7-2012

Perception of Non-Phonological Reduction: A Case Study for Using Experimental Data to Investigate Rule-Based Phonology and Exemplar Theory

Rachael Tatman
College of William and Mary

Perception of Non-Phonological Reduction:
A case study for using experimental data to investigate
Rule-based Phonology and Exemplar Theory

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Art in Linguistics from
The College of William and Mary

by

Rachael Tatman

Accepted for Honors
(Honors, High Honors, Highest Honors)

Erin G. Ament, Director

Christopher A. Ball

Linda A. Lanz

Williamsburg, VA
April 24, 2012

Perception of Non-Phonological Reduction: A Case Study for Using Experimental Data to Investigate Rule-Based Phonology and Exemplar Theory

Rachael Tatman

May 9, 2012

Contents

Chapter 1

Introduction and Background

Phonetics and phonology have had a strained relationship since Trubetzkoy (as cited in (Kohler, 1991)) separated the two, suggesting that phonology is the study of sounds within the framework of the humanities while phonetics is the same study using techniques of the empirical sciences. Recently, criticism of that division, based on evidence from reduction and assimilation, has become increasingly common i.e. (Cohn, 1993; Flemming, 2001; Kohler, 1991, 2000; Mielke et al., 2003), especially within psycholinguistics i.e. (Mitterer, Yoneyama, & Ernestus, 2008; Gaskell, Hare, & Marslen-Wilson, 1995; Fitzpatrick & Wheeldon, 2000), and particularly as new explanations for phenomena are proposed that seem to work equally well within either framework (Cohn, 1993; Flemming, 2001). Phonetics research has generally been quicker to adopt and develop new tools—be it actual equipment, experimental methodology or statistical tools. The artificial divide between phonetics and phonology has hindered phonologists from adopting these tools. While there is certainly a trend against this (consider the formation of the Association for Laboratory Phonology in 2010), it has by no means been universally accepted as the future of the discipline, or even as desirable at all[1]. Regardless, there is certainly deserved support for the use of experimental techniques in the examination of sound and sound systems.

Another recent trend in the way we study sound systems is a focus on perception. Lindblom suggests the reason that linguistic research has historically focused on production was the technical difficulty of gathering reliable perceptual data about language (Hume & Johnson, 2001). The equipment needed for perceptual measurements was bulky, expensive, fragile and difficult to transport. Recent advances in technology, however, mean that even undergraduates can collect and analyze production data. (Consider the ease in collecting reaction time data on a computer instead of using a stereo, piece of paper, pencil and stop-watch.) This new emphasis on perceptual investigations is an important step forwards since there is evidence that input and output phonology are independent of each other (Martin, Lesch, & Bartha, 1999; Fitzpatrick & Wheeldon, 2000). Therefore, studying only articulation would have left our knowledge of phonology crucially incomplete.

Finally, there is yet another strong trend towards the study of connected speech which, after all, forms the vast majority of all speech events. By studying speech exclusively as it occurs or is heard in the laboratory (usually the citation forms of words) linguists "may be missing something important" (Johnson, 2004). Studying connected speech in controlled settings requires new experimental techniques and there has already been quite a bit of work on the problem (Cohn, 2011).

With these three points in mind–the necessity of using empirical and percep-

[1]Consider, for example, the recent and very public debate between Pullum and Brenchley, where one of the main points of contention was the place of empirical evidence in the study of poverty of the stimulus and linguistics as a whole(Pullum, 2011; Brenchley, 2011). We can think of this as an argument between those who favor mathematical models and those who wish to bring an empirical, evidence-based viewpoint into the discussion. Both are valuable and there is certainly no reason why we, as linguists, should not be able to reconcile both methodologies–consider the relationship between theoretical and experimental physics as a model. However, there is currently more resistance to empirical methods, so they require more support to ensure their incorporation.

tual data and of considering conversational speech–this study brings together these methodological philosophies to examine the perception of a phenomenon which has traditionally been described one way in phonetics (using phonetic techniques) and another in phonology. In particular, it seeks to see which current school of phonological theory–rule-based or usage-based–makes predictions which are borne out by empirical evidence.

1.1 Rule-Based Phonology

I am using rule-based phonology here to refer to a variety of formal phonological models: generative phonology (Chomsky & Halle, 1968), autosegmental phonology (Goldsmith, 1979) and Optimality Theory (Prince & Smolensky, 1993). Though each model uses different techniques, they all make the same underlying supposition: that individual sounds (such as the flap) are irrelevant and it is the rules that make the grammar (Hale & Reiss, 2000). Further, these rules should ideally apply across not only one language, but (particularly in the case of Optimality Theory constraints) across all languages.

Rule-based phonology certainly has its uses. It was the earlier model and seeks particularly to model those parts of language use that are innate to humans and therefore universal. The flapping rule discussed below is itself one of the products of rule-based phonology. Where this conceptual model is weak is when rules does not appear to accurately describe language phenomena that nonetheless occur. While Chomsky argued for a model which distinguished between competency and performance and modeled only the former (Chomsky, 1965) the fact remains that rule-based phonology cannot accurately describe all language use.

1.2 Exemplar Theory

Exemplar Theory is the theory that I am using to represent usage-based models. According to Exemplar Theory, listeners keep track of all the linguistic data they are presented with and that information affects their judgments as a tool to deal with variance in speech signals (Johnson, 1997). The more listeners hear a string of sounds, the more likely they are to accept it in both production and perception. This results in self-reenforcing clusters of data points that produce the same speech production patterns as rule-based methods, and deal equally well with the relatively small amount of information children are exposed to(Pierrehumbert, 2003).

While it may seem unfair to group all rule-based models together and only address one usage-based model (I am not looking at cognitive or construction grammar, for instance), Exemplar Theory is uniquely well-suited to deal with non-phonological reduction (see 1.3). In addition, it deals specifically with the mechanisms of perception; most rule-based models are more concerned with production.

1.3 Coronal Stop to Flap Reduction in American English

There is a well-known process of intervocalic alveolar flapping in English (Giegerich, 1992, p. 226), classic examples of which include the homophones of "rider" and "writer" and "atom" and "Adam". (A flap is characterized by a quick motion of the tongue which is similar to the creation of a full stop, but with a much shorter duration.) The most-cited form of the rule is that oral alveolar stops (/t/ and /d/) are reduced to a flap in American English when they occur after a stress vowel and before an unstressed vowel, potentially even when /ɹ/, /d/ or

/n/ occurs after the stressed vowel but before the sound itself (De Jong, 1998)[2]. For example, this process reduces "atom" and "Adam" to a homophonic pair, where both alveolar stops are reduced to flaps. This type of flapping, which happens predictably and in a constrained environment, is what I will refer to as "phonological reduction" from here on.

However, in connected speech coronal stops surface as flaps or approximants a large percentage of the time and in an unpredictable manner (Aguilar, Blecua, Machuca, & Mann, 1993; Warner & Tucker, 2010; Raymond, Dautricourt, & Hume, 2006). So listeners are exposed to a large number of instances of situations where coronal stops do not surface as their phonologically predicted forms. This is what I refer to as "non-phonological reduction", which makes a particularly good test subject for a study. This is in contrast to findings by Warner and Tucker (2010). Their data was collected by having participants call friends on their cellular telephones while in a sound booth and recording the participants' conversations, as well as by using the traditional isolated word lists and a story-reading task. They found that in spontaneous speech, less often in the story-reading task, and very rarely in the word-list task, stop-to-flap and even stop-to-approximant reduction was widely found, but not in a predictable way. As they put it, "Phoneme identity, speech style, word frequency, and perhaps speaker characteristics determine how strongly consonants are reduced, but stress does not." In other words, they discovered a process of reduction that standard phonology cannot account for, and yet which has a high frequency. It is this process that I refer to as "non-phonological" reduction.

To recap, I am using reduction of coronal stops to look at listener's judgments of non-phonological reduction. Non-phonological reduction differs from phonological reduction in several key ways which are summed up in the table

[2]For the purposes of this paper, only stops which were intervocalic were used as examples of flapping.

Table 1: Differences Between Phonological and Non-Phonological Reduction

(Rule-based) Phonological Reduction	Non-phonological Reduction
Occurs predictably	Unpredictable
x -> y	x -> y, z, p, q, r, etc.
Not affected by speech style or rate	Affected by speech style/rate
Occurs in one environment	Occurs in multiple environments.
Rules-based	Cannot be described by rules, statistically describable

below.

The table shows productive characteristics of both phonological and non-phonological reduction. Phonological reduction is the only type of reduction that can be well accounted for using rule-based phonology models; a stop surfacing in the same environment as either a full stop, flap or approximant is extremely difficult to model if your model consists of a single rule. On the other hand, if each instance occurs a predictable percentage of the time, that can be modeled very easily using Exemplar Theory. Rule-based phonological models are intended to model a speaker's knowledge of language, rather than their performance (Chomsky & Halle, 1968). If this is so, then we would expect that listeners, using their linguistic competence as a guide, would always reject an utterance that didn't follow their internal rules. While Exemplar Theory predicts much the same result as rule-based phonology in most instances, reduction is a special case. Because listeners encounter non-phonological reduction so often in their speech environments, they have a high number of exemplars that do not conform to rule-based phonology. If Exemplar Theory is an accurate model, then this should be borne out empirically.

Chapter 2

Experiment 1

This experiment was designed to determine whether listeners found the presence of non-phonologically predicted word-medial stops or flaps troubling enough to affect their goodness judgments of the word. When presented with a word which did not meet their phonologically-conditioned expectations, how do listeners react? If their internal sound systems are in line with what is predicted by Exemplar Theory and phonetic production data, then they should not find a mismatch troubling. If, instead, they are judging utterances against their competencies rather than their performance, as discussed above, then any mismatch between their phonological-expectations should negatively affect their goodness judgment of words. To test this, I looked at participants reactions to tokens in which reduction should and should not occur, according to the flapping rule outlined in section 1.3 above, with and without reduction.

2.1 Methodology

2.1.1 Items

Item Selection To balance the words for frequency, I used frequency data from the English Lexicon Project (Balota et al., 2007). All of my tokens were drawn from items that occurred either 49 or 50 times in the corpus. This balanced the need for a large number of items which fulfilled my phonological criteria with a desire to use words that would hopefully be somewhat familiar to undergraduate college students[1]. From there, the list of possible tokens was further reduced to only include three syllable words. Stress was not controlled for at this point. Next, all words with non-coronal word-medial stops were selected for use as filler items. Words with only one word-medial coronal stop were divided into three categories: 1) those which should undergo stop-to-flap reduction, words where the coronal stop occurred directly after the vowel which bears the primary stress of the word; 2) those which should not undergo stop-to-flap reduction, where the coronal stop occurred intervocalically but not after the vowel which carries primary stress or where the coronal stops occurred after a fricative; and 3) the rest, which were discarded. For examples, see list below and see Appendices 1 through 3 for the complete item list.

Filler: ovenproof - ˈəvənpɹuf

Reduction expected: inedible - ɪnˈɛ[d/ɾ]ɪbəl

Reduction not expected: redirect - ɹidaɪˈɹɛkt

This ensured that effects such a nasalization or lateralization would not interfere with perception of reduction, as coronal stops which occur directly after nasals or laterals are articulatory distinct from those that do not (Zue

[1]A verbal examination of participants after the experiment revealed that they were familiar with most words. Only "Rococo" and "rapacious" were mentioned as troublesome, and both tokens were filler. Future experimental work might be improved by using frequencies based on a conversational corpus, such as the Buckeye corpus (Pitt et al., 2007), to ensure that participants are not distracted by unfamiliar items.

& Laferriere, 1979). This yielded a total of ten each of the reduced and non-reduced groups. A list of all tokens is included in the appendix.

Item Recording These tokens were then recorded in a sound booth by the researcher, a native speaker of American English, on an AT2020 cardioid condenser microphone using a pop filter. The tokens were recorded, without reduction, using Audacity (A. Team, 2011) as a single sound file that was later segmented into individual word files and then renamed by hand in Praat (Boersma, 2001). Each word was then edited to remove pops and other extraneous noise as well as silence. To create reduced tokens, the first word-medial stop (in the filler words) as well as the word-medial coronal stop in the target words was cut so that approximately half of the stop was removed, as well as the release burst. For the coronal stops, this meant that the sound closely resembled that of naturally-produced flaps, as described by(Zue & Laferriere, 1979). Using the same technique on the non-coronals ensured that participants wouldn't be able to single out the non-filler words. All cuts were made at the zero crossing, with every effort made not to create too sharp of a transition in the wave form. The full and clipped versions of "Broadening" can be seen in Figure 2.1 below.

Using the same token and then manipulating it to mimic reduction ensured that no outside factors such as changes in pitch or tempo would affect listener's judgments. The tokens were then played for a number of linguistically-untrained listeners, who were unable to detect any electronic manipulation in the sounds. So there were six groups of tokens, four combination of factors and the two levels of filler words:

- Reduction present, reduction expected = "derided" with a flap; dɪˈɹaɪɾəd
- Reduction present, reduction unexpected = "redirect" with a flap; ɹiɾaɪˈɹɛkt
- Reduction not present, reduction expected = "derided" without a flap;

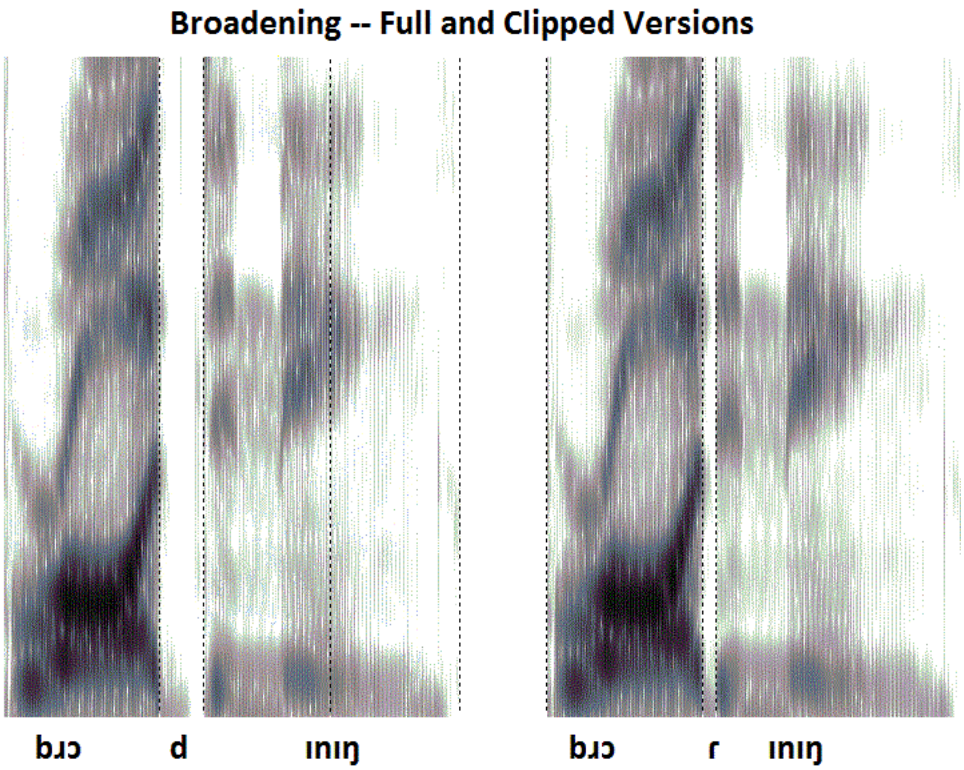

Figure 2.1: Broadening – Full and Clipped Versions

dɪˈɹaɪdəd

- Reduction not present, reduction unexpected = "redirect" without a flap; ɹidaɪˈɹɛkt
- Filler words without reduction = "rapacious"; ɹəˈpeɪʃəs
- Filler words with reduction = "rapacious" with the /p/ shortened ; ɹəˈpeɪʃəs

At this point, all sound files were converted to .mp3 files, as that was the only sound format accepted by the presentation software. Presentation order was created by first completely randomizing the tokens using an on-line list randomization service (Haahr, 2011) and then rearranging the order by hand so that no target (non-filler) words occurred in the first ten items, no target words occurred within four items of each other and no target words occurred next to each other.

2.1.2 Procedure

Participants were tested *en masse* in a classroom setting, so the sounds files were presented as an electronic slide presentation and responses were collected using scantron forms. To present the sound files, Sliderocket was used (S. Team, 2011). The first few slides were instructions, which were also repeated verbally. Participants were informed of their rights, and then told that they would be listening to words and should rate each one on a five point scale, where a five meant that the listener judged that the item "sound[ed] like it was correctly produced by a native speaker of American English" and a one indicated that it did not. Each slide, after the instructions were presented, was white with the item number written in black in the center of the screen, and with a reminder of the scale at the bottom to assist the participants. The target sound played followed by three seconds of silence and then automatically advanced to the next slide. Every ten items, when participants reached the bottom of a column of bubbles, there was a five second break followed by a tone. Participants were exposed to each token with a word-medial coronal stop twice, once with the stop reduced to a flap and once with it fully articulated. Exit questions indicated that this pace was not too fast for participants and that the task was relatively easy.

2.1.2.1 Participants

Participants were twenty students in an introductory linguistics class. All were monolingual native speakers of American English without speech or hearing impediments and over 18 years old. (Four speakers' data sets were discarded because they were not monolingual native speakers of American English without speech or hearing disorders so that, though twenty-four informed participants took the study, only twenty data sets were considered). Those who chose to attend the experiment, whether or not they participated, were given extra credit for their time (a sign-in sheet was provided at the beginning of the session and all those whose names were on it were awarded the extra credit). While no participants deduced the purpose of the experiment, several did mention stops specifically and they were already familiar with the process of flapping in American English. In addition, several indicated that they were not familiar with all of the words in the token set, and that some of the tokens did not sound like real words of English (this was unsurprising as many of the word-medial stops in the filler items were clipped enough to be perceived as different stops). For example, "rapacious" sounded more like "rabacious" due to the reduced length of the closure.

Though data was collected anonymously, each participant was instructed to indicate their gender, native language, monolingual status and whether they had any speech or hearing disorders, as well as to create a unique six digit identification number so that data could be tracked by participant.

2.2 Results

Data was analyzed in SPSS (Norusis & Soci, 1994). Next, the average rating for each item was computed, and that data was then used to run a by-subjects two-way analysis of variance (ANOVA), with expected reduction (reduced or unreduced) and observed reduction (reduced or unreduced) as the factors. There was no main effect of reduction $F(1, 19) = .050$, $p = .825$ and no main effect of expected reduction $F(1, 19) = .219$, $p = .643$. However, there was a highly significant interaction $F(3, 19) = 5.588$, $p = .024$. See Figure 2.2 for a graph of the interaction. When reduction was expected, reduced tokens were judged as better than unreduced ones. When reduction was not expected, unreduced tokens were judged as better than reduced ones.

2.3 Discussion

The results indicate that participants are capable of detecting a mismatch between phonologically predicted and perceived stops. This suggests that phonological rules affect listeners' goodness judgments more than their experience of non-phonological reduction in connected speech. In other words, competence-based phonological rules appear to be perceptually real for listeners. This is in keeping with rule-based phonological models, but somewhat troubling for Exemplar Theory.

Though this interaction is very strong, there is a potential room for error. Since the participants were linguistics students who were familiar with the process of flapping in American English, it is possible that they may have been more sensitive to this sound change and reacted accordingly. In addition, many

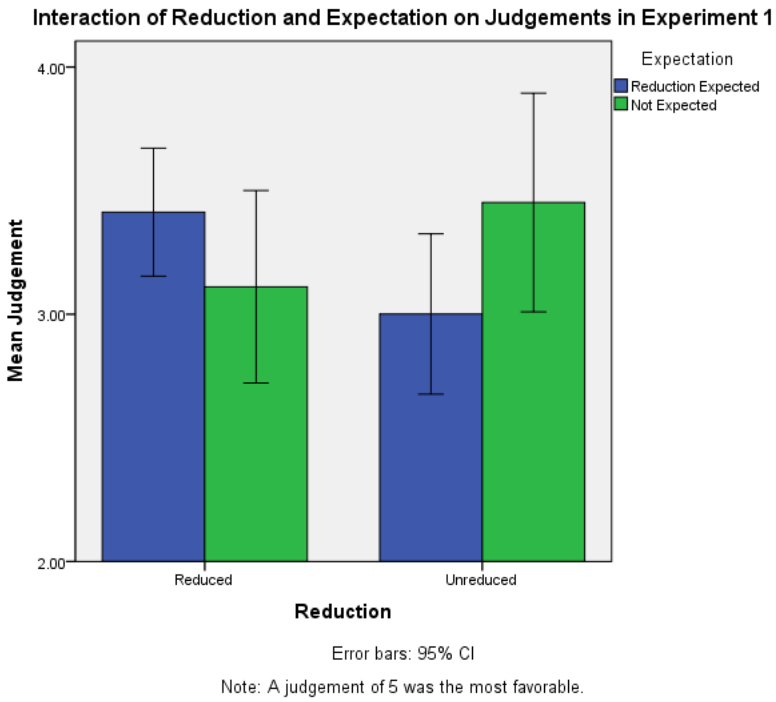

Figure 2.2: Graph of interaction between expected and observed reduction in Experiment 1

participants did not the use full scale, so it was difficult to determine whether they were unsure about their answers or really thought that many of the tokens were marginally acceptable. As you can see in Figure 2.2, even stops occurring where phonologically-predicted flaps should be were still, on average rated no lower than a three on a five-point scale. In order to ensure that these potential methodological problems were not significantly skewing the results, a second experiment was performed.

Chapter 3

Experiment 2

The second experiment was extremely similar to the first, however it was modified somewhat to improve the methodology and ensure that the results of the first experiment were repeatable. The main difference was that the second group of participants were linguistically untrained and were asked to rate the tokens on a two point scale instead of a five point scale. By forcing participants to make a decision one way or the other, I hoped to get at any biases that they might have been withholding due to social stigmas attached to negative responses.

3.1 Methodology

3.1.1 Items and Procedure

The methodology for the second experiment was precisely the same as for the first experiment, using the same items and presentation, except for two things: points two, three and four were removed from the scale (so that a participant could only answer 1 or 5, essentially on a two-point scale) and the judgments were flipped. In the second experiment a judgment of 5 was the least favorable.

This was an attempt to balance any effect of assigning high or low numbers to either judgment. The data collection procedure was also the same as in the first experiment.

3.1.2 Participants

Participants were thirty-three members of a science fiction and fantasy club who arrived before a scheduled meeting in order to participate. Thirty-five participated, but two were not native monolingual speakers of English, resulting in thirty three total participants. Though there was no reward for completing the experiment mentioned beforehand, the author brought cookies to the next meeting. Most subjects were students at the College of William and Mary, and all were between 18 and 28 years old. Again, verbal exit questions revealed that participants did not find the task too difficult, as well as many of the same reactions as the linguistics class. One or two mentioned "the 't' sound" or "the 'd' sound" as being a little strange, but none were familiar with the phonological category of a stop.

3.2 Results

The data were analyzed using the percentage of participants who judged each utterance favorably. Thus, a category might have had a possible maximum acceptance rate of 100% (if all participants judged all items as acceptable) and a minimum acceptance rate of 0% (if all participants judged all items as unacceptable).

A two-way ANOVA was preformed with reduction (reduced vs. not reduced) and expectation (reduction expected vs. reduction not expected) as the factors. There was no main effect of expected reduction $F(1, 32)=2.376$, $p = .132$ and no main effect of actual reduction $F(1, 32)=1.902$, $P = .176$. Once again, there was

Figure 3.1: Graph of interaction between expected and observed reduction in Experiment 2

a strong interaction: $F(3, 32)=0.503$, $p = .004$ (see Figure 3.1). Again, where reduction was expected, reduced utterances were judged as better than unreduced utterances and where reduction was not expected, unreduced utterances were judged as better than reduced utterances.

3.3 Discussion

The interaction in this experiment was precisely the same as it was in the first. Participants found a mismatch between their phonological expectations and the phonetic reality troubling enough that it effected their judgment of the tokens. Because of this we can conclude that this is a robust effect which was not unintentionally created due to any of the factors included above. Again, this suggests that phonology is on some level perceptually real for listeners, whether they realize that it is or not.

3.4 Discussion of Experiments 1 and 2

The results from these experiments are both unsurprising and troubling. On the one hand, they serve as very strong empirical support for the perception of phonology and, specifically, the ability of listeners to determine whether or not an utterance conforms to their phonological expectations. (On the other hand, the fact that listeners *are* capable of determining whether or not a flap should be present is a little unsettling, since they encounter mismatches between their expectations and reality very often in connected speech.) The above experiments suggest very strongly that their goodness judgment of a word becomes less favorable when they encounter a mismatch. Do, therefore, they see all mismatches between phonologically-predicted reduction and observed reduction as speech errors?

The data suggest that this is not the case; that listeners do not consider this type of mismatch an actual error. Note that, in both instances, there was a strong tendency to say that the target utterances were correct. The average judgment for the first and second experiments, respectively, were 3.2 on a five point scale and 62% acceptable, which were both higher than the "null" response of 3 or 50% approval. So in each case, participants were far more likely to judge any given utterance as correct, even if it did deviate from their phonological expectation. This seems to suggest that Exemplar Theory cannot be dismissed; we would expect that listeners are more likely to accept things that they have heard before, and as native English speakers they have heard a large amount of non-phonological reduction. However, the fact that there was an interaction between expectations and reality suggests that listeners do not simply accept all utterances; there is an awareness of “correct” and “incorrect”, which is more consistent with rule-based phonology. This still leaves us with the problem, however, of whether speakers judge these mismatches as “incorrect” in all situations, particularly in connected speech.

It is possible that listener’s judgments were so consistent because the tokens were all produced slowly and without reduction. They are clearly the citation form of the words, and as the citation forms they are presumably expected to conform to all phonological rules–even in an Exemplar Theory model. This may not be the case in connected speech, however. Corpus analysis focusing on “ing” variation and reduction suggests that speech style is extremely important with regard to reduction, with reduction increasing as formality decreases (Abramowicz, 2007). Very informal speech contains the highest amount of reduction, while formal speech (such as the citation forms of words) contains the least. There is evidence that stop reduction works along similar lines (Warner & Tucker, 2010). In order to determine whether listeners find a mismatch between their phonology and the phonetic production of a word unsettling in a connected speech setting, it would be optimal to embed the target tokens in a conversation. There are inherent difficulties in attempting to access phonetic judgments of individual words, let alone individual phonemes, in a conversational setting (Cohn, Fougeron, & Huffman, 2011)but a possible workaround is discussed in the section on future work (See Chapter 5).

So we can conclude that listeners are capable of detecting a mismatch between their phonological expectations when words are produced in their citation forms. The question remains, though, whether they are capable of making that same distinction in connected speech? It seems likely that they may have more difficulty perceiving the mismatch due to the fact that they encounter it much more often; if Exemplar Theory is correct *and* their exemplars take speech style into account.

As mentioned above, the optimal situation for obtaining listeners' judgments of words embedded in a conversational setting is difficult at best. There is, however, a potential work-around. Conversational speech has two main attributes: speed of production and degree of reduction (Dalby & Club, 1986). Since reduction is the phenomenon currently being studied, speech rate was manipulated to approximate connected speech. Embedding the target words in a frame sentence, also produced at the fast rate, would also yield more "conversation-like" items.

Chapter 4

Experiment 3

This experiment attempted to determine to what degree increasing the speed of speech affected participants' judgment. There are three possible outcomes for this experiment.

The first is that changing the speed of the utterance would have no effect whatsoever on the judgments of the participants, as rule-based phonology would predict.

The other, which would be in keeping the Exemplar Theory, is that the rate of production would have an effect on participants' judgments. Exemplar Theory posits that listeners remember everything about an utterance, including its degree of reduction and the speed at which it was produced. Since production data indicates that natural speech is produced more quickly and has a greater rate of reduction (Dalby & Club, 1986; Kohler, 2000; Warner & Tucker, 2010), listeners tend to hear these two traits together more often. Conversely, they also tend not to hear reduction in very slow speech. So fast utterances tend to be reduced and slow utterances tend to be fully articulated. The greater density of data points in these regions means that listeners are more likely

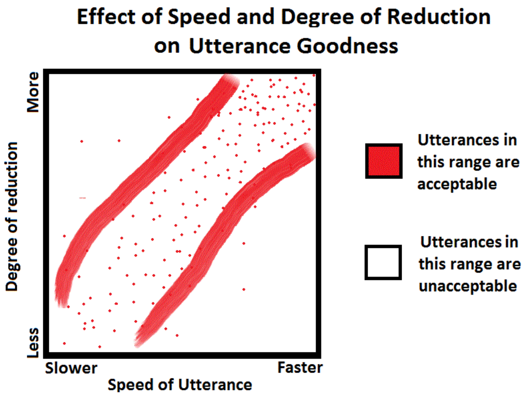

Figure 4.1: Effects of Speed and Degree of Reduction on Utterance Goodness

to accept utterances that fall within them. If this is so, then listeners must also be more likely to reject utterances that fall outside of this region; slow utterances with reduction and fast utterances without it. See Figure 4.1 for a visual representation of this. The lines represent the limits of what a listener would consider acceptable. Slow but reduced utterances would fall in the upper left hand corner and therefore be rejected. Increasing the speed of the tokens moves them to the upper right hand corner, which is within the acceptable limits of this individual's experience.

4.1 Methodology

4.1.1 Items and procedure

The full and reduced items from the previous two experiments were used again in this experiment, as were copies of those items with their duration decreased by approximately 50% to simulate the faster speaking rate found in natural conversation. (Osser & Peng, 1964) found that the average monologue production rate of speakers of American English was 595.7 phonemes per minute, or approximately 10 phonemes per second. The fully-produced tokens were all approximately a second in length and between five (Apache) and nine (adornment) phonemes. A rate twice as fast as their initial rate, approximately 5 phonemes per second, was chosen to mimic the conversational rate. They were individually manipulated in Praat (Boersma, 2001) in order to ensure that the manipulation was not aurally apparent. The data were presented using a Praat multiple forced choice experiment, with participants using a mouse to respond. Participants were asked to rank each utterance on a scale of 1 to 5, with 1 being "poor" and 5 being "good". They were given no other criteria in order to avoid introducing experimenter bias. The experiment was administered in isolated closets using over-ear, sound isolating headphones. Otherwise, the methodology was as in the above experiments.

4.1.2 Participants

Participants were thirty undergraduates at the College of William and Mary currently enrolled in introductory psychology classes. They received class credit for participating. None had a history of hearing or speech disorders and all were native, monolingual speakers of English.

4.2 Results

The data was first analyzed using a three-way ANOVA in order to determine whether there was a three-way interaction between the factors–expectation (reduction expected vs. reduction not expected), reduction (reduced vs. not reduced) and speed (fast vs. slow). There was no three way interaction between reduction, expectation and speed $F(5, 29) = .007$, $p = .933$.

After this was determined, two-way ANOVA's were run to determine whether either the speed/reduction interaction or the reduction/expectations interaction was significant. We would expect, if the Exemplar Theory model posited above was correct, that as speed increased, the preference for reduction would increase. If the rule-based phonology model is the correct one, however, there should be no interaction between speed and judgment. The reduction/expectation interaction is the one that we have observed previously, and we would expect the same interaction here, that those utterances that match expectations with reduction are judged as better than those which did not.

There was an interaction between speed and reduction, $F(1, 29) = 10.159$, $p = .001$ as well as an interaction between reduction and expectations, $F(1, 29) = 6.419$, $p = .011$. These can be seen in the graph Figure 4.2 for the interaction between speed and reduction, and Figure 4.3 for the interaction between reduction and elicitation. Figure 4.2 shows that listeners were more likely to accept reduced speech in a slow setting and more likely to accept unreduced speech at faster speeds. This is troubling, and is discussed at greater length in Chapter 5. On the other hand, reduction was more likely to be accepted when it was expected and rejected when it was not, in keeping with earlier results.

4.3 Discussion

As you can see, the interaction between expectations and reduction was the same as in all previous experiments: subjects preferred tokens which matched their expectations. This was unsurprising. What was surprising was the interaction between speed of utterance and reduction. Neither of the predicted outcomes was the observed one. Instead reduced utterances were actually judged as being worse in the fast condition. This is at odds with what Exemplar Theory would predict since, as explained earlier in Chapter 4, listeners are more likely to have more tokens that are fast and show reduction (even when it's not expected).

It is possible, of course, that because the fast, reduced token were the ones which had undergone the most digital manipulation, listeners honed in on that

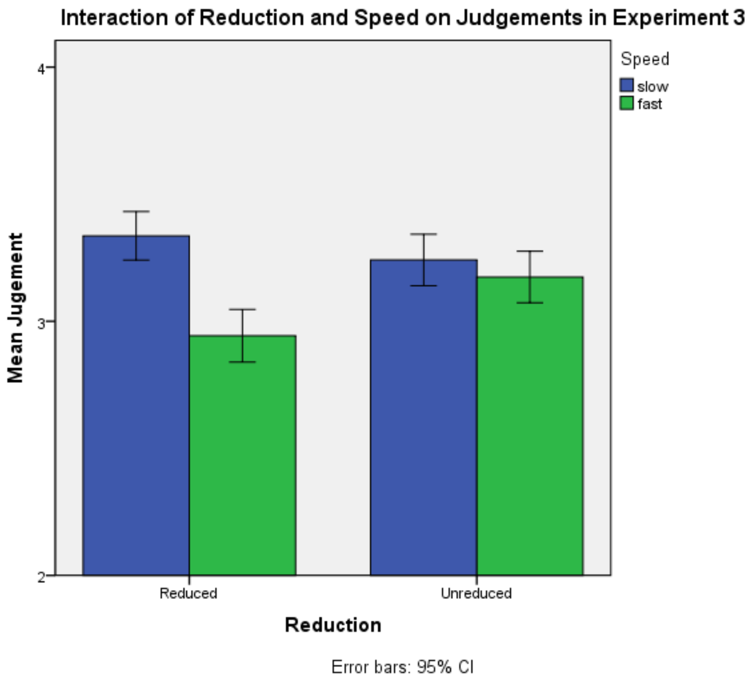

Figure 4.2: Interaction of Reduction and Speed on Judgments on Experiment 3

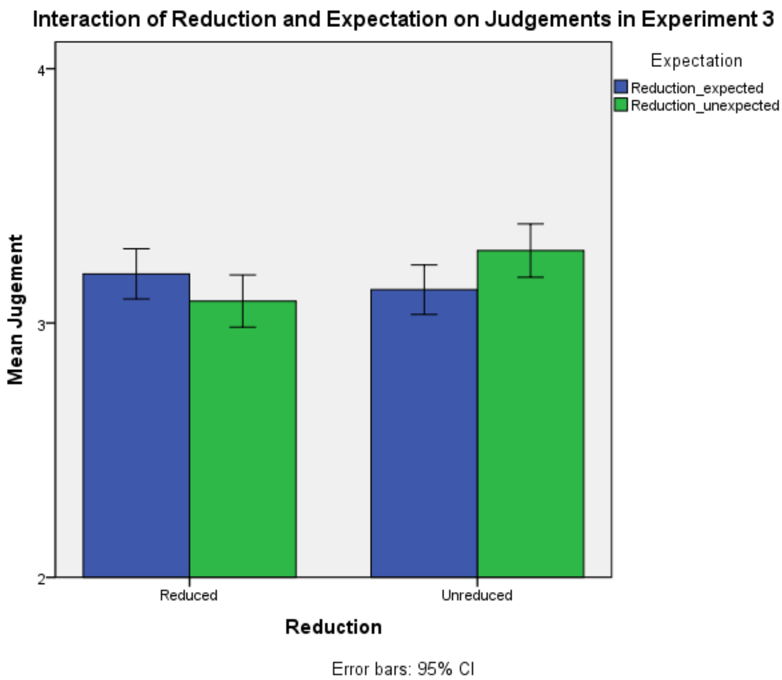

Figure 4.3: Interaction of Reduction and Expectation on Judgments in Experiment 3

manipulation and judged it as less acceptable. If that were the case, however, we might expect them to judge any manipulated token as less good than a similar un-manipulated token. Since they judged reduced tokens as being better than fully articulated tokens in the slow condition, however, we can reject this hypothesis.

Why then would they react in this way? The effect was statistically significant and the sample size large enough that random variation is an unlikely reason for the interaction. The difference must be one between fast and slow speech. What is the process of listening to fast speech like? Obviously, listeners have less time to parse the speech stream, and reduction is more prevalent. This means that listeners must grasp at every bit of information they have to accurately hear what is being said (Mattys, White, & Melhorn, 2005). Without context clues to rely on–since they were hearing these tokens in isolation–listeners were forced to rely on only phonological cues. Reducing those phonological clues meant that they had less speech information to work with, which would make understanding what was said more difficult. This difficulty might have translated to a lower judgment. It was not that they found the way that the word was produced at odds with their internal grammars, but that it interfered with their comprehension. (In other words, they judged slow utterances based on how well they fit their phonological models, and fast utterances on how easy they were to understand.) Further work is needed to determine if the criteria really was different at different speech rates, but it does seem likely.

To summarize: there was an interaction between speed and reduction, so rule-based phonology is not the dominate model here, but Exemplar Theory does not seem to completely describe the data either. A preference for reduced tokens even in the reduced state and, again, the fact that there was a main effect of speed do seem to support Exemplar Theory more than rule-based phonology.

However, there was another force at work on listener's judgments. A greater focus on comprehension at higher speech speeds is one possibility, but that hypothesis requires more experimentation.

Chapter 5

Conclusion

So, to return to the original question, will experiments which use perceptual judgments of connected speech support rule-based phonology or Exemplar Theory? The results of this series of experiments are somewhat mixed. Experiment one clearly showed that participants are capable of differentiating between phonological and non-phonological reduction, lending empirical support to the idea that phonological change is perceptually real for listeners. Experiment two confirmed that this ability is found in non-linguists as well, and that methodological flaws were not responsible for the results of the first experiment. These results could support either Exemplar Theory or rule-based phonology, as both posit that listeners have an internal mechanism for judging sound change. The main difference is how that internal mechanism is formed. In rule-based phonology it is described as a series of switches set in childhood, never to be moved again (Kenstowicz & Kisseberth, 1979). In Exemplar Theory, however, the mechanism is constantly being formed and added to, which allows for speaker's judgments to reflect the speech they encounter everyday (Pierrehumbert, 2003). The former will not allow judgments to be affected by the speed of the utterances (since speed is not part of a competency model), while the latter will. In fact, because Exemplar Theory is statistically based itself, it is very well-suited to dealing with the type of non-phonological reduction discussed here.

Experiment three should have settled the debate. If judgments were statistically based, then reduction in fast speech should not have troubled listeners. If they were rule-based, then speed should have no effect on listener's judgments. On the surface, it would seem that the presence of a main effect of speed ($F(1, 29) = 20.188$, $p < .001$.) on judgments means that rule-based phonology is insufficient to describe the perceptual phonological model underlying listeners' judgments. However, the interaction between speed and reduction was the opposite of what Exemplar Theory would predict. As discussed above, it seems likely that the listeners' criteria for judgment changed based on the speed condition. At the slower speed, they were more concerned with correctness, while at the faster speed comprehensibility seems to be the more important factor. Due to this apparent switching of criteria, it is difficult to tease apart more details of the listener's phonological models. However, the strength of the effect of speed does strengthen the argument for a statistically-based model, i.e.; Exemplar Theory.

Future work would do well to focus on ensuring a consistent judgment strategy on the part of the participants. This could be done through embedding the target words in sentences, thus reducing the amount that comprehension rested only on the sound signal of the individual word. Alternatively, participants could be instructed to focus on how a word sounds. Regardless, this work shows that the traditional divide between phonetics and phonology can be safely closed, even when looking at perceptual data for connected speech. Experiments 1 and 2 support the existence of phonological rules, and Experiment 3 shows that there is still much to learn from experimental data.

References

Abramowicz, Ł. (2007). Sociolinguistics meets exemplar theory: Frequency and recency effects in (ing). *University of Pennsylvania Working Papers in Linguistics*, *13*(2), 3.

Aguilar, L., Blecua, B., Machuca, M., & Mann, R. (1993). Phonetic reduction processes in spontaneous speech. In *Third European conference on speech communication and technology.*

Balota, D., Yap, M., Hutchison, K., Cortese, M., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Brenchley, M. (2011). *Disc: Remarks by Noam Chomsky in London.* http://linguistlist.org/issues/22/22-4650.html.

Chomsky, N. (1965). *Aspects of the theory of syntax* (Vol. 119). The MIT press.

Chomsky, N., & Halle, M. (1968). The sound pattern of English.

Cohn, A. (1993). Nasalisation in English: phonology or phonetics. *Phonology*, *10*(1), 43–81.

Cohn, A. (2011). *The Oxford handbook of laboratory phonology.* Oxford: Oxford Univ. Press.

Cohn, A., Fougeron, C., & Huffman, M. (2011). *The oxford handbook of laboratory phonology.* OUP Oxford.

Dalby, J., & Club, I. U. L. (1986). *Phonetic structure of fast speech in American English.* Unpublished doctoral dissertation, Indiana University.

De Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: flapping revisited. *Journal of Phonetics*, *26*, 283–310.

Fitzpatrick, J., & Wheeldon, L. (2000). Phonology and phonetics in psycholinguistic models of speech perception. *Phonological knowledge: conceptual and empirical issues*, 131–160.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, *18*(01), 7–44.

Gaskell, M., Hare, M., & Marslen-Wilson, W. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, *19*(4), 407–439.

Giegerich, H. (1992). *English phonology: An introduction.* Cambridge Univ Pr.

Goldsmith, J. (1979). The aims of autosegmental phonology. *Current approaches to phonological theory*, 202–222.

Haahr, M. (2011, November). *List randomizer.* Available from http://www.random.org/lists/

Hale, M., & Reiss, C. (2000). Phonology as cognition. *Phonological knowledge: Conceptual and empirical issues*, 161–184.

Hume, E., & Johnson, K. (2001). *The role of speech perception in phonology.* Emerald Group Pub Ltd.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Working papers in linguistics- Ohio State University Department of Linguistics*, 101–113.

Johnson, K. (2004). Massive reduction in conversational American English. In

Spontaneous speech: Data and analysis. proceedings of the 1st session of the 10th international symposium (pp. 29–54).

Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology.* Academic Press San Diego, CA.

Kohler, K. (1991). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica*, *48*(2-4), 180–192.

Kohler, K. (2000). Investigating unscripted speech: Implications for phonetics and phonology. *Phonetica*, *57*(2-4), 85–94.

Martin, R., Lesch, M., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*(1), 3–29.

Mattys, S., White, L., & Melhorn, J. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477.

Mielke, J., et al. (2003). The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica*, *60*(3), 208–229.

Mitterer, H., Yoneyama, K., & Ernestus, M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for/t/-reduction in speech comprehension. *Journal of Memory and Language*, *59*(1), 133–152.

Norusis, M., & Soci, I. S. P. for the. (1994). *Spss advanced statistics 6.1.* SPSS.

Osser, H., & Peng, F. (1964). A cross cultural study of speech rate. *Language and Speech*, *7*(2), 120–125.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, *46*(2-3), 115–154.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). Buckeye corpus of conversational speech (2nd release)[www.

buckeyecorpus. osu. edu] columbus, OH: Department of psychology. *Ohio State University (Distributor).*

Prince, A., & Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. *Optimality Theory: Constraint interaction in generative grammar*.

Pullum, G. (2011). *Disc: Remarks by Noam Chomsky in London.* http://linguistlist.org/issues/22/22-4631.html.

Raymond, W., Dautricourt, R., & Hume, E. (2006). Word-internal/t, d/deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, *18*(1), 55.

Team, A. (2011, November). *Audacity (version 1.3.13 beta).* Available from `http://audacity.sourceforge.net/`

Team, S. (2011, November). *Sliderocket.* Available from `http://www.sliderocket.com/`

Warner, N., & Tucker, B. (2010). "probably, ok, whatever!": Variability in conversational speech stops and flaps. *The Journal of the Acoustical Society of America*, *127*, 1854.

Zue, V., & Laferriere, M. (1979). Acoustic study of medial/t, d/in American English. *The Journal of the Acoustical Society of America*, *66*, 1039.

Appendices

Appendix 1: Words with word-medial stops with phonological stop-to-flap reduction (Reduction Expected)

1. derided - dɪˈɹaɪdəd
2. denoted - dɪˈnoʊtəd
3. curated - kjʊəˈɹeɪtəd
4. collated - kəˈleɪtɪd
5. catalogs - ˈkætəlɒgz
6. broadening - ˈbɹɔːdənɪŋ
7. attitude - ˈætɪtjuːd
8. translating - tɹɑːnsˈleɪtɪŋ
9. inedible - ɪnˈɛdɪbəl
10. impeding - ɪmˈpiːdɪŋ

Appendix 2: Words with word-medial stops without phonological stop-to-flap reduction (Reduction Not Expected)

1. adornment - əˈdɔɹnmənt
2. redirect - ɹidaɪˈɹɛkt
3. redesign - ɹidɪˈzaɪn
4. mystified - ˈmɪstəfaɪd
5. loftily - ˈlɔftɪli
6. listlessly - ˈlɪstlɪsli
7. infested - ɪn’fɛst.əd (syllabification keeps environment from being right for reduction)
8. earnestness - ˈɜɹnɪstnɪs
9. divestment - dɪˈvɛst.mənt (syllabification keeps environment from being right for reduction)
10. bestial - ˈbist.ɪ.əl (syllabification keeps environment from being right for reduction)

Appendix 3: Non-coronal filler items

1. Apache - əˈpætʃi
2. apogee - ˈæpəʊdʒi
3. appealing - əpi’lɪŋ
4. booksellers - ˈbʊksɛləɹzs
5. carcases - ’kɑɹkəsez
6. compelling - kəmˈpɛliŋ
7. conqueror - ˈkɒŋkəɹəɹ
8. disliking - dɪsˈlaɪkiŋ
9. evictions - ɪˈvɪkʃənz
10. lexicons - ˈlɛksɪkənz
11. ovenproof - ˈəvənpɹuf
12. pugnacious - pəgˈneɪʃəs
13. rapacious - ɹəˈpeɪʃəs
14. reclining - ɹəˈklaɪnɪŋ
15. refurbished - ɹiˈfəɹbɪʃəd
16. rococo - ɹoʊˈkoʊkoʊ
17. truculent - ˈtɹʌkjʊlənt
18. worshipful - ˈwɜɹʃɪpfʊl