as the dependent measure. The results indicate that the greater the convergence during shadowing, the greater the acceptance rate of the trained variant (F(1,26)=13.6, p<0.05, r²=0.32).

2.4 General Discussion

Generally, the vocabulary learning experiment methodology elicited perceptual and production changes for both /tw/ pronunciation variants, but the results were less robust for the tsw- variant. Before carrying out the experiment, it was thought that the pronunciation variants were hypothetically possible, but not necessarily attested. During the course of the study, it was discovered that both training variants, plus a third variant, /tfw/, had emerged in variation in common usage, with the retracted variant appearing to be more common among affricate variants. Production measurements showed strong and lasting pronunciation changes for 19 out of 28 participants. Convergence was stronger during shadowing for unfamiliar than familiar words, but after training, familiar untrained words also exhibited pronunciation changes, showing even stronger effects of convergence than trained words for the chw- condition participants. Men were less likely to converge with training stimuli than women, especially when presented with the front tsw- affricate, possibly because men generally showed greater affrication of /tw/ before training. Production changes were less likely overall, and smaller in degree, in
participants who exhibited production of an affricated /tw/ variant prior to the experiment, and were also mitigated by reduction due to repetition effects.

Participants who exhibited significant convergence with training stimuli also exhibited a stronger perceptual learning effect, and non-convergers were less able to generalize learning of the training variant to novel words, but were also generally less accepting of both /tw/ pronunciation variants. Increased acceptance of the training variants in novel words by convergers could be explained by deeper processing during convergence, whether as a result of convergence, or as a requirement for the act of converging. This combined perception-production result paves the way for future research, suggesting that the difference between perceptual adjustment and generalized perceptual learning is a qualitative difference such as that between perceiving and processing. The identification task provided evidence that participants had begun to generalize the pronunciation variant that they heard during training to a new phonological environment. Due to limitations imposed by gender expectations and phonological neighborhood, most of the phoneme boundary shifting for /tu/ occurred along the /t~/ts/ boundary, creating a slightly expanded space for the ts-w- group, relative to the control group, and a shifted and smaller space for the ch-w- group.

Although both the ts-w- and ch-w- stimuli conditioned robust changes in production and perception in a majority of participants, the effects were generally weaker for the ts-w- stimuli, as revealed by a smaller degree of persistence in convergence, and generally less acceptability in the lexical decision task. The differences in responses to the two
training variants have multiple possible explanations. The most obvious explanation is the lack of /ts/ in the phonemic inventory of most English speakers, coupled with the common pronunciation of /tr/ and /tʃ/ clusters with an affricated variant. As mentioned earlier, more tsw- trained participants took exception to the pronunciation variant; therefore, an informal survey of members of the OSU phonetics and psycholinguistics lab groups was conducted to ascertain the goodness of the stimuli as /tw/ words. Some of the tsw- stimuli were judged to contain a sound that could be confused with /sw/ because the fricative portion was much stronger than the burst and because /ts/ is not a phoneme in English. None of the stimuli were judged exclusively to sound like /sw/, but 14% of the stimuli had one or more /sw/ responses. Although the stimuli had also been normed for the current experiment, confusability was not directly addressed in the norming process. Additionally, because of the preference for retraction of /s/ in men, and more front productions for women, the tsw- variant as produced by men is more problematic. The follow-up experiment not only takes into account the delicacy with which these stimuli must be chosen, but also explores the question of how the gender of the talker and the listener may affect perception and adoption of pronunciation variants.

The uneven adoption of the pronunciation variant raises questions about the regularity of sound change actuation. The author takes the position that the actuation of sound change might be considered the ‘straw that broke the camel’s back,’ i.e., the point at which incoming variation combines with accumulated memory traces of the speech of interlocutors and of the speech of oneself in a configuration that changes the acoustic
representation of the target sound to the extent that a change in production ensues. This form of actuation can be an ongoing and repeating series of events, which makes it not different from the diffusion of sound change, as the changes can accumulate to create an eventual phonemic change, or can progress and regress as the nature of the incoming variation changes. The author thus takes the position that sound change is not a single event, but rather a dynamic process, in which slight perceptual and production changes can potentially occur with every conversational interaction, as re-tuning occurs and re-occurs. Over the lifetime of a speaker, many such minute changes can greatly alter the perception and production of an individual or a speech community, and these changes can become regular over long periods of time.

Because the changes in production and perception in the current study are applied by participants to novel words, not heard with the pronunciation variants, it seems reasonable to claim that the production changes that do ensue are the result of regular, phonetically-conditioned sound change. But, in addition to the continuous and discrete variation among words and participants that results, another theoretical issue impacted by this study is that of analogical extension. In the historical linguistics literature, there is a very clear divide between sound change and analogy (cf. Hock, 2003 and Kiparsky, 1995). However, the conditioning environment in this experiment was /t/ before /w/. Seven words containing /t/ plus a following vowel, including /u/, were heard during the lexical decision task, and correctly judged to be words 99.9% of the time, which provided participants with ongoing evidence against generalization of the front or retracted
affricate to a non-/w/ environment. The training stimuli did not give participants evidence of any word-initial /t/ except before /w/. Therefore, the only evidence of a conditioning environment was that of /t/ before /w/. The fact that participants in the two affricated training conditions generalized the change in /tw/ to also be valid for /tu/ in the identification task creates doubt about the division between phonetic conditioning and analogy.

Another facet of the definition of regular sound change that requires further examination on the basis of these experimental results is the assumption that a phonetic conditioning environment “explains” regular sound change. If both ts\textsuperscript{w}- and ch\textsuperscript{w}- are phonoetically natural, as is demonstrated by their historical development from /tw/ in several languages, then what is it about the /tw/ environment that explains these two separate changes? If some language users do not immediately adopt the sound change, or only have limited changes, is the sound change still regular? Must we narrow the definition of regular sound change to a locus of one individual at one time, and what would be gained by doing so? None of the individuals in this experiment demonstrated exceptionless change, but rather had varying degrees of affrication. That the affrication was generally greater in untrained words than in trained words suggests that neither borrowing nor analogy was responsible for the changes: in either case, the trained words would have exhibited the strongest changes. Thus, the changes that were elicited are not uniquely explained by the conditioning environment, nor by exposure to particular lexical items. If borrowing and analogy are eliminated, regular sound change must be the
explanation. But in this mini-sound change, the conditioning environment does not explain the change, and other non-phonetic factors, such as the phonemic inventory containing /ʧ/ but not /ts/, and extra-linguistic factors, such as gender, and other individual differences, influence the degree of change. Additionally, the strict conditioning environment, /tw/, did not limit the perceptual changes from spreading to a similar environment, /tu/. These results uphold a probabilistic, usage-based view of sound change, in which phonetic variation accumulates and delimits the acceptable acoustic and articulatory space of each sound, dynamically and gradually changing based on input.

Methodologically, the most valuable finding of this study is that the actuation of sound change can, in fact, be studied in the laboratory. The experimental conditions are somewhat unnatural in comparison to normal conversational interactions; however, they can be said to simulate a fast-forward version of what might happen when variation in a sound gradually shifts over time. While it is likely that most participants returned to their previously accustomed pronunciation of /tw/ at some point after the experiment, as the pattern of variation in their daily interactions returned to pre-training levels, follow-up studies are planned, to find out if any lasting changes did occur. Regardless of the long-term outcomes, it has been found that a temporary change in perception and production can be elicited by exposing participants to a pronunciation variant and giving them the opportunity to use it during shadowing. The changes that result may be generalized to new words, new talkers, and new conditioning environments, which is as close as we can come at this time to studying the actuation of sound change without a time machine.
Using this methodology, individual differences can be measured, allowing us to learn more about individual adoption of a sound change, and inform us about possible lexical diffusion, frequency effects in sound change, the role of phonetic conditioning, and its generalization, the role of non-phonetic conditioning factors, and the interaction of perception and production.
Chapter 3: Non-linguistic factors in laboratory sound change

3.1 Introduction

The early Neogrammian hypothesis was that sound change relied exclusively on phonetic factors, whether by phonetic conditioning or unconditioned sound shifts. Later Neogrammarians added to this that any other factors, linguistic or extra-linguistic, could not impact, block, or condition sound change. The evidence underlying this premise is that many completed sound changes that exhibit regularity can be described by a phonetic conditioning environment, such as \( t \rightarrow tf / _w \) (where /t/ becomes a palatal affricate when followed by /r/ or /j/), or an unconditioned shift \( t \rightarrow tf \), where every instance of /t/ becomes a palatal affricate. Neogrammarians and their 20\(^{th}\) century counterparts believed that the physical properties of articulation more or less guaranteed regularity of sound change, and that “the narrower and more circumscribed the original context is, the better we can define and determine the likely associated regularity” (Janda and Joseph 2003).

However, the results from chapter 2 demonstrated that even as participants were presented with an innovative variant in a narrow phonetic conditioning environment (\( t \rightarrow ts \) or \( t \rightarrow tf / _w \)), they also generalized the variant to a related environment (\( t \rightarrow ts \) or \( t \rightarrow tf / _u \)), for which they had no evidence that the change should apply, and even
negative evidence in the form of un-altered pronunciations in that environment (i.e., [tu]) and other (i.e., [ti]), which they heard during the lexical decision task. Furthermore, participants in that experiment exhibited differences in perception based on the gender of the talker, which should not happen during phonetically conditioned sound change, according to the Neogrammarian principle that phonetically conditioned sound change must proceed without any cognitive contribution.

The theory of sound change espoused in this thesis proceeds from the assumptions that: 1) general cognitive principles apply to speech production and processing, 2) sound categories are generalizations across stored exemplars which contain additional linguistic and extra-linguistic information, and 3) an individual’s speech perception and production are dynamic. Therefore, the flexible and continuously updated representations of sounds in a language are affected by the changing distributions in the language and the rich traces stored with each of the incoming exemplars. I posit that Neogrammarian sound change is one outcome of a long-term process of sound change, but that all sound change occurs due to the association of phonetic variation with some other co-occurring factor, including phonetic or phonological environment, lexical generalization, talker-specific or general characteristics, register, word class, etc., or the failure to associate the variation to any other source besides the sound category itself. This development takes place in the minds and mouths of individual language users, and may occur over generations, individual lifetimes, or a few years, months, or minutes, as the accumulating evidence of associations strengthens the connection between the “innovative variant” and co-occurring factor. Short-term developments represent such phenomena as perceptual
adjustment (Norris, McQueen, and Cutler 2003), while progressive long-term changes in the individual are likely to be mirrored in their speech community, and may result in a large-scale sound change. Neogrammarian sound change, in which the change exhibits regularity across a speech community (or a subgroup of that community), and across all instances of the conditioning environment, occurs when this phonetically continuous process has shifted the pronunciation norm of all the speakers in a community. Thus Neogrammarian sound change is just one possible outcome of sound change. In this chapter, sound change in the individual will be explored, providing evidence for simultaneous non-phonetic conditioning factors. The interaction of multiple factors elicits changes in speech processing, which illustrate how the phonological system adapts to changes in the incoming distribution of speech exemplars.

3.1.1 Phonology as probabilistic abstraction from richly encoded exemplars

Infants are probabilistic learners, who, even at the earliest stages, form generalizations and associations based on the distributional properties of the input they receive. For a detailed theory of probabilistic learning in language acquisition, and supporting evidence, see Beckman, Munson, and Edwards 2007 and Munson, Beckman, and Edwards 2011. As adults, the categories are more established because of the accumulation of years’ experience, but probabilistic learning, generalizations, and associations continue to re-tune adults’ phonological systems, as evidenced by phenomena such as perceptual learning (e.g., Norris, McQueen, and Cutler 2003), selective adaptation (e.g., Sawusch and Jusczyk 1981), and perceptual drift (e.g., Sancier
and Fowler 1997). Before infants even learn to derive meaning from words, they begin to learn the phonetic probabilities of their language. They learn that some similar-sounding sounds are the same, having a unimodal distribution, but other similar sounds are actually different, having a more bimodal distribution (e.g. Maye et al. 2002 and Maye et al. 2008), and stop attending to contrasts that don’t occur in their native language by around 9 months (Tees and Werker 1984). As children get older, they distinguish more categorically between contrastive sounds, based on the weight of accumulating evidence (e.g., Zlatin & Koenigsknecht 1975). Young language learners learn that some sounds occur more frequently than others, and that some sounds can occur next to each other, and that certain combinations recur and that other combinations never or hardly ever occur. Thus, the phonotactic “rules” or “constraints” are actually just generalizations about the probability of certain patterns of sounds that occur in the language. These qualities, and more, emerge from the distributions of sounds in the ambient language.

In order to map their own productions onto the acoustic space of their adult caregivers, the child must extrapolate from the acoustic patterns of Daddy’s production and Mommy’s production to a unique articulatory space belonging to the child, with very different acoustic parameters. Size relations for Daddy > Mommy > baby is one of the earliest relationships analogized to other domains by young children (3-4 year olds) (Goswami 1995), but the mapping of multiple acoustic signatures onto discrete sound categories is mastered even earlier, as evidenced by even younger children’s speech production and comprehension. It has been hypothesized that imitative gestures in infants require some basic analogical reasoning, in which the corresponding body parts and
gestures of the adult must be matched to the smaller physical space of the infant (Meltzoff and Moore 1997), which can be seen in infants as young as 3 days old. However, the process of learning language specific sound categories may take the form of a self-organizing system, in which local acoustic maxima across a continuous acoustic space are generalized to form category mappings, onto which not only the infant’s own productions, but also those of adults and other children are mapped, which then are fine-tuned and used as the basis for further abstraction (e.g., Kuhl and Meltzoff 1996). Thus, probabilistic learning, associative learning, analogical reasoning, and generalization from continuous input to abstract categories are imperative to the construction and maintenance of the phonological system, beginning in infancy, and continuing throughout adulthood.

In an exemplar account of phonological acquisition and re-tuning, the acoustic and articulatory phonetic properties of sounds do not exist in a vacuum, and do not need to be filtered to remove extra-linguistic information. Rather, contextual information is processed alongside the phonetic properties, which can assist in speech processing by giving additional information such as talker-specific characteristics, rate of speech, register, situational information, and so on. Having this information increases the likelihood of accurate identification of words and sounds, and can increase the speed at which processing takes place.

As already mentioned in the previous chapters, listeners are accustomed to variation. In fact, infants must learn to parse variation as a part of the acquisition process, in order
to be able to make the kinds of cross-modal inferences about "vowel categories" that Kuhl & Meltzoff (1996) show infants making as young as 4 months. Language users come to know and expect that the particular combination of sounds will affect the realization of the sounds in a word or an utterance (coarticulation) (e.g., Mann & Repp 1980, Kawasaki 1986, Beddor 2009, Connine and Darnieder 2009). They also know that some other facts may affect the realization of the sounds, such as rate of speech or register, which may increase or decrease coarticulation or change the acoustic characteristics of each sound (e.g., Summerfield 1981, Miller and Volaitis 1989, Dupoux and Green 1996). They also know that facts about the talker may affect the realization of the sound, from physiological, such as size and age and sex, to social, such as gender, dialect, and social affiliation (e.g., Mann and Repp 1980; Kraljic, Brennan and Samuel 2008; Maye, Aslin, & Tanenhaus 2008). These different kinds of associations, among others, might be made to help identify what sound is intended by a talker, and how that matches up to the sounds in that word as the listener is familiar with it. These associations are learned by experience, as the listener stores up different realizations of each word in varying contexts, spoken by various talkers. Generalizations that may emerge from these associations help to make language processing more efficient, but they may also form the basis of sound change.

3.1.2 An exemplar account of sound change

A “phonetically conditioned” sound change may begin as normal coarticulation. Presumably, this is what is meant by Janda and Joseph (2003) as the “Big Bang” of sound
change. It is generally agreed upon that any change in pronunciation that is not attributable to borrowing or lexical or morphological analogy, is a result of a process that starts with phonetic variation, including a range of coarticulatory processes. As the variation progresses, for example, in the co-production of sounds, as in phonetically conditioned sound change, there may be a number of generalizations that may emerge from the pairing of sound variation and associated factors. For example, a pronunciation variant may become associated with a particular word or class of words, or with a particular talker characteristic, such as gender or social class. Janda and Joseph (2003) hypothesize that “speakers’ imposition of phonological and sociolinguistic conditions” during the “spread” or expansion phase of sound change is separate from the initiation of sound change, rather than co-occurring. I would argue that associations of phonetic variation with phonetic or phonological conditioning environments, words, word classes, physical or sociolinguistic characteristics of talkers, or any other co-occurring factors, may occur simultaneously. I would also argue that because all kinds of sound change have their roots in phonetic variation, phonetic conditioning is not the kind of explanation or guarantee of regularity imagined by Janda and Joseph (2003), when they said, “purely phonetic environments guarantee that a change will be applicable whenever its most general conditions are met, since phonetic environments, almost by definition, are maximally general” (209). This hypothesis is based on a linguistic theory that holds that the acoustic and/or articulatory phonetics of speech can be isolated from other co-occurring linguistic and extra-linguistic information in the speech signal and, more generally, from cognitive processes, as stated by the Neogrammarians:
Aller Lautwandel, soweit er **mechanisch** vor sich geht, vollzieht sich nach **ausnahmslosen** Gesetzen, d.h. die Richtung der Lautbewegung ist in allen Angehörigen einer Sprachgenossenschaft, ausser dem Fall, dass die Dialektspaltung eintritt stetts dieselbe, und all Wörter, in denen der der Lautbewegung unterworfene Laut unter gleiche Verhältnisse erscheint, werden ohne ausnahme von der Änderung ergriffen. (Wilbur 1977: xiii, citing Brugmann’s Preface from Morphologische Untersuchungen)

[E]very sound change, as long as it proceeds **physically**, is implemented according to **exceptionless** rules, i.e., the direction of sound change is in every member of a speech community, unless dialect split is caused by the change, and all words in which the sound appears in the same conditioning environment, will, without exception, be subject to the change] (my translation, my emphasis).

More recently, it has been discovered that talkers and listeners do make use of additional information in the speech signal, for social and linguistic reasons, probably encoding this information alongside the phonetic information for each exemplar (e.g., Sidaras, Alexander, and Nygaard 2009). For example, experience with a specific talker helps listeners to identify words spoken by that talker, and experience with talkers of different dialects aids processing of speech from new talkers from those dialects (e.g., Clopper and Bradlow 2008). And, while sometimes misperception of coarticulation is a springboard for sound change (e.g., Ohala 1971, 1993, etc.), at other times speakers make use of the information as a cue or predictor of upcoming sounds in a word, which may lead to enhancement of coarticulatory effects, and to sound change as coarticulation becomes disassociated with the conditioning environment and attributed directly to the sound (Beddor et al. 1986, Beddor 2009, Lindblom et al. 1995, Guion 1996, 1998, etc.).
The theory of sound change espoused in this thesis holds that language users make use of any information that might enable them to make better predictions about incoming language data, just as any intelligent species makes use of any information that enables them to make better predictions about the world around them. The basis for sound change is phonetic variation, and the “Big Bang,” or inception, of sound change is the learned association of phonetic variation to some other co-occurring factor. The development of this association occurs as the accumulation of variation and accumulated evidence of association strengthens the connection between the “innovative variant” and co-occurring factor. Contrary to the Neogrammarian principle, I argue that all sound change is necessarily cognitive, but that this does not make it inherently unpredictable. Rather, the predictions must be based on language-specific factors, in addition to so-called universals.

The experiment in chapter 2 provided evidence that generalization could take place concurrently with phonetic conditioning, that even though the environment was /tw/, /tu/ was also affected, even in the face of negative evidence. That experiment also contributed to the body of research showing that non-linguistic factors can and do influence perception of sound categories. For example, whether a talker is presented as male or female influences whether listeners identify an ambiguous sound as /s/ or /ʃ/ (Mann and Repp 1980; Johnson 1981; Strand and Johnson 1996). Similarly, in the experiment in chapter 2, the boundaries for /ts/~t/ and /tʃ/~t/ varied depending on the gender of the talker, in addition to training condition. Not only that, but convergence was less robust
among male participants, illustrating gender differences in adoption of an innovative variant (as in, e.g., Pardo 2006).

In order to provide further evidence of non-phonetic conditioning occurring concurrently with phonetic conditioning, the experiment in chapter 2 was repeated with some changes, including training conditions in which participants were exposed to two different /tw/ variants, split by talker gender. That is, one group heard only men saying $tsw$- and only women saying $chw$- and another group heard only women saying $tsw$- and only men saying $chw$-, with the expectation that participants would develop a gender-conditioned sound split for /tw/. The group that heard only women saying $tsw$- and only men saying $chw$-, is called the *socially expected*/*traditional* group, because other studies (such as with /s/ and /ʃ/, for example) have shown that males are expected to have more retracted pronunciations, and females more anterior (e.g. Strand 1999, Mays 2010). The training condition with only men saying $tsw$- and only women saying $chw$- could be thought of as the *socially unexpected*/*non-traditional* group, as the relationship between men’s and women’s productions is reversed.

As shown in the first experiment, both $tsw$- and $chw$- are appropriate and natural outcomes for a phonetically conditioned sound change resulting from coarticulation of /t+w/. $chw$- is analogically comparable to the outcomes of /t+r/ and /t+j/, and, as such, is an easier variant for listeners to process and for shadowers to produce. This fact alone provides evidence that language-specific factors and analogical processing can affect phonetically conditioned sound change. But the next experiment, in setting up two
potential sound splits, tests whether evidence of a sound split is enough to overcome preference for one variant, and whether evidence of a sound split can condition sound change on the basis of gender. It examines whether there is a difference in the outcome of perceptual learning depending on whether the split runs parallel to an analogically similar, and thus more expected, distributional difference between men and women (as in /s/ and /ʃ/), or if the split runs contrary to the known gender-patterned relationships in sounds.

If a split in the perceptual adaptation of sounds on the basis of gender were possible, it could provide evidence for a gender-conditioned sound change. A gender-conditioned sound change could be extremely regular, if adopted by all members of a language community, on the basis of physical sex alone. Or it could be incredibly complicated as different levels of masculinity and femininity were expressed through the use of these sound variants, differing in degrees of fronting and retraction or affrication or lack thereof, probably making use of different distributions depending on the interlocutors or function of discourse. This experiment only looks at the perception of a statistically regular split between men and women, to see if a generalization between two different pronunciation variants, each with an associated gender, can develop. If the different split training groups show differences in their phonological perception of the target sound, this provides further evidence that non-phonetic factors may influence sound change, or they may at least influence perceptual adjustment, which is a phenomenon that is closely related to sound change, as it re-tunes the phonological system.
3.2 Methods

The procedure for the training in this experiment is structured nearly the same as in the experiment in chapter 2. Differences include the complete omission of en- vocabulary items, so that only vi- and tw- words were heard during training. The testing portion of the experiment is also different, in that in this experiment, eye tracking technology was used, whereas in the experiment in chapter 2, participants went through a lexical decision task and an identification task.

3.2.1 Participants

Participants received participation credit in partial fulfillment of a course requirement for introductory linguistics classes or a cash honorarium for their participation. A total of 176 students participated in the eye tracking experiment, with 30 excluded from the results for experimental error or too much missing data from the eye tracker (e.g., the eye tracker was unable to get a clear lock on the participant’s gaze due to various factors such as dirty contacts or glasses, long eyelashes, participant closing eyes too often or too long). Another 24 participants were excluded as non-native English speakers, and 10 participants were excluded because they averaged around 25% or less looks to the target for more than two kinds of target, which means they did not perform better than someone randomly selecting one of the words or non-words on the screen. The remaining 112 participants included 50 men and 73 women between the ages of 18 and 36 (mean=22.3, sd=4.0).
A second version of the experiment, experiment 2b, which used the same lexical decision and identification tasks as the experiment in chapter 2, was taken by an additional 114 participants, who also received partial class credit. 24 participants were excluded as non-native English speakers. 7 participants were excluded for experimental errors. An additional 10 were removed from the identification task results because they failed to make use of the range of choices (e.g., did not identify any form of tsu as “tsu”). The remaining 73 participants’ responses to the identification task alone were included in this analysis, to provide corroboration of the patterns found in the eye tracking study.

3.2.2 Materials

The stimuli used in training were taken from the same stimuli used in the experiment in chapter 2, plus some that were not previously used, but were produced by the same talkers. The training stimuli included 30 vi- words and 30 tw- words (with both pronunciation variants), spoken by 3 male and 3 female talkers, as in chapter 2. All stimuli were natural speech produced by trained students from the Linguistics and/or Psychology Departments at the Ohio State University.

The eye tracking testing stimuli were recorded by 4 new talkers (2m, 2f), using a shadowing procedure. Before recording, the talkers were trained to be able to pronounce both the front and retracted /tw/ variants. Talkers were presented with stimuli one at a time over headphones, along with orthographic representations, in which the variant pronunciations were written out using <chw> for the retracted variant (e.g., chwin), and <tsw> for the front variant (e.g., tswin). All of the stimuli that were presented for
shadowing were pronounced by the author and recorded to be used as exemplars for test stimuli, so that the pronunciations would be more uniform, especially in duration. Talkers were instructed to repeat the word as closely as possible to the original pronunciation. These second recordings were used as test stimuli in the present experiment. Two different versions (one of each /tw/ variant produced, with each pair produced only by a male or a female talker) of 60 tw- words and non-words were used as test stimuli, and 60 sw- and 60 ch- words and non-words, also evenly divided among talkers, were also used as stimuli. Two versions by two different talkers of 55 vi- words and non-words were used as filler stimuli. Each word or non-word was pronounced by two talkers of the same gender, but the total number of tokens was split evenly across men and women. (For example, two different male talkers said the word violin, but two female talkers said the word violet, and so on.) 50 f- and 50 b- words and non-words, evenly divided among the talkers, were also used as fillers.

All of the words were recorded at 44,100 Hz, then segmented in Praat (Boersma and Weeninck 2005) and normalized to 70dB. Unlike the experiment in chapter 2, the stimuli were not downsampled, but rather presented at a 44,100 Hz sampling rate. The tsw- stimuli, in particular, contained a high amplitude band at frequencies reaching above 11,025 Hz. Because of the difficulties with the tsw- variant in chapter 2, it was thought that this additional information might be helpful for some participants, and the E-Prime version used in this experiment allowed a higher sampling rate.
3.2.3 Procedure

As in the first experiment, participants first completed a familiarity rating task, in which participants were shown an orthographic representation on a computer screen, one word at a time, tw- and vi- words, plus words containing initial tV-, tr-, str-, ts- (tsu) and ch- (chew), and medial tw-, tr-, and str-. Participants were instructed to say each word out loud and rate how familiar they were with that word by clicking the box next to the appropriate rating:

5. very familiar – I know this word and use it
4. somewhat familiar – I know this word, but I may or may not use it myself
3. neither familiar nor unfamiliar – I may know this word, but do not use it
2. somewhat unfamiliar – I may have heard this word before, but have never used it
1. very unfamiliar – I have never heard or used this word before

The productions were recorded using Audacity via a desktop microphone, at a 44,100 Hz sampling rate. All experimental stimuli were presented and mouse click responses were recorded using E-prime 1.2. Auditory stimuli were presented over headphones.

The second task was the perceptual training. Participants were randomly assigned to one of four different conditions. As in the previous experiment, the front affricate training group heard /tw/ words pronounced with the front affricated [tw] variant, referred to as the tsw- group. The retracted affricate group heard words pronounced with the retracted affricated [fhw] variant, referred to as the chw- group. There were two split-training
conditions. One group heard only male talkers producing the $chw$- variant, and only females pronouncing the $tsw$- variant. The other split-training group heard only female talkers producing the $chw$- variant, and only males pronouncing the $tsw$- variant. Based on the results in the previous experiment and other research showing that males are expected to have more retracted pronunciations, and females fronter ones (e.g. Strand 1999), the male-$chw$- group is called the socially expected/ traditional group, and the male-$tsw$- group is called the socially unexpected/ non-traditional group. An untrained group did not receive any training before completing the eye tracking task.

Participants were told that they would learn words from two different sections of the dictionary, including $vi$- and $tw$-, and that words they already knew would also be presented, to aid in learning by associating new words with familiar ones. Training was divided into blocks, with $vi$- words presented first, followed by $tw$- words. Participants were instructed to say the word out loud every time they heard it pronounced over the headphones, but that anything that was only printed on the screen should be read silently. They first saw the word on the screen in isolation, and heard it pronounced. On the next screen, they saw the word at the top of the screen, and heard it pronounced again, followed by a definition in the middle of the screen. Upon proceeding to the next slide, they heard the word again as they saw an image illustrating the definition. After 5 words, the entire procedure was repeated except the definitions were replaced by sentences using the words in context, followed by pictures illustrating the sentences, and each of these 5-word sub-blocks was followed by a quiz, in which a definition was provided, and participants had to choose from one of 4 words. Each word was pronounced by a model
talker three times in a definition trial, and 3 times by a talker of another gender for the sentence trial, for a total of 6 times, by 2 different talkers. The words were to be pronounced out loud by the participant three times on each trial.

The training section in this experiment was shorter than the training section in the first experiment due to the omission of the en- words. Participant feedback suggested that learning 45 new words was too much for one experiment. Other changes included the addition of a picture for each sentence and definition, to entertain as well as illustrate the concepts contained in the definition or sentence. Quizzes were increased in frequency and predictability, to every 5 words, as opposed to being sporadically offered every 5-10 words. This way, participants only had to remember 5 words at a time, at least 2 of which were familiar. The overall number of training words for each ‘dictionary section’ was increased from 28 each to 30 each, to allow for the even distribution of quizzes. Additionally, the tsw- training words were subjected to a normative procedure, in which faculty and students from the Phonies and Psycholinguistics lab groups at OSU were asked to report whether each word sounded like it started with an sw- or tw-. Only recordings that passed this procedure (one or fewer sw- ratings out of 20) were used in training.

After training, participants filled out a background questionnaire, were offered a break, and either completed the testing half of the experiment as explained in chapter 2, with lexical decision, post-test production, and identification tasks, or else participated in the eye tracking task. Eye-tracking participants were instructed to sit in front of the Tobii
eye tracker and computer screen. Each trial displayed 4 written choices, including words and non-words, in black type on an off-white screen; an example slide is shown in Fig. 3.1. Participants were instructed that after hearing a stimulus, they should select the word or non-word that they had heard by clicking on the appropriate target with a mouse. The location of their gaze was also recorded about every 17-20 ms. Targets beginning with *sw*- or *ch*- always had a *tw*- competitor. Targets beginning with *tw*- had *ch*- competitors half the time, and *sw*- competitors the other half of the time. During filler trials, targets beginning with *f*- or *b*- always had a *v*- competitor. Targets beginning with *v*- had *f*- competitors half the time, and *b*- competitors the other half of the time. All targets and competitors included words and non-words in mixed trials. Competitors were as closely matched to the targets as phonotactics would allow (i.e., if the target and competitor were both words, they formed a minimal pair). Every 18 trials, participants were shown a slide with 4 images, and were instructed to click on the picture that was different from the other 3 pictures. These trials were designed to give participants a mental as well as visual break from the task. Every 36 trials, participants were given the option of taking a quick break to take a sip of water and/or stretch in their chair. After 225 word/non-word trials, participants were required to take a break while the experiment administrator started the second half of the experiment using a second program file, and they were encouraged to get out of the chair for a minute or two during this time. The second block of 225 words proceeded in exactly the same manner. The order of presentation was randomized and then corrected to ensure that no two consecutive trials contained the same initial sound. Additionally, no *tw*- or *v*- stimulus was repeated twice in the same block, the
pronunciation variants were evenly distributed across the two blocks, and \( f \)-, \( b \)-, \( ch \)-, and \( sw \)- stimuli and competitors were evenly divided across the two blocks.

Figure 3.1: An example of four word and non-word choices participants saw on the computer screen during eye tracking trials.

3.3 Eye Tracking Results

3.3.1 Mouse click identification

The accuracy, or endorsement rate, of target stimuli from mouse clicks was converted from proportions (Table 3.1) to rationalized arcsine units (RAU), and subjected to a 3-way items ANOVA with talker sex and /tw/ variant as between items factors, and training condition as a within-items factor, and stimulus as random error. A main effect of /tw/ variant (\( F(1,116)=43.3, p<0.05 \)) shows that \( chw \)- had higher endorsement rates than \( tsw \)-. A main effect of talker sex (\( F(1,116)=6.5, p<0.05 \)) reveals that the male talkers received higher endorsement rates for both /tw/ variants. A main effect of training condition (\( F(3,348)=14.0, p<0.05 \)) suggests general differences across training groups, with the split-condition groups performing better than the single-variant trained groups. An
interaction between talker sex and /tw/ variant (F(1,116)=4.8, p<0.05) points us to the
difference in endorsement rates for male and female talkers’ tsw- productions in Table 3,
in which the male talkers’ tsw- productions receive somewhat greater endorsement than
the females’ (t(40)=2.5, p=0.015).

Finally, an interaction of training condition and /tw/ variant (F(3,348)=14.6, p<0.0)
reveals the effects of training on the perception of /tw/ variants. Bonferroni corrected
pairwise comparisons show that the tsw- trained group had a greater overall endorsement
rate of the tsw- pronunciation variant as instances of /tw/ than the chw- trained group
(t(59)=7.3, p<0.0083), and the traditional (men chw- women tsw-) group also had more
endorsements for tsw- than the chw- trained group (t(59)=7.7, p<0.0083), and than the
non-traditional (women chw- men tsw-) group (t(59)=4.0, p<0.0083), but not
significantly more than the tsw- group (t(59)=1.7, p=0.09). The non-traditional group had
a non-significantly greater endorsement rate of tsw- than the tsw- group (t(59)=2.2,
p=0.03), and a significantly greater endorsement rate of tsw- than the chw- group
(t(59)=5.2, p<0.0083). Both single-variant trained groups more-or-less equally endorsed
the chw- variant, but the non-traditional (women chw- men tsw-) group had more
endorsements for chw- than the traditional (men chw- women tsw-) group (t(59)=3.6,
p<0.0083), than the tsw- trained group (t(59)=4.3, p<0.0083), and even than the chw-
trained group (t(59)=4.1, p<0.0083).
Table 3.1: Percentage endorsement of /tw/ pronunciation by training condition. Participants in the tsw- trained group endorsed more words and non-words with the tsw- variant as instances of /tw/, and participants trained on both variants generally endorsed more of both /tw/ variants than the participants trained on a single variant.

<table>
<thead>
<tr>
<th>/tw/ pronunciation variant</th>
<th>training condition</th>
<th>all chw-trained</th>
<th>all tsw-trained</th>
<th>traditional male chw-female tsw-</th>
<th>non-traditional male tsw-female chw-</th>
</tr>
</thead>
<tbody>
<tr>
<td>chw- male talker</td>
<td>96.5</td>
<td>96.1</td>
<td>97.9</td>
<td>97.8</td>
<td></td>
</tr>
<tr>
<td>chw- female talker</td>
<td>95.4</td>
<td>95.8</td>
<td>98.2</td>
<td>98.9</td>
<td></td>
</tr>
<tr>
<td>tsw- male talker</td>
<td>85.5</td>
<td>95.2</td>
<td>95.5</td>
<td>93.8</td>
<td></td>
</tr>
<tr>
<td>tsw- female talker</td>
<td>73.2</td>
<td>82.0</td>
<td>88.0</td>
<td>82.5</td>
<td></td>
</tr>
<tr>
<td>average chw-</td>
<td>96.0</td>
<td>96.0</td>
<td>98.0</td>
<td>98.3</td>
<td></td>
</tr>
<tr>
<td>average tsw-</td>
<td>79.6</td>
<td>88.8</td>
<td>91.9</td>
<td>88.3</td>
<td></td>
</tr>
</tbody>
</table>

In summary, the tsw- trained group endorsed more of the tsw- variants than the chw- trained group, as expected, but the he most interesting pattern is that the traditional trained group (which only heard tsw- when pronounced by a woman) also endorsed more tsw- variants than the chw- group and the non-traditional group, while the non-traditional group (which only heard chw- when pronounced by a woman) had higher endorsement rates for the chw- variant. It is unclear what this result means. The rates of endorsement for the chw- variant as instances of /tw/ were almost at ceiling because the competitors did not contain a w, so that any initial confusion was corrected by analysis of the spelled forms of the words. This may partially explain why the chw- group did not perform significantly better than the tsw- group on the chw- variant. The eye tracking data in the
next section yields more detailed information about possible confusion between target and competitor before participants eventually clicked on the target.

3.3.2 \textit{Gaze Location Results}

The eye tracking results are shown in Figures 3.2-3.17. Figures 3.2-3.5 illustrate how each training group performed on each /tw/ variant, in relation to each other, so that overall differences between groups can be seen at a glance, though each figure will not be remarked upon individually. The results are broken down and discussed by the single-variant groups (Figs. 3.6-3.9), and the split-training groups (Figs. 3.10-3.13), which are also broken down further by the gender of the participants (Figs 3.14-3.17). While this method of presentation is somewhat redundant, it is difficult to see each of the comparison groups individually in Figs. 3.2-3.5, and while it is important to be able to see the overall relationship of each training group to the other, it is also important to see the relationship between groups (i.e., front or retracted) within types of training (single or split), and between genders in the split-training group.

Each graph displays looks to the target versus the competitor over time, beginning with the onset of the stimulus, and ending 1500 ms thereafter, by which time most participants had made their selection. The measurement of identification for each time sample is the log odds of adjusted looks to the target versus the competitor, which is calculated as the natural log of the looks to the target (1=yes, 0=no) plus 0.5 (as a continuity correction) over looks to the competitor plus 0.5, as in equation (1).
(1) \log \text{odds} = \ln \left( \frac{0.94 + 0.5}{0.06 + 0.5} \right)

For example, if a participant was looking at 94% of targets for a particular set of stimuli at a particular time, while looking at the competitors 6% of that time, the adjusted log odds would be calculated as in (2):

(2) \quad 0.944 = \ln \left( \frac{0.94 + 0.5}{0.06 + 0.5} \right)

Thus, if the value on the y-axis of the graphs reads 0.944, that represents around 94% looks to the target, and 0.0 means that participants were looking at the target at the same rate as the competitor. The additional “0.5” was added to avoid having “0” appear as the numerator, or especially the denominator (e.g., Fleiss 1981, Haldane 1967, or Gart and Zweifel 1967).

All of the statistical analyses were performed using item means collapsed over the time samples from 0 - 400 ms (Table 3.2), and from 400 - 1200 ms (Table 3.3), which is when any meaningful differences appeared. By collapsing the time element, a general measure of certainty was captured, both in how quickly participants looked at the target, resulting in a higher mean proportion of looks, and the overall correct identification of the target by the end of the trial. Item means were used because the duration of each stimulus was different, the longest lasting 998 ms, the shortest being 242 ms. Because
there were missing trials by subject, such as when the eye tracker did not register a fixation (and these could be on longer or shorter stimuli), the subject means were highly variable. By comparing across items, differences in stimulus duration became less relevant as the focus was on any difference in response to the same stimulus, across the different training conditions.

Figure 3.2: Looks to tw- target relative to looks to ch- competitor, for chw- stimulus, female talker. The chw- training group had more looks to the target than the tsw- training group.
Figure 3.3: Looks to $tw$- target relative to looks to $ch$- competitor, for $chw$- stimulus, male talker. The $chw$- training group had more looks to the target than the $chw$- training group. The traditional split training condition had more looks to the target than the other training groups, though all groups accepted this variant more readily compared to the other variants.

Figure 3.4: Looks to $tw$- target relative to looks to $sw$- competitor, for $tsw$- stimulus, female talker. The $tsw$- training group had more looks to the target than the $chw$- training group. The traditional group (trained hearing women say $tsw$-) had more looks to the target (even more so than the $tsw$- group) than the non-traditional group (heard women saying $chw$-) and the all $chw$- trained group.
Figure 3.5: Looks to tw-target relative to looks to sw-competitor, for tsw-stimulus, male talker. The tsw-training group had more looks to the target than the chw-training group. The traditional group (trained hearing women say tsw-) also had more looks to the target than the non-traditional group (heard men saying tsw-). Both split groups had more looks to the target than the retracted (all chw-trained group).
Table 3.2: 4-way repeated measures ANOVA by items, averaged over time samples from 0-400 ms (20 samples) after onset of stimulus.

<table>
<thead>
<tr>
<th>Factors</th>
<th>df</th>
<th>F</th>
<th>sum of squares</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>variant</td>
<td>1</td>
<td>0.024</td>
<td>0.007</td>
<td>0.876</td>
</tr>
<tr>
<td>talker sex</td>
<td>1</td>
<td>0.047</td>
<td>0.013</td>
<td>0.829</td>
</tr>
<tr>
<td>condition</td>
<td>3</td>
<td>1.677</td>
<td>0.153</td>
<td>0.172</td>
</tr>
<tr>
<td>subject sex</td>
<td>1</td>
<td>5.471</td>
<td>0.197</td>
<td>0.021 *</td>
</tr>
<tr>
<td>variant*talker sex</td>
<td>1</td>
<td>0.066</td>
<td>0.18</td>
<td>0.797</td>
</tr>
<tr>
<td>variant*condition</td>
<td>3</td>
<td>0.048</td>
<td>0.004</td>
<td>0.986</td>
</tr>
<tr>
<td>talker sex*condition</td>
<td>3</td>
<td>2.410</td>
<td>0.220</td>
<td>0.067</td>
</tr>
<tr>
<td>variant*subject sex</td>
<td>1</td>
<td>1.369</td>
<td>0.049</td>
<td>0.244</td>
</tr>
<tr>
<td>talker sex*subject sex</td>
<td>1</td>
<td>0.899</td>
<td>0.032</td>
<td>0.345</td>
</tr>
<tr>
<td>condition*subject sex</td>
<td>3</td>
<td>0.763</td>
<td>0.076</td>
<td>0.516</td>
</tr>
<tr>
<td>variant<em>talker sex</em>condition</td>
<td>3</td>
<td>0.714</td>
<td>0.065</td>
<td>0.544</td>
</tr>
<tr>
<td>variant<em>talker sex</em>subject sex</td>
<td>1</td>
<td>0.006</td>
<td>0.000</td>
<td>0.939</td>
</tr>
<tr>
<td>variant<em>condition</em>subject sex</td>
<td>3</td>
<td>1.310</td>
<td>0.130</td>
<td>0.271</td>
</tr>
<tr>
<td>talker sex<em>condition</em>subject sex</td>
<td>3</td>
<td>0.927</td>
<td>0.092</td>
<td>0.428</td>
</tr>
<tr>
<td>variant* talker sex<em>condition</em>subject sex</td>
<td>3</td>
<td>0.936</td>
<td>0.093</td>
<td>0.423</td>
</tr>
<tr>
<td>Residuals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (item)</td>
<td>112</td>
<td>30.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (item*condition)</td>
<td>336</td>
<td>10.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (item*subject sex)</td>
<td>112</td>
<td>4.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (item<em>condition</em>subject sex)</td>
<td>336</td>
<td>11.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3: 4-way repeated measures ANOVA by items averaged over 400-1200ms after onset of stimulus.

<table>
<thead>
<tr>
<th>Factors</th>
<th>df</th>
<th>F</th>
<th>sum of squares</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>variant</td>
<td>1</td>
<td>10.55</td>
<td>2.194</td>
<td>0.0015 *</td>
</tr>
<tr>
<td>talker sex</td>
<td>1</td>
<td>13.93</td>
<td>2.896</td>
<td>0.0003 *</td>
</tr>
<tr>
<td>condition</td>
<td>3</td>
<td>21.41</td>
<td>1.523</td>
<td>0.00000 *</td>
</tr>
<tr>
<td>subject sex</td>
<td>1</td>
<td>57.99</td>
<td>1.066</td>
<td>0.00000 *</td>
</tr>
<tr>
<td>variant*talker sex</td>
<td>1</td>
<td>1.71</td>
<td>0.356</td>
<td>0.193</td>
</tr>
<tr>
<td>variant*condition</td>
<td>3</td>
<td>7.71</td>
<td>0.549</td>
<td>0.0005 *</td>
</tr>
<tr>
<td>talker sex*condition</td>
<td>3</td>
<td>0.96</td>
<td>0.068</td>
<td>0.414</td>
</tr>
<tr>
<td>variant*subject sex</td>
<td>1</td>
<td>2.32</td>
<td>0.043</td>
<td>0.130</td>
</tr>
<tr>
<td>talker sex*subject sex</td>
<td>1</td>
<td>0.01</td>
<td>0.000</td>
<td>0.913</td>
</tr>
<tr>
<td>condition*subject sex</td>
<td>3</td>
<td>3.66</td>
<td>0.279</td>
<td>0.013 *</td>
</tr>
<tr>
<td>variant<em>talker sex</em>condition</td>
<td>3</td>
<td>2.03</td>
<td>0.144</td>
<td>0.109</td>
</tr>
<tr>
<td>variant<em>talker sex</em>subject sex</td>
<td>1</td>
<td>0.16</td>
<td>0.003</td>
<td>0.913</td>
</tr>
<tr>
<td>variant<em>condition</em>subject sex</td>
<td>3</td>
<td>4.07</td>
<td>0.311</td>
<td>0.007 *</td>
</tr>
<tr>
<td>talker sex<em>condition</em>subject sex</td>
<td>3</td>
<td>0.16</td>
<td>0.012</td>
<td>0.927</td>
</tr>
<tr>
<td>variant* talker sex<em>condition</em>subject sex</td>
<td>3</td>
<td>1.83</td>
<td>0.140</td>
<td>0.142</td>
</tr>
</tbody>
</table>

Residuals

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Error (item)</td>
<td>112</td>
<td>23.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (item*condition)</td>
<td>336</td>
<td>7.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (item*subject sex)</td>
<td>112</td>
<td>2.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (item<em>condition</em>subject sex)</td>
<td>336</td>
<td>8.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.3.2.1 Results – single-training condition groups

The results for the retracted (chw-) and front (tsw-) training conditions were similar to those in the first experiment. The tsw- group accepted the front variant, spoken by the male talkers (Fig.3.6) more readily than the control group, as defined by more looks to the target (t(29)=5.02, p<0.0125) and somewhat more than the chw- trained group for the
female talkers (Fig. 3.7; t(28)=2.4, p=0.02)\(^8\). The \textit{chw-} group accepted the retracted variant, by both male (Fig. 3.8) and female talkers (Fig. 3.9), defined by average looks to target being significantly greater than the control group (male talker: t(29)=2.8, p<0.0125; female talker: t(28)=2.4, p=0.02). The \textit{chw-} variant, especially as spoken by men, was widely endorsed by all groups, including the control group, which reached over 80\% looks to the target, so even though the front and retracted groups did not differ significantly in their acceptance rate or looks to the target, it was not an effect (or lack of effect) of training, but rather of the prior acceptability of this variant, and clear orthographic difference between targets and competitors.

This result replicates the findings in chapter 2, that perceptual adjustment did occur, with the front trained group generally accepting more \textit{tsw-} tokens as instances of /tw/, and the retracted trained group generally accepting more \textit{chw-} tokens as instances of /tw/. The additional information gained by being able to see how quickly and how confidently the target was identified, shows us that even when the final identification rates were about the same (as in Fig. 3.7 and 3.9), the group that was trained on the variant was faster to identify the target, suggesting a stronger connection between the pronunciation variant and the associated /tw/ category.

While the results from this and the experiment in chapter 2 cannot be directly compared, changes to the training section (reduction of number of words to learn, inclusion of illustrations to bring home the point as well as entertain, and replacement of

---

\(^8\) All alpha values comparing training conditions should be corrected for multiple comparisons, e.g., \(\alpha/4=0.0125\), but p-values that would be significant by the less conservative measure are also reported.
some of the \textit{tsw}-training stimuli) may have increased the degree of perceptual learning, especially for the \textit{tsw}-condition, which exhibited a robust effect of training in this experiment.

Figure 3.6: Looks to \textit{tw}-target, relative to \textit{sw}-competitor, male talker. The front trained (\textit{tsw}-) group is much more accepting (faster to look to target and more confident overall) of the \textit{tsw}-variant, as spoken by men, than the untrained or \textit{chw}-trained groups.
Figure 3.7: Looks to *tw*- target, relative to *sw*- competitor, female talker. The front trained (*tsw*- ) group is more accepting of the *tsw*- variant, as spoken by women, than the untrained or *chw*- trained groups, though the rates are low compared to other variant+gender combinations.

Figure 3.8: Looks to *tw*- target, relative to *ch*- competitor, male talker. The retracted trained (*chw*- ) group is more accepting of the *chw*- variant, as spoken by women, than the untrained or group; although all trained groups endorse male *chw*- at a very high rate, and even the untrained group looks to the *tw*- target much more often than the competitor.
3.3.2.2 Results – participant gender in single-training conditions

As can be seen in the ANOVA in Table 3.3, participant gender was the most significant main effect, and had significant interactions with condition and with variant by condition. In the single-training groups, female participants had overall more looks to the targets for both /tw/ variants than male participants (Figs. 3.10-13). The interaction of participant gender with /tw/ variant can be seen as female participants’ having especially more looks to the target for the /tsw-/ variant than men (Figs. 3.10 and 3.11). The interaction of participant gender, variant, and condition can be seen as front-trained women have more looks to the target for /tsw/- stimuli, while retracted-trained women have more looks to the target for /chw/- stimuli.
Figure 3.10: Looks to \( t_w \)-target, relative to \( t_s \)-competitor, male talker. The front trained (\( t_s w \)-) women are more accepting of the \( t_s w \)-variant spoken by men, than the other groups, including front trained men.

Figure 3.11: Looks to \( t_w \)-target, relative to \( t_s \)-competitor, female talker. Women in both groups had more looks to the target than men in either group. Of these, the front trained women had the most looks to the target.
Figure 3.12: Looks to $tw$- target, relative to $ch$- competitor, female talker. The retracted trained ($chw$-) women have more looks to the target than the men.

Figure 3.13: Looks to $tw$- target, relative to $ch$- competitor, male talker. The retracted trained ($chw$-) group is on average more accepting of the $chw$- variant spoken by men, but the women in both groups have slightly more looks to the target than the men.
3.3.2.3 Results – split-training condition groups

The eye tracking graphs are laid out for the split-training groups in parallel to the single-trained conditions, in Figs. 3.14-3.17. As a reminder, the expected outcome of training for the gender-split training groups was that participants would find it easier to identify targets for the gender+variant combinations that they heard during training. For example, the traditional group heard men saying chw- and women saying tsw-, so it was expected that they would be more confident and faster to identify chw- as an instance of /tw/ when spoken by men, and tsw- as an instance of /tw/ when spoken by women, but that the reverse distribution would be more difficult (i.e., less expected). The results that emerged as expected are ready acceptance (more looks to target than the untrained group) by the non-traditional group of the chw- variant spoken by women (Fig. 3.14; t(28)=2.8, p<0.0125), and of the tsw- variant spoken by men (Fig. 3.15; t(29)=6.0, p<0.0125), and ready acceptance by the traditional group of the chw- variant spoken by men (Fig. 3.16; t(29)=7.5, p<0.0125), and of the tsw- variant spoken by women (Fig. 3.17; t(28)=3.7, p<0.0125). Thus, the split-training groups did endorse the gender+variant combinations that they were trained on. It was also expected that participants would show greater confusion (shown by fewer looks to the target) for gender+variant combinations that were contrary to training; however, non-acceptance only occurred for the non-traditional group on the female tsw- variant (Fig. 3.17). The traditional trained group quickly identified all four kinds of target, including those not found in training.
Figure 3.14: Looks to tw- target relative to looks to ch- competitor, for chw- stimulus, female talker. Both split-training conditions had more looks to the target than the untrained group.

Figure 3.15: Looks to tw- target relative to looks to sw- competitor, for ts w- stimulus, male talker. Both split-training groups had more looks to the target than the untrained group.
Figure 3.16: Looks to $tw$-target relative to looks to $ch$-competitor, for $chw$-stimulus, male talker. Both split-training groups had more looks to the target than the untrained group.

Figure 3.17: Looks to $tw$-target relative to looks to $sw$-competitor, for $tsw$-stimulus, female talker. The traditional group (trained hearing women say $tsw$) had more looks to the target than the non-traditional group (heard men saying $tsw$) and the untrained group.
3.3.2.4 Results – participant gender in split-training condition

The split training groups were subdivided into male and female participants to visualize the effect of and interactions with participant gender (Figs. 3.18-3.21). The general overview of these four graphs shows that it was female participants who were responsible for the high degree of confidence in identifying both trained and untrained gender+variant combinations in the traditional group. It is as though they analyzed the gender-split distribution of variants as indicative of two endpoints of a single continuous distribution, in which some very front and very retracted affricated variants were possible, without regard to gender, but also including intermediate forms.

The *chw-* variant elicited a similar response when pronounced by women (Fig. 3.18) or men (Fig. 3.19), with female participants in the traditional group having more slightly more looks to the targets than their male counterparts (for female *chw-* \((t(28)=2.7, p=0.01)\), for male *chw-* \((t(29)=2.4, p=0.03)\)), and the non-traditional men and women having insignificantly fewer looks than the traditional trained women. The *tsw-* variant, spoken by women (Fig. 3.20), elicited more-or-less the expected response, with low acceptance by both men and women in the non-traditional group, and greater acceptance by men and women in the traditional group; however, women in the traditional group still had somewhat (but not significantly) more looks to the target in their identification of this variant than traditional men.
Figure 3.18: Looks to tw- target relative to looks to ch- competitor, for chw- stimulus, female talker. While the non-traditional group was trained on female chw-, and men and women in that group performed about the same, men in the traditional group had fewer looks to target, while the women in the traditional group had more looks to the targets.

Figure 3.19: Looks to tw- target relative to looks to ch- competitor, for chw- stimulus, male talker. Men and women in the non-traditional group performed about the same as men in the traditional group (who heard male chw-), but women in the traditional group had more looks to the targets.
Figure 3.20: Looks to \( tw^- \) target relative to looks to \( sw^- \) competitor, for \( tsw^- \) stimulus, female talker. Generally, men and women in the non-traditional group were less accepting of the female \( tsw^- \) than men and women in the traditional group, but women in the traditional group had more looks to the target than men in that group.

Figure 3.21: Looks to \( tw^- \) target relative to looks to \( sw^- \) competitor, for \( tsw^- \) stimulus, male talker. The women in both the traditional and non-traditional group were more accepting of male produced \( tsw^- \) than the men in either group. The traditional trained females had the most looks to the targets.
3.3.3 Discussion of Experiment 2a

The expected outcome of training for the gender-split training groups was that participants would find it easier to identify targets for the gender+variant combinations that they heard during training. For example, the traditional group heard men saying *chw*- and women saying *tsw*-; so it was expected that they would be faster to identify *chw*- as an instance of /tw/ when spoken by men, and *tsw*- as an instance of /tw/ when spoken by women, but that the reverse distribution would be more difficult (i.e., less expected). This pattern would be a very likely outcome if participants came to the experiment with no prior experience of the pronunciation variants, and especially no experience with variation of pronunciation in that category at all, and no analogous (front versus retracted) patterns within the language. However, as might have been expected, the results show that phonetic variation does not exist within a vacuum, but rather is language-specific, and is affected by language users’ prior experience with the language. Each participant learned a pattern, but the associations made by participants were not as simple as they were designed to be.

In some general ways, participants responded as expected during the experiment, with the front and retracted trained groups accepting more of the front (*tsw*) and retracted (*chw*) stimuli, respectively, as instances of /tw/ in the eye tracking identification task. The *chw*- targets spoken by the male talker were generally widely accepted. The *tsw*- targets when spoken by a female talker, showed a split by training condition among men and women, with the traditional group (who heard women using the *tsw*- variant), more
readily identifying the stimulus as a /tw/ target. In this way, the traditional trained group responded most like expected. The only unexpected response from the traditional group was the female participants in that group’s confident identification of male *tsw*- and female *chw*- as well. Perhaps training on a distribution that maintains an exaggerated relationship between gendered productions makes more variability in those distributions acceptable. The traditional men responded as expected, with better identification of /tw/ targets that fit the training profile (men *chw*, women *tsw*) than non-traditional pairings. The non-traditional trained women were confident at recognizing the trained combination (men *tsw*, women *chw*), but were also quick to recognize male *chw*, as everyone did, but not female *tsw*. It seems that men’s pronunciation of the retracted variant was universally acceptable, while the front variant was less acceptable for men and for women, which is possibly attributable to its not being as frequent in the ambient language. The non-traditional men performed the least like expected. Rather than associate the two variants with the gender of the trainers, they seem to have disregarded the *tsw* variant outright, but were quick to endorse the *chw* variant for both genders. In each training condition, women had more looks to the appropriate /tw/ targets than men. It is unknown whether this was because the women were better at identifying and incorporating patterns of variation, or were less perceptive of the associations of each variant, and thus generally more accepting of pronunciation variants. It does bring to mind the findings of women being more likely to accommodate in certain situations (e.g., Pardo 2006), often taking the lead in early sound change in progress (e.g., Labov 1984), and the greater prevalence of autistic characteristics in men than in women, and women
with low autistic quotient actually having a reduced ability to parse out the source of variation (Yu 2010).

3.4 Results of Experiment 2b

3.4.1 Identification task

The results of the identification task are very similar to the eye tracking results, and are also verification of the results from the identification task in chapter 2. As can be seen in Figures 3.22-3.25, all four training groups endorsed a greater range of variants as instances of /tu/ than the untrained group (significant effect of condition \( F(4,104) = 6.1, \ p<0.0125 \)), as was the case in the experiment in chapter 2. Another parallel finding is that these front and retracted trained groups also exhibited shifting of their /tu/ perceptual space as a result of training on variant pronunciations of /tw/. Another similarity to the ID task in chapter 2 is the limitation imposed by the interaction of gender and sound category: the perceptual space for the male talker, who is expected to be more retracted, has less room for expansion along the two~chew boundary, while the perceptual space for the female talker, who is expected to have a more front pronunciation, has more flexibility along the two~chew boundary, but slightly less wiggle room at the two~tsu boundary. Also note that the two~tsu boundary has a less steep slope for both talkers because /ts/ is not a contrastive sound in English.
Figure 3.22: Results of identification task for male talker, with female listeners only, split by training condition. The retracted trained group accepted more retracted affrication and less front affrication, while the front trained group accepted more front affrication and less retracted affrication, relative to each other. Both split-training groups exhibited a widened category with shifted boundaries on each side of /tu/. With the exception of one outlier at the point of 50% between two~tsu, the traditional group (men chw-) endorsed fewer front tokens as instances of /t/.

Figure 3.23: Results of identification task for male talker, with male listeners only, split by training condition. The retracted trained group accepted more retracted affrication and less front affrication, while the front trained group accepted more front affrication and less retracted affrication, relative to each other. The socially traditional trained group (men chw-) endorsed more front affricated tokens than the group that had heard men using the front variant during training, which parallels the finding in the eye tracking task, in which men in the non-traditional condition disregarded the tsw- variant.
Figure 3.24: Results of identification task for the female talker, with female listeners, split by training condition. All of the groups endorse a greater range of affricated variants as instances of /tu/ than the untrained group, but there are no between group differences.

Figure 3.25: Results of identification task for the female talker, with male listeners. The retracted trained group endorsed more retracted variants as /tu/ than the other groups, resulting in a more retracted two~chew boundary. The traditional group (men chw~, women tsw-) endorsed more front affricated variants than the non-traditional group, resulting in a more advanced two~tsu boundary. Non-traditional male listeners performed similarly to those in the eye tracking experiment, in rejecting the front affricated variant.
Just as in the eye tracking experiment, the non-traditional trained men (men tsw-, women chw-) seem to have disregarded the tsw- variant altogether, evidenced by the lack of expansion along the two~tsu boundary, relative to the untrained group. However, the men and women in the single training conditions performed similarly. The greatest differences can be seen for the male talker, in which the two~tsu and the two~chew boundaries are complementarily affected by the front and retracted training. That is, the entire /tu/ category has shifted in two opposite directions based on training. Responses to the female talker show general broadening of the category for all but the untrained group and the non-traditional group.

3.5 Discussion And Conclusions

While the split-gender training did not have the expected effect, in which the trained variant+gender combination would elicit the fastest and most accurate identification responses, there was an effect of training for each trained condition. The front (all tsw-) and retracted (all chw-) conditions responded similarly to the groups in the first experiment in chapter 2, with greater acceptance of the trained variant, and a corresponding shift in the perceptual space of /tu/ (though only for the male talker, in this case). In the ID task, the socially traditional (or analogous) male chw-, female tsw- trained group exhibited a less steep boundary between two~chew for the male talker, indicating more ch-like tokens in their /tu/ representation. The socially non-traditional male tsw-, female chw- trained group exhibited a less steep boundary between two~tsu for the male talker, indicating more ts-like tokens in their /tu/ representation. Differences
in the eye tracking results between the two split-training conditions, and differences between the single-variant groups and the split-training conditions for the different genders suggest that the men in the traditional (male \textit{chw-}, female \textit{tsw-}) group successfully underwent a gender-conditioned sound split. Women in the traditional group accepted both training variants for both genders, and so did not exhibit a split, but rather allowed greater variation within both the /tw/ and /tu/ categories, without regard to gender. Women in the non-traditional group (male \textit{tsw-}, female \textit{chw-}) exhibited a split in which men could use either variant, but women were restricted to the more generally acceptable \textit{chw-} variant. Men in the non-traditional group experienced a sound shift in that they only endorsed \textit{chw-} for both men and women. So, all in all, there was category widening, category shifting, and category splitting.

The interaction of the analogical similarity of \textit{chw-} to the historical outcome of /tr/ and /tj/, along with the social weighting of front and retracted pronunciations of /s/, and by extension, /t/, doesn’t allow a quick transformation of the /tw/ category using these two unequal variants. The \textit{chw-} variant will continue to be more readily accepted by talkers of both genders as long as the /tr/ and /tj/ analogs exist for both genders. It would require many more exemplars to accumulate, or an explicit social judgment to be made, before the analogical pattern of men exhibiting more retracted productions than women could be overridden. Of course, it remains to be seen just how many more exemplars, or exactly what overt social marking would be required to change the pattern.
As laid out in the introduction, it was believed that if associations to a non-linguistic parameter, such as the gender of the talker, could condition perceptual adjustment in two directions, that would provide evidence for an exemplar theory of sound change, in which associations, single or multiple, can be formed with pronunciation variants, which can guide the trajectory of sound change. While the evidence in this chapter is not terribly clear, there is nevertheless evidence for the above hypothesis, as well as corroborating evidence for the findings in chapter 2, which support the methodology of researching sound change in the laboratory. Once again, the results of the identification task provided evidence that generalization could take place at the same time as phonetic conditioning occurred, that even though the conditioning environment was /tw/, /tu/ was also affected. The experiment also showed that non-linguistic factors can and do influence perception of sound categories, such that the location and slope of the category boundaries for /ts/~t/ and /ʧ~/~t/ varied depending on the gender of the talker, the gender of the participant, and the training condition. An exemplar theory of sound change, in which variation in pronunciation can come to be associated with phonetic and/or non-phonetic factors, is supported by the kinds of changes exhibited by both the single-variant and the split-training participants.
Chapter 4: Dynamics of perception and production

4.1 Introduction

We are just beginning to understand how dynamic perception and speech processing really are. At one time, it was thought that once a language user’s phonological system was acquired, it remained unchanged throughout his or her lifetime\(^9\), but it has since been demonstrated that patterns of speech production, at least, do change over a person’s lifetime (e.g., Bowie 2000, Harrington 2006, Sankoff and Blondeau 2007), and that the phonological system, in fact, rapidly adapts to changes in input (e.g., Eimas and Corbit 1973, Norris, McQueen, and Cutler 2003). Some of these effects have been explored in experimental paradigms. Selective adaptation to speech sounds (e.g., Eimas and Corbit 1973) occurs when repeated exposure to a speech sound reduces the number of potential targets that are acceptable as instances of that sound, thus narrowing the category. Selective adaptation can be applied to various dimensions, such as place of articulation or VOT, and may be generalized across phonemes sharing that dimension (e.g., Eimas and Corbit 1973, Miller 1981, Samuel 1986). Perceptual adjustment in speech (e.g., Norris, McQueen, and Cutler 2003) occurs when repeated exposure to an ambiguous speech

---

\(^9\) Thus sound change, in the views of structuralists and generativists of the last century, occurred due to differences between the phonological system of the adult and that acquired by the child. See, e.g., Chomsky and Halle 1968, and Bloomfield 1933.
sound creates a shift in the acoustic-perceptual boundary between two contrastive sounds, thus increasing the range of acoustic space from which one of the categories draws its tokens, at the expense of the other category. This type of change may involves a shift in only the one boundary where the ambiguous targets lie, thus the category expands along one dimension to include the area under the new boundary (as the single-trained groups do in the ID task in chapters 2 and 3), or else the overall area (or volume, etc.) of perceptual space belonging to the target category stays the same, resulting in a shift of the entire category, and corresponding shifts along multiple boundaries (which might have occurred in the experiments in chapters 2 and 3 prior to the additional effects of testing). Perceptual adjustment can also occur along multiple dimensions, such as place of articulation or VOT, and may be generalized to other phonemes that share the adjusted feature (e.g., Norris et al. 2003, Kraljic and Samuel 2006). A third form of re-tuning the perceptual system has not (to my knowledge) been as widely discussed in the literature (however, cf. “range effects” found in Keating Mikos and Ganong 1981, and Clayards et al. 2008), but is likely to have been demonstrated in many experimental situations that include training a listener with variable speech sounds, during which the listener’s perceptual category expands to include more variability on both (or multiple) sides of the category’s perceptual space. These changes may occur as a result of this incoming variability, or by the expectation of variability.

This chapter lays out a theoretical description of the nature of phonological representations and how they are mapped onto an acoustic space, and especially details how this acoustic-phonological perceptual space adapts to new input. The known types of
changes to phonological representations are discussed under a probabilistic/exemplar theoretical account, and the ways these phonological changes may contribute to sound change is explored. Evidence is presented for the previously unexplored type of perceptual re-tuning, which will be referred to as perceptual broadening. This chapter will also explore some related short-term changes in production whose lack of persistence can be explained by social factors, providing one answer to the question ‘when does variation NOT lead to sound change?’

4.1.1 Mapping abstract representations onto a perceptual space

I take the position in this thesis that a sound category can be viewed as an abstraction that is mapped onto a multidimensional acoustic space and which emerges from a grouping of similar exemplars that share acoustic (and/or articulatory) characteristics. This abstraction may have different configurations stored in memory, as the result of accumulations of exemplars sorted by linguistic information, talker characteristics, situations, or other relevant information, but it is also sensitive to changes in the input and can adapt quickly to the different probabilistic distributions of incoming exemplars, as evidenced by changes such as perceptual adjustment or selective adaptation. While these representations are robustly multidimensional, for illustrative purposes, the graphs in this chapter are two-dimensional, to illustrate changes occurring along one or two dimensions in one of the many possible representations.

Probabilistic and/or exemplar models can account for dynamic changes to the mapping of categories on an acoustic space in a way that mirrors observed experimental
outcomes such as selective adaptation and perceptual adjustment (see, e.g., Kleinschmidt and Jaeger 2012). One way of achieving this dynamic mapping is that the training stimuli are weighted for recency and the perceptual space is temporarily modified according to the distribution of these recent exemplars, as illustrated in the hypothetical distributions in Figures 4.1-4.3. Kleinschmidt and Jaeger (2012) use Bayesian Belief Updating, which similarly allows the parameters of the category to change based on the probabilistic distribution of recent input.

These concepts can be easily demonstrated using hypothetical data, such as the manipulated VOT of voiceless stops, as in Eimas and Corbit 1973 and Kraljic and Samuel 2006. Selective adaptation to speech sounds (Fig. 4.1) can be visualized as a process of making the category more selective about what may count as an instance of the sound in question. During selective adaptation training, listeners are exposed to a sound with little or no variability. In order to increase efficiency of processing, the category can shrink to include a much smaller pool of exemplars. After selective adaptation training, listeners endorse a much smaller range of variants as belonging to a particular category than they did before training. While encountering extremely limited variability seems unlikely to occur outside of an experimental setting, perceptual narrowing may also occur when a small subset of a category that encompasses a large perceptual space splits off from the rest of the category.

Perceptual adjustment training (Fig. 4.2) exposes listeners to a range of exemplars that lie at one edge or just beyond that boundary of the category. To accommodate the incoming variation, the affected boundary must be adjusted to fit the distribution of new
exemplars, often creating the potential for overlap with a neighboring category, which can also shift, or if it does not, increases the range of acoustic space in which both categories are in competition. After perceptual adjustment training, listeners endorse more tokens that were previously ambiguous or belonging to a neighboring category as exemplars of the target. Perceptual adjustment may be a mechanism that occurs during sound shifts, as variation in a sound category leans in one direction, pulling the category mean along with it.

Perceptual broadening (Fig. 4.3) reduces the strength of the perceptual magnet effect (that is the warping of the perceptual space around the median pronunciation) because of the dispersion of newly accumulated exemplars. In normal conversational interaction, a variety of words with a variety of sounds are uttered, that, as a sample, should be theoretically similar to the token distribution of sounds in a given language. The longer the conversation, and the more conversational topics, the closer the sample should be to the population of sounds in a language\(^\text{10}\). But in an experimental environment, in which individual words are presented as training or testing stimuli, the distribution of sounds is different. For example, experimental word lists are generally composed of content words, especially in their base forms (i.e., lacking declension or conjugation), so sounds that are more frequently encountered in function words or certain morphemes are underrepresented in experiments of this type. Additionally, when one of the

\(^{10}\) This is my logical assumption, and is discussed in Beckman et al (forthcoming) *Analyzing the sounds of languages* chapter 5; however, it is an empirical question that has not actually been demonstrated, to the best of my knowledge.
manipulations includes exposing participants to a certain sound, that sound experiences a boost in exposure relative to the exposure encountered in a normal conversational setting. During perceptual broadening training, listeners are exposed to a wide range of variation. The breadth of variation, while within the range normally experienced in speech over a long period of time, is greater than that normally encountered in a conversational interaction, and creates the expectation of a broader category, as a more widely dispersed range of exemplars are activated in close temporal proximity to one another. As the perceptual space of the category increases to accommodate more variation, the probability of the more central exemplars’ being activated decreases (relative to the previous distribution), while the probability of activation of the more extreme exemplars increases. After perceptual broadening training, listeners endorse a greater range of variants in multiple directions as instances of that category, as illustrated in section 4.3. The phenomenon of perceptual broadening could explain the loss of non-native contrasts during early language acquisition (Werker and Tees 2002), as well as more temporary re-tuning in an experimental paradigm, and could describe the re-shaping of individuals’ phonological spaces during phonological mergers or push chain shifts. The first experiment addresses perceptual broadening that results from exposure to variation.

Another condition under which perceptual broadening may take place is when a listener is uncertain about the pattern of incoming speech. When a listener encounters an unfamiliar accent, the listener must be prepared to adapt to unexpected variation, which may extend to or beyond the boundaries of previously experienced categories. If a listener recognizes a certain dialect or idiolect, he or she can rely on stored exemplars
associated with that dialect or idiolect. Experience with any of the types of input mentioned above that result in perceptual changes could bring about the same kind of phonological changes if the listener anticipates that those settings would be useful. For instance, listeners in a perceptual adjustment experiment retained changes to their perceptual space after 25 minutes of silent activity (Kraljic and Samuel 2005), anticipating a potential on-going need for the adjusted categories. Similarly, it is my hypothesis that inexperience with a variety may also allow widespread pre-activation of the category’s extreme values in order to prepare for incoming exemplars across a wider perceptual space than usual. This kind of pre-activation relies on abstract predictions about the nature of the speech about to be processed, just as generalization of perceptual learning to different speakers does (e.g., Kraljic and Samuel, 2006). If the variety is unfamiliar, all, or a widely dispersed subset of exemplars should be prepared for activation. This state of preparation allows sound category boundaries to rapidly shift along the dimensions that seem most variable in the incoming speech signal. The listener attends to the dimension of speech that varies the most from the expected pattern, expanding the sound category along those dimensions to prepare for incoming exemplars. When we give a pre-test, for example, a lexical decision task, as part of an experimental paradigm, we may actually be influencing how participants will perceive the following training stimuli, and the weirder the forms in the pre-test, the more likely it may be that participants will form an expectation about variability along the dimensions being tested. Additionally, information about the talker(s) or stimuli that is given to or withheld from participants may affect how they sort the incoming variability.
It has been demonstrated that listener expectations and perceptions about a talker can influence the way that talker’s speech is perceived (e.g., Hay, Nolan, and Drager 2006, Niedzielski 1993, Babel 2012). Thus, if a listener knows ahead of time that he will be listening to a talker of a specific familiar variety, the listener can prepare his phonological space to adapt to a known configuration, relying on stored exemplars associated with that variety. But what happens when a listener is unfamiliar with a talker’s dialect and has no available representation for what the new accent should sound like? It is my position that when a listener encounters an unfamiliar accent, or has some reason to believe that the talker’s speech might somehow surpass previously encountered variability, the listener should be more prepared to adapt to unexpected variation, thus calling on all stored exemplars in a category to be at the ready. This state of preparation allows sound category boundaries to rapidly expand along the dimensions that seem most variable in the incoming speech signal. For example, participants in the experiments reported by Maye, Aslin, and Tanenhaus (2008) demonstrated an increase in endorsement rates not only for the stimuli that were shifted in the direction of training (front-vowel lowered), but also in the opposite direction (front-vowel raised)\(^{11}\), which was replicated by Weatherholtz (2012), suggesting an increase in category size possibly resulting from a lack of familiarity with the idiolect (or a weirdness factor) combined with exposure during lexical decision tasks to both kinds of variation. It is logical that in encountering an unknown accent, in order to be able to adapt quickly, the perceptual boundaries would

\(^{11}\) While Maye et al. claimed that the relative weakness of the increase in acceptance of front-vowel raised items demonstrated that the effects were not due to perceptual broadening, the increase was nonetheless significant both by items and by subjects, suggesting that some perceptual broadening did occur.
need to be flexible. Enlarging the perceptual space for an apparently variable sound category would allow more possible matches to known lexical items, which would facilitate communication, though at some processing cost (as in Clopper, forthcoming). The second experiment examines this kind of perceptual broadening.

Figure 4.1: Selective adaptation is a kind of perceptual sharpening that results from accumulation of exemplars in a very tightly focused perceptual space. Hypothetical distribution based on Eimas and Corbit 1973. White bars represent the probabilities of exemplars occurring in that acoustic space of the perceptual category before training. Grey bars represent the distribution of the incoming stimuli. The adjusted perceptual space would be calculated from the combination of these two distributions, with additional weighting of the more recently acquired exemplars.

Figure 4.2: Perceptual adjustment is a kind of shift in the perceptual space that results from the accumulation of exemplars in a different distribution from that normally experienced. Hypothetical distribution based on Kraljic and Samuel 2006. White bars represent the probabilities of exemplars occurring in that acoustic space of the perceptual category before training. Grey bars represent the distribution of the incoming stimuli. The adjusted perceptual space would be calculated from the combination of these two distributions, with additional weighting of the more recently acquired exemplars.
Figure 4.3: Perceptual broadening is a broadening of the perceptual space resulting from the widely dispersed accumulation of recent exemplars. Hypothetical distribution based on participants’ baseline distribution from Eimas and Corbit 1973. White bars represent the probabilities of exemplars occurring in that acoustic space of the perceptual category before training. Grey bars represent the distribution of the incoming stimuli. The adjusted perceptual space would be calculated from the combination of these two distributions, with additional weighting of the more recently acquired exemplars.

4.2 Experiment 3

The data illustrating perceptual broadening from exposure to variation are taken from filler trials in the experiment described in chapter 3, but for the sake of clarity will be
referred to as experiment 3. A complete description of the experimental design can be found there, but a brief recapitulation of the relevant points is given here.

4.2.1 Stimuli

Two instances each of 30 /v/ words were used for training. 50 /f/ (as in *fat*) and 50 /b/ (as in *bat*) initial words and non-words plus two instances each of 50 /v/ (as in *vat*) words and non-words were used as test stimuli. /tw/ words were also part of the experimental stimuli, but are not relevant to this analysis. Six talkers (3 male, 3 female) produced the training stimuli, and four new talkers (2 male, 2 female) produced the test stimuli. All sound files were digitally recorded in a soundproofed recording chamber at a sampling rate of 44,100 Hz, using Audacity. The files were normalized to 70dB for presentation over headphones as training and testing stimuli.

4.2.2 Participants and Procedure

Participants (here, a subset of 24 from one training condition (the *chw*- condition): 14 female, 10 male) underwent training, in which they heard and said 30 English words beginning with /v/ (plus 30 tw- words, not discussed here). Each word was spoken by one male and one female talker, and each of these tokens was played three times for the participant over the course of training, for 3 repetitions of 60 unique tokens of word-initial /v/. An additional 10 control group participants (6 female, 4 male) did not receive training before proceeding to the testing section. For testing, participants were seated in front of a Tobii eye tracker with a computer screen, which for each trial displayed 4 written choices including words and non-words; an example slide is shown in Fig. 4.4.
After hearing a stimulus, they selected the word or non-word that they had heard by clicking on the appropriate target with a mouse. The location of their gaze was also recorded about every 17-20 ms. Targets beginning with /f/ or /b/ always had a /v/ competitor. Targets beginning with /v/ had /b/ competitors half the time, and /f/ competitors the other half of the time. Targets and competitors included words and non-words, and because of the distribution of the language, could not be completely evenly distributed in each category; although, attempts were made to make the distributions of words and non-words as targets and competitors as even as possible.

Figure 4.4: An example of four word and non-word choices participants saw on the computer screen during eye tracking trials.

4.2.3 Acoustic Analysis

While all tokens were judged prior to use in the experiment to sound like good examples of the words and non-words they were intended to represent, acoustic analysis revealed variation in duration, amplitude, strength of voicing, and strength of frication. Voiced fricatives are articulatorily complex to produce, with competing requirements in
air pressure at the glottis and place of constriction resulting in a large degree of variation. When the sub-glottal air pressure is stronger than the oral air pressure, voicing is strong, but the frication is reduced; when the oral air pressure is greater, frication is strong and voicing is suppressed (Stevens 1998: p. 465); when the constriction becomes too narrow but sub-glottal pressure remains strong, the voiced fricative becomes more like a voiced stop (Stevens 1998: p. 379). Therefore, the difference between manner of articulation in voiced stops, voiced fricatives, and voiceless fricatives is one of degree rather than a categorical difference in gestures.

4.2.4 Test Stimuli

To illustrate the gradual nature of the differences between /b/, /v/, and /f/, the test stimuli are graphed in Fig. 4.5. While /b/ is generally bilabial and /f/ and /v/ are generally realized as labiodental, there is no contrast between labials and labiodentals in English; therefore, manner of articulation is more important than place in distinguishing among these three sounds, but acoustic analysis reveals some degree of overlap, resulting in a continuous distribution. The /v/ training stimuli are compared in Fig. 4.6. Measurements were averaged over the entire duration of the stop/fricative by calculating the mean across six overlapping windows across the entire segment. Center of gravity (COG) illustrates the contribution of different frequency components, where more frication increases the COG, while voicing lowers it. Tokens with both voicing and frication have more intermediate values. Spectral slope isolates the voicing component (0-500Hz) and subtracts it from the higher frequency amplitude (1000-22050Hz). The more negative the
slop, the greater the contribution of voicing. These two measurements illustrate the gradual change from stop to voiced fricative to voiceless fricative. Generally, /f/ tokens had a slope of -20 to 0dB and COG of 3000-7000Hz. /b/ tokens’ slope ranged from -50 to -35dB, and their COG was generally below 1000Hz. The /v/ tokens were intermediate to these, with a slope averaging from -40 to -20dB, and a COG concentrated between 250-2000Hz.

Figure 4.5: Distribution of /f/, /v/, and /b/ test stimuli, showing normal amounts of overlap of /v/, /b/, and /f/.

4.2.5 Training Stimuli

The 60 /v/ tokens used in training were analyzed to see how much variation participants were exposed to (Fig. 4.6), along both center of gravity (~voicing) and
spectral slope (~manner). COG measurements indicate a number of primarily voiceless tokens, with COG exceeding 2000Hz, but the majority of tokens had some degree of voicing, with COG around 1000Hz or below. Slope measurements show several stop-like tokens, with very strong voicing components and little or no frication.

Figure 4.6: Distribution of /v/ training stimuli. Note the overlapping distribution with /f/ and /b/ test stimuli in Fig.4.5.

Out of the types of phonological changes described in the introduction, exposure to this distribution of variation should cause perceptual broadening of participants’ /v/ category, so that they identify more voiceless (/f/) and stop-like (/b/) sounds as instances of /v/, which would result in lower accuracy for /f/ and /b/, and fewer looks to the target relative to the competitor.
4.2.6  Accuracy (Mouse Click) Results

Accuracy results for the trained group are listed in the top four rows in Table 4.1. An ANOVA on item means split by condition, and target and competitor sound and word or non-word status, showed a main effect of target sound \( (F(3,193)=26.0, p<0.05) \), in that /v/ targets were most accurately identified, even though the distribution of /v/ overlapped with the perceptual space for /b/ and /f/. A main effect of target word status \( (F(1,193)=45.1, p<0.05) \) demonstrated that word targets were more accurately identified than non-word targets. A main effect of competitor word status \( (F(1,193)=6.2, p<0.05) \) showed that targets with non-word competitors were more accurately identified than targets with word competitors. The combination of these two results, which is supported by a marginal interaction between target and competitor word status \( (F(1,194)=3.8, p=0.053) \), illustrates the Ganong (1980) effect, in that non-word targets with real word competitors were least accurately identified and that real word targets with non-word competitors were generally more accurately identified. However, a significant interaction between target word status and target sound \( (F(3,193)=6.3, p=0.053) \) demonstrated that /v/ targets were nearly immune to this effect, having consistently higher accuracy rates even when the target was a non-word, while /f/ and /b/ targets suffered from reduced accuracy when they were non-words.
Table 4.1: Accuracy rates for /f/, /b/, and /v/ stimuli, split by word status (whether the targets and competitors were words or non-words) and training condition. Trained participants favored /v/ over /b/ and /f/ in almost all conditions.

<table>
<thead>
<tr>
<th>training condition</th>
<th>target~competitor</th>
<th>b target~v comp</th>
<th>v target~b comp</th>
<th>f target~v comp</th>
<th>v target~f comp</th>
</tr>
</thead>
<tbody>
<tr>
<td>trained</td>
<td>non~non</td>
<td>0.728</td>
<td>0.922</td>
<td>0.946</td>
<td>0.979</td>
</tr>
<tr>
<td></td>
<td>word~non</td>
<td>0.911</td>
<td>0.985</td>
<td>0.981</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>non~word</td>
<td>0.576</td>
<td>0.903</td>
<td>0.759</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>word~word</td>
<td>0.898</td>
<td>0.994</td>
<td>0.913</td>
<td>0.971</td>
</tr>
<tr>
<td>no training</td>
<td>non~non</td>
<td>0.865</td>
<td>0.943</td>
<td>0.945</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>word~non</td>
<td>0.983</td>
<td>0.976</td>
<td>0.994</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>non~word</td>
<td>0.858</td>
<td>0.945</td>
<td>0.896</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>word~word</td>
<td>0.994</td>
<td>1.00</td>
<td>0.971</td>
<td>0.976</td>
</tr>
</tbody>
</table>

Accuracy results for the untrained control group are shown in the bottom half of Table 4.1. Control group participants did not exhibit the same asymmetries as the trained group; therefore, they were much more accurate at identifying /f/ and /b/ targets because they did not misidentify them as instances of /v/. Paired t-tests comparing item means across conditions for /b/ stimuli (t(49)=6.3, p<0.05) and /f/ stimuli (t(49)=3.7, p<0.05) confirm this analysis: /b/ and /f/ targets were more accurately identified by untrained participants, who did not undergo perceptual broadening as a result of training. T-tests comparing item means of /v/ stimuli were not significantly different across conditions.

4.2.7 Eye tracking results

The different groups’ on-line processing of /f/ and /b/ and /v/ shows a number of interesting details about the effects of perceptual broadening. Figures 4.7-4.10 display the log odds of looks to the target divided by looks to the competitor over time, beginning with the onset of the stimulus. In order to statistically analyze the data, time was
collapsed into two bins, one for the first 500 ms beginning with and immediately following the onset of the stimulus, and from 500 ms to 1500 ms after stimulus onset. Then mean log odds over each time bin was calculated for each stimulus item. Because the stimuli were of varying durations, a by-items ANOVA was calculated for each time bin, in which item means were compared across training conditions, with sound pair (v-target~f- competitor, etc.) as a between-items factor, and training condition as a within-items factor, and stimulus as the random error term. The ANOVA of the first time bin revealed no main effects or interactions. The second time bin showed a main effect of sound pair (F(3,204)=3.7, p<0.05) and training condition (F(1,204)=50.0, p<0.05), and an interaction between sound and training (F(3,204)=9.96, p<0.05). Post-hoc t-tests reveal that untrained participants had more looks to the target than trained participants for /b/ stimuli (t(49)=6.7, p<0.05; Fig. 4.7), and /f/ stimuli (t(48)=5.0, p<0.05; Fig. 4.8). The /v/ targets with /f/ competitors (Fig. 4.9) and /b/ competitors (Fig. 4.10) did not show any differences between training groups.

\footnote{One of the /v/ stimuli had to removed from analysis of the eye tracking results due to coding errors, though it was able to be corrected in the mouse click results.}
Figure 4.7: Trained and untrained participants log-odds looks to /b/ targets vs. /v/ competitors. The untrained group exhibited more looks to the target.

Figure 4.8: Trained and untrained participants log-odds looks to /f/ targets vs. /v/ competitors. The untrained group had more looks to the target.
Figure 4.9: Trained and untrained participants log-odds looks to /v/ targets vs. /b/ competitors. There is no significant difference between training groups in looks to target.

Figure 4.10: Trained and untrained participants log-odds looks to /v/ targets vs. /f/ competitors. There is no significant difference between groups in looks to target.
4.2.8 Experiment 3 Discussion

As predicted by a model of perceptual broadening, trained participants performed the same as untrained participants on /v/ stimuli with /b/ and /f/ competitors, but trained participants had fewer looks to the targets than the untrained participants as well as reduced accuracy on mouse click identification for /b/ and /f/ stimuli with /v/ competitors. This response pattern appears to be the inverse of the effect found for selective adaptation, in which training with reduced variability leads to reduction of the acoustic perceptual space that is mapped to a particular category. Experiment 3 shows that training with variable stimuli leads to a broader mapping for that sound category, at the expense of neighboring categories.

It is possible that phonological priming could be responsible for these effects; however, if this were the case, perceptual adjustment would also be considered a result of phonological priming, as one sound category is continuously activated by hearing a number of words containing that sound. In fact, in a system in which sound categories are abstract generalizations that emerge from the lexicon, we might expect that rapid accumulation of a number of exemplars would increase the resting activation level of the entire sound category. However, it does not necessarily follow that the perceptual boundaries of that sound category would be affected by simple activation. The evidence that perceptual learning and perceptual broadening change the mapping of sound categories to a different perceptual space indicates a more dynamic process than simple priming. Additionally, one would expect much faster looks to /v/ targets for trained
participants than untrained participants in this experiment, if there were priming effects, but it would seem that any benefit of exposure to /v/ words during training was canceled out, possibly due to processing costs associated with broader categories (cf. Clopper, forthcoming). In addition to this evidence, experiment 4 shows that expectation of variability can also elicit perceptual broadening, beyond that elicited by variability in training or testing stimuli alone, and that the effect on production is more limited in scope than that obtained during perceptual adjustment in chapter 2.

4.3 Experiment 4

In addition to the kind of perceptual broadening that results from extensive and variable input, perceptual broadening can also occur when a listener has only the expectation that the input may be broadly variable. When a listener anticipates that the speech they hear will be different from what they are familiar with, they may expand speech sound categories along the dimensions in which they notice variability.

The research question guiding this experiment was to find out how participants would respond to the variability introduced in the laboratory sound change paradigm if they suspected that the variability was specifically attributable to a stigmatized (yet for many, unfamiliar) language variety or dialect. While it is usually the case that listeners try to understand talkers, and make appropriate adjustments to their perceptual mappings, it has also been shown that listeners may make adjustments that are counterproductive. For example, Rubin (1992) demonstrated that when students heard a native speaker’s voice and believed the talker was Asian, they reported that the talker sounded “more foreign”
than when students heard the same voice and believed the talker was white. Three possible outcomes could result from exposure to an unfamiliar yet stigmatized dialect: a) lack of perceptual adaptation, indicating the unwillingness to meet the speaker in the middle, as observed by e.g., Rubin (1992); b) perceptual adaptation fitting the distribution of variability exhibited by the talker, as exhibited by most of the participants in the experiment described in chapter 2; or c) confusion about how to handle the incoming variation, and subsequent perceptual broadening along the dimension(s) of variability exhibited by the talker, as in Maye et al. 2008.

While listeners must adapt their perceptual space to handle differences between language varieties and different accents if they want to communicate successfully, and may even converge in speech production (e.g., Kim, Horton, and Bradlow 2011, but cf. Bourhis and Giles 1977, among others), it seems unlikely that these kinds of interactions can encourage rapid sound change, unless the innovative talker has a higher prestige level than the listener (as in Labov’s (1966 and later) “change from above”). If immediate language change were a natural and unavoidable outcome of casual dialect contact, we would already have seen massive dialect leveling in this new global age in which more and more people have the opportunity to meet people from outside of their dialect region, and access to media brings many accents and dialects into people’s homes. However, we do know that sound change from below can eventually reach the upper crust of society, as in the vowel shift toward the vernacular observed in Queen Elizabeth (Harrington 2006). Therefore convergence was also measured for the participants in the current experiment, and the results of this production study suggest that because participants could attribute
pronunciation differences to a specific (and stigmatized) dialect, this encouraged a lack of persistence in convergence that was in strong contrast to the standard trained group.

4.3.1 Materials

The stimuli for this experiment were the same recordings described in chapter 2 for the retracted affricated (chw-) training condition. All of the test stimuli for the lexical decision and identification tasks were also the same as described in chapter 2.

4.3.2 Participants and Procedure

19 students (6 m, 13 f) participated in the Appalachian condition of this experiment, as part of a course requirement. Results from the other conditions from the experiment in chapter 2 are reviewed here for comparison. The procedure and stimuli were identical to the experiment in chapter 2, except that the definitions and sentences were written on the screen in Appalachian-style diction. For example, whereas the experiment in chapter 2 explained the term viator as follows: “a traveller, a wayfarer,” the definition for the Appalachian condition read: “somebody who wanders around a lot and don’t never come home.” Additionally, participants were told in the instructions preceding the experiment that “the unfamiliar vocabulary words you will learn are archaic, obsolete, or obscure forms which may be retained in geographically and socially isolated communities, such as those in the Appalachian region.” Thus while the Appalachian condition listeners heard the exact same stimuli as the retracted trained group in the experiment in chapter 2, the possibility that the talkers were Appalachian was present in their minds, which undoubtedly colored their perception of the stimuli. In fact, one participant adopted
vowels that are actually used in Appalachian pronunciations: monophthongal /ai/ and a raised diphthongal /æ/.

4.3.3 Lexical Decision Results

In comparison to the standard retracted trained group, the Appalachian trained group performed much worse on the unfamiliar trained words with the chw- pronunciation variant (t(40)=2.6, p<0.05). However, they performed as well as the standard chw-trained group on familiar chw- words, and endorsed as many as, if not slightly more of, the familiar untrained tsw- words than the tsw- trained group, as can be seen in Table 4.2. The Appalachian group only had difficulty remembering the unfamiliar tw- words. Because the definitions and sentences were written with Appalachian diction, they may not have been as clear, and without valid semantic processing, the words could not end up in long-term memory. However, the Appalachian group performed as well as the other groups on the unfamiliar en- and vi- words, which suggests that they processed words with a pronunciation variant differently from the other words, and that memory encoding was not the issue because they did learn the unfamiliar en- and vi- words. Though they heard the chw- variant during training, the Appalachian group endorsed familiar words with the tsw- variant at a greater rate than the other training groups, as well as endorsing familiar words with the chw- variant as well as the standard chw- trained group. These results suggest that perceptual broadening occurred, in that not only did they incorporate the variant they were trained on into their phonological systems, but also the variant that
they did not hear during training. The diagnosis of perceptual broadening is also supported by their performance on the identification task.

Table 4.2: Lexical decision accuracy rates (endorsement rates) for familiar and unfamiliar, trained and untrained words, featuring /tw/ pronunciation variants, and vi- and en- words for comparison.

<table>
<thead>
<tr>
<th>Training group</th>
<th>tsw-variant</th>
<th>chw-variant</th>
<th>vi- and en-words</th>
<th>tsw-variant</th>
<th>chw-variant</th>
<th>vi- and en-words</th>
<th>tsw-variant</th>
<th>chw-variant</th>
<th>vi- and en-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group [tw] trained</td>
<td>79.3%</td>
<td>75.3%</td>
<td>81.8%</td>
<td>92.6%</td>
<td>96.2%</td>
<td>97.5%</td>
<td>78.9%</td>
<td>84.2%</td>
<td>97.6%</td>
</tr>
<tr>
<td>Front (tsw-) trained</td>
<td>79.3%</td>
<td>64.2%</td>
<td>82.9%</td>
<td>94.3%</td>
<td>98.6%</td>
<td>96.7%</td>
<td>89.4%</td>
<td>84.1%</td>
<td>96.8%</td>
</tr>
<tr>
<td>Retracted (chw-) trained</td>
<td>72.5%</td>
<td>78.4%</td>
<td>83.7%</td>
<td>96.9%</td>
<td>97.6%</td>
<td>96.7%</td>
<td>78.7%</td>
<td>88.0%</td>
<td>98.3%</td>
</tr>
<tr>
<td>Appalachian (chw-) trained</td>
<td><strong>68.3%</strong></td>
<td><strong>59.2%</strong></td>
<td><strong>82.8%</strong></td>
<td><strong>95.0%</strong></td>
<td><strong>98.6%</strong></td>
<td><strong>95.5%</strong></td>
<td><strong>92.5%</strong></td>
<td><strong>87.5%</strong></td>
<td><strong>93.6%</strong></td>
</tr>
</tbody>
</table>

4.3.4 Identification Results

The identification (ID) task results (Figs. 4.11 and 4.12) show some perceptual broadening of the /tu/ category for all trained participants, most especially the Appalachian trained group, while untrained participants exhibit a narrower perceptual space for /tu/. A 3-way ANOVA across subject means for each stimulus revealed a main effect of training condition (F(3,88)=7.9, p<0.05), with different levels of /t/ endorsement across training groups, a main effect of stimulus composition (F(12,1056)=583, p<0.05), because responses differed depending on which point along the continuum was presented...
(50% tsu, etc), and an interaction between training and stimulus (F(36,1056)=2.6, p<0.05), showing shifting for some groups, and an interaction between stimulus and talker gender (F(12,1056)=82, p<0.05), as the male distribution was centered more in the retracted space (Fig. 4.11) and the female’s was centered more in the front space (Fig. 4.12).

In follow-up comparisons, /t/ responses were averaged over both continuua and both talkers for each participant (because training condition did not interact with stimulus and talker gender), and subjected to t-tests with Bonferroni corrections (α/n). Plain [tw] trained participants showed only a marginal increase in /tu/ perceptual space, relative to the untrained group, edging into tsu and chew space (t(31)=1.9, p=0.066). Appalachian condition chw- trained participants showed an increase into both tsu and chew space, marginally greater than the plain [tw] group (t(39)=2.05, p=0.047), and significantly greater than the untrained group (t(35)=3.7, p<0.0125), and the chw- group(t(28)=3.4, p<0.0125), but on average, not significantly greater than the tsw- group.

The perceptual space is less flexible at the chew~two boundary because /tu/ competes with /tʃu/, so the largest adjustments occur to the two~tsu boundary, where /ts/ is not phonemic. For the male talker /tu/ is already allowed to be very retracted, while the female’s production is expected to be more fronted, even relative to the average productions of these two talkers, as noted in chapter 2. Therefore, it is interesting to note that the Appalachian condition participants have the most extended boundary, even where gender and phonetic contrast impose limitations.
Figure 4.11: Identification results for stimuli along two continuua produced by a male talker, from *tsu* to *two*, and from *two* to *chew*, as in chapter 2, but with the addition of the Appalachian retracted trained group, shown with square points.

Figure 4.12: Identification results for stimuli along two continuua produced by a female talker, from *tsu* to *two*, and from *two* to *chew*, as in chapter 2, but with the addition of the Appalachian retracted trained group, shown with square points.
4.3.5 Discussion of Perception Results

The results from the ID task in chapter 2 can be interpreted in light of the mechanism responsible for selective adaptation, perceptual adjustment, and perceptual broadening. The \textit{chw}- group experienced a shift of the \textit{two~chew} boundary, but had a similar \textit{two~tsu} boundary relative to the untrained group. The \textit{tsw}- group experienced a shift of the \textit{two~tsu} boundary, but had a similar \textit{two~chew} boundary relative to the plain [tw] trained group. These two results suggest a shift of one boundary, as expected for the early stages of perceptual broadening (while continued exposure could possibly lead to a shift of both boundaries). The plain [tw] trained group had expanded boundaries relative to the untrained group, suggesting that the lexical decision task acted as a training for the [tw] group on both \textit{tw}- variants, also affecting participants’ /tu/ category. That the \textit{tsw}- and \textit{chw}- groups showed assymetrical shifting of boundaries and acceptance rates during the lexical decision task, suggests that the initial training biased them somehow against accepting as many of the other variant, while the plain [tw] group showed an expanded distribution along both boundaries relative to the untrained group, and the Appalachian \textit{chw}- condition showed an even more expanded distribution. The perceptual broadening exhibited by the Appalachian condition suggests that their training didn’t inhibit adoption of other variants as it did with the \textit{chw}- group, but rather increased their ability to incorporate new variation into the relevant sound category, and to apply it to new conditions. This result suggests that generally increased category flexibility may be a marker of perceptual broadening.
4.3.6 Production Results

Chapter 2 demonstrated that perceptual learning may have some effect on production at least for a short time after shadowing. While the Appalachian condition participants exhibited a change in perception, it was unclear whether they would also exhibit any lasting changes in production because, not only did they exhibit perceptual broadening rather than a shift, because they may have held the belief that they were hearing talkers of a stigmatized dialect, they had a good reason to maintain a distinction based on dialect, and to avoid convergence during shadowing (as in e.g., Rubin 1992), but especially to avoid prolonged pronunciation changes that would align them with a stigmatized variety.

A subset of the Appalachian condition participants (4 male, 4 female) were randomly selected for acoustic analysis to find out whether they converged to the trainers’ chw-productions during shadowing and after the experiment. As in the results for chapter 2, the acoustic distance between participants’ productions and the training stimuli was calculated by normalizing each participant’s measurements for rise time, slope, and center of gravity against the chw-training condition tokens, as shown in equation (1). This step resulted in 3 z-scored values for each word, one each for rise time, slope, and center of gravity.

\[ (1) \quad z_j = \frac{x_j - \bar{x}}{sd} \]
Then a Euclidean distance function combined the three z-scores, as in equation (2).

\[
(2) \quad \sqrt{slope^2 + rise^2 + centroid^2}
\]

The resulting measurement represents a normalized measure of acoustic distance between participants’ and trainers’ /tw/ pronunciations. Larger numbers indicate greater distances between participants’ and trainers’ productions.

Figure 4.15 shows the Appalachian and standard chw- trained groups’ Euclidean distance from training stimuli before, during, and after training. The standard chw-trained group generally converged to a much stronger degree than the Appalachian group, with 10 out of 14 standard chw- participants, and only 4 out of 8 Appalachian participants exhibiting significant convergence (defined by t-test between baseline and shadowing at p<0.05) during shadowing. The standard chw- group also generally persisted in their convergence, even affricating words that were not heard during shadowing, while everyone in the Appalachian group returned to baseline productions after training and showed no significant difference from baseline in trained or untrained words. This pattern is consistent with the expectation that socially marked exemplars are processed differently, so as not to have a lasting influence on production.
Figure 4.13: Appalachian trained participants showed less convergence during shadowing and no persistence after training.

4.4 Conclusions and Discussion

The results from experiment 3 show an effect of bias in selecting /v/ words and non-words resulting from rapid exposure to a wide range of variation in /v/ stimuli. While the variation of /v/ tokens was within the normal parameters of speech, the recency of exposure gave more weight to the extreme variants than would normally be accorded, thus broadening the perceptual space of /v/. When /v/ words and non-words were the target stimuli, accuracy rates approached ceiling for both groups, and no processing
benefit was observed for the trained group. When /f/ and /b/ words and non-words were the target stimuli, the trained group’s accuracy rates and looks to the appropriate target fell, as they perceived more /v/ tokens than were presented to them. Untrained participants did not exhibit this kind of bias, and were more accurate and faster than the trained group at identifying /f/ and /b/ targets. Additional experiments using less variable stimuli may help to confirm that this effect is not due to simple phonological priming, but rather is an instance of perceptual broadening.

The results from experiment 4 show an effect of general broadening of the /tw/ and related /tu/ category to include multiple kinds of affrication for the Appalachian trained participants. While the standard tsw- and chw- groups showed evidence of a boundary shift consistent with perceptual learning, the Appalachian group extended their boundaries on both sides of the /t/ distribution to include more exemplars as instances of /tw/ and /tu/. The Appalachian trained participants had the greatest expectation of variation and therefore accepted more affricated variants into their /t/ representation for these talkers. Like their standard chw- trained counterparts in chapter 2, the Appalachian group generally showed convergence during shadowing, though not as strongly, but unlike the standard group, the Appalachian group on the whole did not persist in their pronunciation changes after the experiment. In order to facilitate understanding, this group greatly expanded their perceptual space and even altered their productions during exposure to the training stimuli, but demonstrated a socially motivated divergence of production and perception as the perceptual changes persisted up to the end of the experiment, but production quickly returned to normal before the final task. Not only
does this effect mirror what may occur during cross-dialect communication (e.g., Hay, Nolan, and Drager 2006), or cross-generational communication during sound-change-in-progress (e.g., Hay, Warren, and Drager 2006), but it also reveals the importance of encouraging participants in a laboratory sound-change experiment to think of the training stimuli as exemplars of their native variety.

Perceptual broadening appears to be a mechanism in which an increase in the variability of a speech sound, or an expectation of variability, results in an increase in the perceptual space for that sound. Perceptual broadening can easily be explained by probabilistic and exemplar models of speech perception, as the complement to selective adaptation. The effect of variation in training stimuli on the perceptual system of participants should be taken into account when designing experiments, for it appears that control groups that are trained using “normal” or “natural” stimuli may still adapt their perceptual systems to those stimuli, possibly altering what we think of as a “baseline” or “normal” measurement, thus affecting the interpretation of results that are compared to this control group. Some form of perceptual broadening could potentially be used to elicit and study perceptual mergers in the laboratory, and models of perceptual broadening have the potential to describe what happens to the phonological system of an individual who is exposed to a range of variation across categories, and to predict when their own categories will remain distinct, and at what point they may merge.
Chapter 5: Conclusions

The goal of this dissertation was to outline a general model of sound change in the individual, informed by empirical data, and based on general cognitive principles. The framework underlying this theory is probabilistic and usage-based, making use of exemplars or phonetically rich memory representations of sounds that include contextual details and talker specific information. The basic theory posits that sound change begins when the observed variation in a sound is no longer normally distributed around the stored central values for that sound. That is, when the nature of the normal phonetic variation leans to one direction or another, the sound category may shift to reflect the changing nature of the incoming sounds (as in, e.g., Hockett 1965). Experiments that elicit changes in the perceptual mapping of the phonological system, such as perceptual learning (e.g., Norris, McQueen, and Cutler 2003), selective adaptation (e.g., Eimas and Corbit 1973), or perceptual broadening (e.g., Keating, Mikos, & Ganong 1981), provide evidence that listeners’ phonological systems are sensitive to changes in the distribution of sounds in a language. Experiments that elicit changes in production based on phonetic characteristics of an interlocutor’s speech, such as in chapter 2, provide evidence that differences in the distribution of variation can also affect speech production. These findings provide evidence in support of Hockett’s (1965) hypothesis. As an additional
(perhaps obvious) point of clarification, I would like to mention that the interaction of language users is a necessary component, in that without the input from many other members of the speech community, there would be no distribution of sounds and no variation (and no language, when you get down to it). Because of this multiplicity of input, however, each individual in a community will have a unique phonological distribution.

Greater nuance is needed in a model, to get to the roots of the “actuation problem” (Weinreich, Labov, and Herzog 1968), that is, if shifting variation is the source, what is the cause of the shifting, and what turns variation into sound change? Additionally, the following findings complicate the picture: 1) While convergence and perceptual changes were somewhat linked in chapter 2, not all participants exhibited convergence or perceptual learning, and we need a theory that also accounts for the lack of participation in a sound change. 2) Other researchers have found asymmetries between perception and production changes (e.g., Kraljic, Brennan and Samuel 2008). 3) Production changes did not persist after shadowing a talker of a stigmatized dialect in chapter 4, and convergence has been found to be variable depending on one’s feelings toward one’s interlocutor (e.g. Rubin 1992, Babel 2010). 4) Chapter 3 illustrated that listener and talker gender can also affect changes in perception. These findings suggest that shifting variation is not enough to determine sound change, even if a cause for the shifting can be apprehended.

The second premise of the general model was that a pronunciation variant could become associated with any number of co-occurring factors, phonetic or not. The above
problems with individual differences in perception and production can be partially explained by this second premise. The differences in patterns of association in the conditioned splits in chapter 3 suggest that multiple possible associations may sometimes contribute to variation rather than resolving it, as occurred for the traditional trained females, because at first, the category may broaden before splitting; although, the men in the traditional category successfully made the association between talker gender and /tw/ variant. The fact that participants showed different degrees of adoption of the /tw/ variants across different words showed that association may be possible across single lexical items, but the fact that most participants in chapter 2 generalized pronunciation and perception changes to new words not heard during training, suggests that sound changes may begin in a few lexical items and spread to other words by way of abstraction to the /tw/ sound category level. The fact that some changes in perception also spread to a related category, /tu/, suggests an even higher level of abstraction. It is unknown whether concomitant changes to /t/ in other environments occurred, at an even higher level of abstraction, but that may be explored in future research.

While modern linguists agree that sound change begins with variation in sound, this variation exists in every language at every time, and is not an explanation of sound change, as framed by Weinreich, Labov, and Herzog (1968) as the “actuation problem” and was complained about at least as early as Schuchardt (1885). A truly explanatory account of sound change must explain what caused the transition from variation to sound change. For explanatory factors, we must look deeper, into the minds of language users, to find out what the variation means to them. Variation can have social meaning,
grammatical meaning, lexical meaning, pragmatic meaning… Variation can be assigned any meaning that gives it purpose or helps a language user make sense of it. Whatever co-occurring factor, or combination of factors, is most salient to a language user will make the most sense of the variation, and will guide the direction of change for that individual. When multiple language users in a community come to the same association, then the change is likely to take hold.

Which brings us to the third principle of my general model of sound change, which is that all sound change (and truly all language change) is cognitively based. The differences between males’ and females’ responses to variation, and the differences between the Appalachian and standard trained groups also suggest that there is a non-automatic cognitive component to perception and production changes, at least for non-categorical variation. Of course, even mere coarticulation requires planning of the motor gestures required for consecutive speech sounds, and requires a delicate balance between linguistic and social information. Learning to associate a pronunciation variant with a co-occurring factor, such as a phonetic or phonological environment, utilizes the same cognitive processes as those required to form associations with certain words, a talker characteristic, prosodic position, or grammatical category. This power of association is the driving force behind sound change. Coarticulation or other phonetic variation does not develop into a sound change unless some association is retained after the original conditioning factors become irrelevant.
Experiment 1 demonstrated that early sound change in the individual could be recreated in the laboratory, that by exposure to aggressive variation in a sound, participants could be made to change their perception and production of that sound. The perception changes could be extended to new talkers, and production changes could be extended to new words and even different, but related conditioning environments. Phonetically conditioned sound change not only affected sounds in specific conditioning environments, but could also spread to related environments, by the extension of phonologically analogical changes from /tw/ to /tu/. The possibility that phonological analogy played a part in phonetically conditioned sound change was also suggested in that convergence to and perceptual adaptation to chw- was more robust than convergence to and perceptual adaptation to tsw-, as the former variant exhibits a stronger analogical tie to the outcomes of /tr/ and /tj/; although, the social effect of gender may have been partly responsible for some men’s reluctance to accept the front variant.

The second experiment duplicated the results from the first experiment with some tweaking to the design and stimuli. Results from the second experiment also demonstrated that differences in sound change may be subject to social factors, as in the difference between male and female productions, and analogical considerations, as in the parallel development of /tr/ and /tj/. The most interesting finding, in my opinion, in the second chapter was how participants in the split training condition had the strongest effect of training for the stimuli produced by women in their training condition. The traditional group showed high endorsement rates especially for tsw-, relative to the other groups, and the non-traditional group had high endorsement rates for chw-, and less for
A question that deserves follow-up is whether this effect was in any way related to the claim that women tend to lead in gradual sound change (Labov 1990). If listeners held the belief that women used innovative variants more often (and these participants were linguistics students), might they pay more attention to the forms used by women? Additionally, while the participants were past the generally accepted critical period of acquisition, they were still quite young, and it would be interesting to see what generalizations may be made by younger and older speakers. Labov (2007) hypothesizes that “structural patterns are not as likely to be diffused because adults do not learn and reproduce linguistic forms, rules and constraints with the accuracy and speed that children display” (352). So, it is possible that these participants did not form the associations that were intended for them to make, if it is truly more difficult for adults to learn new linguistic patterns. However, this is an empirical question that remains to be answered. Whether the transmission (or incrementation) of a sound change only occurs in early language acquisition, while the spread from adult to adult is an entirely different mechanism, or whether there is some continuum between the two endpoints of the same process of organizing and re-organizing the phonological system, is a question that has yet to be resolved.

The third chapter laid out a theoretical explanation for perceptual adjustment and selective adaptation, and provided evidence for a third kind of phonological change, perceptual broadening, incited by either a change in input patterns or expectation of the listener. These related mechanisms of retuning the phonological system, by altering the mapping of sound categories to certain acoustic-perceptual dimensions, can be compared
to the reorganization that might occur during a language user’s lifetime during a sound change in progress. That production changes were not lasting in the Appalachian condition, or among men in the /sw/ condition, suggests that social factors (e.g., in-group, dialect, gender) may impact the adoption of sound change during the lifetime at least as much as age.

If this thesis has successfully argued that non-phonetic associations may be made at the same time as phonetic associations, then we must reanalyze how we apply the comparative method. Traditionally, the assumption was that a sound change had to follow through to completion before analogical changes could take place, following the Neogrammarian principle that sound change is blind to cognitive factors. However, if sound change is not blind to cognitive factors, we can expect to see much more complex interactions of factors conditioning sound change. This does not free us from the requirement of hunting down an explanation for seeming irregularities, but rather requires us to look at the whole language as an inter-connected system, in which social factors, pragmatic usage, frequency and familiarity of words and sound combinations, syntactic and prosodic structures all interact in the mind of the talker and listener. As some of the best linguists in the world have asserted, in order to be a good historical linguist, you have to be a good linguist.

Why should we assume that all sound change should be regular, when most of the time this is not the case? Either way, the historical linguist looks for regularity to be able to form correspondence sets, and scrutinizes every instance to be sure it is not a
borrowing or analogical form. And what do we do if we find a correspondence in some words, but not others? We must postulate that that set of words must have formed a class of some sort, and look to find out why they patterned together, and at the same time, look for a reason why the others did not pattern that way. Coming at it from the Neogrammarian principle, an explanation of why the corresponding forms pattern together is unnecessary beyond the statement of some conditioning environment. If it is simply a ‘given’ that sound change is regular, then there is nothing more to explain. There was some sound change X –> Y, and all of these forms of class Z were affected by it, but an explanation must be found for each exceptional form, or class of exceptions, if that is possible. However, if we take the position that regularity in sound change is also exceptional, an accident of history, then not only must we look for the reasons for the exceptional forms (why they didn’t change, or why they underwent different changes), but also why the regular sound change took place, and what must have happened to allow regularity in the outcome.

For example, the outcome of Grimm’s Law in Proto-Germanic was considered to be very, but not completely, regular, until Verner’s Law was discovered. Much fuss was made about Verner’s Law, including investigation into not only the historical conditioning environment, and the probable state of the language at that time, but also the articulatory factors that were likely contributors to voicing after an unstressed syllable, and so on. But much less thought went into why Grimm’s Law applied in the first place, and why it was so regular (and if this had anything to do with the 2nd Lautverschiebung or similar affrication processes in English). I’m not saying that this line of inquiry hasn’t
been followed, but it has enjoyed limited success. Though, note that some researchers have recently made very good claims for a reversal of Grimm’s and Verner’s Laws (Euler & Badenheuer 2009; Kroonen 2011).

Thus, by insisting that regularity is not necessarily the expected outcome of any sound change, phonetically-conditioned or otherwise, it forces us to explain not only irregularities, but also regularities. I, for one, am interested in just what combination of factors it really does take in order for (large-scale) regularity to ensue from a (small-scale) sound change. And it seems to me that until we understand how sound change operates in the individual and in the community, we will not be able to explain how we get wide-spread Neogrammarian sound change. One day, we may even be able to form predictions, similar to meteorological predictions, in which the current state of variation in the language combined with its past currents may predict whether or not a sound change may take place, and what direction it may take. To be certain, the day is far off when meteorologists are considered to be more often right than wrong, and the prediction of language change is probably further off still. But the goal of this dissertation is to pave the way for that day.
Partly cloudy with a chance of sound change

Figure 5.1: Partly cloudy with a chance of sound change. One day we may be able to predict sound change as well as we predict the weather.


Lyell, Charles. (1830). The principles of geology: *being an attempt to explain the former changes of the Earth's surface, by reference to causes now in operation*. Volume 1. London: John Murray.


Norris, Dennis, James McQueen, & Anne Cutler. (2003). Perceptual learning in speech. Cognitive Psychology, 47: 204-238.


