Title
Phonetic Attention and Predictability: How Context Shapes Exemplars and Guides Sound Change

Author
Manker, Jonathan T.

Publication Date
2017-07-01

Phonetic Attention and Predictability:
How Context Shapes Exemplars and Guides Sound Change

By

Jonathan Taylor Manker

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Keith Johnson, Chair
Professor Susan Lin
Professor Andrew Garrett
Professor Darya Kavitskaya

Summer 2017

Phonetic Attention and Predictability:
How Context Shapes Exemplars and Guides Sound Change

Jonathan Taylor Manker

Abstract

Phonetic Attention and Predictability:
How Context Shapes Exemplars and Guides Sound Change

by

Jonathan Taylor Manker

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Keith Johnson, Chair

In this dissertation, I investigate how word predictability in context modulates the listener's attention to phonetic details, and how this in turn affects sound change. Three sets of experiments are designed to investigate these questions: In the first set of experiments, involving discriminability tasks, I demonstrate that (1) contextual predictability affects *speech perception*, and that listeners attend more to the phonetic details of unpredictable speech. In the second set of experiments I use the phonetic accommodation paradigm to show that (2) the effect of contextual predictability on speech perception in turn affects *speech production*. This by itself suggests relevance in sound change. In the third set of experiments I apply the model to a specific example of sound change: the reduction of function words. Using an error detection task I show that (3) listeners attend to the details of content words more than function words (with all other variables controlled for) which is linked to their differences in contextual predictability. I then propose a two-step model of sound change involving the propagation of contextually-modulated variation with a perceptual (rather than production) bias followed by the acquisition of new variants.

The results build and expand on several strands of literature which have not been fully connected previously. The findings for the effect of predictability on speech perception corroborate a number of past experiments showing that higher level linguistic information can have the effect of aiding speech recognition (Miller, Heise & Lichten 1951, Pollack & Pickett 1963), perceptually restoring missing information (Warren 1970, Marslen-Wilson, & Welsh 1978, Samuel 1981), or generally diverting attention from the raw auditory signal (Cole, Jakimik, & Cooper 1978, Ganong 1980). Additionally, this research considers dual-processing models of speech perception (Norris & Cutler 1979, Lindblom et al. 1995, Hickok and Poeppel 2004, 2007) in a broader context, considering how word predictability and expectancy modulate the type of listening used. The findings also add to the literature on exemplar theory (Johnson 1997, Pierrehumbert 2002, Goldinger 2007), particularly to hybrid models including both abstractions and exemplar clouds within the lexicon. Finally, I propose a new model of perception-based sound change driven by contextual predictability that can account for cross-linguistically common patterns of function word and morpheme reduction (Bell et al. 2001, Jurafsky et al. 2001, Beckman 1998) that does not rely on teleological production-based accounts of reduction (Lindblom 1990, Alyett & Turk 2004).

To my family, my parents Donn and Pam Manker

and my brother Christopher Manker

Table of Contents

List of Figures

List of Tables

Acknowledgements

The experience of writing a dissertation has been less an endeavor of creation than one of exploration, and, when fate should allow it, discovery. Often the words and ideas presented herein come not as fruits of labor but as sleep on a restless night, when the seemingly intangible is realized in the effort and struggle of non-action. And while this experience, and the moments of enlightenment which allowed me a fleeting glimpse at something profound in the architecture of reality, have been my own, I am deeply grateful for the guidance, support, and assistance I have received from a number of my advisors, colleagues, students, friends, and family.

I would first and foremost like to thank my advisor Keith Johnson, who encouraged me to broaden my research and explore new horizons once distant and arcane to me. His advising cultivated in me an eagerness for experimentation and the courage to question and hypothesize---and even the courage to fail in ways which have led to further growth as an academic. I would also like to thank the other members of my committee: Andrew Garrett, for advising me to make sometimes difficult changes to my work that always resulted in improvement; Susan Lin, for her encouragement and enthusiasm for my research as well as insights on key aspects of my experiments; and Dasha Kavitskaya, whose similar interests to mine, covering multiple areas of linguistics (phonetics, phonology, sound change, and field work), served as a model. I also owe certain insights to discussions with Susanne Gahl and Sharon Inkelas.

Additionally, I would like to thank many of the individuals in UC Berkeley's PhonLab, Phonology Phorum, and the rest of the graduate student body, my cohort in particular, for their constructive questions, comments, and insights that helped to strengthen my work as well as serve as a source of encouragement during the inevitable crests and troughs of life as a graduate student.

Furthermore, I want to give a special thanks to the undergraduate student researchers from the LRAP program who assisted with running experiments, collecting, and organizing data. Ashley Song and Kyuseok Lee were my first students who assisted (and put up with) my first pilot experiments as I was learning the ropes in the PhonLab. Libby Perfitt was an enormous help with my first breakthrough experiment, while Sofea Dil and Steven Ho were invaluable for their assistance in a marathon semester of experiments that expanded the findings of my dissertation.

Finally, I wish to thank my family, my parents Donn and Pam Manker and my brother Christopher Manker. They provided much needed emotional support during this herculean task, without which this dissertation would not be possible.

Chapter 1
Introduction

1.1. Contextual Predictability

Consider the following sentences containing blanks:

(1a) Kings and queens live in _____.

(1b) The woman suddenly saw the _____.

If hearing sentence (1a), the listener might have some pretty good guesses about what word is about to follow; 'castles' seems likely, although even someone expecting 'palaces' would not be surprised if the word turns out to be 'castles.' In this case we would say the word in the blank (1a) is *semantically (contextually) predictable*. The predictability is semantic because the meaning of the preceding part of the sentence is what helps determine likely candidates for the upcoming word. This is not the case in sentence (1b); just about any English concrete noun could plausibly fit in the blank, and thus whatever word ends up occurring in this blank would be considered *contextually unpredictable* by comparison. Imagine that either of the sentences was pronounced in a noisy environment. Assuming the listener heard all the previous words, if the word in the blank was partially masked by noise, which should we expect would be more easily identified? My hypothesis, based on many similar observations from the literature detailed in the following section, would be the word in sentence (1a). The listener can rely on knowledge of his language and the real world in sentence (1a) but will have to rely entirely on the auditory signal to determine the word in (1b). Thus, the hypothesis (built on the findings of the literature detailed in the next section) is that listeners will attend to the phonetic details of *unpredictable* speech more so than for *predictable* speech.

Now consider the examples below:

(2a) The cat is sleeping ___ the box.

(2b) There is a ___ in the box.

In sentence (2a), the word in the blank might also be considered predictable. It could be 'in,' 'on,' 'by' 'above,' 'under' or possibly a few other locative pronouns. What is most responsible for making the word predictable here--- by ruling out thousands of possible English words--- is not the semantics of this sentence, but the fact that the surrounding syntactic structure determines

the word in the blank must be a member of the preposition class. Prepositions, as functional, closed-class words have a restricted set of members. Thus, we would call the potential word in (2a) *syntactically (contextually) predictable*. Compare this to sentence (2b). The syntax determines the word in the blank must be a noun. However, there are thousands of possible words that would be grammatically acceptable in this blank, including most nouns besides some ruled out for semantic reasons, such as abstract nouns or items too large to fit in boxes. Thus, the word in (2b) would be, at least relatively speaking, unpredictable compared to (2a). This is a different kind of predictability than shown in (1a) or (1b), but I hypothesize it will affect the perception of speech in the same way: speakers will pay more attention to the phonetic details of the unpredictable word in (2b). If this is in fact shown to be the case, it will suggest listeners attend to the phonetic details of open-class, content words more than they do for closed-class function words. This could then be a relevant mechanism in the spread of sound change, and could account for the distinct, reductive phonological changes that tend to target more predictable function words. The following sections will review the literature relevant to these observations.

1.2. **Top-down processing***: Top-down processing aids in speech perception but impedes attention to the auditory signal*

Research beginning in the latter half of the 20th century began to reveal a number of phenomena which suggest that listeners do not always process all the details of the auditory signal and convert it into a string of phonemes before word recognition occurs. This was the prevailing view of the structuralists (Bloomfield 1933) who proposed that linguistic levels (phonetics, phonology, morphology, etc.) are completely separate, where the output of one level is unidirectionally fed into the next. These principles lend themselves to a model of speech perception driven by bottom-up processing, that is, listeners perceive the auditory signal, translate it into a string of phones, identify the intended phonemes, string these together into morphemes, and so on, until a complete, meaningful utterance is recovered from the speech signal. A model including top-down processing, however, suggests that meaning, context, or any higher level knowledge of one's language could be used to aid in determining the intended sounds of the speech signal. This higher level linguistic knowledge could include awareness of phonotactic rules--- what strings of sounds are legal in a given language--- knowing what words are real in a given language, or considering what strings of words make sense, semantically and pragmatically, in a given context. While top-down processing is shown to aid in speech perception and word recognition, at least when an appropriate amount of context is available, it also results in lower awareness and attention of the details of the auditory signal.

Several phenomena began to be observed beginning in the 1950s, finding experimental evidence of the role of contextual information and how it influences the perception and attention to the speech signal. For example, Miller, Heise & Lichten (1951) showed that words were better identified in noise when occurring in sentence contexts. This would indicate that having additional sentential context aids the recognition of words, and hints at the idea that sounds might be restored in some way if they are not actually perceived in the speech signal due to noise. The research of Liberman et al. (1957, 1967) reveals vast variability in the speech signal and the overlapping of

phonemes in connected speech, suggesting there is no one-to-one mapping between acoustic signals and phonemes. Pollack & Pickett (1963) demonstrated this problem by excising words from their original sentential contexts asking listeners to identify them, revealing the difficulty listeners had with word recognition in this situation--- only about 50% of words were recognized without their original surrounding contexts, though 70% of words were identified with only one or two additional words from context. These findings highlight the problems of recovering the intended signal in purely bottom-up model of speech perception, and the need to incorporate the effects of higher level knowledge into models of speech perception. Around this time, Halle & Stevens's (1959, 1962) "analysis by synthesis" model of speech recognition proposed that listeners used acoustic cues to generate hypotheses of the intended words of the speech signal, in turn comparing these to the raw auditory input until speech recognition is achieved. This is one of the earliest models that considered that listeners simultaneously used higher level linguistic information, such as considering what words are real words, and what utterances would make sense in a given context to decipher the speech signal, incorporating top-down processing into the model. Thus, these studies show a balance between the reliance on the auditory signal and contextual information with the suggestion that details not even present in the speech signal could be subconsciously restored--- which results in less veridical representations of the speech signal stored and activated by the listener.

Later research continued to uncover experimental evidence showing that higher level linguistic information influences the perception of acoustic information through a variety of psycholinguistic phenomena. Warren (1970) demonstrated the *phoneme restoration effect* by which listeners may seemingly hallucinate a speech sound that fits with the broader context. Warren replaced single phonemes within words with non-speech sounds such as a cough (such as replacing the /s/ in 'legislature'). Subjects were unable to determine the location of the replaced sound when asked; when told the [s] sound had been replaced with a cough, subjects still had trouble noticing it. Warren's findings suggest that listeners had identified the word based on the beginning of the word and/or the surrounding context, and had in some way restored the sound from their own knowledge of the sounds that should occur in the word. Samuel (1981) also showed the relevance of sentence-level contextual predictability on the phoneme restoration effect. In his experiments, he asked subjects to determine whether a noise was added to a word or if it was replacing a sound. Subjects were more likely to hear a sound as being added when the context biased the listener to expect a certain word. In these cases, phonemic restoration had occurred, and subjects seem to have perceived sounds that were not actually present in the auditory signal. Samuel concluded from these results that "restoration is a function of context; the greater the context, the greater the expectation, the greater the restoration" (p. 481).

Another phenomenon showing the effect of context on speech perception was that of *fluent restoration*. In the fluent restoration paradigm, subjects were asked to shadow speech that they heard--- that is, repeat the words exactly as they were heard, and in some cases these words contained phonemic errors. Marslen-Wilson & Welsh (1978) presented subjects with passages containing target words with errors in either the first or third syllables of words, and found that subjects were more likely to restore the sounds in the third syllables as opposed to the first syllables. The argument is, once again, that context guides the processing of the auditory signal--- three syllables into a word, listeners may already have identified the word, and the intended or

correct sounds may be perceptually restored based on the expectations established through higher level linguistic information--- in this case, what words exist in a particular language or which ones would make sense in a given context.

Error detection experiments also show the use of higher level linguistic information in the perception of the speech signal and its ability to divert attention from the details of the auditory signal. Cole, Jakimik, and Cooper (1978) asked subjects to indicate when they noticed errors that involved changing a single phoneme of the intended word (such as 'boy' pronounced as 'poy'). The results showed that subjects more reliably detected errors in word onsets as opposed to in word final position. Nevertheless, subjects had a slower reaction time in determining errors in initial syllables as opposed to final syllables. This suggests that word final errors may have gone unnoticed, as higher level linguistic information overrode the auditory signal. At the same time, subjects' phonetic processing faculties may have been less taxed by having an expectation of the upcoming word, allowing them to identify errors more quickly when they noticed them at all.

The "Ganong" effect (Ganong 1980) is yet another psycholinguistic phenomenon that demonstrates the usage of higher level linguistic knowledge in influencing the perception of the auditory signal. In this experiment, listeners were presented with acoustically manipulated word stimuli that fell along an phonetic continuum between two minimal pairs, such as 'tash'-'dash' or 'task'-'dask.' In each pair of words, one was a real word ('dash,' and 'task') while the other in the pair was a phonotactically legal pseudoword. Subjects were then asked to identify the sound the words began with. Stimuli containing an ambiguous [t] / [d] segment occurring halfway along the continuum showed a bias in being identified as the real word member of the pair--- thus subjects were biased towards hearing 'dash' and 'task' even when the auditory signal was completely ambiguous. The results of Ganong's research show that listeners actively use their language's lexicon to guide speech perception. Perhaps one might say the listener shoehorns the auditory signal to yield something maximally meaningful, even if the fit is imperfect but close enough. In any case, this contextual information results in the activation of an abstract representation of speech that may not be faithful to the acoustic details present in the signal.

Newer models of speech perception and word recognition during this time took into account these experimental findings and expanded the role of top-down processing of the speech signal. Marslen-Wilson & Welsh's (1978, 1987) cohort model of lexical retrieval proposes that listeners process the speech signal from left to right (chronologically), using the lexicon to develop and revise groups of candidate words until a word is recognized. For example, if the onset and first vowel of a word heard in speech are [kæ], listeners will compile a cohort of words which begin with this sequence: {'cat', 'cab,' 'caterpillar,' 'cacophony,' etc.}. Once the next sound is heard, [kæt], the set of words will further be diminished. If the string of sounds ends up being something like [kætəp] there will be only a single possible English root word matching this string, 'caterpillar,' and at this point, called the uniqueness point, recognition will occur as there are no other possible candidates. This model can account for Marslen-Wilson and Welsh's (1978) findings for fluent restoration, such that speakers were more likely to restore mistakes in the third syllables of words, since it may be that listeners do not rely on the auditory signal after the point of recognition, as precise phonetic attention is less important at that point.

McClelland & Elman's (1986) TRACE model went further in considering the role of top-down processing on speech perception. Their model included three layers of processing: features,

phonemes, and words. Perceptual pathways linked the different levels: bottom-up processing could occur when collections of features from the feature level are used to activate phonemic representations in the phoneme level, followed by strings of phonemes activating words in the word layer of processing. TRACE also includes connections from the word to phoneme level, which is essentially a top-down interaction. Should contextual information inform the listener about the likelihood of word's identity--- either before or after it is heard--- the word, and its phonemic representation, can activate or revise the phonemic level. This helps account for cases in which contextual information causes word onsets to be misperceived (such as the bias in hearing Ganong's ambigious [t/d] onsets as the member of the pair that forms real words with the following segments) which are harder to explain in classic versions of the cohort model. Additionally, TRACE is well adapted to predicting the influence of higher level linguistic information on speech perception beyond knowledge of the lexicon, such as syntactic and semantic information that biases the listeners towards certain more plausible words.

A review of the literature shows a long and well-established body of research which demonstrates the effect of top-down processing in speech perception. The task of perceiving speech is not achieved merely through processing the raw auditory signal, but rather this process occurs "through an interaction of sound and knowledge" (Cole & Jakimik 1980:136). The goal of speech perception is not to record every acoustic detail of speech faithfully, but rather, to extract meaning. Given the variability of the acoustic signal, due to individual differences as well as the coarticulation and reduction of fluent speech, top-down processing is effective in achieving this goal, though may have consequences in the maintenance of stable acoustic representations of words over time.

1.3. **Dual modes of listening***: Listeners attend less to sound when they attend more to meaning*

Many studies have described two distinct mechanisms which humans use in recognizing speech. The first mechanism or mode of listening involves directly processing the raw auditory signal, in some way extracting speech sounds, and in turn identifying words. The second mode of listening focuses on extracting meaning rather than sound; entire words are activated in the mind of the listener as opposed to sounds. This second mode of listening is considered to be more common in regular communication (Lindblom et al. 1995) and is prone to top-down processing phenomena such as phoneme restoration. Similar to the effect of context and top-down processing, the literature reveals that when listeners are more closely focused on extracting meaning from the speech signal that they will attend less to acoustic details.

The need for describing two distinct speech modes may have its origins in Liberman et al. (1967) which noted the profound acoustic variability that occurs in fluent speech and the difficulty listeners would have in normalizing the speech signal. Liberman proposed that there was a "special decoder" for "phoneme perception" (p. 431) which involved extracting gestures, rather than stable acoustic representations, from the speech signal. Liberman's proposed 'speech mode' involves a tight connection between perception and production, such that the same neural mechanisms used to produce sounds are also involved in identifying them.

Norris & Cutler (1979) showed that phoneme identification and word identification was achieved through distinct mechanisms. They point to Morton & Long (1976) who show that when

words are highly predictable, lexical access is quicker, and phoneme monitoring will occur *after* lexical access. Here one could imagine that the actual details of the auditory signal may be largely ignored in favor of activating abstract representations of the intended sounds. In contrast, Foss, Harwood, and Blank (1980) found that reaction time in identifying phonemes is not dependent on word frequency, suggesting that phoneme identification can occur *before* lexical access. Norris & Cutler claim the results are not inconsistent, but suggest different instructions in the experiment may have caused the difference. When subjects are required to comprehend and recall sentences, they may focus their attention on meaning and will extract phonemic information after lexical recognition, whereas in a pure phoneme detection task, subjects will listen only for sounds and may not activate lexical representations until after phoneme identification, if at all. Klatt's (1979) model of speech perception also includes segmental and non-segmental means of recognizing words. Using LAFS (lexical access from spectra), the listener matches entire words with stored acoustic forms of those words, while using SCRIBER, listeners can extract speech segments from the auditory signal in order to identify the forms of new words. Klatt says LAFS is probably the mechanism used in "normal conversation speech" (p. 308) despite the presence of both models. LAFS would also require less attention to the veridical details present in the auditory signal.

Another model described in terms of distinct sound and meaning mechanisms for listening to speech is that of Lindblom et al. (1995). Building on observations made by Ohala (1981, 1983), Lindblom proposed two modes of listening which may be active under different circumstances. The 'what' mode is the standard means of listening to speech, in which listeners focus on the content and meaning of what is being said. In this mode, listeners make full use of all higher level knowledge (phonology, syntax, semantics) for aiding word recognition. In the 'how' mode, speakers are more concerned with the phonetic detail of speech, where these details might include phenomena such as coarticulation and reduction. Ohala (1981) proposed that situations in which a single speaker may adopt a new variant may arise from misperception when a listener fails to undo a coarticulatory effect, and does not reconstruct the speaker's intended form. Building on this, Lindblom et al. (1995) suggest this could be what occurs when a listener processes speech in the "how" rather than the "what" mode, and the auditory signal is processed in its raw, literal form, causing a listener to fail to undo a coarticulatory effect.

More recent literature in neurolinguistics has provided evidence of different neural pathways that might be involved in 'what' and 'how' listening. Hickok and Poeppel (2004, 2007) define two streams of processing involved in what they refer to as *speech recognition* and *speech perception.* The first stage of processing the auditory signal is bilateral and occurs in the superior temporal gyrus, however, further processing progresses along one of two streams. The *ventral* stream involves projection ventro-laterally toward the inferior posterior temporal cortex bilaterally. It is here that lexical retrieval, sound to semantic mapping, occurs (*speech recognition*), interfacing with various regions of the brain where conceptual information (visual, motor, etc.) is stored. The *dorsal* stream diverges from the superior temporal gyrus, projecting dorso-posteriorally in the region of the posterior Sylvian fissure and progressing finally towards the frontal regions, primarily oriented on the left side. The dorsal stream is involved both in motor planning as well as the sublexical processing of auditory input (*speech perception*) such as phoneme identification and rhyming tasks; this dual function suggests a link between production and perception and mirrors the proposals of Lindblom's (1967) motor theory of speech perception.

The ventral stream is roughly analogous to Lindblom's 'what' mode, while the dorsal stream corresponds to the 'how' mode.

The literature establishes the likelihood that speech may be processed in either a mode that focuses on sounds or a mode that focuses on words and meaning. It is not clear under what circumstances each mode is used, though it seems likely that the 'what' / ventral mode could only be used in the presence of abundant context that may aid the listener in identifying massively reduced forms of words. By using contextual information to activate word forms, a more abstracted and less veridical form of the word would be stored in memory. In contrast, the 'how' / dorsal mode would probably be used in the absence of context, when the listener must rely entirely on the auditory signal; this could be the case when listening to words in isolation or perhaps when sentence context fails to provide any meaningful information that would aid in word recognition. In any case, effects of top-down processing--- such as restoring missing sounds or identifying phonemes *after* the activation of lexical forms, could only occur under the conditions of 'what' listening. Given the connection established between the presence of context and effects that occur during 'what' listening, and the reverse in 'how' listening, the results of the experiments in the current study may provide new clues for the activation of different listening modes. This may give insight as to whether different modes are used for different speech contexts--- such as in sentence context or when words are in isolation, lacking context--- or if different modes of listening are activated online during the perception of speech, as a function of the degree of context available for identifying each word.

1.4. **Exemplar Theory***: Listeners store instances of words in memory but context determines the veridicality or abstractness of exemplars*

A primary question of concern in this dissertation is what details of the auditory signal are stored in memory. On opposing sides are abstractionist models which suggest listeners reduce the auditory signal to a string of phonemes, and pure exemplar models which suggest most or all acoustic detail is retained in memory. However, more recently proposed models suggest that high level knowledge of one's language may effectively act to filter the auditory signal, whereby the veridicality or abstractness of given exemplars may be a function of the contextual information available.

Until the last few decades, it has been assumed in traditional models of linguistic organization including structural (Bloomfield 1933) and generative linguistics (Chomsky & Halle 1968) that words are stored in the speaker's mind exclusively as abstract representations consisting of strings of phonemes. Phenomena such as allophony, coarticulation, and reduction resulted from the output of phonological rules or from unintended consequences of the limitations of the human articulatory system. The driving principle behind this system was economy and efficiency of memory; in the same way that segments are economical by re-using articulatory gestures (features), word representations stored in memory drew from a minimal set of contrastive sounds. The listener in turn had the job of decoding the speech signal--- undoing the rules that resulted in the phonetic output and treating speaker variation as noise, with the goal of extracting the intended phonological forms of words from the messy acoustics.

Liberman (1967), as previously discussed, noted the lack of invariance of words in fluent speech and the difficulty listeners would have extracting the intended abstract representations of words from this sea of variety. Thus, his proposal that listeners perceived articulatory gestures rather than the acoustic signal in its raw state was an attempt to address the problem of decoding the speech signal. Another line of research argued that vowels are perceived based on the ratios of vowel formants rather than the absolute frequencies of the formants themselves (Peterson 1961, Sussman 1986, Miller 1989). Others considered that vowel perception might also involve cues such as duration and the vowel's trajectory (Lehiste & Metzger 1973, Hillenbrand & Nearey 1999). Much of this research thus proposed complex algorithms by which the extraction of abstract representations from the speech signal might be achieved.

More recently, others have pointed to factors that suggest that words may not be stored exclusively as abstract representations, and that extracting such forms from the speech signal may be unnecessary. These findings point to exemplar models of word storage, whereby individual instances or traces of words are stored in "clouds" of examples within memory. The results of Goldinger (1996) are particularly convincing with respect to this point. His experiments showed that subjects were better at remembering whether a word had been heard previously in the experiment when the words were spoken with the same voice. Following the traditional expectations of decoding the speech signal, listeners would strip away all variation in words, including speaker and voice information, and all that would be stored concerning a word's identity would be its abstracted phonological form. However, since subjects showed they recalled words significantly better when repeated with the same voice that had previously spoken the word, this suggests that voice cues are stored alongside linguistic information. Furthermore, it suggests that individual instances of words are stored in memory.

Johnson (1997) argues that listeners do not strip away variation in order to decode the speech signal, but instead actually use that information in speech recognition. Listeners make perceptual judgments about a speaker's identity, such as gender, and use this information to guide the recognition of particular words. A number of studies suggest that speech might be normalized relative to a speaker's gender (May 1976, Mann and Repp 1980) and also that perceptual boundaries between two sound categories (such as different vowel phonemes) may be different when one believes he is listening to men as opposed to women (Johnson 1990). However, Johnson's (1997) argument is that prior experience with a wide range of both male and female voices may aid perception rather than be a hurdle to overcome with complex perceptual algorithms for gender normalization. This is later supported by the findings of Strand (2000) showing that word recognition occurs more rapidly when listening to voices that align with gender stereotypes.

Other evidence points to individual words having independent phonetic histories, where different words undergo different rates of change. Traditional models of sound change upheld the view that words were stored as abstract strings of phonemes. The Neogrammarians argued that sound change was *lexically abrupt* and *phonetically gradual* (Labov 1981, concerning the Neogrammarian hypothesis) which is essentially the concept of *regular sound change*. That is, an abstracted phonological unit, X, would change gradually over time but *in all words containing that sound*. Later accounts proposed the existence of change that was *lexically gradual* but

phonetically abrupt (Wang 1969), often referred to as *lexical diffusion*. In this way, one phoneme in a word may abruptly become an entirely different phoneme or allophone, and such a process spreads slowly throughout the vocabulary until possibly, though not necessarily, affecting all words containing that sound. Either of these cases does not contradict the existence of abstract phonological forms. For regular sound change, an abstracted unit such as the phoneme /p/ merely changes in its precise phonetic characteristics as it occurs in all words, whereas as in lexical diffusion, a word that is /pater/ one day may become /fater/ the next for a given speaker, whereas this change does not immediately affect other words containing /p/ in the same environment.

More recent work, however, has pointed to the existence of *phonetically gradual* and *lexically gradual* changes (Hooper 1976, Bybee 2001, 2002, Phillips 2006). These accounts typically appeal to word frequency as an important motivator in the spread of sound changes. Word frequency itself is a characteristic specific to individual words, and suggests that words, rather than just sounds, can have their own phonetic histories. Usually reductive type changes (lenition, deletion, vowel centralization, etc.) target high frequency words, whereas analogical changes (such as lexical diffusion) may target low frequency words (Hooper 1976). Recent research in usage based models considers the concept of frequency in a "favorable context" for a given sound change (Brown 2013, Bybee 2014). If a word tends to occur in a particular phonetic environment that might condition a change, such a change is more likely to be adopted in all instances of the word. Additionally, the findings of the phonetic accommodation paradigm (Goldinger 1998, Shockley, Sabadini, & Fowler 2004, Nielsen 2011) show that extremely subtle, sublexical changes can occur in specific words over a short period of time in a lab setting. This literature will be reviewed more in depth in chapter three.

The existence of lexically and phonetically gradual changes, effects of word frequency, and the findings of the phonetic accommodation paradigm are inconsistent with models of the lexicon that only include abstract phonological forms of words. Abstraction-only models are incompatible with the concept that stored word forms could contain their own fine-grained phonetic information. Exemplar models, on the other hand, are well-equipped to accommodate this type of information. In exemplar models, each word contains a cloud of particular instances of those words that the listener has perceived in her past linguistic experience. These instances, or exemplars, may also contain voice and socio-indexical information about the speaker who uttered a particular word. This model better explains how listeners store knowledge about expectations of how other speakers sound (based on gender, vocal tract size, social identity, etc.) and how listeners may use this information to improve speech perception.

Exemplar models have a few weaknesses. Johnson (1997) points out the issue of the "head-filling-up problem," whereby the model may require listeners to possess an almost limitless amount of memory. This aspect of the exemplar model is particularly at odds with the older concept of memory efficiency, whereby the phonetic output was derived from the abstract phonological representation through the application of rules and unavoidable limitations of an individual's articulatory system. While Johnson rejects a "pure" exemplar model, where every single phonetic detail of every single word is stored in memory, he also points to studies of human visual memory (Standing, Conezio, and Haber 1970) that suggest human brain can store a large

amount of information and that there is little need to assume memory efficiency should be a guiding principle in models of the lexicon. Additionally, abstractionist models are better equipped to handle the spread of phonological changes, particularly when they show regularity. For example, the generalization that /t/ becomes [ɾ] in intervocalic position is easily captured with a phonological rule that affects all instances of /t/ in all words in this environment. Pierrehumbert (2002) therefore proposes a hybrid model which contains exemplar clouds of specific phonemes, including word exemplars containing a particular sound. This account better models the application of phonological rules, but at the same time is better at handling exceptions and the irregularities of diffusion.

More recent work has also suggested that listeners do not simply store all the details of the raw auditory signal, and may in some way filter out some details due to higher level linguistic knowledge and experience. Goldinger (2007) proposes that "each stored exemplar is actually a product of perceptual input combined with prior knowledge…" (p. 50), and argues against pure exemplar or abstractionist models. Pierrehumbert (2006) shows that speakers attend more to informative socio-indexical features, suggesting some acoustic information may not always be stored in memory. Maye (2007) demonstrates that a listener's exemplars are shaped by past linguistic experience, such that one's L1 effectively trains the listener to notice certain phonetic details and disregard others.

The hypothesis of this dissertation is that listeners may disregard phonetic details when there is more contextual information that makes words predictable. Thus, when words are predictable from prior context, or words are strongly confirmed by subsequent context, phonetic details of those words may either filter some details from being processed at all, or may cause faster decay of those details once the words are confirmed. The results of the experiments in this dissertation will corroborate and expand on aspects of the architecture of exemplar/abstractionist hybrid models of lexical storage.

1.5. **The role of perception and production in sound change***: The spread of sound change is rooted in perceptual biases*

A primary goal of the current study is to provide an account for certain types of sound changes, most notably the reductive sound changes that occur in function words to the exclusion of content words. Here I will propose a radically different account from previous ones, many of which invoke the role of the speaker in sound change as well as the phenomenon of word frequency. On the contrary, I propose these changes result from indirect effects resulting from listener misperception that occur due to the phenomenon of contextual predictability (see chapter 5 for the full account).

Many theories of reduction and coarticulation in particular appeal to the limitations of the articulatory system and the general concept of "ease of articulation." That is to say, sounds that are more difficult to make, or are more difficult to make in proximity to each other may eventually change to become articulatorily easier in some way. Zipf (1949:543) declares that "each individual will adopt a course of action that will involve the expenditure of the probably least average of his work (by definition, *least effort)*," while Martinet (1960:167) applies the observation more

specifically to the balance between the needs of the listener and the speaker as "[t]he permanent conflict between man's communicative needs and tendency to reduce to a minimum his mental and physical activity." Without a doubt, the literature can confirm the strong tendencies of certain variation to arise due to articulatory limitations. It is fairly intuitive, for example, that two adjacent consonants of different places of articulation would be simpler to pronounce if one assimilated to the place of the other. Less intuitive might be the fact that voiced fricatives or voiced velar sounds may be more difficult to produce (as opposed to their voiceless counterparts) due to the aerodynamic voicing constraint (Ohala 1983). In any case, this variation that arises does have its origin in speaker-oriented articulatory limitations.

All of this research, however, only points to reasons for the existence of phonetic variation and the results that tend to occur for articulatory reasons. More difficult to explain is how this variation becomes phonologized or lexicalized--- that is, when speakers intend to produce what once began as accidental variation. While many sound changes and phonological patterns are based on tendencies that result from limitations of the articulatory system, there is no reason to suggest that the speaker ever intends to change her language or in any way is less than perfectly faithful in producing speech exactly as it was originally transmitted and learned. To corroborate this point, despite the tendencies for certain phonological patterns to emerge due to articulatory reasons, this unintended variation can in fact be controlled by the speaker. Keating (1985) points out that many aspects of what may have been considered "universal phonetics," such as the lengthening of vowels before voiced segments, in fact show wide variation among the world's languages and are even absent in some. In other words, speakers are capable of resisting the slide into articulatorily "simpler" forms.

Some studies have pointed to the phenomenon of contextual predictability as a force that shapes language, but argue that the speaker deliberately alters the speech signal in anticipation of the needs of the listener. Jurafsky et al. (2001), for example, shows that words that are more predictable in context show significantly more vowel reduction and shorter word duration. This finding seems to align with Lindblom's formal "H&H" (hyper and hypo) theory, which proposes that "[w]hen output constraints dominate, we expect to see hyperforms, whereas with system constraints dominating, hypospeech will be observed" (Lindblom 1990: 413). That is to say, speakers purposefully yield to the law of minimal effort by reducing speech which is predictable, and use more energy to produce hyperarticulated forms of words when they are less predictable. To add to that, the speaker keeps "a running estimate of the listener's needs for explicit signal information on a moment-to-moment basis and then adapts the production of the utterance elements… to those needs" (Lindblom et al. 1995: 5). A similar idea is proposed by Alyett & Turk (2004), who state that "prosodic prominence is a linguistic means of achieving smooth signal redundancy. Prosodic prominence increases… with unpredictable sections of speech" (31). Here they point to a specific example of sound change--- the development of the prosodic system--- that could result from the speaker purposefully altering the speech signal in accordance with judgments of the listener's needs. The important thing to note here is the focus on the actions of the speaker. Such a hypothesis could predict that hyper- and hypoarticulation could result in say, a phonologized prosodic system whereby root syllables, being less predictable, are more prominent, while affixes, being more predictable, avoid stress; or whereby content words resist erosion, but

function words readily reduce. According to this account, this would result from the actions of the speaker eventually becoming phonologized, with at best a passive role in sound change assigned to the listener.

An alternative analysis suggests a production-based account for this reduction that does not involve the speaker making purposeful articulatory decisions in anticipation of the needs of the listener. Gahl, Yao, & Johnson (2012) argue that this reduction is the result of the ease of lexical access. That is, when words are predictable, they are primed not only for the listener but the speaker as well, and are more quickly and more easily retrieved from the lexicon. The speed of retrieval results in some amount of reduction. While at first this account may seem to be one that merely draws a different conclusion from the same results, they show there is also more reduction in words with high neighborhood densities. For a word to have a high neighborhood density means there are many other words in the lexicon that are phonologically similar, differing by only a single phoneme (thus 'hat' has a high neighborhood density--- 'hack,' 'hot, 'hit,' 'cat,' 'mat,' etc. are all phonologically similar, while 'shampoo' has a low neighborhood density). Words that have high neighborhood densities are shown to be retrieved more quickly, and also undergo more reduction than words of low neighborhood density. This is in spite of the fact a higher neighborhood density should mean more difficulty in word recognition on the part of the listener, since there are more phonologically similar words. Gahl, Yao, & Johnson's (2012) account of predictability-modulated reduction provides a less teleological role for the speaker but also serves to make this explanation a less likely source for the reduction that occurs in function words and morphemes. Because this online reduction occurs to any words that are easily retrieved from the lexicon, if it resulted in language-wide, phonologized sound changes we would expect to see it occurring in words with high neighborhood densities as well, but this does not in fact occur.

Bybee & Hopper (2001) offer another explanation that could account for reduction in function words (or potentially morphemes as well) which identifies speaker-oriented effects of word frequency as a possible cause. They claim that "the origins of reduction are in the automatization of neuro-motor sequences which comes about with repetition. This automatization involves the reduction of the magnitude of articulatory gestures and the increased overlap of these gestures" (p. 11). While this provides a less teleological account of how reductive sound changes could develop, it may suggest that the adoption of the reduced forms as the intended, phonological targets in the lexicon are due to speaker-oriented reduction that readily becomes a part of the language over time. If this reduction became phonologized in an uninhibited manner as soon as it occurred, we might expect the "massive reduction" that Garrett & Johnson (2012) warn against, which is not actually attested. This by itself suggests there must be some other mechanism at work. Additionally, the resulting patterns of sound change do not always occur strictly in the most frequent environments. On the contrary, many of the reductive changes that occur in function words and morphemes are categorical, such that they occur to all phonologically eligible forms in particular word classes (for example, initial /ð/ developing in determiners, pronouns, and function adverbs or the reduced forms of auxiliaries). Otherwise, if frequency were the only relevant factor, we might expect /ð/ voicing in the word 'think' which is more frequent than the word 'these.' The account proposed in this dissertation does not argue against the relevance of frequency in sound

change--- which arguably is one factor in 'word predictability'--- but rather considers the need to understand the effects of contextual predictability to account for certain types of sound change.

Ohala (1981) discusses the issue of phonologization in sound change--- that is, the observation that the inclination for certain variation to occur due to articulatory considerations does not mean that that variation will become the intended goals of speakers, at least without some other mechanism at work. He identifies listener misperception as the ultimate means by which new variation will become phonologized. He describes *hypocorrection* and *hypercorrection* as two means by which listener perception might occur. Either of these assume that under normal conditions, the listener has certain knowledge or expectations about how coarticulation affects proximate sounds, for example that a back vowel like /u/ might be fronted before a coronal consonant such as /t/. *Hypocorrection* occurs when speakers fail to undo the coarticulatory effect and "reconstruct" the intend form as /ut/, believing the intended form was something like [yt]. *Hypercorrection* is the opposite, when a listener assumes the similarity of nearby sounds is due to coarticulation, when in fact it is not. Ohala proposes that when these processes occur, a "mini-sound change" has taken place, whereby what was once articulatory variation now becomes the intended articulatory plan for the listener turned speaker.

Garrett & Johnson (2012) argue that such misperception is likely to be exceedingly rare or else we would once again expect "massive reduction" whereas there is in fact a resistant to change, even for those changes which are extremely common. They therefore argue in favor of a 'speech' mode of listening commonly used for processing fluent speech which typically ignores the retrieval of raw articulatory detail from the auditory signal instead focusing on word recognition and extracting meaning. Another issue arising without further amendments to Ohala's original theory of misperception is the question of why a fluent speaker of a language would readily accept a novel word form--- one that he perceived due to under-accounting for the effects of coarticulation--- into his lexicon. It is not yet clear what effect perceiving non-native phonetic characteristics or outright speech errors has on one's production of speech, although in the case of the former, the answer may lie in attitudes towards the speaker or other sociolinguistic factors (Babel 2010).

In this dissertation I will propose a model of sound change accounting for phenomena such as function word reduction which will consider the roles of contextual predictability and listener perception. This account will attempt to address many of the issues identified by the previous literature. First of all, I will avoid explanations which invoke teleology, such that speakers are in any way trying to simplify or make the language "better" in any way. Following Ohala (1981:197), I agree that we should "assume only that speaker and hearer are interested in communicating and will pronounce words only as they have heard them (or think they have heard them) pronounced by others." Secondly, my account will not rely on claims that speakers show a bias in reducing words of higher contextual predictability or frequency, even unintentionally; however, the evidence I provide does not dispute these findings (such as those of Gahl, Yao, and Johnson 2012 or Bybee & Hopper 2001). While these phenomena may have some facilitative effect on the listener-oriented model of contextual predictability I propose, they are unable to account for the sound changes in question on their own. The attested phonological patterns of reduction that are

common to function words can occur in my own model even if we assume an equal amount of variation occurring in all words.

1.6. Structure of this dissertation

This dissertation will be organized as follows: Chapters two, three, and four will include the background, methodology, results, and discussion for three sets of experiments which explore various aspects of the questions and hypotheses proposed in this introduction, while chapter five will include the discussion which is relevant to the findings of all three chapters.

In chapter two, I consider how *speech perception* is influenced by semantic predictability, where I hypothesize that words that are more predictable based on the meanings of surrounding words will receive less phonetic attention. I use a discriminability task in which subjects are presented with sentences containing either predictable or unpredictable words based on the preceding context, along with a repetition of the target word, and they are asked to determine if the word was exactly the same or different. Subjects demonstrate better discriminability when listening to unpredictable words, showing they are attending more to the phonetic details of these words. A second experiment tests the effect with predictability based on *subsequent* context, but yields no effect.

In chapter three, I consider how this perceptual phenomenon affects *speech production*. All the experiments in this chapter use the phonetic accommodation paradigm to show the effect of perception on production. A preliminary experiment considers whether listeners accommodate more to words in isolation as opposed to in sentence context. The two main experiments of this chapter find that listeners show a greater degree of imitation or accommodation to words which are unpredictable based on preceding sentence context, however no effect is found from subsequent sentence context. Additionally, I show that the effect is strongest when no instruction to imitate is given, whereas telling subjects to imitate seems to override contextually-modulated phonetic attention.

In the fourth chapter, I investigate whether the results of chapters two and three are also applicable in the case of *syntactically predictable* words which is relevant in answering how context and phonetic attention might affect the reductive sound changes occurring to function words. Here I use an error detection test which reveals listeners pay more attention to syntactically unpredictable content words than they do to syntactically predictable function words, which suggests the phenomenon could be relevant in shaping their phonetic trajectories.

Finally in chapter five I discuss how these findings add to the literature concerning top-down speech perception, dual models of speech perception, and exemplar theory. Additionally, I lay out a proposal for a model of sound change by which variants are selected or ignored due to the influence of contextual predictability. The resulting model accounts for function word reduction and other changes through the consequences of biases in listener perception.

Chapter 2

Speech Perception and Semantic Contextual Predictability

2.1. Introduction

A review of the literature in chapter one has shown that word recognition is achieved through both processing the raw auditory signal as well as higher level linguistic information---knowledge of a language's phonemic inventory, phonotactics, syntax, semantic and pragmatic context, etc. As demonstrated through experiments in phoneme restoration (Warren 1970, Samuel 1981) and fluent restoration (Marslen-Wilson & Welsh 1978), higher level information can even override the auditory signal (or in some way divert the listener's attention from noticing the phonetic details), suggesting that this phenomenon may work in some way to shape the phonetic details of exemplars that are stored in memory. In this chapter I will discuss the results of two experiments which show that (1) listeners attend more closely to the phonetic details (i.e., the auditory signal as a whole) of words that are contextually unpredictable, rather than predictable, from the preceding context while (2) the same difference is not found for words that differ in predictability from the subsequent context. This suggests that listeners are storing finer, and perhaps more precise acoustic details in the exemplars of words that are unpredictable.

Contextual predictability is the likelihood that the listener can determine a particular word based on the syntactic, semantic, and pragmatic content surrounding that word, and is independent of the acoustic cues present. That is to say, a word's predictability in a given context would be the same regardless of whether that word was spoken clearly, severely reduced, or even omitted entirely. Furthermore, we will distinguish between *preceding* contextual predictability and *subsequent* contextual predictability, with two separate experiments in order to compare their effects on the attention to phonetic detail. Table 2.1 shows examples of target words (in bold) that appear in sentences which are both predictable and unpredictable based on both preceding and subsequent contexts.

Chapters 2 and 3 will involve experiments dealing with *semantic* contextual predictability, as opposed to *syntactic* contextual predictability which will be explored in chapter 4. Semantic

Table 2.1: Predictable and unpredictable target words based on preceding and subsequent context

	predictable	unpredictable
preceding	(a) When there is a blackout we light **candles**.	(b) The man is looking at the **candles**.
subsequent	(c) The **candles** melted wax on the birthday cake.	(d) The **candles** are lying on the kitchen floor.

contextual predictability means a word is more predictable because the meaning of word is congruous with the meanings of words preceding and/or following it. More technically, we might say that these surrounding words *prime* the word in question, having mutual associations and shared properties (Collins & Loftus 1975) such that a listener or speaker is likely to think of these words as having a semantic connection. For example, in the sentences given in table 2.1, all display semantic contextual predictability for the target words in bold. In sentence (a) the noun 'blackout' and the verb 'light' will prime 'candles'. In sentence (c), the verb 'melted,' and the nouns 'wax' and 'birthday cake' will prime 'candles,' but in this case these priming words follow the target word, rather than precede it. In (b) or (d), none of the words would have any semantic association with 'candles,' while thousands of other possible nouns would make sense semantically and pragmatically in this position. *Syntactic* contextual predictability, the focus of chapter 4, concerns the fact that different syntactic positions (e.g., head of a determiner phrase vs. the complement) are far more limited in the number of possible lexical items that occur in these positions, which will affect the predictability of a given word.

Previous literature as a whole suggests a hypothesis that listeners will pay more attention to the phonetic details of *unpredictable* speech as opposed to *predictable* speech--- perhaps less processing or storage of these details is needed if the pool of possible competitor words is sufficiently diminished, particularly by the preceding context. Nevertheless, results from past research yield inconsistent findings that fail to yield a conclusive account of the effect of contextual predictability on phonetic attention. For example, Marslen-Wilson and Welsh (1978), using the fluent restoration paradigm, asked subjects to listen to sentences and repeat exactly what they heard. Target words contained errors, such as 'travedy' instead of 'tragedy.' In some cases, fluent restoration occurred, such that subjects shadowed the target words but produced the correct phonemic form without hesitating or replicating the error in their own speech. Fluent restoration was found to be significantly more likely in the 3rd syllable of words as opposed to the 1st syllable. This is considered to be a word-level effect, where the predictability of the 3rd syllable is derived from the fact that the first two syllables have already severely diminished the pool of possible English words (if not diminished to a single candidate in which case word recognition has already occurred). Secondly, they found that listeners were also more likely to restore the correct forms fluently when the target words were predictable from the preceding context. For example, subjects were more likely to notice an error in the word *misfortune* within the sentence "It was his *misfortune* that they were stationary," where the target word is unpredictable, as opposed to the word *cigarette* in the sentence "Still, he wanted to smoke a *cigarette*," where the target word is predictable. While this is also a top-down effect, it is based on the semantic and syntactic context

rather than word-level predictability, and parallels the goals of the current experiments. While Marslen-Wilson & Welsh (1978) found greater restoration for contextually constrained (predictable) words, they found no contextual-based difference in error detection, when subjects were asked to identify words containing errors (but not shadow them in their own speech).

The findings of Samuel (1981) also suggest a complicated situation. His experiments investigated the interaction between attention to the auditory signal and top-down effects by testing the strength of the phoneme restoration effect in a variety of contexts. In this phenomenon, first demonstrated in Warren (1970), a phoneme in a word may be replaced by noise and subjects may be asked to judge whether a particular word was intact; in many cases, subjects fail to notice the missing phoneme, having perceptually restored it. In a general discussion, Samuel (1981) states that "restoration is a function of context; the greater the context, the greater the expectation, the greater the restoration" (481). This is shown to be the case with restoration at the word-level, where more restoration occurred for longer words as opposed to shorter words.

However, in another experiment Samuel (1981) considered sentence-level contextual predictability, the focus of this chapter. Here, subjects were asked whether a noise replaced a sound in a word or if the sound was added to the word. Secondly, they were asked to identify a target word. All target words occurred in minimal pairs such as 'battle' vs. 'batter.' Subjects heard these words in contexts that would bias them towards one or the other, such as 'The soldier's thoughts of the dangerous [batter/battle],' while the auditory signal was actually ambiguous in the replacement stimuli, where a noise would completely replace the minimally different sound (thus [bæɾ*]). In these cases, when subjects responded that the noise was added rather than replaced, phonemic restoration occurred (subjects believed they had heard a phoneme simultaneously with a noise, when there was only noise). Among these instances, subjects were shown to be more biased to hearing the predictable word rather than the unexpected word. While this was the result he expected to find, Samuel also found that subjects showed greater discriminability in determining that a sound was added or replaced when words were *predictable*, rather than unpredictable, suggesting that in some way listeners were paying more attention to the phonetic cues of predictable words. Samuel reasoned that this greater acoustic accuracy suggests that "[w]hen preceding context makes a word predictable, the load on the perceptual system decreases. This apparently leaves more processing capacity available for the fine level of acoustic analysis needed to discriminate added and replacement items" (491).

Some literature has also addressed whether word predictability based on subsequent context should have the same effect as with preceding context. In an article following up on the initial findings on phoneme restoration, Warren & Sherman (1974) found a similar effect of phoneme restoration based on subsequent context. In this set of experiments, listeners heard sentences such as "George waited for the deli[ ]ery of his new color TV," in which a sound was removed and replaced by noise. In this case, listeners would not be able to predict whether a [v] or perhaps a [b] were to follow the initial portion, "George waited for the deli…" since the following word could be 'delivery' or 'deliberations.' However, in these cases, subjects still had difficulty in identifying which sound had been replaced, much as in the earlier studies concerning preceding context. Explaining this phenomenon, they state that subjects "listening to a sentence

may store the auditory input in some partially processed form until confirmation of a particular verbal organization is achieved" (155). This suggests that restoration may occur post-perceptually, as the stored auditory input is altered in the listener's memory after subsequent context is considered. Likewise, Kawashima et al. (1998) found a strong effect of phonemic restoration in Japanese when sentence initial words were predictable based on subsequent context.

Additionally, Connine et al. (1991) considered the effect of subsequent context on word recognition. In their study, they asked subjects to identity the initial sounds in words which had been manipulated to have ambiguous phonetic cues. Each word was part of a minimal pair, such as 'tent' and 'dent,' while the initial alveolar segment was altered to have an ambiguous VOT duration. In two different conditions, biasing context was provided either three syllables after the target word, the 'near' condition, or six syllables later, the 'far' condition. Subjects showed a significantly higher rate of being perceptually influenced by the subsequent biasing context when within three syllables as opposed to six. Their analysis suggests a short window of about one second in which subsequent context will influence the perception of acoustic information. Szostak & Pitt (2013) conducted a similar study but following the phoneme restoration, rather than phoneme identification paradigm. Their results suggest that the window of bias may be more than a second, although determined that the simpler syntactic constructions they used in comparison to that of Connine et al. (1991) may have freed up more processing for the phonetic-semantic interface.

The current study builds on the findings from the phoneme and fluent restoration literature in several ways. First of all, one primary difference will be to investigate the perception of subphonemic acoustic details, rather than the presence or absence of particular phonemes. This will be more relevant for considering how phonetic details are stored in exemplars, and how this is affected by a word's predictability. Because of this, the previous literature simultaneously makes two predictions: while the literature has shown phoneme level restorative effects which are stronger in predictable speech, Samuel's (1981) finding that subjects were better at discriminating whether a sound was added to or replacing a phoneme for predictable words might suggest more attention to subphonemic details in predictable speech since there is more processing capacity available for noticing these details. Additionally, observing a bias in the storage of subphonemic details based on word predictability will be more directly relevant to questions in sound change, particular when phonetically gradual changes take place. Secondly, the current experiment will follow a new methodology that will take into account more recent models of speech perception, in particular Lindblom et al.'s (1995) 'what' and 'how' modes or Hickok and Poeppel's (2004, 2007) neurolinguistic findings, in order to induce a natural manner of processing speech relevant for understanding sound change as it occurs in typical linguistics settings. Lastly, in the discussion and analysis of these results I will consider their applicability not only to speech perception but to linguistic structure and sound change.

2.2. **Experiment #2.1:** Preceding Contextual Predictability and Phonetic Attention

The purpose of the first experiment (#2.1) is to confirm or reject the hypothesis that listeners pay more attention to the phonetic details of words which are unpredictable based on preceding context rather than predictable words. Subjects will listen to sentences containing target

words in predictable or unpredictable contexts, followed by a repetition of the target word which may or may not be phonetically altered. Subjects are then asked to determine whether the repeated word sounded the same as when they heard the word in sentence context.

2.2.1. Methodology

2.2.1.1. Stimuli

Sixty target words were chosen which would occur in either predictable or unpredictable sentence contexts. In order to constrain the phonetic cues that would be the focus of the listener's discrimination, all target words were two-syllable k-initial words with initial stress; furthermore, the initial /k/ always preceded a vowel and did not occur in a consonant cluster. Thus, all target words took the shape /ˈkVC(C)(C)V(C)(C)/ such as 'copper,' 'kitten,' and 'quarter.'

The target words occurred in sentences containing context that made them either predictable or unpredictable based on the preceding context. In these sentences the target word always occurred finally in the sentence (whether predictable or unpredictable). The sentences averaged 10 syllables in length. In order to ensure that listeners would in fact be able to predict the target words based on their preceding contexts, a Cloze test (Taylor 1953) was administered via the online crowdsourcing application Amazon Mechanical Turk. Other Cloze tests administered in a similar fashion have shown comparable results to those obtained within a lab setting (Schnoebelen and Kuperman 2010). Subjects were compensated $2 for a 15 minute task which included all 60 sentences with target words believed to be predictable from the surrounding context. Additionally, 40 filler sentences were included to prevent subjects realizing the /k/ initial phonological pattern of the target words. Participants read each sentence on the screen, with a blank indicating the target word, as shown below in (3):

(3) Kings and queens live in _____.

Subjects were instructed to fill the blank with the first word that came to mind that would be both sensible and grammatical, which they typed in a text box below the sentence. Overall, between nine and 34 of the 34 subjects were able to guess each of the 60 predictable target words (see Appendix B for a full list of the stimuli and their predictability values). While misspellings and word form variations were accepted, synonyms or related words were not included.

All target and filler sentences were read off a computer monitor in a casual style by a native speaker of northern Californian English in his 20s. The sentences were recorded as .wav files at a sampling rate of 22.1 KHz in a sound booth in the UC Berkeley PhonLab using an AKG C3000 microphone.

2.2.1.2. Stimuli manipulation

The listening task involved hearing sentences with target words in either a predictable or unpredictable, followed by a repetition of the target word, either the same or different as when heard in context, at which point subjects were asked to judge whether it sounded exactly the same as when heard in the sentence or not. The sentences themselves were left unmodified except for that fact that the recording of the predictable word in its context was copied and pasted into the

unpredictable counterpart sentence, replacing the original recording of the word. This was done in order to ensure that the same recording of the same word was heard, with only the surrounding context being different (in case the model unknowingly hyperarticulated the unpredictable words for the needs of listener, as theorized in Lindblom et al. 1995, Alyett & Turk 2004, etc.).

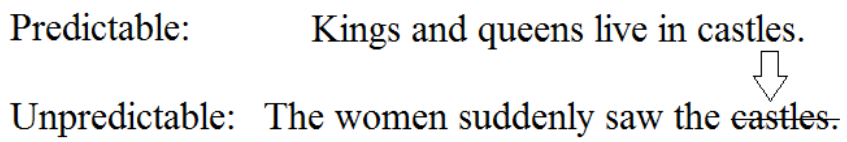

Figure 2.1: Stimuli manipulation, copying predictable target words into unpredictable sentences.

Sentences were presented in their entirety, followed by about one second of silence, 1.5 seconds of static (white noise), and another 0.5 seconds of silence before hearing the target word repeated (thus, an approximately 3 second delay between the end of the target word in sentence context and its repetition in isolation). In half of the trials, the repeated word was left unchanged. In the other half, the target word underwent two slight acoustic modifications: 1) The VOT of the initial /k/ sound was doubled to a minimum duration of 100 ms. Additionally, the pitch of the first syllable was raised by approximately 20 Hz. Combined, these two enhancements gave a sense of stronger prominence to the first syllable. These acoustic manipulations were achieved using the pitch and duration manipulation tools in Praat v. 6.0.14 (Boersma and Weenink 2014). This yielded what were, to the impressions of our team, natural sounding stimuli, lacking any noticeable breaks between manipulated and unmanipulated sections of speech.

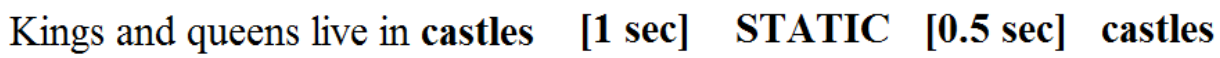

Figure 2.2: Sequence of events heard in each stimulus of discrimination task

2.2.1.3. Procedure and Subject Groups

The experiment consisted of a single block of 80 trials. Of these 80 trials, only 30 were target trials of the sort described above. Of these 30, 15 of the sentences contained target words in predictable environments, while the other 15 featured *different* target words in unpredictable environments. In this way, no subject ever heard a target word more than once during the session, which avoided priming the word and making it potentially more predictable.

The remaining 50 out of 80 trials were fillers of different sorts. Twenty-five of these were phonetic comparison type questions like the target stimuli, however they did not feature /k/ initial words in sentence final position. Because of this, subjects would not be able to focus exclusively on one word since they would not know which word would be repeated. Additionally, 25 other filler questions were included of a different type. For these stimuli, a sentence would be heard

(e.g., “Jupiter is the largest planet in the solar system.”) but no word was repeated; instead, it was followed by a content type question, for example “What is the largest planet in the solar system?” or “What was the fourth word in the preceding sentence?” The purpose of these questions was to attempt to induce a more natural listening style--- such as listening in Lindblom et al.’s (1995) ‘what’ mode. Subjects did not know whether they would be asked a phonetic comparison question or content question until after they heard the sentence; thus they would be disinclined to listen in a less natural ‘how’ mode, or listening for sound rather than meaning. Since our hypothesis asserts that the semantic context may be modulating attention to phonetic detail, it is more likely that this effect would be stronger when listening in the ‘what’ or meaning-focused mode, and listening in the ‘how’ mode might override the phenomenon we are attempting to observe.

The subjects were divided into two larger counterbalanced groups such that one group would hear half of the stimuli (including both predictable and unpredictable target words) while the other group heard the same target words but in reverse predictability environments. That is to say, if group A heard ‘cabins’ in a predictable sentence, and ‘college,’ in an unpredictable sentence, then group A’ heard ‘cabins’ in an unpredictable sentence, and ‘college,’ in a predictable sentence. Because of this, no subject ever heard the same target word twice (to avoid priming it and making it more predictable). Since all 60 target words occurred in both predictability environments (though were heard by different subjects), variables such as frequency and phonological characteristics of the words were kept identical into the two word groups over all subjects (predictable vs. unpredictable).

Additionally, due to the size of the target word list and the length of each run of the experiment, these two larger groups were broken into two smaller groups that each encountered only half of the total stimuli. Thus, of the 120 stimuli (60 target words occurring in contexts that were both predictable *and* unpredictable) each group only heard 30 (15 predictable and 15 unpredictable) for a total of four groups: A and B, and the counterbalanced predictability-reversed A’ and B’ groups. In all there were twenty subjects in each of the four groups for a total of 80.

Table 2.2: Four subject groups

Group A (20 subjects): heard ‘cabins’ in predictable sentence, ‘college’ in unpredictable, etc.	Group B (20 subjects): heard ‘canvas’ in predictable sentence, ‘carpet’ in unpredictable, etc.
Group A’ (20 subjects): heard ‘cabins’ in unpredictable sentence, ‘college’ in predictable, etc.	Group B’ (20 subjects): heard ‘canvas’ in unpredictable sentence, ‘carpet’ in predictable, etc.

2.2.1.4. Equipment and Subject Recruitment

As with the Cloze test, this experiment was conducted online using Amazon Mechanical Turk, which has been shown to be an effective means of running speech perception type

experiments (Yu & Lee 2014). The experiment itself was run using the online survey writing software site SurveyGizmo. Participants were asked to wear headphones and take the experiment in an environment free from distractions and noise. Before officially beginning, the subjects took a short demonstration to ensure that they understood the nature of the task, specifically that they were listening for subtle acoustic differences in the word repetitions, rather than mismatched words or phoneme-level errors.

Subjects were recruited at random using Mechanical Turk, although they were required to be located in the United States and had been granted "masters" status (this being awarded to regular, reliable workers through Mechanical Turk; this requirement was added with the hopes of recruiting a reliable pool of subjects). Subjects provided informed consent and upon completion of the task they were compensated $3 for the approximately 20-25 minute experiment.

2.2.2. Results

2.2.2.1. Statistical Analysis

For each of the 30 target stimuli for each subject, the only data collected was a categorical yes/no answer in response to the question "Does this isolated word sound exactly the same as it did when heard in the sentence?" Additionally, the data included the sequential trial number of each stimuli as it occurred in the experiment (between 1 and 80).

The results were analyzed using the *d'* statistic which measures a subject's discriminability or sensitivity in detecting a particular characteristic. In this case, the measurement shows the subjects ability to detect the similarity of the target word in the sentence as opposed to the target word repeated in isolation. In detecting whether two things are the same or different, subjects have two ways of being correct and two ways of being incorrect. If the subject correctly noticed that the two stimuli were the same, it is a *hit*, whereas if she correctly noticed they were not the same, it is referred to as a *correct rejection*. If the subject did not notice two stimuli were the same, and thought the word sounded different in repetition, this is a *miss*, whereas if the word was different when repeated but the subject did not notice, it is referred to as a *false*. *D'* considers not only the accuracy of being correct but also the response bias: some subjects are prone to misses (perhaps failing to notice the similarity of the stimuli due to the absence of the surrounding words) or false alarms (not listening at a close enough level to reliably detect similarity) such that most of their responses are either "yes" or mostly only "no." Thus, a subject who responded "yes" to all stimuli would have a 100% hit rate (always answering correctly when the stimuli are the same), but also a 100% false alarm rate (hearing similarity when it is not present), in which case he or she would have a particularly low *d'* score.

In this experiment, I seek to determine whether subjects show better discriminability of *unpredictable* target words as opposed to *predictable* ones. In this case, *d'* of both predictable and unpredictable trials was calculated for each of the 80 subjects. A t-test was then conducted in order to determine whether discriminability was better for predictable or unpredictable words over all subjects. This was done as opposed to calculating over all subjects in order to account for the random effect of individual variation among subjects.

Additionally, it should be noted that *d'* values are less accurate for extreme hit and miss rates, when these approach 100% or 0% (Stanislaw & Todorov 1999). This is a somewhat common occurrence in our data since there are only 15 predictable and 15 unpredictable trials for each subject. Thus, we used the loglinear approach as proposed in Hautus (1995). Using this method, 0.5 is added to both the number of hits and false alarms, and a value of 1 is added to both the number of same and different trials (thus, 16).

2.2.2.2. Results: *d'*

The experiment was administered 81 times, with one subject being removed due to missing the content questions consistently, many of which served as attention checks. Each of these 80 subjects was presented with a total of 30 target stimuli in which the target words were heard in context and then repeated in isolation, and were asked to judge whether the repeated word was acoustically identical. Thus there were 2400 total responses with exactly half containing contextually predictable target words, with the other half containing unpredictable words. Of those 2400 responses, there were 535 total errors, for a 22.3% error rate. This is far better than chance (being 50% with just two possible answers, "yes" or "no") which indicates the subjects understood the task but enough errors were induced to analyze whether there was a predictability-based bias.

Of these 535 errors, 291 (54.4%) were made when the target word was predictable, while subjects made just 244 (45.6%) errors when the target word was unpredictable, suggesting a bias towards making more errors with predictable target words. Overall, subjects also showed a higher false alarm rate as opposed to miss rate. Of the 535 errors, there were 313 (58.5%) false alarms (subjects believed they heard similarity that was not present) in comparison to 222 (41.5%) misses (subjects failed to notice the similarity of the stimuli). Within the misses and false alarms, a similar ratio of predictable to unpredictable errors holds. Of the 313 false alarms, 170 (54.3%) were made in response to predictable words, with 143 (45.7%) for unpredictable words. Of the 222 misses, 121 were for predictable words (54.5%) as opposed to 101 for unpredictable target words (45.5%). Thus, subjects showed almost the exactly same pattern of bias for both misses and false alarms.

Table 2.3: Errors: false alarm, miss, and total error rate for predictable and unpredictable stimuli. Percentages indicate the percent among the total number of false alarms, misses, etc. for all predictable or unpredictable stimuli showing more errors for predictable stimuli.

	false alarms	misses	total errors
predictable	170 (54.3%)	121 (54.4%)	291 (54.4%)
unpredictable	143 (45.7%)	101 (45.5%)	244 (45.6%)

Table 2.4: All responses for predictable and unpredictable stimuli

	hits	correct rejections	false alarms	misses
predictable	479	430	170	121
unpredictable	499	457	143	101

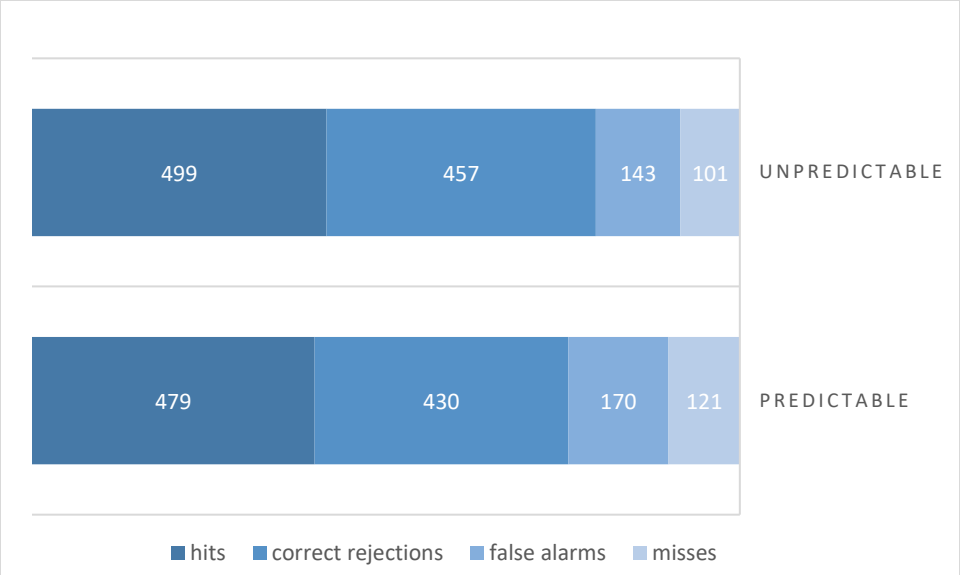

Figure 2.3: Responses according to predictability of stimuli

In order to judge statistical significance, *d'* for both predictable and unpredictable target word discrimination was calculated for all subjects, and a t-test was administered to compare the results. The mean *d'* score for the discrimination of predictable target words was 2.37 compared to 2.54 for unpredictable target words, $p = 0.0165$, indicating that subjects did in fact exhibit better discrimination of the unpredictable words in sentence vs. isolated context as opposed to the same words in predictable contexts.

2.2.2.3. Results: Mixed-effects regression model

In addition to analyzing the data with the *d'* statistic, a mixed-effect regression model was used to consider the effect of both predictability and trial number. The model was run in R using the *lmer()* function within the lme4 package. The response variable was ANSWER, which was either "correct" or "incorrect," depending on whether the subject's judgment of similarity or difference was correct. Predictor variables included PREDICTABILITY (whether the target was "predictable" or "unpredictable"), ORDER (the random trial number with the experiment, included to consider the change in accuracy over time), GROUP (A-D) referring to which subject group the token occurred in, and SAME, which was either "yes" or "no" and referred to the whether or not the target was actually different or not. Random effect intercepts included SUBJECT and TARGETWORD.

As with the *d'* statistic, PREDICTABILITY was significant ($p = 0.012$), with a similar p-value to what was found for *d'*. The mixed effects regression model also found SAME to be a significant predictor ($\chi = 5.4085$, $df = 1$, $p = 0.02004$) of whether or not the subject was correct (ANSWER). This is an indicator of the fact that there were more false alarms than misses (that is to say, when SAME was "yes," the subjects' ANSWER is more likely to be "incorrect"). ORDER was not

significant, which suggests that subjects did not improve or decline in their accuracy of distinguishing the target words in context versus isolation throughout the course of the experiment.

Table 2.5: Model output for effect of predictability (preceding context) on discriminability

Response: ANSWER ~ PREDICTABILITY * GROUP * ORDER * SAME + (1|SUBJECT) + (1|WORD)			
Effect	β	t	p
(Intercept)	1.234	28.876	<0.001***
predictability: unpred	-0.039	-2.513	0.0121*
same: y	-0.077	-2.332	0.0233*
order	-0.0001	-0.335	0.7377
group: B	0.062	1.196	0.2341

2.3. **Experiment #2.2:** Subsequent Contextual Predictability and Phonetic Attention

2.3.1. Methodology

A second version of the experiment, #2.2, was designed and conducted in order to determine whether a similar perceptual effect would occur when words were predictable based on subsequent context. While largely following the previous experiment (#2.1) considering preceding context, practical considerations led to some methodological differences.

2.3.1.1. Stimuli

The same 60 /k/ initial target words were re-used to explore the effect of subsequent context. The carrier sentences, however, were entirely different, though often contained similar content words that would prime the target word. When considering preceding context, the target word was always the last word in the sentence, so there was no concern about controlling for the amount of time or amount of phonetic material between the target word and its repetition. This *is* a concern when dealing with subsequent context, however, since the target words were kept close to the beginning of their carrier sentences. Thus, the predictable and unpredictable sentences pairs always had the same number of syllables preceding and following the target word. The target word always began between the second and fifth syllables of the sentence, and the remaining portion of the sentence was kept to no more than ten additional syllables.

A Cloze test was once again administered via Amazon Mechanical Turk to check the predictability of the words in their contexts. Subjects read the sentences which contained both (limited) preceding context as well as the subsequent context, with the target word blanked out. Because of the decision to keep the same word list, some of the target words proved difficult to make predictable based on subsequent context (the original words in some cases were chosen based on their ability to be predictable based on preceding context). Nevertheless, the stimuli were improved until the average predictability was roughly equal to that of the preceding context stimuli. On average, 21/30 (70%) of participants were able to guess the word in the blank based

on the subsequent context, compared to 73.1% predictability averaged over all subjects for the preceding context sentence list.

The target and filler sentences in this experiment were read by a different model than in the first experiment (#2.1), however the second model was also a native speaker of Californian English in his 20s. The speech style was notably more formal and disconnected than that of the first model. This was due to the unforeseen increase in difficulty in cutting and pasting the target words in a natural sounding manner (see Stimuli manipulation below) when the words occurred in sentence medial position.

2.3.1.2. Stimuli manipulation

The stimuli were manipulated and presented identically to how they were in experiment #1. The predictable version of the word was once again copied and pasted into the unpredictable sentence so the same exact recording of the word would be heard but with different surrounding context. This did prove difficult due to the higher amount of coarticulation inherent with sentence-medial words. However, natural sounding stimuli were achieved by instructing the model speaker to read the sentences in a more formal and moderately less connected style. Once again, the sentences were heard followed by a one second delay, 1.5 seconds of white noise, and another 0.5 seconds of silence before the target word was repeated. Again, the target word was either repeated unaltered from its original form within the sentence, or the VOT was doubled to a minimum of 100 ms and the pitch of the first syllable was raised by 20 Hz.

2.3.1.3. Procedure and Subject Groups

The structure of the second experiment was identical to the first, with a single block of 80 trials, 30 of these being target words, of which half were predictable and the other half were unpredictable. The other 50 trials included 25 other phonetic comparison questions with repeated words in other sentence positions (in this case also initially or finally) and which were not /k/ initial. Another 25 trials included content questions in order to deduce a natural mode of listening.

The 80 subjects were once again divided into four groups of 20, with two groups each receiving half of the word list, with the additional two groups counterbalanced such that they heard the same target words in opposite predictability environments (see previous table 2.2).

2.3.1.4. Equipment

As with the first experiment, the subjects for the second experiment were recruited using Amazon Mechanical Turk (with the requirements that they be located in United States and have Master's qualification) while the survey was run using the survey writing site SurveyGizmo. Participants were asked to use their own headphones and take the experiment in a quiet location. The subjects were compensated $3 for the 20-25 minute experiment.

2.3.2. Results

The second experiment (#2.2), now testing the effect of subsequent contextual predictability on phonetic attention, was conducted with 80 subjects via Amazon Mechanical Turk. Once again, there was a total of 2400 responses for target stimuli (80 subjects * 30 stimuli), half

of these having contextually predictable target words with the other half having contextually unpredictable target words. Of the 2400 responses, there were 633 errors (both misses and false alarms) for a total error rate of 26.4%, slightly higher than that of the first experiment (#2.1) which had a 22.3% error rate. Nevertheless, this still indicates that subjects were correct at a far better rate than chance (50%).

Of these 633 errors, 308 (48.7%) were made in response to predictable target words, compared to 325 errors (51.3%) made when the target words were unpredictable. As with the first experiment (#2.1), there were more false alarms (hearing similarity not present) than misses (failing to notice similarity), with 328 misses and 305 false alarms. Among the misses and false alarms, the percentages of predictable and unpredictable errors remained similar to the total number of errors, with a slight bias towards more errors for unpredictable target words (see table 2.6).

Table 2.6: Experiment #2.2: Miss, false alarm, and total error rate for predictable and unpredictable stimuli. Percentages indicate the percent among the total number of false alarms, misses, etc. for all predictable or unpredictable stimuli showing more errors for predictable stimuli (subsequent context).

	false alarms	misses	total
predictable	148 (48.5%)	160 (48.9%)	308 (48.7%)
unpredictable	157 (51.5%)	168 (51.2%)	325 (51.3%)

Table 2.7: All responses for predictable and unpredictable stimuli (subsequent context)

	hits	correct rejections	false alarms	misses
predictable	439	453	148	160
unpredictable	427	441	157	168

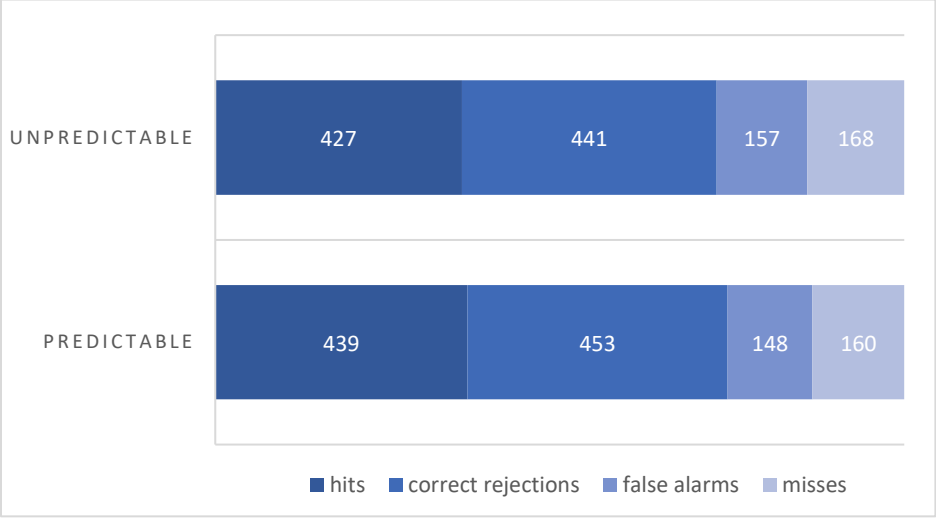

Figure 2.4: Responses according to predictability of stimuli (subsequent context)

The error rate, both in misses and false alarms, is actually higher for unpredictable target words as opposed to predictable ones, which was the opposite result from experiment #2.1, however statistical significance was not achieved. *D'* was once again used to determine statistical significance and was calculated for each of the 80 subjects for both predictable and unpredictable target words. A paired t-test was administered to compare the performance in these two groups. The mean *d'* value for predictable words was 2.28 compared to 2.22 for unpredictable words with a p-value of 0.205, indicating there was no significant difference in behavior in the two groups, suggesting that predictability based on subsequent context did not have the same effect.

The data was also analyzed with a mixed-effects regression model identical to that used in the first experiment (#2.1). The model considered the effect of the predictor variables PREDICTABILITY, GROUP (Subject groups A-D), SAME (whether or not the stimulus was the same when repeated), and ORDER (the random trial number within the experiment) on the response variable ANSWER (whether the subject's answer was correct or incorrect), with random effect variables SUBJECT and TARGETWORD. As with the *d'* statistic, the mixed-effects model showed no significant effect of PREDICTABILITY on whether the subject was more likely to be correct or incorrect in comparing target words when repeated in isolation ($p = 0.368$). This further suggests that subsequent context does not have the same effect on perception as preceding context.

Table 2.8: Model output for effect of predictability (subsequent context) on discriminability

Response: ANSWER ~ PREDICTABILITY * GROUP * ORDER * SAME + (1|SUBJECT) + (1|WORD)			
Effect	β	t	p
(Intercept)	1.246	27.340	<0.001***
predictability: unpred	0.0152	0.900	0.368
same: y	0.0212	0.551	0.584
order	0.0003	0.743	0.457
group: B	-0.0042	-0.340	0.735

2.4. Discussion

The results show that context interacts with speech perception on the phonetic level. When preceding context led listeners to anticipate an upcoming word, they showed a decreased ability to recognize the acoustic details of that word. This aligns well with previous observations in the literature showing a diminished degree of attention to phonemic level details in both fluent restoration (Marslen-Wilson & Welsh 1978) and some of the observations for phonemic restoration (Samuel 1981).

It is not clear why Samuel (1981) found better discrimination of added vs. replaced noises for predictable words in sentence contexts, however this result seems to be an outlier considering the entire body of literature. Samuel surmised that listeners might have more processing capacity available when an upcoming word is predictable, such that they can devote more attention to

acoustic details. One difference is that Samuel's study was concerned with attention at the phoneme level (presence or absence of a phoneme) while the current study considered only sub-phonemic details. However, fluent restoration studies (Marslen-Wilson & Welsh 1978, etc.) also consider the listener's attention to phonemic details, with results aligning with the current findings. One more likely possibility is that the nature of the added/replaced task may have induced phonetic listening (such as Lindblom et al.'s (1995) 'how' mode) as opposed to more a more natural mode of listening for meaning (Lindblom's 'what' mode). My own method of mixing content questions with phonetic comparison questions may have kept the subjects processing sentences for meaning rather than sound--- except in the case of unpredictable speech, where the listener must rely entirely on the auditory signal. Repetition of phonetic tasks may override (or even reverse) the natural mechanisms of listening to speech in context. This could also explain the null result found for

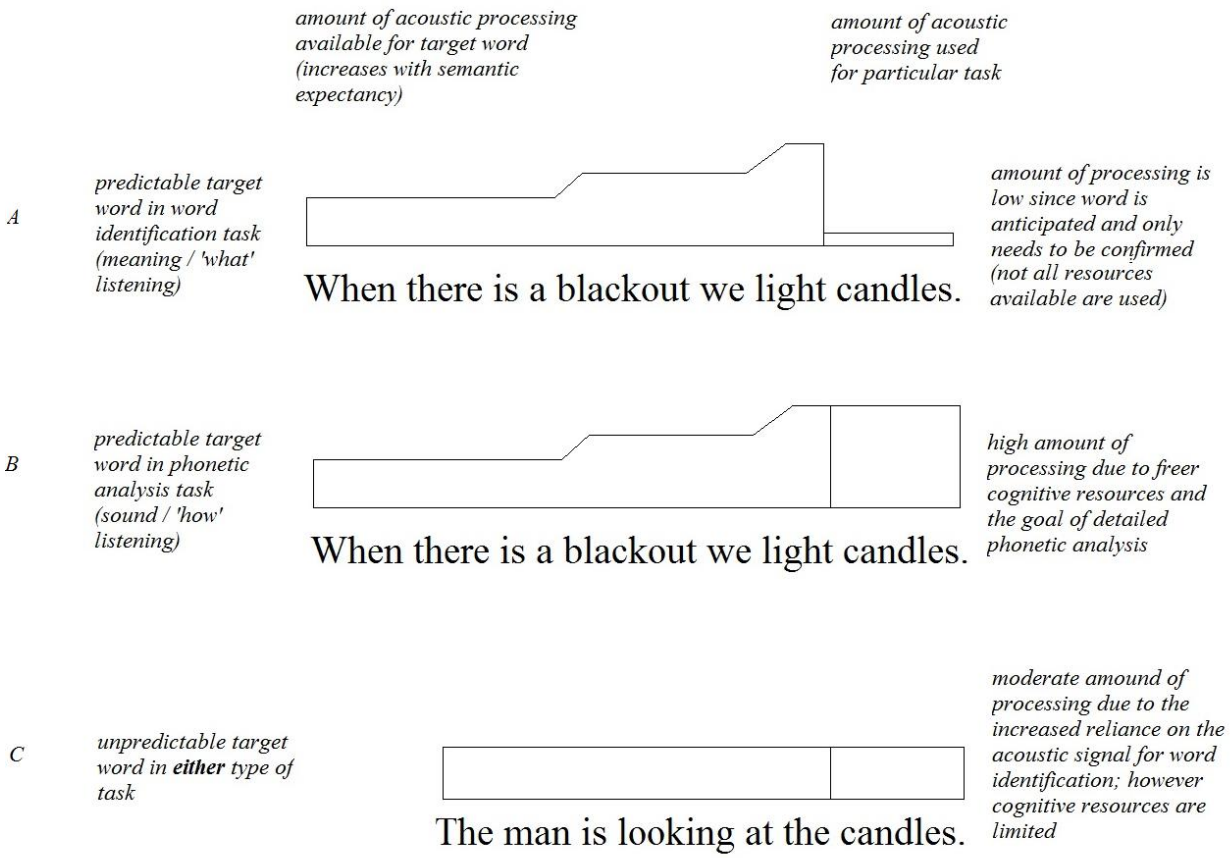

Figure 2.5.: Target word predictability and the cognitive resources available for acoustic processing: *A and B show the difference in processing predictable words depending on the nature of the listening task; While the resources available for acoustic processing are freed up because of the word's predictability may be higher, little is needed for word confirmation (A) as opposed to intentional attention to the auditory signal (B). In contrast, fewer resources are available for unpredictable speech, as word recognition will be more cognitively taxing (C); all of these resources are used however for identifying the word, regardless of the nature of the task.*

predictability modulated differences in Marslen-Wilson & Welsh's (1978) error detection study, where subjects may have been listening more closely to the auditory signal than they would have in a fluent restoration study. Future research is needed to further explore these hypotheses. These possibilities are modeled in the figure 2.5.

Perhaps the most interesting result obtained in the current study is the difference in how preceding and subsequent contextual predictability affected speech perception. While preceding contextual predictability caused listeners to pay less attention to phonetic details, there was no difference in attention to these details when words were predictable based on subsequent context. This result could shed light on whether this effect is *perceptual* or *post-perceptual*. If this phenomenon were perceptual, it would mean there is a difference in the actual sensory input that is stored and processed by the linguistic faculties in the brain, while a post-perceptual phenomenon would involve differences in memory and how the input changes differently after some (even small) amount of time. In other words, if this phenomenon is rooted in perception, it would be the result of the brain in some way priming itself to receive, process, and store fewer phonetic details, and this could only occur when the *preceding* context informs the listener that a word is likely to be predictable; listeners would not be able to determine a word is going to be predictable from the subsequent context until after that context is provided. On the other hand, if this phenomenon were post-perceptual, it could mean that the listener is paying attention to the phonetic details of both predictable and unpredictable speech equally but once receiving strong enough contextual confirmation that the word is in fact what the listener believes it to be, whether based on preceding or subsequent context, that the phonetic details rapidly fade from memory as it becomes unnecessary to store these details once the listener is confident he or she has identified a particular word. Given that this phenomenon only occurred in the first experiment (#2.1), when the target words were predictable based on preceding context, and not when words were predictable based on subsequent context, this suggests the phenomenon is perceptual rather than post-perceptual.

However, there is reason to doubt whether the results are accurately indicating that this process only occurs with preceding context and is thus perceptual, particularly when considering the findings of previous literature. Warren & Sherman (1974), Kawashima et al. (1998), Connine et al. (1991), and Szostak & Pitt (2013) all found evidence of subsequent context affecting the perception of preceding words, much in the way that preceding context might affect subsequent words. Why then did subjects exhibit different behavior in our phonetic comparison task when target words were predictable based on subsequent context?

The first possibility is an additional confounding variable that distinguishes the two experiments. In the first experiment (#2.1), the target word always occurred in sentence-final position and thus was always the last word heard before the pause, static and repetition of the target word. While subjects did not know the last word would be always be the target word, the final word was likely to have maintained a stronger impression as the most recent word that was heard. In contrast, in experiment 2.2, the target word was never sentence-final and there were always words following it, with as many as ten intervening syllables before the pause, static, and repetition of the word. Not only was there more phonetic material before the word was heard again, there would have been a longer duration, both being factors which may have caused the target words to

decay in memory more profoundly than those in experiment 2.1. This still leaves the question of why predictability would not have the same effect after a longer delay (that is, even if the details are diminished for both predictable and unpredictable speech, it still seems more details should remain for unpredictable speech if more were stored to begin with). One possibility is that while contextually unpredictable words result in more phonetically detailed exemplars, those details might fade more rapidly since listeners are unable to abstract the details (relying on abstract representations of other instances of that word or of the particular acoustic features exhibited in that production of the word).

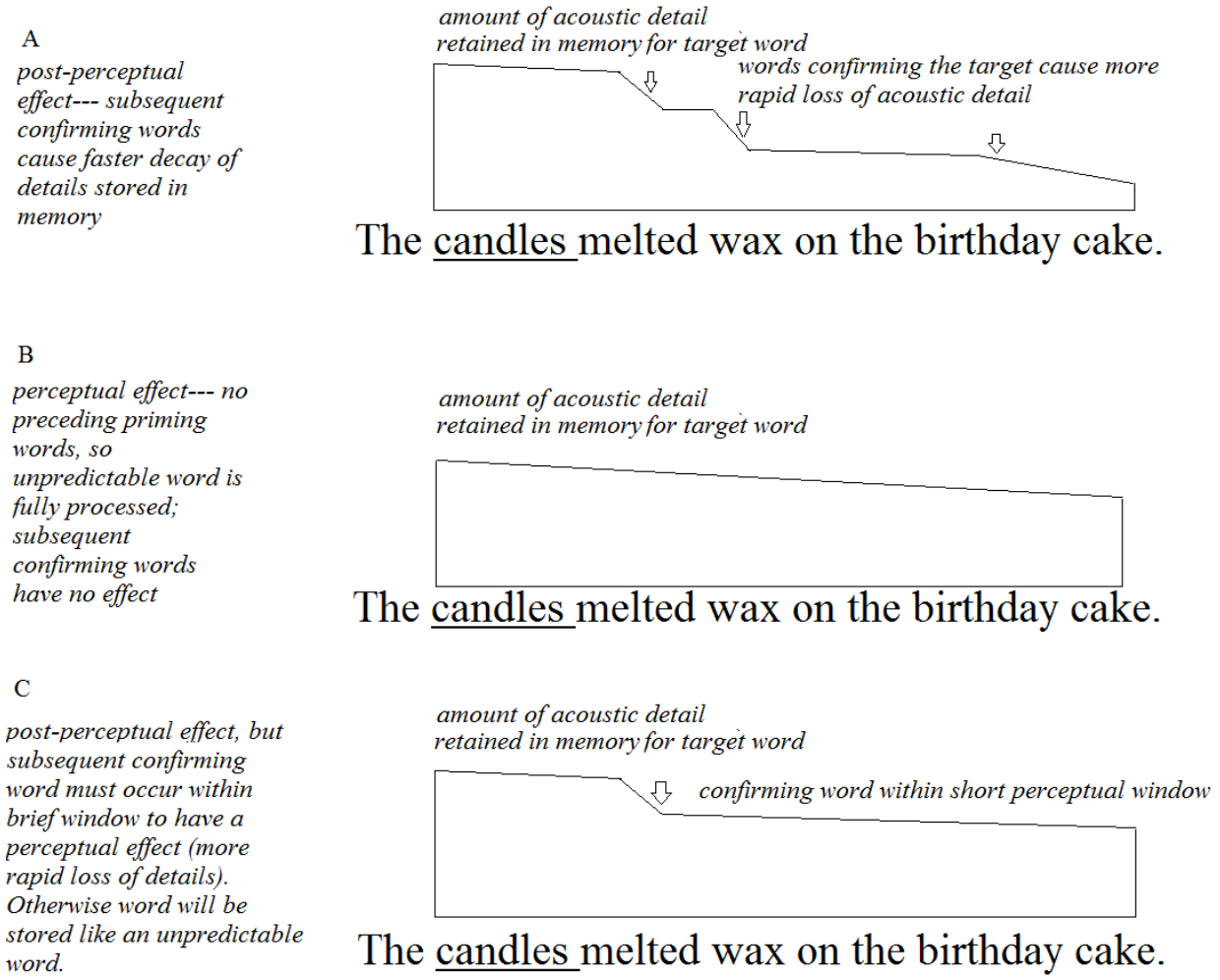

Figure 2.6.: Decay in phonetic detail of words made predictable by subsequent context: *In (A), the effect is post-perceptual where potentially any subsequent word that confirms the target word results in a more rapid loss of the phonetic details of that word (and a move to more abstracted detail). In (B) the effect is perceptual, the amount of details stored is a function of preceding contextual predictability entirely (with some loss gradually occurring with time). In (C), the effect is post-perceptual but the window for this effect is small; Target words lacking a confirming word immediately following them may pattern as unpredictable words would.*

The second possibility is that suggested by the findings of Connine et al. (1991) and Szostak & Pitt (2013). While these studies find perceptual effects on words based on subsequent context, there are limitations to this phenomenon. Connine et al. (1991) specifically find that there is a window of time, perhaps about one second, during which subsequent context will act to confirm the identity of a word and thus leave it susceptible to phoneme misidentification or in the case of Szostak & Pitt (2013), phoneme restoration (though they suggest the window may be longer than one second). In the second experiment (#2.2), the subsequent context that acted to make the preceding word predictable was not kept a constant distance from the target word. Post-hoc analysis of the stimuli in our own study proved difficult to determine precisely where the confirming context begins. For example, in the stimuli 'This castle was built for kings and queens,' it is not clear whether 'built' would confirm the word 'castle,' or if the predictability of the word is more dependent on the priming effect of 'kings' and 'queens,' which would more drastically reduce the number of competing words which might logically occur in place of 'castles.' These various possibilities are modeled in figure 2.6.

These results show that some amount of misperception (or less precise perception and/or storage) has occurred particularly in words that are predictable based on preceding context. This suggests an interface between semantics and phonetics which works to shape the exemplars that are stored in memory. Borrowing details from Pierrehumbert's (2002, 201) hybrid model of speech perception, in which she proposes the existence of both abstract representations and exemplars, might allow for a model which can accommodate the current findings. When context fails to provide clues confirming a word's identity, a listener's acoustic impression of a word may contain more raw acoustic details. For example, in a case in which the listener must rely entirely on the auditory signal, if he hears a potentially ambiguous word that could be 'tie' or 'die,' he may store a fairly precise measurement of the initial consonant's VOT (e.g., '30 ms'). Otherwise, if the context confirms the word can only be 'tie,' the stored "exemplar" would merely contain an abstracted form of the phoneme /t/, perhaps as much based on prior exemplars of /t/ as the production just encountered. If administered a phonetic comparison task such as the one conducted in the current experiments, the subject could at least partially reconstruct the VOT of this instance of the word 'tie' by drawing from either other exemplars of this word or even a purely abstract phoneme /t/.

Despite these findings, there remain additional questions as to how this might affect sound change. While the results show that listeners are storing less precise phonetic details of words that are predictable from preceding context, it is not clear whether these details will in turn affect the production of those words. That is to say, is the difference in perception of predictable and unpredictable speech profound enough to shape the exemplar cloud in a way that results in detectable phonetic drift? This next question will be explored in chapter 3, as I consider how semantic contextual predictability influences not only the perception of speech, but its production, as listeners in turn become speakers.

Chapter 3

Phonetic Accommodation and Semantic Contextual Predictability

3.1. Introduction

The results of chapter two support the literature reviewed in the introduction, showing that listeners pay more attention to words that are unpredictable in context than those that are predictable. This was demonstrated by presenting subjects with the same recordings of the same words in either predictable or unpredictable contexts, repeating the words in isolation, and asking the subjects to determine if they sounded the same or different when repeated. Subjects demonstrated better perception of the phonetic similarities and differences when the words were unpredictable in context. While the effect was shown to be significant when words were predictable based on preceding context, the same was not found when predictability was based on subsequent context. Thus, speech perception is shown to be modulated by contextual predictability, and this likely affects the acoustic details that listeners store in exemplar memories.

In the present chapter, the results concerning semantic contextual predictability and speech perception will be extended in order to understand whether this bias in speech perception results in a bias in speech production. If it is not only the case that listeners' perception is influenced by contextual predictability, but that perception in turn influences production, these findings will become directly applicable to sound change. In this chapter I will detail the results of two larger experiments which consider how context--- generally speaking in the first experiment, and more specifically contextual predictability in the second--- interact with the listener's attention to phonetic details. The experiments use the phonetic accommodation or imitation paradigm (Goldinger 1998, etc.) as a tool to determine what phonetic details subjects perceive after hearing target words in a variety of contexts, and how that affects their production of those words. The first experiment (#3.1) compares the degree of accommodation in isolated phrase vs. sentence context. The second experiment (#3.2) considers how the variable of word predictability within the context of a sentence influences the degree of accommodation, both for preceding and subsequent context. As with chapter two, the results suggest listeners attend more closely to sub-

phonemic details of pronunciation in low-context and low predictability environments, but build on these findings in demonstrating its effect on speech production.

3.2. **Background**: Phonetic Accommodation

Phonetic accommodation, or imitation, is a phenomenon whereby a speakers' pronunciation is subtly, and often subconsciously, influenced by perceiving the phonetic details of other people's speech. The phonetic accommodation paradigm is rooted in exemplar-based theories of speech (Johnson 1997, Pierrehumbert 2002) which propose that individual tokens of words are stored in memory. Recognizing words may be achieved without the need to normalize speech to abstracted, phonological forms of words (though exemplars may be used alongside abstracted forms, as proposed by Pierrehumbert 2002). Goldinger (1996) found that listeners use details of a speaker's voice to aid in speech recognition, rather than discarding these details for the purpose of normalization.

Sancier & Fowler (1997) was an early study which examined the commonly held observation that a person's speech may change after exposure to speech of another dialect or language. Their study observed the speech of a bilingual speaker of Brazilian Portuguese and English after spending prolonged amounts of time in either Brazil or the United States. After hearing the relatively longer VOT (voice onset time) of American English stops for several months, the subject's Brazilian English stops were produced with significantly longer VOT than normal as well. Goldinger (1998) was among the earliest phonetic accommodation studies which attempted to induce the phenomenon in the lab over short amounts of time. Subjects were asked to produce words before and after hearing the same words produced by a model speaker. If speakers were shifting their speech in the direction of the model, it would indicate that subjects were processing and storing the phonetic details, at least for a short amount of time. To judge the similarity of the subjects' productions to that of the model, other subjects were given an AXB task in which they were asked whether a given subject's pre-stimulus (A) production or post-stimulus production (B) was closer to that of the stimulus (X). Goldinger found that speakers did in fact sound more like the model after hearing the model, suggesting they were storing more than just the abstracted word form, even when not given instructions to imitate or give extra attention to the auditory signal. Additionally, Goldinger found that immediate repetition yielded a stronger imitation effect than delayed repetition, suggesting a particular exemplar will become less influential over time. Also, low frequency words are better imitated than higher frequency words, perhaps due to that fact that speakers have more stored exemplars of high frequency words which compete with the stimulus. Finally, Goldinger showed that multiple repetitions of the stimulus also yielded a stronger imitation effect, presumably adding to the listener's pool of exemplars.

The years following Goldinger's study have witnessed a blossoming of phonetic accommodation studies, in both methodology and scope. Goldinger's (1998) study utilized the AXB perception task for determining similarity to the stimulus, which has the strength of being able to account for similarity along any number of dimensions and for any phonetic cues that the judges happen to notice. However, this model has weaknesses in being impressionistic and imprecise in determining just what cues are being imitated by the subjects. Later studies involved lab-induced imitation in which researchers obtained quantitative measurements of specific

linguistic features and were able to pinpoint the features subjects were imitating. Shockley, Sabadini, & Fowler (2004) presented listeners with stimulus words with artificially lengthened VOT, and found that subjects stored this subphonemic information in their exemplars, evidenced by a significant and measurable increase in VOT after hearing the stimulus. Other studies have shown imitation to other phonetic features such as vowel quality (Tilsen 2009, Honorof et al. 2011, Babel 2010, Babel 2012), F0 (Mixdorff et al. 2012, Heath 2014), and vowel nasalization (Zellou et al. 2013).

Other studies of phonetic accommodation investigated various aspects of the mechanisms behind the phenomenon. Pickering & Garrod (2004) proposed the *interactive alignment* account in which speakers automatically "align" at all linguistic levels--- from phonetics to semantics--- for the purpose of facilitating communication in dialogue. Shockley, Sabadini, & Fowler (2004) propose that humans have "a fundamental disposition to imitate" (p. 422) suggesting the process of phonetic accommodation is automatic and occurs without any intention to mimic the speech that we hear. Lewandowski (2012) found that speakers accommodated even when explicitly told *not* to sound like their interlocutors. Nevertheless, others have explored mechanisms that may impede the process. Babel (2010) found that subjects' attitudes towards a particular group of people and their dialect influenced the phenomenon of accommodation. In this case, New Zealanders who held negative opinions of Australians showed less convergence to the model's Australian English, while those holding more positive views showed a greater degree of accommodation. Babel (2012) showed that physical attractiveness may also play a role in phonetic accommodation. Additionally, she found that subjects accommodated more to low rather than high vowels, suggesting specificity in which phonetic characteristics were accommodated. Pardo (2006) found that in conversational pairs, both the gender and conversation role (the one either giving or receiving the instructions) also facilitated or impeded the process of accommodation.

Phonetic accommodation has been shown to be sensitive to and influenced by phonological factors. Nielsen (2011) considered the relationship between exemplars and abstracted representations of sounds. Her study found that hearing target words with lengthened VOT affected the production of other words that had not been heard as part of the stimulus. That is to say, hearing /p/ produced with lengthened VOT in certain words affected other instances of /p/, and even new productions of /k/ (being another voiceless aspirated stop), though to a lesser effect, suggesting the presence of both exemplar and abstracted representations. Such results suggest an extra dimension of complexity in the model not considered in earlier studies of phonetic accommodation.

Furthermore, Nielsen (2011) showed that speakers accommodated lengthened VOT of voiceless stops, but did not accommodate reduced VOT of voiceless stops. The restriction may be rooted in phonological considerations. VOT is a salient phonetic cue which distinguishes voiced and voiceless stops, having relatively shorter and longer VOT respectively. Lengthening the VOT of voiceless stops may help enhance their distinctiveness compared to their voiced stop counterparts. On the other hand, shortening the VOT of voiceless stops causes them to become more like voiced stops, encroaching on the perceptual territory of voiced stop VOT. Mitterer &

Ernestus (2008) also found sensitivity of phonetic accommodation to a language's phonological system. Their study found that subjects would not accommodate to categorical phonological differences such as the type of trill used in Dutch--- alveolar or uvular--- suggesting unintentional accommodation occurs mostly along gradient, phonetic dimensions. Overall, research in the area of phonetic accommodation in the last decade has revealed a number of factors--- social and phonological--- which may facilitate or impede the phenomenon from occurring.

The vast majority of phonetic accommodation studies have only observed the phenomenon in the production of isolated words. Isolated words are devoid of any additional context and their predictability would be a function of their frequency. As a result, few studies have considered how context and predictability, or other top down factors, interact with phonetic accommodation. One exception is Nye & Fowler (2003), which investigated how "order of approximation" to English influenced the degree of accommodation. Sentences of nonce words were constructed that range in how closely they resembled English, primarily in phonotactics. Higher orders of approximation included many words that closely resemble English words, whereas lower orders included words that grossly violated English phonotactic rules (lax vowels in open syllables, words lacking vowels, etc.). The results of the study showed higher degrees of imitation to lower orders of approximation to English. This suggested that speakers drew more heavily from their own linguistic experience whenever possible, indicating an interaction with higher level linguistic knowledge (phonotactics, the lexicon, etc.). This observation harmonizes with Goldinger's findings that subjects accommodate more to words of lower frequency, suggesting the new exemplars have less competition with the older exemplars. In any case, Nye & Fowler (2003) used an AXB perception task to determine the similarity of the pre- and post-stimulus productions to the model. Thus, it is unclear what features and what words were being accommodated within the sentence framework, and whether the structure and semantics supplied by the sentence interacted with the degree of accommodation.

Thus, results from previous phonetic accommodation studies reveal some findings hinting at an interaction between higher level knowledge of the language and the occurrence or degree of imitation observed (particularly related to word frequency). In any case, the literature on phonetic accommodation illustrates something very similar to what is found in general models of speech perception (Cole & Jakimik 1980) as well as phoneme restoration (Warren 1970, Samuel 1981) and fluent restoration (Marslen-Wilson and Welsh 1978). Additionally, recent work in exemplar theory, the theoretical framework on which much of phonetic accommodation is based, has also led to findings that higher level linguistic information can modulate attention to details, affecting what is stored in memory. Goldinger (2007) proposes that "each stored exemplar is actually a product of perceptual input combined with prior knowledge…" (p. 50) suggesting that exemplars are not merely raw acoustic data, but are filtered by a listener's cognitive experience. Goldinger's proposal may be reflected in the findings of Pierrehumbert (2006) as well, which shows that speakers attend more to informative socio-indexical features. Likewise, Maye (2007) demonstrates that a listener's exemplars are shaped by past linguistic experience, where one's L1 has trained the listener to cue in to specific phonetic details. These studies show the relative flexibility of exemplars, which can be molded and shaped by past experience and expectation. Context, in any form, thus allows the possibility for expectancy and predictability of speech, while

words in isolation provide none; without context, listeners are left to rely entirely on the auditory signal and may store more accurate representations of this signal in memory.

Thus, based on findings of chapter two and the literature thus far reviewed, my hypothesis is that listeners will attend to phonetic details more strongly in words as opposed to sentences, and by extension, when words are unpredictable in a given sentence context. Listeners may utilize something along the lines of Lindblom et al.'s (1995) "what" or ventral mode of speech--- attending to meaning--- in the presence of context and higher level information, and "how" or dorsal mode--- attending to the auditory signal--- in its absence, when exemplars may be stored more as raw, unfiltered data. The listening mode may then be modulated by the structural context of the utterance, whether in isolation vs. when within sentences, or the contextual predictability. The experiments in this study address these questions, considering how inducing different listening modes through the presence or absence of context can lead to different degrees of phonetic attention and thus phonetic accommodation.

3.3. **Experiment #3.1**: Structural Context and Phonetic Attention

3.3.1. Purpose and Hypothesis

The first experiment (#3.1) largely follows the design of imitation experiments such as Shockley, Sabadini, and Fowler (2003) and Nielsen (2011), which showed subjects imitated lengthened VOT after hearing the stimulus. However, this experiment compares the effect of hearing the target word within a context-free phrase[1] ('the *pelican'*) as opposed to being embedded in a sentence ('The *pelican* is flying over the beach'). I hypothesize that sentential context, full of higher level syntactic and semantic information, will result in more top-down processing such that listeners will attend to subphonemic details less. Thus, more phonetic accommodation towards the model's VOT is expected when listening to isolated phrases as opposed to those within sentences.

3.3.2. Method

3.3.2.1. Stimuli

All target words in this experiment begin with the target phoneme /p/ (see Appendix A for a full list of the stimuli used in this experiment). All target words had initial stress, were from one to three syllables in length with the first syllable following a CVC(C) structure, avoiding any clusters with /p/. Following the observations of Goldinger (1998), I avoided words of very high frequency in order to encourage phonetic accommodation. The model for the experiment was a female speaker of Australian English. While most of the potential subjects were likely to be speakers of American English, we felt the use of a speaker with some phonetic differences from the subjects might be more likely to induce more robust effects of phonetic accommodation.

[1] Phrases were used instead of isolated words because the experiment was originally designed for investigating whether part of speech is a variable in impeding the carry-over imitation effect (from target to novel words). Thus, the stimuli needed a preceding word to unambiguously mark the part of speech--- such as the verb 'to pin' as opposed to the potential noun 'the pin'.

The target sentences were recorded in a sound booth in the UC Berkeley PhonLab and were read at a natural pace. After recording, the VOT of the initial /p/ sounds of the target words were artificially lengthened to twice their initial length, or to a minimum of 120 ms in the case the tokens did not pass this threshold by doubling their initial VOT. A stable portion of the VOT, beyond the burst but before any hint of voicing, was copied from the original token, and inserted between the burst and onset of voicing. Some care had to be taken to ensure the tokens were natural sounding, which in some cases required revision. Splicing at zero-crossings improved the naturalness of the tokens.

3.3.2.2. Procedure

The experiment consisted of three blocks: (1) baseline reading; (2) immediate shadowing; (3) post-exposure reading, lasting approximately 20 minutes. The experiment was conducted in a sound booth in UC Berkeley's PhonLab, where subjects were recorded with an AKG C3000 microphone (recorded as .wav files at a sampling rate of 22.1 KHz) and heard stimuli via AKG K271 headphones, adjusted to a comfortable volume. The experiment was run on a Lenovo ThinkCentre desktop computer using a template Python script developed for administering imitation and accommodation experiments.

As shown in Table 3.1, the first block, the baseline reading, consisted of 60 two-word phrases, both nouns and verbs, read off the screen in a random order. The noun phrases were of the form *determiner* + *noun* (e.g., 'the pelican,' 'the porch, 'the pantry') while the verb (infinitive) phrases were of the form 'to' + *verb* (e.g., 'to perish,' 'to publish,' etc.). Of the 60 words in the baseline reading, 20 were target /p/ noun phrases that would end up being heard in immediate shadowing block, along with 20 additional "novel" /p/ noun phrases and 20 "novel" verb infinitive phrases which would not be heard in the immediate shadowing block. I used this method in order to replicate the carryover effect found in Nielsen (2011), where the effect of VOT lengthening in the target words was carried over to words that had not been heard as part of the stimulus, but began with the same or similar phoneme. In the second block, the immediate shadowing, participants in two different conditions heard either (A) 20 target sentence stimuli containing the target phrases, played twice in random order for a total of 40 repetitions, or (B) 20 stimuli consisting only of the target noun phrases, also played twice in random order for a total of 40 repetitions. In the case of condition (A), the subjects were told to listen for the /p/ initial word in the sentence, and repeat only that word and the word immediately before it (the determiner). For condition (B), the target phrases heard in the stimulus were extracted from the original recordings of the entire sentences, such that subjects in condition (A) and (B) both heard the exact same recording of the exact same target phrase, but either in isolation or in sentence context. The third block, the post-exposure reading, was identical to the first block but allowed for measuring the effect of the stimulus on the target words. Crucially, the only difference between conditions A and B was the structural context of the stimulus heard in the immediate shadowing block, where the phrase was heard inside a sentence for condition A and in isolation in condition B; the production tasks were the same for both conditions and in all three tasks.

3.3.2.3. Subjects

Condition A consisted of 11 subjects (8 female, 3 male, an additional subject's data being eliminated due to a recording malfunction) while condition B consisted of 12 subjects (9 female, 3 male). Subjects were recruited primarily from UC Berkeley's undergraduate student population

Table 3.1: Experiment #3.1 groups and blocks

condition	baseline reading		immediate shadowing	post-exposure reading	
A	**isolated phrases**	20 target NPs ('the pelican')	heard **sentences** consisting of the 20 target NPs: '*The pelican* is flying over the beach.' repeated **isolated phrases**: '*The pelican*.'	**isolated phrases**	20 target NPs ('the pelican')
		20 "novel" NPs ('a partner')			20 "novel" NPs ('a partner')
		20 "novel" VPs ('to perish')			20 "novel" VPs ('to perish')
B	**isolated phrases**	20 target NPs ('the pelican')	heard **isolated phrases** consisting of the 20 target NPs: '*The pelican'* repeated **isolated phrases**: '*The pelican.'*	**isolated phrases**	20 target NPs ('the pelican')
		20 "novel" NPs ('a partner')			20 "novel" NPs ('a partner')
		20 "novel" VPs ('to perish')			20 "novel" VPs ('to perish')

and were all fluent, native speakers of English with no history of speech or hearing disorders. Subjects provided informed consent and were compensated $5 for their participation.

3.3.3. Baseline to Immediate Shadowing

3.3.3.1. Measurements

VOT was measured using a hybrid method which combined automated VOT measurements with thorough quality control. A python script using the pyalign subprocess (Sprouse & Johnson 2016), utilizing the Penn Phonetics Lab Forced Aligner (Yuan & Liberman 2008), created textgrids for each .wav file for their analysis in Praat, v. 6.0.14 (Boersma & Weenink 2014). Another Python script was used to identify the location of the burst for the /p/ in each target word. A new textgrid was produced for each .wav file including point tiers for the location of the burst and the onset of the vowel. Finally, each textgrid-aligned .wav file was inspected by the researchers using a Praat script for rapid editing of textgrids. Aberrant burst or vowel onset measurements could be corrected in the textgrids by dragging the point tier to the correct location. VOT was calculated as the difference between the burst and vowel onset.

3.3.3.2. Statistical Analysis

In order to quantify the effect of structural context on phonetic accommodation, a mixed-effects regression model was fitted to the data in R using *lmer()* function in the lme4 package. The response variable, RELVOTDIFF, was the difference in VOT divided by word duration from baseline to immediate shadowing, in order to normalize for rate ($vot_{post\text{-}exposure}$ / $worddur_{post\text{-}exposure}$ – $vot_{baseline}$ / $worddur_{baseline}$), as it was observed that subjects tended to speak more quickly as the experiment progressed. SUBJECT and TARGETPHRASE were included as random intercepts in the model in order to account for variation among subjects and in behavior towards specific phrases. The model included the fixed effect CONDITION, which was either A (shadowing sentences) or B (shadowing isolated phrases).

3.3.3.3. Results

Analysis of the results indicates that CONDITION, whether or not the target phrase was heard in sentence context or in isolation, was only significant at a liberal alpha level of 0.1 (for condition B, $\beta = 0.0112$, $t = 1.66$, $p = 0.09687$). Over all subjects in condition A (sentence listeners), VOT decreased by 6.29 ms on average (median value of 6.8 ms) as shown in Table 3.2, with some decrease observed even in the relative duration of VOT to word duration (an average -0.25% decrease). On the contrary, over all subjects in condition B (isolation listeners), VOT showed only the slightest amount of increase (mean = 0.06 ms, median 0.51 ms), with a relative lengthening of VOT to vowel duration (0.87% increase). Thus, for all subjects this suggests a slight decrease in VOT for sentence listeners, and a slight increase in VOT for isolation listeners, although with an only marginally significant difference.

Table 3.2: VOT increase from baseline to shadowing by group

	VOT increase mean	VOT increase median	RelVOT mean (vot/wordlength * 100)	RelVOTdiff median (vot/wordlength * 100)
A (sentence listeners)	-6.29 ms	-6.80 ms	-0.25%	-0.41%
B (isolation listeners)	+0.06 ms	+0.51 ms	+0.87%	+0.85%

A clearer picture emerges when looking at changes by subjects. Figures 3.1 and 3.2 show the individual and group variability in change from baseline to immediate shadowing VOT and VOT:WordLength (relativized for speech rate) respectively. In Figure 3.1, we see that most of subjects in condition A (sentence listeners) had lower VOT in the immediate shadowing block, while most of the subjects in condition B (isolation listeners) had higher VOT in the immediate shadowing block. Figure 3.2, showing the ratios of VOT to word length, which normalized for any change in speech rate, shows a similar situation. In general, more subjects showed an increase in VOT:WL ratio compared to VOT length alone, suggesting subjects were in fact speaking quicker by the second block. While nearly half in condition A (sentence listeners) do show an

increase in relative duration of VOT to word length, Figure 3.2 shows that all but two of the subjects in condition B (isolation listeners) showed an increase in VOT:WL. These two outlier subjects in condition B, curiously enough, showed a greater decrease in VOT and VOT:WL than any subjects in either group. The presence of these outliers greatly decreases the significance of the effect observed in the data, which otherwise shows a regular difference between subjects' degree of accommodation between the two conditions, with greater accommodation when hearing an isolated stimulus as opposed to a stimulus within a sentence. Figure 3.3 shows the spread of the data by subject.

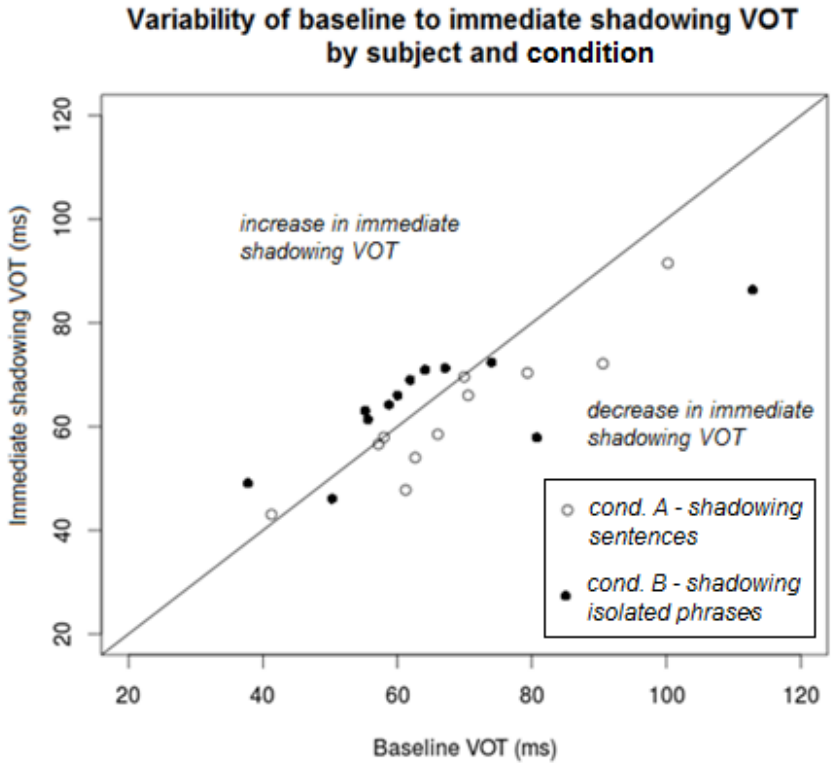

Figure 3.1: Variability of baseline to immediate shadowing VOT by subject and condition: Each dot represents a subjects' mean baseline VOT compared to their mean immediate shadowing VOT. Subjects falling to the left of the line indicate an increase in VOT from the baseline to the immediate shadowing task, while subjects falling to the right of the line indicate a decrease in VOT.

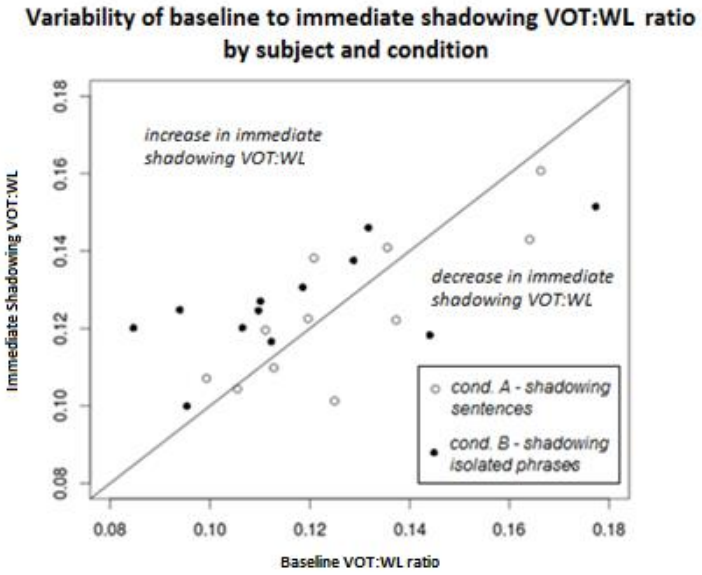

Figure 3.2: Variability of baseline to immediate shadowing VOT to word length ratio by subject and condition: Each dot represents a subjects' mean baseline VOT:WL ratio compared to their mean immediate shadowing VOT:WL ratio. Subjects falling to the left of the line indicate an increase in VOT relative to word length from the baseline to the immediate shadowing task, while subjects falling to the right of the line indicate a decrease in VOT:WL.

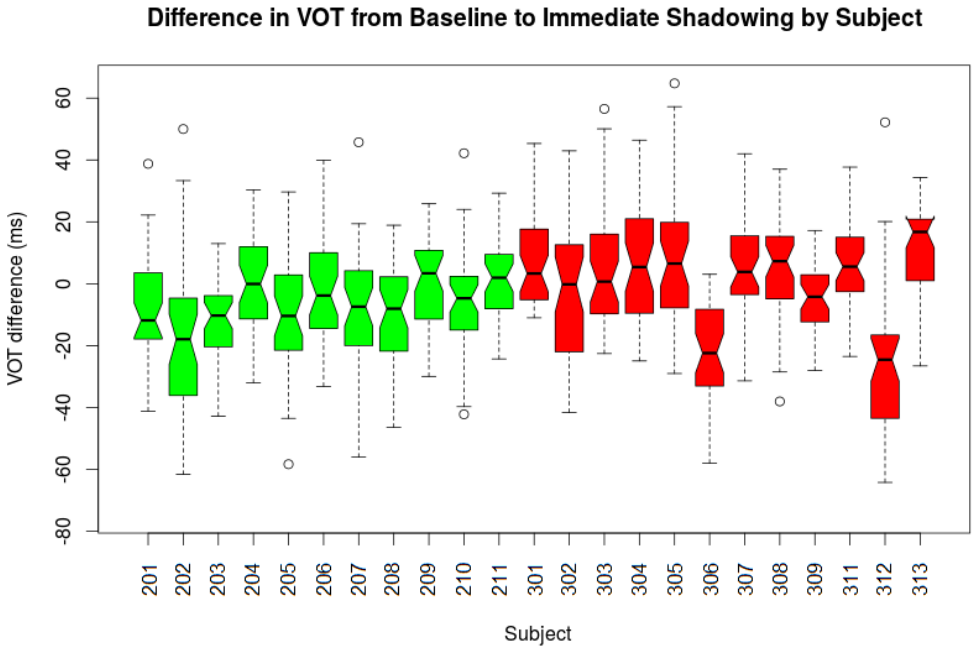

Figure 3.3: Box plot showing spread of VOT difference by subject (Higher difference means more accommodation to the model in the immediate shadowing task). Note the potential outliers, subjects 306 & 312.

3.3.4. Baseline to Post-Exposure

Along with the change in baseline compared to the immediate repetition shadowing, the change in VOT from the baseline to post-exposure reading was also considered. Following the findings of Goldinger (1998) Nielsen (2011), new exemplars may persist even after a shadowing task, and those exemplars may affect other words that share similar sounds. For analyzing this carry-over effect, and whether the difference still held between the two group conditions, a mixed-effects regression model was once again fitted to the data with R's *lmer()* function in the lme4 package. As with before, the response variable RELVOTDIFF was the difference in VOT divided by word duration from baseline to post-exposure, to control for changes in speech rate ($vot_{postexposure}$ / $wl_{postexposure}$ − $vot_{baseline}$ / $wl_{baseline}$). SUBJECT and TARGETPHRASE were included as random intercepts in the model, while CONDITION, A (sentence listeners) or B (isolation listeners) was the fixed effect.

3.3.4.1. Results

Surprisingly, most subjects showed not only a decrease in VOT going from baseline to post-exposure reading, but a decrease in VOT relative to word duration. While a faster speech rate might be expected, it is not clear why the VOT of /p/ was actually shorter with respect to word duration after hearing the stimulus. For all words, condition A (sentence listeners) subjects showed a larger mean decrease of -0.77% in the VOT to word duration ratio, while those in condition B (isolation listeners) showed a decrease of -0.44% in VOT:WL ratio. For just the subset of 20 words heard as the stimuli in the immediate shadowing block, condition A subjects showed a mean VOT:WL decrease of -0.93% compared to a decrease of -0.37% for condition B subjects. While this suggests longer VOT for the isolation listeners, as expected, the mixed-effects model reveals no significant difference between the two conditions (all words: $p = 0.5674$; only stimuli words subset: $p = 0.2958$).

3.3.5. Discussion of Experiment #3.1

The results from the first experiment (#3.1) found marginally significant differences in the degree of phonetic accommodation that occurred among subjects listening to speech within different structural contexts. Subjects hearing a target word within a sentence showed little to no accommodation of lengthened VOT, while subjects hearing the target word within an isolated phrase showed a small degree of accommodation. Two outlier subjects in the isolated phrase listening condition (B) exhibited the strongest decrease in VOT of all subjects, which strongly affected the significance of the effect. Given the regularity of the effect for the rest of the subjects in this condition, the distribution of these subjects may actually be bimodal, while the two outliers represent phonetic divergence from, rather than convergence to the model speaker. Following the findings of Babel (2010), this divergence may have been caused by the attitudes of these two subjects towards Australian speakers such as our model. While I had hoped using a non-local speaker would result in a stronger effect of phonetic accommodation, a speaker of a local dialect may have avoided social impediments of imitation for the subject.

A particularly unexpected occurrence in this experiment was the general decrease in VOT following the baseline reading, both in the immediate shadowing and post-exposure blocks. Some of this decrease was likely due to rate change, as subjects may have become tired throughout the course of the experiment, or were repeating increasingly primed and familiar phrases. However, the results suggest a decrease in VOT:WL ratio, particularly in the final block of the experiment in both conditions. It is possible this was the result of the overuse of the /p/ gesture, which was not dispersed among a variety of other filler phrases beginning with other sounds. While I had anticipated a stronger accommodation effect with a more overt phonetic target (keeping every target word beginning with /p/ with no fillers), it may have resulted in gestural fatigue, obscuring the degree of accommodation.

The results of this experiment leave open the possibility of multiple analyses. One possibility is that the presence of syntactic structure, requiring a certain amount of processing in order to group constituents and assign semantic roles for arguments, etc., resulted in decreased attention to phonetic details. Additionally, the presence of meaningful semantic context may aid the listener in word recognition by supplying top-down information, such that less reliance on the auditory signal is necessary. If either of these is the case, from these results it is not clear that the mere presence of syntactic structure and semantic context has the effect of "flipping a switch" from a mode of listening focused on meaning vs. one focused on sound. Rather, the phonetic attention given to particular words may be a direct function of the amount of context available to that particular word. The present experiment did not control for contextual predictability, with words occurring in a variety of locations within the sentence, with presumably a range in how much context aided in word recognition. For some of the stimuli, the target word was the first noun phrase in the sentence (subject), where for others it was the final (object). Based on the results from the speech perception experiments in chapter two (#2.1 and #2.2), preceding and subsequent contextual predictability may work differently in their interaction with phonetic attention. For some sentences, the target word was fairly predictable ('The old man smokes a pipe') whereas in other cases it was not ('The boy watched the pandas'). The next series of experiments controls for these variables in order to examine the particular phenomenon of contextual predictability.

Another possible analysis is that, rather than being caused by the mere presence of structure or context, that simply having longer segments of speech to listen to decreased the phonetic attention given to a particular feature of a particular target word. Since listeners in condition A heard the target within an entire sentence, there were far more details for them to attend to, while condition B had much shorter utterances to process. Unfortunately, this confounding variable may be impossible to control for, as more syntactic structure will usually result in more phonetic information.

3.4. Experiment #3.2: Phonetic Accommodation and Contextual Predictability

3.4.1. Purpose and hypothesis

The first experiment (#3.1) yielded only marginally significant results in showing a difference in phonetic attention when listening to sentences versus isolated phrases. This may have been due to various errors in design or confounding factors, or may point to the possibility that it is predictability specifically, and not merely the presence of syntactic and semantic content that modulates a listener's attention to phonetic detail. The second experiment (#3.2) considers the effect of semantic contextual predictability and parallels the discriminability experiment in chapter two, using similar stimuli but following the phonetic accommodation paradigm.

Thus, once again we have sentences such as 'When there is a blackout we light *candles*,' in which the final word is more predictable than in the sentence 'The man is looking at the *candles*.' The word 'candles' in the first sentence is primed by 'blackout,' and 'light' immediately before it. As with the discriminability experiments in chapter two (#2.1 and #2.2), two separate versions of the experiment will consider the effects of preceding and subsequent contextual predictability. Thus, in the sentence 'The *candles* melted wax on the birthday cake,' *candles* is predictable but

Table 3.3: Examples of predictable and unpredictable target words based on preceding and subsequent context

	predictable	unpredictable
preceding	(a) When there is a blackout we light **candles**.	(b) The man is looking at the **candles**.
subsequent	(c) The **candles** melted wax on the birthday cake.	(d) The **candles** are lying on the kitchen floor.

only due to the following words, 'melted,' 'wax,' and 'cake.' While the results in chapter two seemed to suggest there was no significant effect of subsequent context on the attention to phonetic details, the phonetic accommodation task will provide more gradient results that might detect small biases in perception, which may help confirm or reject the results of the discrimination task.

Following the observations of the literature, which mostly shows an effect of context diverting attention from phonetic details, I expect more attention to be given to the phonetic details of a word when there is no context to indicate or confirm what the intended sounds should be. A strong effect may indicate that Lindblom et al.'s (1995) modes of listening--- whether focused on meaning or sound--- may be activate or deactivated online in accordance with a given word's predictability. The second experiment utilizes the phenomenon of phonetic accommodation once again to determine what phonetic details the listeners has attended to. Thus, a higher degree of accommodation is expected for unpredictable words, where less accommodation is expected for

words that are predictable based on preceding context. Based on the results of chapter two, no difference in imitation is expected for predictable and unpredictable words based on subsequent context.

3.4.2. Experiment #3.2a: Preceding Context

3.4.2.1. Method

3.4.2.1.1. Stimuli

The target words and the predictable and unpredictable frame sentences are identical to those used in the discrimination experiment (#2.1) in chapter two. All target words began with the phoneme /k/ and were two syllables with initial stress. Target words occurred as the final word in all sentences, such that predictability could be determined entirely from the preceding context. The sentences containing the /k/ initial target words averaged approximately 10 syllables in length. The model speaker for this experiment was a male speaker of American English from northern California. The stimuli were recorded in a sound booth in a casual speaking style.

Each /k/ target word had two features digitally manipulated. The VOT of the /k/ was approximately doubled to a minimum of 100 ms. In addition, the pitch of the first syllable was raised by approximately 20 Hz, overall giving a sense of stronger prominence to the first syllable. For this experiment, acoustic manipulations were achieved using the pitch and duration manipulation tools in Praat v. 6.0.14 (Boersma and Weenink 2014). By our own impressions this was a more successful means of creating natural sounding lengthened VOT than the copy-and-paste technique used in first experiment in this chapter (#3.1, imitation of words in isolation vs. in sentences). Lastly, the manipulated words were copied from their occurrence in the predictable sentences and pasted into the unpredictable contexts. By doing this, listeners would hear the exact same recording of the word (e.g., 'cabins') in either a predictable or unpredictable context. The model speaker may have pronounced the words subtly different based on the context (e.g., hyperarticulated when in an unpredictable context to serve the needs of the listener, as theorized in Lindblom 1995, Alyett & Turk 2004). Thus, this copying technique assured that listeners would not hear a hypo- or hyperarticulated form corresponding with the contextual environment.

3.4.2.1.2. Measuring Predictability

The same target words and sentences were used as in the discrimination experiments in chapter two (#2.1 and #2.2). These were subject to predictability judgments made via an online survey conducted using Amazon Mechanical Turk. Subjects were compensated $2 for the 15 minute survey which included all 60 sentences with target words believed to be predictable from the surrounding context, along with 40 filler sentences to prevent subjects realizing most of the words began with /k/ sounds. Participants read each sentence off the screen, with a blank indicating the target word, as shown below in example (3.1):

(3.1) Kings and queens live in _____.

Subjects were asked to fill in the blank with the first word they thought of, which must be grammatical and would be a word others would likely think of as well. Overall, all 60 target

predictable words were correctly guessed by between nine and 34 out of 34 total subjects, and were always the most commonly guessed words for each sentence (see Appendix B for a full list of the stimuli and their predictability values). As the target words and sentences were identical to those in the discrimination tasks in chapter two, the same responses were used to confirm the predictability of the target stimuli for this experiment.

3.4.2.1.3. Procedure and Subject Groups

The entire experiment was run twice with 20 subjects each. For both experiments, the subjects' task was to repeat sentences presented aurally. In the first version of the experiment, subjects were given no instruction to imitate, as is the practice in the phonetic accommodation paradigm (Goldinger 1998, etc.). In the second version, subjects were told to imitate, or "sound more like the model speaker in some way." Following the findings from the literature and the experiments in chapter two which suggest that higher level contextual information results in a decrease in phonetic attention, it is hypothesized that this effect might be overridden when listeners are asked to tune in phonetic details, and thus predictable and unpredictable speech might show no difference in perception or processing. Normally, the task of listening to sentences can rely on contextual information to store the words in memory, and only abstract forms of words would be necessary to store in memory if context predicts and/or confirms the identity of the target words---- this following Lindblom et al.'s (1995) 'what' mode of listening, which is centered on extracting and processing meaning rather than sound. When asked to remember the phonetic details and reproduce them, this may induce the 'how' mode of listening, where regardless of whether the abstracted word can be remembered and reproduced, the particular exemplars in full acoustic detail need to be stored in order to imitate those details.

Among the 20 subjects in each experiment, 10 subjects formed counterbalanced groups A and A' as shown in Table 3.4. For all groups in the study, the experiment consisted of a single block in which the subjects were asked to shadow (repeat immediately after hearing) 100 sentences, consisting of 60 sentences containing target words. The typical phonetic accommodation design involves a baseline reading followed by the stimulus and ending with a post-exposure reading, in order to assess a subject's change in speech after hearing the stimulus. However, this was impossible in the current experiment since a baseline reading of all the stimuli would quite possibly neutralize the effect of unpredictability, with all the target words being potentially predictable (depending on how well the subject retains the stimuli in memory). In lieu of the baseline reading, measures of accommodation are based on whether subjects' productions of the predictable or unpredictable target words are closer to the model.

Furthermore, of the 60 target words, group A heard 30 of these in a predictable context and 30 in an unpredictable context. Group B heard the same 60 target words, but the 30 that were predictable in group A were heard in an unpredictable context in group B, while the 30 unpredictable words in group A were heard in a predictable context. Counterbalancing the groups was necessary in case there were biases in the words selected in a particular group, such that, for example, the 30 predictable words just happened to be of a significantly different frequency than the unpredictable words, or if there were differences in neighborhood densities, etc. Thus if both groups behaved similarly, it is a good indication of sensitivity to contextual predictability as

opposed to other possible variables. Additionally, this approach meant that no subject heard the same target word twice (in both predictable and unpredictable environments) such that this would potentially prime the target words and affect their predictability when heard the second time (alongside not having a baseline reading, which could also affect predictability). All groups also heard 40 filler stimuli sentences which contained no final word with an initial /k/ sound and were not in any way manipulated, to obscure the prominent final /k/ word pattern at least partially.

Table 3.4: Conditions, groups, and blocks for experiment #3.2a

	No instruction to imitate	Told to imitate
Group A	30 predictable: "The pioneers made log *cabins*."	30 predictable: "The pioneers made log *cabins*."
	30 unpredictable: "The first thing Mary saw was the *coffins*."	30 unpredictable: "The first thing Mary saw was the *coffins*."
	40 fillers	40 fillers
Group A' – counterbalanced (reverse predictability)	30 predictable: "The vampires are sleeping in *coffins*."	30 predictable: "The vampires are sleeping in *coffins*."
	30 unpredictable: "Joe turned and saw the *cabins*."	30 unpredictable: "Joe turned and saw the *cabins*."
	40 fillers	40 fillers

3.4.2.1.4. Subjects

Both experiments had a total of 20 subjects consisting of two counterbalanced groups of 10 subjects. The 20 subjects given no instruction to imitate included 8 males and 12 females, while the 20 subjects told to imitate included 9 males and 11 females. Most of the subjects were UC Berkeley undergraduates. Subjects provided informed consent and were compensated $5 for their participation.

3.4.2.2. Results

3.4.2.2.1. Measurements

A similar method to what was used for measuring the data in the first phonetic accommodation experiment (#3.1) was also used for the second experiment (#3.2). The boundaries for the burst, vowel onset, and end of the vowel in the first syllable were generated and then adjusted in a Praat textgrid, and durations were extracted using a Python script. Pitch measurement was achieved using a Python script using the Entropic Signal Processing System (a package of UNIX environment speech processing and analysis tools) routine get_f0 (Talkin & Lin

1996) which was fed the vowel onset and end values from the textgrid. These boundaries were used as endpoints for measuring the mean pitch of the vowel in the target syllable.

3.4.2.2.2. Statistical Analysis

A mixed-effects regression model was used to determine the effect of contextual predictability on phonetic accommodation. The model was run in R using the *lmer()* function in the lme4 package. Response variables included VOTDIFF which was the difference in VOT between the model and the subject for each token ($vot_{subj} - vot_{model}$), RELVOTDIFF[2] which was the difference in VOT in relation to the vowel duration between the model and subject, thus normalized for rate ($vot_{subj}/vd_{subj} - vot_{model}/vd_{model}$), as well as RELPITCHDIFF which was the difference in pitch between the model and subject, in this case being the pitch of the target syllable (the first syllable of the target words which had lengthened VOT) divided by the average pitch of the entire target utterance ($target.pitch_{subj}/utterance.pitch_{subj} - target.pitch_{model}/utterance.pitch_{model}$). This was necessary given the model speaker was male, and had a lower speaking voice than most of the female subjects. All of these variables involve subtraction of the model's values from the subject's values in order to show whether the subject exceeded (i.e., a positive difference) or fell short of (i.e., a negative difference) the model's performance.

Random effects in the model included SUBJECT and TARGETWORD; only random intercepts were included for both. Models including random slopes of SUBJECT and TARGETWORD for PREDICTABILITY were often degenerate with a correlation between the intercept and slope equal to 1 and did not contribute to the model. Fixed effects included PREDICTABILITY, whether or not the target word was predictable or unpredictable, which is hypothesized to be a significant predictor variable of the degree of imitation of the model; ORDER, the number of the trial within the experiment from 1 to 100, which is included to observe whether there is change in the degree of imitation over the course of the experiment; and GROUP (A or B, with counterbalanced target sentence lists) which was included to control for any differences in behavior with respect to the different sets of stimuli being used. Results for the no instruction to imitate and the told to imitate conditions were analyzed separately.

3.4.2.2.3. VOT: No instruction to imitate

When given no instruction to imitate, the independent variable PREDICTABILITY showed a significant effect for both response variables, VOTDIFF and RELVOTDIFF, (VOTDIFF : $\beta = 0.0049$, $t = 5.444$, $p = <0.0001$, RELVOTDIFF: $\beta = 0.0714$, $t = 4.969$, $p = <0.0001$) with results for all tested predictor variable for RELVOTDIFF shown in Table 3.6. The mean and median of VOTDIFF and RELVOTDIFF according to the predictability conditions can be seen in Table 3.5. VOTDIFF is shown to be greater (further from zero) for predictable words and lower (closer to zero) for unpredictable words. This means there is greater difference from the model in predictable words, and that subjects were closer to the model's VOT when shadowing contextually unpredictable

[2] RelVOTDiff in this experiment normalized VOT to vowel length, rather than word length as in the first imitation experiment, due to technical reasons involving the scripts that extracted this information from the textgrids.

words. The same pattern occurs with RELVOTDIFF where the lower percentages for unpredictable words indicate the subjects were closer in imitating the VOT to vowel duration ratio of the model.

The lack of any ORDER effect suggests subjects are not getting closer to or further from the model's VOT throughout the course of the experiment. The lack of any GROUP effect suggests the two counterbalanced groups (10 subjects each, where the predictability of the target words was reversed) were not significantly different in any way that might have affected the significance of the PREDICTABILITY effect.

3.4.2.2.4. VOT: Told to imitate

When told to imitate, subjects' VOT averaged 14.77 ms closer to the model over all subjects as opposed to when given no instruction to imitate, which is an expected result suggesting the subjects were in fact trying to sound like the model. Unlike when given no instruction to imitate, however, PREDICTABILITY was not a significant predictor of VOTDIFF (unpredictable words, $\beta = 0.0018$, $t = 1.358$, $p = 0.175$) or RELVOTDIFF (unpredictable words, $\beta = 0.0197$, $t = 1.072$, $p = 0.284$). On the contrary, there was an effect of ORDER on both VOTDIFF ($\beta = 0.0001$, $t = 5.601$, $p = <0.001$) and RELVOTDIFF ($\beta = 0.0016$, $t = 5.075$, $p = <0.001$), with the positive coefficients indicating a positive correlation between ORDER and both VOTDIFF and RELVOTDIFF, suggesting the subjects' VOT became more like the model's over the course of the experiment.

A post-hoc exploration of the data, which involved a model containing all interactions of the variables with predictability, did find a significant interaction between PREDICTABILITY and ORDER (for both response variables, VOTDIFF and RELVOTDIFF, $\chi = 4.1961$ df = 1 , p= 0.041 *). This suggests that imitation of predictable versus unpredictable words may change in different ways over time. While future research is needed to confirm this effect, figure 3.5 models the subjects' VOT difference from the model over each quartile of the experiment. This suggests that subjects' VOT became more like the model's over time but only for *unpredictable* target words, where no to little change occurred for *predictable* target words. Looking at only the fourth quartile of the experiment we find a significant difference in VOT imitation for predictable and unpredictable words ($\beta = 0.008$, $t = 2.757$, $p = 0.006$).

Density distributions for both conditions for both predictable and unpredictable words are shown in Figure 3.4.

Table 3.5: VOTDIFF and RELVOTDIFF means and medians by condition and predictability

		NO INSTRUCTION TO IMITATE		TOLD TO IMITATE	
		Predictable words	Unpredictable words	Predictable words	Unpredictable words
VOTDIFF	*mean*	-59.0 ms	-54.1 ms	-42.8 ms	-40.7 ms
	median	-58.6 ms	-52.0 ms	-45.4 ms	-41.9 ms
RELVOTDIFF	*mean*	-56.98%	-50.15%	-45.65%	-43.23%
	median	-55.49 %	-49.09%	-45.76%	-44.08%

Table 3.6: Model output for effects for response variable RELVOTDIFF[3]

Response: RELVOTDIFF ~ PREDICTABILITY + GROUP + ORDER + (PREDICTABILITY*ORDER) + (1|SUBJECT) + (1|WORD)						
	NO INSTRUCTION TO IMITATE			TOLD TO IMITATE		
Effect	β	t	p	β	t	p
(Intercept)	-0.5486	-7.253	<0.0001	-0.5559	-8.276	<0.0001
predictability:unpred	0.0714	4.969	<0.0001	0.0197	1.072	0.284
group:B	0.0030	0.340	0.738	0.0476	0.652	0.523
order	<0.0001	0.075	0.941	0.0016	5.075	<0.0001

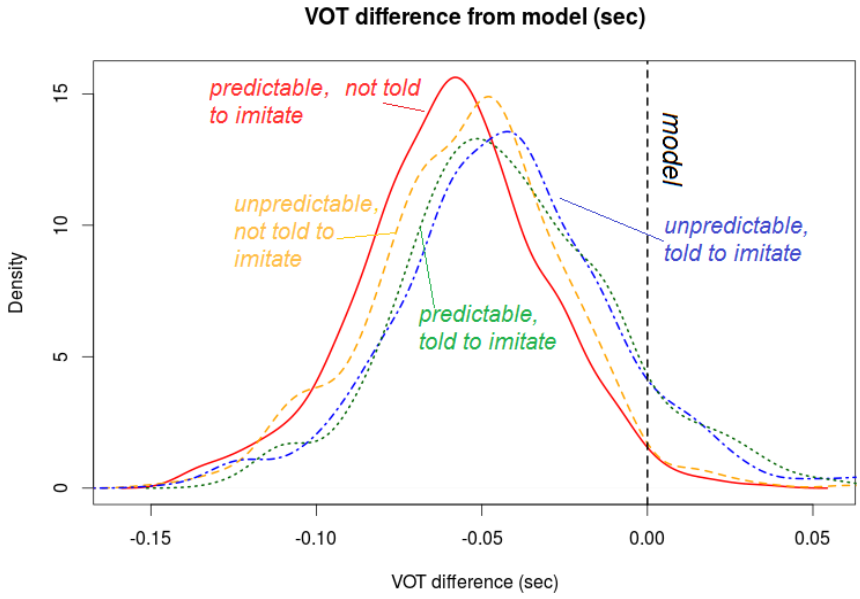

Figure 3.4: Density distributions by condition and predictability

[3] Table is not included for the response variable VOTDIFF, which had the same pattern of significance as RELVOTDIFF.

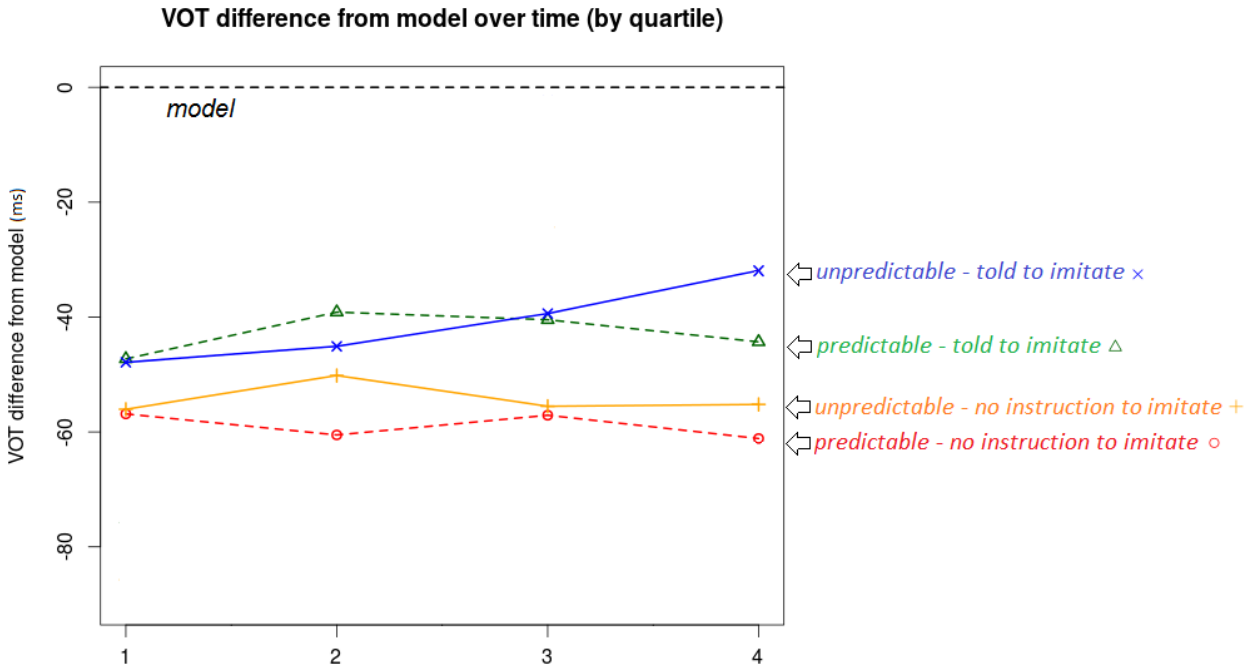

Figure 3.5: VOT difference from the model over time, from the 1st to 4th quartiles of the experiment (#3.2a), in each of the four instructional and predictability environments.

3.4.2.2.5. Pitch

The results for pitch will be analyzed using separate models for men and women. This decision is based on two factors. First of all, gender has been shown to influence accommodation (Namy et al. 2002, Pardo 2006, etc.) and F0 in particular can facilitate accommodation (Babel & Bulatov 2011). Perhaps more importantly, however, is the fact that men and women have different natural pitch ranges and because the model speaker was male, men and women may utilize different strategies for imitating pitch. Men, having a similar range to the male model, may have been imitating absolute pitch while women were imitating relative pitch. A model including both men and women but with GENDER included as an independent variable shows significant effects of GENDER as well as in interactions with other variables (e.g., with no instruction to imitate the effect on RELPITCHDIFF of GENDER, $\chi = 5.6091$ df = 1 p = 0.018 *, or GENDER by ORDER, $\chi = 4.9551$ df = 1 p= 0.026 *). Thus, RELPITCHDIFF will be considered separately for men and women. The statistical model includes the same predictor variables as with VOT---PREDICTABILITY, GROUP, and ORDER, as well as random intercepts for both WORD and SUBJECT.

3.4.2.2.5.1. Pitch: by gender, no instruction to imitate

When analyzing pitch separately for both men and women when given no instruction to imitate, both genders still show a strongly significant effect of PREDICTABILITY on the degree of pitch imitation (Tables 3.8 and 3.9). Table 3.7 shows the mean and median values for RELPITCHDIFF for both men and women. Here, 0% would indicate having a target syllable pitch

to utterance pitch ratio identical to the model, while negative ratios indicate the target pitch was not as high relative to the rest of the utterance as it was for the model speaker. When given no instruction to imitate, both genders show a higher RELPITCHDIFF in unpredictable words (and closer to model), which indicates the target syllable was pronounced with higher pitch. Women actually displayed positive values for RELPITCHDIFF for unpredictable words; this suggests they may have overshot the model's pitch contour for target unpredictable words. Men show a significant effect of ORDER on RELPITCHDIFF, while women do not, suggesting that the men's imitation of pitch became closer to the model's over time.

3.4.2.2.5.2. Pitch: by gender, told to imitate

When told to imitate, PREDICTABILITY is significant only for women, but remains not quite significant for men (tables 3.8 and 3.9). In any case, the difference in imitation between predictable and unpredictable words is greater when not told to imitate, while target syllables of unpredictable words are only slightly closer to the contour displayed by the model (higher, closer to 0%) when told to imitate (for both men and women). Men show no significant effect of ORDER when told to imitate, while women do; this ORDER effect was strongly significant for VOT imitation when told to imitate. Density distributions for men and women's pitch differences from the model for both conditions in predictable and unpredictable words are shown in Figures 3.6 and 3.7 respectively.

Table 3.7: Mean and median values for RELPITCHDIFF by condition and predictability

		NO INSTRUCTION TO IMITATE		TOLD TO IMITATE	
		Predictable words	Unpredictable words	Predictable words	Unpredictable words
RELPITCHDIFF (MEN)	*mean*	-23.82%	-17.00%	-9.06%	-7.18%
	median	-19.83%	-15.83%	-6.37%	-4.64%
RELPITCHDIFF (WOMEN)	*mean*	-3.55%	6.88%	-2.30%	3.50%
	median	-5.01%	9.66%	0.06%	4.15%

Table 3.8: Linear mixed-effects model for response variable RELPITCHDIFF (MEN)

Response: RELPITCHDIFF (MEN)~ PREDICTABILITY * GROUP * ORDER * GENDER + (1|SUBJECT) + (1|WORD)						
	NO INSTRUCTION TO IMITATE			TOLD TO IMITATE		
	β	t	p	β	t	p
(Intercept)	-0.2116	-2.583	0.036*	-0.1189	-3.504	0.0025**
predictability:unpred	0.0509	3.636	0.0003***	-0.0192	1.564	0.1185
group:B	0.0045	0.050	0.9621	0.0272	0.683	0.5165
order	-0.0006	-2.663	0.0081**	0.0003	1.142	0.2541

Table 3.9: Linear mixed-effects model for response variable RELPITCHDIFF (WOMEN)

Response: RELPITCHDIFF (WOMEN)~ PREDICTABILITY * GROUP * ORDER * GENDER + (1|SUBJECT) + (1|WORD)						
	NO INSTRUCTION TO IMITATE			TOLD TO IMITATE		
	β	t	p	β	t	p
(Intercept)	0.0115	0.188	0.8537	-0.0224	-0.340	0.7401
predictability:unpred	0.1053	5.059	<0.0001	0.0534	3.320	0.0009***
group:B	-0.2150	-2.263	0.0472*	-0.0636	-0.763	0.4650
order	0.0005	1.287	0.1985	-0.0007	2.448	0.0147*

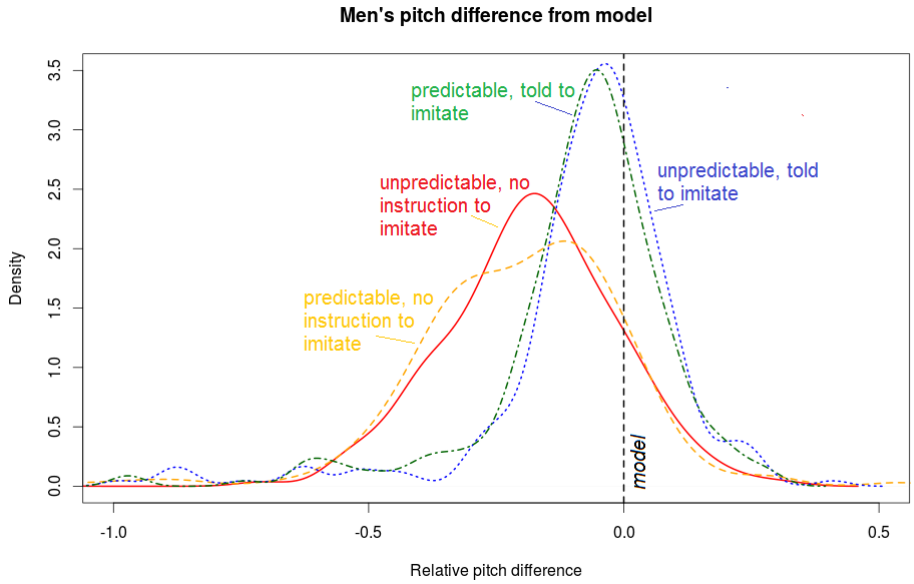

Figure 3.6: Density distributions for men's pitch difference from model by condition and predictability

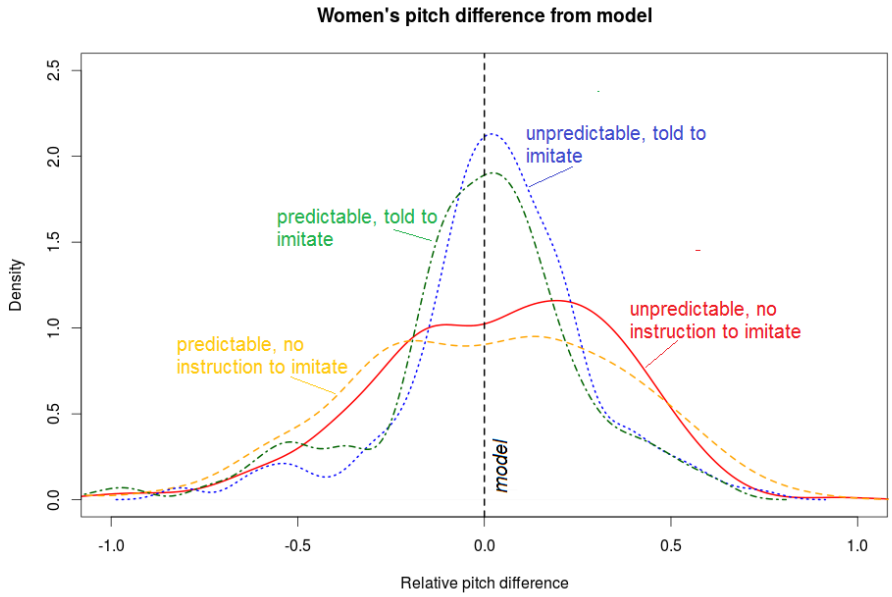

Figure 3.7: Density distributions for women's pitch difference from model by condition and predictability

3.4.2.2.6. Vowel Imitation

One weakness in the design of this experiment is the lack of a baseline and post-exposure reading to confirm that subjects were in fact converging towards the model. Instead, I consider only whether the subjects were closer to the model in their productions of predictable or unpredictable words, and if a significant difference occurs, it can only be the result of convergence towards the model. One counterargument could be that subjects were merely hyperarticulating the unpredictable words and not necessarily showing accommodation at all (following, e.g., Lindblom's (1990) H & H theory). Such hyperarticulation could result in longer VOT and higher pitch for unpredictable words. However, an examination of the data in both conditions reveals there is no significant difference in the difference in vowel duration from the model (no instruction to imitate, $\beta = -0.0008$, $t = -0.606$, $p = 0.545$, told to imitate, $\beta = -0.0007$, $t = -0.450$, $p = 0.6528$) where we would expect hyperarticulation would result in increased vowel duration as well. Only the artificially exaggerated features--- VOT and pitch--- show effects of predictability, supporting the analysis that this difference is due to accommodation and not predictability-modulated hyperarticulation. Thus, despite having no baseline or post-exposure reading, the results establish the fact that convergence is indeed occurring.

3.4.2.3. Predictability and imitation condition

The results suggest differences in the two conditions, whether subjects are given no instruction to imitate or are told explicitly to imitate, particularly in the independent variables PREDICTABILITY and ORDER. Tables 3.10, 3.11, and 3.12 each show the results of mixed-effects

models indicating the differences in these effects as they occurred in the two experimental conditions.

Table 3.10: Linear mixed-effects model for response variable RELVOTDIFF

Response: RELVOTDIFF ~ PREDICTABILITY + ORDER + CONDITION + CONDITION * PREDICTABILITY + CONDITION * ORDER + (1 | SUBJECT) + (1 | WORD)			
	β	t	p
(Intercept)	0.5283	8.364	<0.001***
predictability:unpred	-0.0194	-1.176	0.2398
order	-0.0016	-5.488	<0.001***
conditionB	0.0416	0.590	0.5577
predictabilityunpredictable:conditionB	-0.0524	-2.254	0.0243*
order:conditionB	0.0015	3.765	<0.001***

Table 3.11: Linear mixed-effects model for response variable RELPITCHDIFF(MEN)

Response: RELPITCHDIFF(MEN) ~ PREDICTABILITY + ORDER + CONDITION + CONDITION * PREDICTABILITY + CONDITION * ORDER + (1 | SUBJECT) + (1 | WORD)			
	β	t	p
(Intercept)	0.1079	3.116	0.004**
predictability:unpred	-0.0189	-1.605	0.1087
order	-0.0003	-1.273	0.2032
conditionB	0.0998	2.287	0.0321*
predictabilityunpredictable:conditionB	-0.0274	-1.502	0.1334
order:conditionB	0.0008	2.747	0.0061*

Table 3.12: Linear mixed-effects model for response variable RELPITCHDIFF(WOMEN)

Response: RELPITCHDIFF(WOMEN) ~ PREDICTABILITY + ORDER + CONDITION + CONDITION * PREDICTABILITY + CONDITION * ORDER + (1 | SUBJECT) + (1 | WORD)			
	β	t	p
(Intercept)	0.0553	1.007	0.3211
predictability:unpred	-0.0570	-2.962	0.0031**
order	-0.0006	-1.890	0.0589.
conditionB	0.0032	0.044	0.9648
predictabilityunpredictable:conditionB	0.0552	-2.052	0.0404*
order:conditionB	0.0003	0.608	0.5436

The results of the mixed-effects models confirm that subjects are influenced by PREDICTABILITY and ORDER in different ways in the two instructional conditions. PREDICTABILITY

has a significantly stronger effect on the degree of imitation when given no instruction to imitate as is shown from the significant interaction with experimental condition for RELVOTDIFF and RELPITCHDIFF (women). ORDER, in contrast, is more influential on the degree of accommodation when told to imitate, and is shown in its significant interaction with CONDITION for RELVOTDIFF, RELPITCHDIFF (men) and RELPITCHDIFF(women).

3.4.3. **Experiment #3.2b**: Subsequent Context

3.4.3.1. Method

Paralleling the first version of the experiment (#3.2a) which considered only the effect of preceding contextual predictability on phonetic accommodation, the second version considered the effect of subsequent contextual predictability. While using the same experimental framework and set of target words, there were a few methodological differences that arose due to practical considerations. The same target sentences and recordings were used as in the subsequent context discrimination task in chapter two.

3.4.3.1.1. Stimuli

The stimuli consisted of the same set of 60, /k/ initial target words used in the preceding contextual predictability experiment (#3.2a). The target sentences in which these words were embedded were re-written such that the target words were always unpredictable from the preceding words, but were either predictable or unpredictable based on the following words. Unlike in the preceding context version, where target words were always sentence final, the amount of time and phonetic material between the stimulus and when the subject repeated the sentence needed to be controlled for. Thus, the target words in each predictable-unpredictable sentence always had the same number of syllables before and after the target word. While not an absolute requirement, the target word always occurred near the beginning of the sentence, beginning somewhere between the 2nd and 5th syllables of each sentence, and the remainder of each carrier sentence was no more than ten additional syllables.

As with the preceding context stimuli, a Cloze test was conducted using Amazon Mechanical Turk in order to confirm the predictable target words were in fact judged to be predictable based on the surrounding semantic content. For this test, for each stimuli subjects read the entire sentence off the screen with a blank in place of the target word. Subjects were asked to guess word that would make sense and be grammatical in the blank. Because the decision to test subsequent contextual predictability was made after running the first experiment concerning preceding context, some of the target words proved difficult to make predictable based on subsequent context. Despite this, with some revision based on responses from the survey, the average predictability of all target words was similar to that of the preceding context experiment, with participants guessing the words correctly in 70% of the trials compared to 73.1% for the preceding context stimuli.

Also, due to the availability of the model who read the preceding context stimuli, a second model, also a speaker of Californian English in his 20s, read the subsequent context stimuli. Additionally it should be noted that his style was somewhat more formal than that of the first

model. Because the target words occurred in sentence medial position, rather than sentence final, it was more challenging to copy and paste the stimuli into different sentences (see Stimuli manipulation below) while maintaining natural sounding stimuli. As a result, the linguistically informed model was instructed to speak more formally with less coarticulation.

3.4.3.1.2. Stimuli manipulation

The exact same method was used for digitally manipulating the stimuli as with the preceding context phonetic accommodation experiment (#3.1). The target words as recorded in their predictable context sentences were copied and pasted into their corresponding unpredictable sentences, replacing the original recordings of the words in the unpredictable sentences. This assured that subjects heard the exact same recording of the same word but with different surrounding context which made the word predictable or unpredictable. This process was more difficult with sentence medial words in the subsequent context experiment (#3.2b), and thus the model was instructed to speak in a natural but more formal or disconnected style, to avoid a degree of coarticulation that would remain after digitally replacing words. As with the preceding context experiment (#3.2a) , the VOT of the initial /k/ sounds of the target words was doubled to a minimum of 100 ms and the pitch of the first syllable was raised by about 20 Hz. The resulting stimuli sounded natural according to the judgments of the research team.

3.4.3.1.3. Procedure and Subject Groups

The exact same experimental procedure was followed for the subsequent context version of the experiment (see section 3.4.2.1.3.). Subjects were presented with 100 stimuli, 30 with predictable target words, 30 with different unpredictable target words, and 40 filler sentences. Two counterbalanced groups were presented with half of the stimuli each in order to avoid having subjects hear both predictable and unpredictable stimuli containing the same target words. In all cases, subjects were asked to repeat each sentence as they heard them. In the first of two conditional groups, subjects were given no instruction to imitate, while in the other group they were told explicitly to sound more like the model in some way. In all, there were 20 subjects (14 females, 6 males) given no instruction to imitate (10 in each of the two counterbalanced groups) and 20 subjects (15 females, 5 males) told to imitate (10 in each of the two counterbalanced groups) for a total of 40 subjects. Subjects provided informed consent and were compensated $5.

3.4.3.2. Results

3.4.3.2.1. Measurements

The data was measured precisely the same way as for the preceding context version, where start and stop points were generated and adjusted in Praat, followed by the extraction of duration and pitch information with separate scripts.

3.4.3.2.2. Statistical Analysis

The same mixed-effects regression model was again used to analyze the effect of subsequent context on phonetic accommodation. VOTDIFF indicated the difference in VOT from the model for the target words' initial /k/ sounds, while RELVOTDIFF indicated the difference in

VOT from the model normalized for rate of speech. RELPITCHDIFF indicated the subject's difference in pitch from the model of the first syllable of the target word relative to the mean pitch of the utterance. Negative values indicate undershoot of the particular feature, where subjects failed to achieve as great of VOT to speech rate ratio or pitch of the target syllable to utterance pitch ratio as the model.

Fixed effects in the model again included PREDICTABILITY (whether the target word was in a predictable or unpredictable context), ORDER (the trial number of stimulus within the experiment, out of 100), and GROUP (which of the two counterbalanced groups the subject was in). Random effects included intercepts for SUBJECT and WORD. The two instructional conditions (told to imitate and no instruction to imitate) were again considered separately.

3.4.3.2.3. VOT: No instruction to imitate

When given no instruction to imitate, there was no significant effect of the predictor variable PREDICTABILITY on RELVOTDIFF (unpredictable words, $\beta = -0.041$, $t = -1.92$, $p = 0.055$). While this may be marginally significant, it suggests the subjects were further from the model for unpredictable words. Also, the effect on VOTDIFF was not significant (unpredictable words, $\beta = -0.00001$, $t = 0.011$, $p = 0.992$). Additionally, no effect was found on ORDER ($\beta = -0.0002$, $t = -0.623$, $p = 0.533$) The mean and median values for both response variables, VOTDIFF and RELVOTDIFF, can be seen in the table 3.13. There is no clear pattern suggesting subjects were closer to model in imitating VOT for unpredictable words.

Table 3.13: Mean and median values of VOTDIFF and RELVOTDIFF in both instructional conditions, subsequent context

		NO INSTRUCTION TO IMITATE		TOLD TO IMITATE	
		Predictable words	Unpredictable words	Predictable words	Unpredictable words
VOTDIFF	*mean*	-75.5 ms	-74.0 ms	-57.2 ms	-55.2 ms
	median	-73.7 ms	-74.0 ms	-57.7 ms	-55.3 ms
RELVOTDIFF	*mean*	-60.0%	-61.5%	-54.7%	-53.5%
	median	-52.4 %	-52.4%	-46.8%	-44.5%

3.4.3.2.4. VOT: Told to imitate

As with when given no instruction to imitate, subjects who were told to imitate also showed no effect of word PREDICTABILITY ($\beta = 0.0097$, $t = 0.495$, $p = 0.6204$) on RELVOTDIFF. There was however a significant effect of ORDER ($\beta = 0.0007$, $t = 2.027$, $p = 0.043$*), suggesting subjects' VOT became closer to the model's over time. Mean and median values for can also be seen in table 3.13. The density plot showing VOTDIFF in both instructional conditions can be seen in figure 3.8.

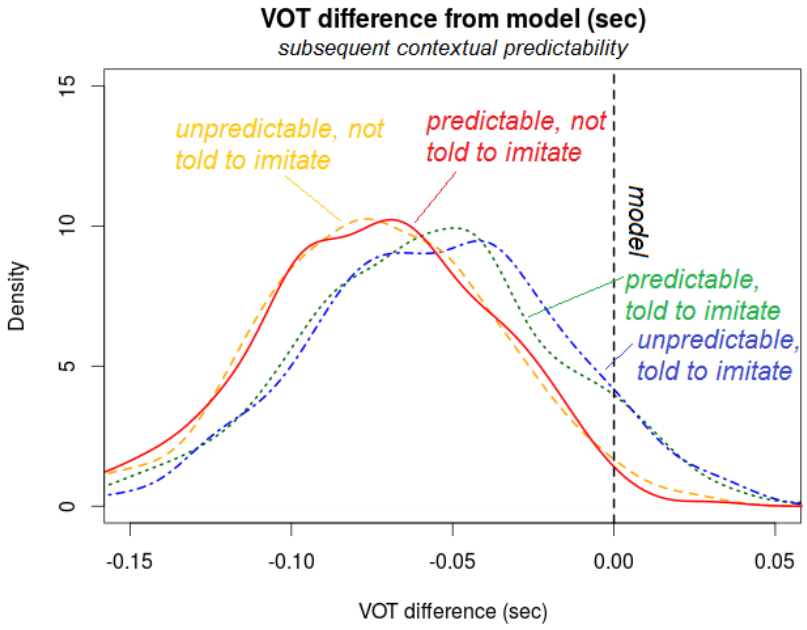

Figure 3.8 Density distributions for VOT difference from model by condition and predictability, subsequent context

3.4.3.2.5. **Pitch:** Men

When given no instruction to imitate, men showed no effect of word PREDICTABILITY (unpredictable words, β = -0.0037, t = -0.433, p = 0.666) on the imitation of pitch (RELPITCHDIFF). There was a significant of ORDER (β = 0.0005, t = 2.925, p = 0.004 **), however.

When told to imitate, PREDICTABILITY again was not significant (unpredictable words, β = 0.00938, t = 0.765, p = 0.445), however ORDER (β = 0.0008, t = 3.320, p = 0.001 **) was. Mean and median values of RELPITCHDIFF in both instructional conditions for predictable and unpredictable target words are shown in table 3.14, with a density plot in figure 3.9.

Table 3.14: Mean and median values of RELPITCHDIFF (men) in both instructional conditions, subsequent context

		NO INSTRUCTION TO IMITATE		TOLD TO IMITATE	
		Predictable words	Unpredictable words	Predictable words	Unpredictable words
RELPITCHDIFF	*mean*	-36.7%	-37.6%	-11.3%	-10.7%
	median	-35.9%	-36.3%	-10.4%	-8.4%

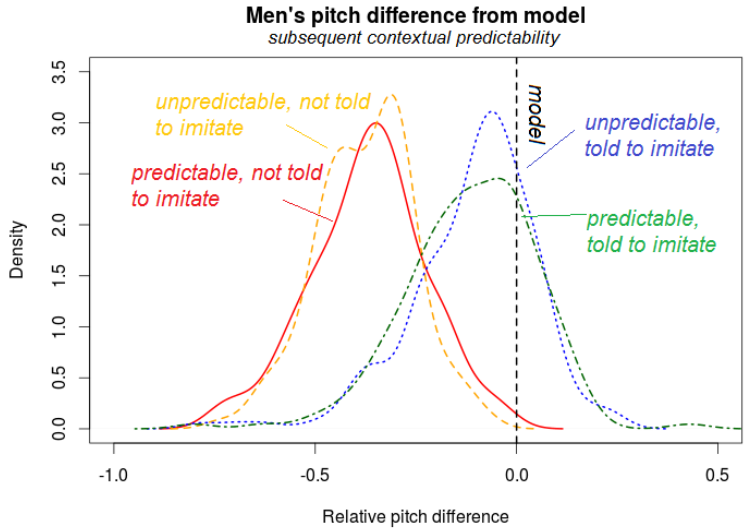

Figure 3.9 Density distributions for men's pitch difference from model by condition and predictability, subsequent context

3.4.3.2.6. **Pitch:** Women

When given no instruction to imitate, women showed no effect of word PREDICTABILITY (unpredictable words, β = 0.0264, t = 1.548, p = 0.122), or ORDER ((β = 0.0006, t = 1.885, p = 0.0598) on RELPITCHDIFF.

When given the instruction to imitate, PREDICTABILITY (unpredictable words, β = 0.0134, t = 0.943, p = 0.346) was once again not significant, however ORDER (β = 0.0005, t = 2.089, p = 0.037 *) was. Mean and median values of RELPITCHDIFF in both instructional conditions for predictable and unpredictable target words are shown in table 3.15, with a density plot in figure 3.10. Surprisingly, women's relative pitch of the target syllables was actually greater than when told to imitate. A similar scenario occurred in the preceding context experiment (#3.2a) but to a lesser extent, which may have reflected the use of a different model who may have spoken with more monotone intonation in the subsequent context experiment (#3.2b).

Table 3.15: Mean and median values of RELPITCHDIFF (men) in both instructional conditions, subsequent context

		NO INSTRUCTION TO IMITATE		TOLD TO IMITATE	
		Predictable words	Unpredictable words	Predictable words	Unpredictable words
RELPITCHDIFF	*mean*	22.8%	25.8%	7.4%	8.6%
	median	24.5%	27.0%	0.3%	2.1%

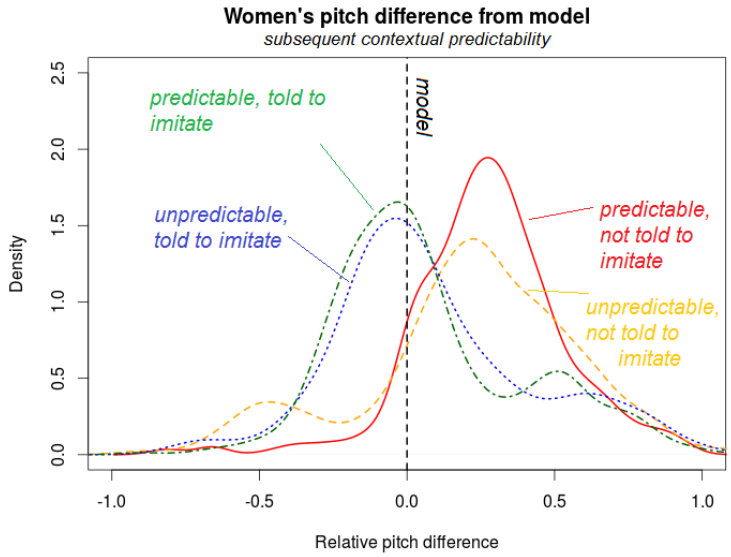

Figure 3.10 Density distributions for women's pitch difference from model by condition and predictability, subsequent context

3.5. Discussion

Experiment #3.1 considered whether stronger imitation would occur when listening to isolated speech as opposed to sentences. The results were marginally significant in suggesting this was the case, however there were several confounding variables (syntax, semantics, total length of sentences vs. isolated phrases) making it unclear what was causing the difference in the two groups. Both parts of the second experiment (#3.2) considered only the effect of the contextual predictability of words within sentences. When given no instruction to imitate, the predictability of the target word based on preceding context was found to be a significant factor in the degree of imitation, where unpredictable words were more closely imitated than their predictable counterparts. In contrast, predictability was *not* significant when told to imitate, however exploratory results suggest a significant effect of a predictability by order interaction when told to imitate, such that only unpredictable words were becoming pronounced more like the model's over time. As for the effect of subsequent context, no significant effects of word predictability were found on the listener's attention to phonetic details. This result mirrors the findings for the discrimination experiments of chapter two (#2.1 and #2.2), in that listeners showed greater attention to words that were predictable from preceding context but no effect was found for subsequent context.

3.5.1. Results from experiment #3.1 and relevance to phonetic accommodation

The results in the first phonetic accommodation experiment (#3.1), comparing imitation of words in isolation vs. words in context, yielded suggestive, but ultimately inconclusive, results. The hypothesis was that perhaps the mere presence of sentential context would in some way de-activate phonetic listening ('how' mode) in favor of semantic listening ('what' mode). However,

a number of variables were not controlled for, so for example, the target words were sometimes predictable and sometimes unpredictable--- and as unpredictable as they would have been in isolation. The comparison of these two experiments--- the first showing results at the threshold of a liberal metric of significance ($p < .1$), as opposed to the second, which investigated the effect of word predictability specifically, which yielded a small but highly significant effect ($p < 0.001$)--- suggests that perhaps that phonetic attention is a function of predictability in particular, whereas the mere presence of complex syntactic and semantic information in need of processing, etc., has less of an effect. Nevertheless, future experiments can address these other contributions of context and how they modulate phonetic attention.

Additionally, the results of both sets of experiments in this chapter add new insight to the phenomenon of phonetic accommodation. As with the findings of Pardo (2006) and Babel (2010, 2012), who found social conditions modulating accommodation, as well as Nye & Fowler (2003) and Nielsen (2011) who found elements of linguistic structure also facilitating or impeding accommodation, the present study finds an interaction between higher level linguistic information and accommodation, in particular contextual predictability. While my findings show that accommodation does in fact occur in words placed in a sentence context, it is not yet clear whether accommodation fails to occur at all in certain conditions, such as when shadowing predictable words. In any case, accommodation does not seem to be a purely automatic process, but interacts in complex ways with social factors and linguistic information at many structural levels.

3.5.2. **Preceding Context:** Not told to imitate vs. Told to imitate

In the phonetic accommodation studies in this chapter, subjects were either given no instructions to imitate or were told explicitly to imitate. The results showed a stronger effect of predictability when no instruction to imitate was given, while the difference in predictability was less profound and non-significant when told to imitate. These findings suggest that when given no instruction to imitate, this likely induced a more natural style of listening to sentences, and given the abundance of syntactic and semantic context, this would suggest speakers could recognize and process the sentences using Lindblom et. al's (1995) 'what' mode, in which speakers focus on extracting meaning as opposed to sound. However, when the context fails to give clues that might help identify a subsequent word, listeners process more phonetic details of these unpredictable words. As a result, listeners store more veridical acoustic details of unpredictable words, which also is reflected in the speech of these listeners turned speakers.

When told to imitate, subjects were closer to the model in VOT and pitch as opposed to when given no instruction to imitate (whether for target predictable or unpredictable words). However, there was no significant effect of predictability. The closeness of the subjects to the model indicates subjects were listening more closely to the details of the auditory signal than they would need simply for identifying and repeating the words in a sentence. This suggests subjects were listening in something like Lindblom et. al's (1995) 'how' mode--- focusing more on the sounds as opposed to extracting meaning. Thus, subjects may have been able to override the more natural 'what' listening mode which may have resulted in a reduced usage of context clues for identifying words--- meaning word predictability biases did not occur. The stronger order effects seen in purposeful imitation suggest that speakers were in fact trying to sound like the model and

became better at doing so after hearing and producing more speech. The predictability by order interaction found in further exploration of the data, which was the only effect of predictability observed in the told-to-imitate condition--- is a more surprising result. While more research is needed to confirm this effect, it is possible that being told to imitate induces listening more in the 'how' mode, and it is possible that over time in this particular experiment that subjects became accustomed to hearing words with initial stress. For predictable words, not only was the word was predictable, but the stress pattern may have become predictable as well. In imitating the predictable words, subjects merely had to make sure the word was in fact the word they anticipated, and that the initial syllable was stressed, and then in their own production applied their own abstracted notion of stress rather than imitating the phonetic details of the particular exemplar. On the other hand, when hearing unpredictable words, the word itself could not be anticipated (even if the stress pattern was predictable). Thus, more thorough phonetic processing would occur in order to identify the word, resulting in more precise imitation of the actual perceived exemplar. Subjects would become no better at anticipating unpredictable words throughout the course of the experiment, unlike with the learnable predictability of the stress pattern, yielding the difference in behavior observed for predictable and unpredictable words over time.

3.5.3. Relevance to Sound Change

The results of the imitation experiments (#3.2a and #3.2b) corroborate and expand on the findings of the discrimination experiments (#2.1 and #2.2) in chapter two. In both sets of experiments it was found that predictability based on preceding context affected the perception of phonetic details, such that less attention was given to the phonetic details of words that were predictable, and more to words that were unpredictable. This suggests that the exemplars of predictable instances of words contain either fewer or less precise details than those of unpredictable words. It is also possible that some of the details stored in the instances of predictable words are actually projected, reconstructed, or even overwritten by abstract representations--- such as the phonemic forms of words or even specific phonemes, and possibly other non-linguistic information such as general vocal features of particular speakers stored from previous experiences with those speakers. The phonetic accommodation experiments in the current chapter, however, expand on these findings by showing that not only is what is stored in memory different for predictable and unpredictable words, but as these details are available for use in speech production, this perceptual bias may also be reflected in speech production. Thus, this strongly indicates that these findings are relevant to sound change, as we might even say our experiment induced a mini-sound over the course of 20 minutes.

Just how then would such a sound change, based on a word's predictability, be implemented? Broadly speaking, the suggestion here is not that particular instances of words, used in conjunction with different words, would somehow undergo lexical splits. In this scenario, we might except to see cases in which, for example, a /p/ > /f/ spirantization might occur in a word such as 'copper' but only when coming after a word like 'penny' which might prime the word 'copper.' Although some elements of Lindblom's H&H theory (1990) might predict similar situations, where more predictable instances of words may undergo some degree of reduction, this is a phenomenon which is claimed to be driven by choices in online speech production in order to

meet the needs of the listener, and not in a way that affects the stored representation of words. For our findings to be relevant to this type of reduction, it would have to be built on a claim that there are separate exemplar clouds for each word *in every possible context*. Thus there would be have to be a distinct exemplar cloud for "productions of 'copper' following the word 'penny'" from which speakers are drawing during speech production, which seems highly unlikely.

Another possibility for how this phenomenon could affect widespread sound change is that every word has its own "global predictability" setting, much like has been suggested for word frequency (Bybee 2001, Phillips 2006) which in turn acts in some way to shape the exemplar cloud, such that high frequency words may (like high predictability words) be given less phonetic attention. Unlike word frequency, however, determining a word's "global predictability" would be quite challenging. Individual words, as demonstrated in all of the stimuli in the phonetic accommodation experiments in this chapter, can usually occur freely in environments that are either contextually predictable or unpredictable. It is possible, however, that a large-scale usage-based study might reveal that some words are actually more likely to occur in predictable contexts than unpredictable contexts, and thus, as a whole listeners are paying less attention to their phonetic details. It is possible that some words, for syntactic reasons like word order, might tend to occur more often at the ends of sentences and thus may be more likely to be predictable, but examples of globally predictable words, based purely on semantic considerations (the likelihood of occurring after words associated with or priming that word) are difficult to imagine. If words do in fact have their own unique "global predictability" in some sense, this attribute may guide the lexical diffusion of a sound change like word frequency.

Drawing from both of these ideas--- the possibility of separate exemplar clouds for words in occurring in conjunction with other words, and the likelihood of a word usually occurring in a predictable environment--- leads to one phenomenon in sound change that is a plausible candidate for being influenced by word predictability. Obscured compounds are words which were historically compounds, derived from two separate words, that have become phonetically eroded such that original parts of the words are no longer recognizable--- in form or meaning (Brugmann 1892). The result is a word that is no longer morphologically analyzable. Examples of these include 'hussy' < 'house wife,' 'gossip,' < 'God's sibb," and 'daisy,' < 'day's eye' (Minkova & Stockwell 2005:274). In some sense there was, phonetically speaking, a split which occurred in the these words depending on what the preceding word was, with 'eye' pronounced as [i] following 'day's' and as [aɪ] elsewhere. Without a doubt, new exemplar clouds eventually developed that were specific to these new words, as the separate elements of 'day's eye' become stored in a single cloud for the resulting word 'daisy' rather than in separate clouds for 'day('s)' and 'eye.' Phillips (2006:47) claims that the development of these obscured compounds is "surely due at least in part because of their high frequency." She adds to this that not all of the examples given in Minkova & Stockwell (2005:274) are common, such as [bosən] < 'boatswain,' but were at least frequent among the specialized group that used them (pg. 213). This explanation still feels somewhat unsatisfying, however, given the unclear application of the concept of word frequency as it may apply to different groups of people in different settings. Perhaps word predictability can explain the other part of this phenomenon. In some sense, these words are "frequent," particularly in the fact that the separate elements of the obscured compounds frequently co-occur; that is to say, the

second element in particular becomes predictable over time. Thus, as the compounds evolve from productive co-occurrences of two words ('a flower that is like a day's eye'), to compounds that take on a specific standard meaning but with analyzable parts ('that flower is called a day's-eye'), the compound may establish its own, independent exemplar cloud, where the version of 'eye' following 'daisy' is more predictable than the isolated version, and becomes subjected to its own set of sound changes that shape the newly independent exemplar cloud in a distinct way.

While these experiments focus on the predictability of words, the literature suggests that predictability would have a similar effect on predictable morphemes within words or even specific parts of words such as segments nearing the ends of words. For example, Marslen-Wilson & Welsh (1978) show that speakers were more likely to notice errors in onsets rather than in codas. Judging from the results of the experiments in this chapter, I would expect to see more precise imitation of root words as opposed to suffixes, as well as of onset segments as opposed to codas. These hypotheses should be further explored in future research. Nevertheless, the extension of these findings may also provide alternate explanations for the adoption of reductive and inventory reducing type sound changes that target suffixes and coda position segments in particular. While coda position within words may receive less phonetic attention, the regularity of changes that often occur there--- such as final devoicing--- may call into question whether predictability could be a factor in driving this change. For example, the codas of some words that are longer or have low neighborhood densities (such as the /t/ in 'alphabet'; there are no other word roots beginning with /ælfəbɛ/) might be more predictable than those in words that are shorter or have higher neighborhood densities (such as the /t/ in 'mat,' where there are many other /mæ/ words such that the /t/ might not be predictable). Thus, an interesting line of future research would be to investigate whether the diffusion of coda reductive type changes correlates with word length or neighborhood density, which could suggest the influence of word predictability.

One crucial aspect of the proposal that word, morpheme, or segment predictability might affect sound change has not yet been addressed. While the experimental findings show lower phonetic attention to words that are predictable, and that this in turn affects production in the adoption of new variants, this still leaves the question of how this difference in attention results in sound change. In fact, we are faced with a paradox: while I have suggested in the current discussion that *predictable* words in compounds may be susceptible to change, it was in fact *unpredictable* words that showed the most change in the imitation experiments. That is to say, by effectively inducing a sound change that disproportionately affected unpredictable words, we might expect quite the reverse case, where predictable words in compounds resist change. This issue is equally important and relevant to case of sound change that occurs in syntactically predictable speech as functional words and morphemes, the subject of chapter four. A model that accounts for the observed patterns in the literature and the experimental results of all three chapters is proposed in detail in chapter 5.

Chapter 4

Syntactic Predictability and Phonetic Attention

4.1. Introduction

The previous two chapters showed experimental evidence that word predictability based on semantic context is influential in directing phonetic attention, shaping a word's exemplar cloud, and guiding mini-sound changes. However, the applications of these lab-induced phonetic alterations to permanent, language-wide sound change is less clear, with word predictability possibly playing a role in lexical diffusion, particularly in the development of obscured compounds. The current chapter will investigate a different type of predictability, which will be referred to as *syntactic predictability*, which concerns the likelihood of a word being determined if its class is determined by the surrounding syntax; if the target word is of a closed class with few members, it will be more predictable than if it were in an open class with many more competing lexical items, even in the complete absence of semantic context to anticipate or confirm a word's identity. This type of predictability is transparently relevant to the function-content word dichotomy, whereby function words are members of closed classes and content words are members of open classes. Furthermore, function and content words, and generally other word and morpheme categories with contrasting syntactic predictabilities, display profoundly different phonological patterns. The result is often what appears to be class-specific phonological erosion, targeting words of high syntactic probability (such as functional words and morphemes). This chapter will investigate whether these class-specific patterns of sound change and erosion may in fact have their origins in a bias in phonetic attention that results from the categorically distinct syntactic predictability inherent to classes such as function and content words.

4.1.1. Syntactic predictability

The type of predictability which was the central focus of chapters 2 and 3 was predictability determined by the semantic congruency of surrounding words, such that the target word is primed by close association of these surrounding words. In example (4a), reprinted from chapter 2, we see that the target word, 'candles,' is primed by the words 'blackout' and 'light.' While there may be other logical words that could maintain semantic congruency in this context (e.g. 'fires'),

'candles' and other words related to light are primed, and as they are anticipated by the listener, are predictable. While other words might be grammatical ('baseballs,' 'grasshoppers,' etc.) they would result in nonsensical utterances and would not be primed by the semantic context, and would thus be unpredictable.

4a) When there is a blackout, we light **candles**.

4b) The man is looking at the **candles**.

In contrast, syntactically predictable words are not necessarily primed through semantic association with other words in context. Many word categories may be predictable within sentences based on the co-occurrence of words of other categories in accordance with a language's phrase structure rules. For example, in terms of linear syntactic structure, a determiner is usually predictable by the eventual presence of a noun, and in reverse, a noun is usually predictable by the preceding presence of a determiner. In terms of hierarchical syntactic structure, the presence of a determiner phrase will imply the presence of a determiner as the head of the phrase, and of a noun (and its modifiers) as the complement of the phrase. Consider the examples in 5a and 5b, in which the only possible lexical item that can fit in the blank in 5a is a noun, and in 5b, a determiner.

5a) Joe turned and saw the ____.

5b) Joe turned and saw __ cabin.

In both of these examples, the category of the word in the blank is predictable. As this narrows down the number of possible lexical competitors to include only members of that category, let us now consider the set of possible words that could go in either of these sentences. In (5a), without any additional context, possible words might include {chair, door, car, cat, hippopotamus, cloud, water…} with an almost endless number of possibilities, spanning most members of the noun category (except perhaps for various abstract nouns such as 'idea' or 'happiness,' etc., which are actually *semantically* unpredictable in this case). In contrast, the set of possible words that could fit in the determiner slot in 5b is limited to about twenty members, such as articles {the, a, an}, demonstratives {this, that}, a handful of possessives {my, your, our, etc.}, quantifiers {some, each}, and numbers {one}. Technically, there is nothing about the semantic context of 5b that is limiting the possible lexical candidates--- 'the cabin' would be as plausible as 'one cabin' or 'our cabin.' However, the odds of correctly guessing the word in the blank in 5b) might be around 1 in 20 compared to 1 in a million (or infinity if we consider the productive nature of nouns in English) in example 5a. Thus, we consider the predictable nature of the target word in 5b) to be syntactic rather than semantic.

4.1.2. Syntactic Predictability and Function vs. Content Word/Morpheme Classes

Of the dozen or so word categories in English these can be grouped into two broad classes referred to as function and content words. For the purposes of this study, the single linguistic criterion used for distinguishing which class a particular word category should fall under is that of productivity, whether a word category is open or closed, which closely aligns with the total number of members of a particular word class. In English, new nouns, verbs, adjectives, and adverbs can and are regularly created as novel lexemes in the language, and these classes represent minimally

thousands of lexical items each. On the other hand, new function words cannot easily be created, only evolving slowly throughout time, and each of these classes contains only dozens of members each. While there are a number of other criteria for organizing word categories into classes, such as the degree and type of reduction (Altenberg 1987), the criterion of productivity is the most relevant characteristic for syntactic predictability since it correlates well with the number of lexical items in a given category--- that is, open class word categories tend to have many members, whereas closed class words tend to have few members.

Table 4.1: Content and function word categories

Content words (open)	Function words (closed)
nouns	determiners
verbs	pronouns
adjectives	auxiliary verbs
adverbs	prepositions
	conjunctions

Additionally, it should be noted that the phenomena to be examined in this chapter are almost certainly just as relevant for function vs. content *morphemes* in addition to words. That is to say, for example, some morphological (as opposed to syntactic) positions within words may be occupied by closed, functional morphemes with a limited number of possible members that could grammatically fill a given position. For example, the derivational morphemes that could act as prefixes or suffixes for a given verb root in English would be limited to a set of a few dozen individual morphemes (re-, un-, over-, under-, pre-, etc.) , and likewise syntactic (or in that case, morphological) predictability would be relevant in these cases as well. Thus, the word *morphosyntactic predictability* may also be used in this more general sense, however the experiment in this chapter will deal exclusively with the syntactic predictability of word classes.

4.1.3. **Syntactic Predictability:** Phonological Patterning and Cognitive Behavior

Previous literature is rife with examples of distinct phonological patterns for word categories of contrastive syntactic predictabilities such as function vs. content word classes and the affix vs. root distinction. Meanwhile, psycholinguistic and neurological studies have mirrored these findings showing distinct behaviors exhibited by subjects in both the perception and production of similar word classes of differing syntactic predictability. The nature of these differences in many ways suggests behavior consistent with that described in both the literature and experimental findings of the previous two chapters concerning the relationship between semantic predictability and phonetic attention, which builds a hypothesis that syntactic predictability, leading to less attention to phonetic details for predictable words, may be one root cause of much of the behavioral and phonological duality between these word classes.

4.1.3.1. Phonological Duality and Syntactic Predictability

Among the world's languages, word categories of significantly different syntactic predictabilities, which can often be grouped into the larger function vs. content, open vs. closed,

or affix vs. root classes, often display differences in phonological patterns that reflect divergent outcomes in historical changes through either regular sound change or lexical diffusion. Most often these class-specific subphonologies are characterized by a reduction of allowable contrasts, vowel reduction, and lenition for function words (i.e., categories with relatively higher syntactic predictability) and the maintenance of contrasts and a resistance to lenition for content words (categories with lower syntactic predictability).

The example which will perhaps be the most familiar to the reader is that of the development of the voiced interdental fricative /ð/ in word initial position in English. Historically, English had no voiced fricatives, but borrowings from French led to the adoption of /v/ and /ʒ/ and these as well as /ð/ resulted from regular sound change of voiceless fricatives becoming voiced in intervocalic position. However, in word initial position, /θ/ weakened to /ð/ but only for several function word categories; not a single noun, verb, or adjective in English begins with the /ð/ sound. The weakening to /ð/ occurred in pronouns ('thou,' 'thy'), articles ('the'), demonstratives ('this,' 'that'), and functional adverbs ('then,' 'though'), whereas nouns ('thing, 'thistle,' 'thorn'), verbs ('think,' 'throw'), adjectives ('thick,' 'thin'), content adverbs ('thickly,' 'thinly') and prepositions ('through') remained resistant. Prepositions are often, though not always, described as function words, and they will be treated as such in this chapter based on their relatively high syntactic predictability. In any case, it holds that this lenitive sound change targeted only lexical items of functional classes, particularly among the categories with the highest syntactic predictability (i.e., there are more prepositions than articles in English).

A fairly similar case occurred in the development of voiced fricatives from voiceless ones in many Athabascan languages, such that lenition targeted morphemes of high morphosyntactic predictability. In the Hän language of eastern Alaska, for example, voiced fricatives developed from voiceless ones in syllable onset position only in the conjugation and qualifier prefix zones within the verb template, regardless of the phonological environment (whether word initial or intervocalic) (Manker 2015). Meanwhile, voiceless fricatives remained as such in the disjunct and pronominal prefix zones. The difference between conjugation and qualifier prefixes compared to disjunct and pronominal prefixes lies in the morphosyntactic predictability of the two categories--- there are only a handful of conjugation and qualifier prefixes, compared to several dozen disjunct and pronominal prefixes. Thus, with fewer competing morphological items, conjugation and qualifier prefixes are more predictable based on this fact alone.

DERIVATIONAL-THEMATIC ZONE	#	PRONOMINAL ZONE	%	QUALIFIER ZONE	CONJUGATION ZONE	**STEM**	SUFFIXES

Fig. 4.1: Athabascan Verb Zones (Reduced from Jetté & Jones 2000 for Koyukon)

Furthermore, in English, only function words can avoid being stressed entirely (though not all function words are stressless), whereas even monosyllabic content words must contain at least one stressed syllable. The lack of stress can often be accompanied by optional reduction that does not affect content words. Compare the following examples of function-content homophones: 'inn'

can only be pronounced as [ɪn], with primary stress, where 'in,' the preposition, can occur in the reduced form [n̩]; 'will,' as a noun can only be pronounced as [wɪl] whereas the auxiliary verb 'will' can be realized as [l̩], forming a contraction with the preceding word. Such optional reduction and contraction is common in many other languages, including German (Kabak & Schiering 2006). Experimental evidence (Jurafsky et al. 2001, Bell et al. 2003) also shows a higher tendency for reduction in function words in English, while Shi et al. (2005) find that the effect holds for some aspects of reduction even when word frequency is controlled for.

Another common pattern cross-linguistically is a dichotomy between root and affix phonology. Affixes, in particular inflectional affixes, are functional morphemes which tend to have far fewer members in their set than roots, and as a result are also highly morphosyntactically predictable. Quite often, phonological inventories available to affixes are subsets of those available to roots, with a reduced set of contrasts (Willerman 1994). Additionally, affix sub-inventories tend to consist primarily of less marked sounds. For example, in Classical Arabic, affixes do not contain pharyngeal consonants but they are permitted in roots (McCarthy 2005). In Lushootseed, a Salish language, ejective sounds are prohibited in affixes, but commonly occur in roots (Bates et al. 1994). Vowel reduction can also be specific to affixes. In Hän (Athabascan), the phone [ə] only occurs in affixes; additionally, long vowels are not permitted in affixes (Manker 2012). For many other examples of root-affix phonological asymmetries, refer to Beckman (1997, 1998) and Urbanczyk (2011).

4.1.3.2. Salience and Syntactic Predictability

Several studies have addressed the attention and salience of different classes of words and morphemes of divergent morphosyntactic predictabilities. These studies show a common theme indicating that words and morphemes of lower morphosyntactic predictability--- open class, content, root morphemes, etc., are in some way more salient, while functional, closed class, syntactically more predictable morphemes are less salient.

Jarvella & Meijers (1983) looked specifically at differences in verb stem and affix processing in Dutch (Following the analysis of the current chapter, verb stems would be morphosyntactically less predictable due to having far more members than affixes). In their study, they found that subjects displayed a shorter response time in identifying similarity of words sharing a stem rather than sharing the same inflectional form (for example, tense). They argue that words are accessed based on their stems, with affixes being accessed only secondarily. Because of this, stems are considered to be the more salient portion of a word as opposed to its affixes.

The literature concerning language acquisition discusses the lateness of learning inflectional affixes in particular (van de Craats et al. 2000, Polišenká 2010, Penke 2012) and explains this in terms of the salience of this morphosyntactic class compared to others (derivational affixes and roots). In particular, affixes are shown to have lower phonological salience and in English tend to have shorter phonological forms than roots, sometimes existing as a single consonant in a coda cluster. Of course, the suggestion here is that the lower phonological salience is due to affixes' phonologically reduced forms, and not necessarily a characteristic of being a

member of a functional morpheme class. Nevertheless, the two variables here, the grammatical category and phonological reduction, are impossible to separate from these observations alone.

Bates and Goodman (1997) also point to the lateness of acquiring grammatical words and morphemes, indicating they are less salient and "hard to perceive" (pg. 18). Ultimately, however, they consider that this is a result of grammatical words and morphemes being highly frequent and predictable, leading to phonetic reduction on the part of speakers, rather than being fundamentally less salient or even that they should be considered separate from the lexicon and content words. Likewise, Cutler & Foss (1977) showed that reaction time in identifying the initial segment of content and function words was similar when sentence stress was controlled for, suggesting while function words are typically less acoustically salient, this difference in salience may not necessarily suggest fundamental differences in how the words are processed, even when in sentence context.

While several studies have pointed out differences in attention to, or salience of words of different morphosyntactic categories, in some of these studies it is not clear what dimensions of "salience" or "prominence" are being considered--- whether semantic or conceptual salience, or specifically in the salience of their phonetic characteristics, which is the primary focus of our study. Additionally, the confounding factor of salience in production with salience in perception obscures precisely what phenomenon is at work. If, as suggested by Bates & Goodman (1997), perceptual salience of content words is merely derivative of acoustic salience, then the lower perceptual salience of syntactically predictable words would simply be the result of their phonetic reduction. The research in this chapter will attempt to isolate these variables.

4.1.3.3. Neurological behavior and Syntactic Predictability

Other studies have considered differences in event related potentials (ERPs) for open (content, low syntactic predictability) and closed class (functional, high syntactic predictability) morphemes. Neville et al. (1992) found different electrophysiological responses for open and closed class words in sentence contexts. Closed class function words displayed an earlier N280 component (a negative potential about 280 ms after a stimulus) with a greater amplitude. This was strongest on the anterior temporal and temporal locations in the left hemisphere. Open class content words had a stronger N350 component, located on the posterior regions of both hemispheres. She concluded that such results contribute to an understanding of the "different functional subsystems" utilized when processing words of different classes.

Kutas (1997) discusses some of Neville et al.'s (1992) findings which display a stronger N400 component (N350 is considered a member of the larger N400 component class) for open class words. While Neville had noticed some relationship between N400, frequency, anomalous or unexpected semantic information, Kutas goes as far as saying that stronger N400 components were indicative of "semantic expectancy and its consequences for online processing and not lexical class" (pg. 385). Additionally, Kutas (1997) showed that N400 was typically diminished in

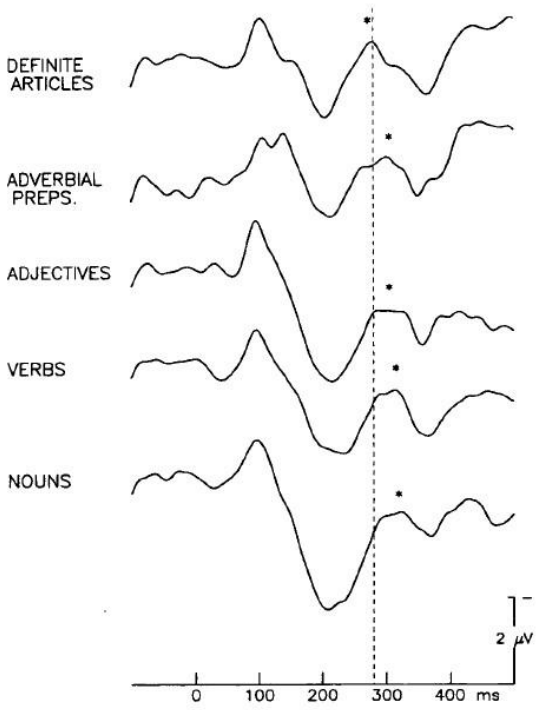

Figure 4.2: Averaged ERPs measured at the left frontal region for five lexical classes, from Kutas (1997)

amplitude near the end of sentences, where there was more syntactic and semantic context. Furthermore, multiple repetitions of open class content words could elicit a reduced N400 component. As for N280, Kutas showed that its peak latency correlated more with word length and word frequency and the higher frequency and shorter length of closed class function words was the cause of the stronger N280 component. Thus, Kutas concluded, given these different latencies (about 280 ms for closed class words and 330 ms for open class words) for what she calls lexical processing negativity, as well as semantic expectancy, that the scalp distributions for open and closed class words are "remarkably similar" (pg. 387).

Overall, this suggests that word expectancy or predictability results in different neural responses, with the activation of different areas of the brain depending on whether words are predictable--- syntactically or semantically--- or not. In this case, the condition of predictability is not an inherent property of the word, as we might claim if the same words always received the same electrophysiological responses regardless of context, number of repetitions, etc. Rather, this duality is dependent on the syntactic and semantic context. The observation that the reduced N400 response was a sign of "semantic expectancy" suggests there must be a separate type of expectancy or predictability active for closed class words within context, pointing to evidence of syntactic predictability as an independent factor. While these results do not directly consider phonetic attention, it does show contrastive neural behavior for words of different syntactic predictabilities, perhaps reminiscent of the dual stream model of speech perception (Hickok and Poeppel 2004,

2007), which proposes different neural pathways when listening primarily for meaning as opposed to listening primarily to the auditory signal (see 1.3).

4.1.3.4. Aphasia Studies and Dual Processing Models

Studies of aphasia patients have also led to observations of contrastive behaviors for words of different syntactic predictabilities such as function and content words. Researchers for long have observed patients with Broca's aphasia who are *agrammatic*, displaying a sort of telegraphic speech consisting of content words but few or no function words. For example, Gardner and Zurif (1975) observed Broca's patients who showed a higher error rate when reading grammatical particles (such as 'or,' 'at,' etc.) compared to picturable nouns. Andreewsky & Serron's (1975) study presented French words that have both functional and content homographs (*mais* as either the conjunction 'but,' [mɛ] or the noun 'corn' [mais]) to Broca's patients, finding they were usually pronounced as if open class words. Such observations led to Bradley's (1978) model of a dual lexicon, with a separate lexicon for roots and another lexicon for affixes.

Later studies, however, questioned the need for two separate lexicons, and rather proposed dual *processing* of open and closed class words. Bradley's (1978) findings were based on the fact that normal subjects showed more errors in producing low frequency content words, but that frequency did not have an effect on the production of function words. By contrast, for aphasics, word frequency was relevant for the production of both open and closed class words. Bradley interpreted the results as indicating that for normal speakers, the insensitivity to frequency for closed class words implied a separate lexicon. However, Biassou et al. (1997) and others were unable to replicate Bradley's (1978) results. Biassou et al. (1997) argued that the results were consistent with a single lexicon, and that two routes for processing were involved after lexical retrieval.

These findings once again point to dual behavior for words of contrastive syntactic predictabilities, such as content and function words. While most of the literature on aphasia concerns production, it may indicate different mechanisms of storage and retrieval that may also be relevant for speech perception. The findings in chapter 3 concerning the perception-production link, such that differences in perception based on semantic predictability immediately resulted in differences in production as shown in the phonetic accommodation experiments, may support this assertion.

4.2. **Experiment #4.1:** Phonetic Attention and Syntactic Predictability: Content and Function Words

The question of consideration in this chapter is whether the attention to phonetic details is modulated by syntactic predictability. The first observation we can draw from previous literature is that the perception and production of words of high vs. low syntactic predictability is sharply distinct. Secondly, the literature suggests that words and morphemes of low morphosyntactic

predictability are in some way (conceptually, perceptually, acoustically) more salient. The third observation is the alignment of these findings with the divergent phonological patterns that tend to emerge in high vs. low predictability word classes. Tying these observations together suggests that listeners may also pay less attention to the phonetic details of syntactically predictable words, and in this way act to shape the exemplar clouds in ways that lead to the observed phonological patterns.

The experiment in this chapter (#4.1) will test the hypothesis that listeners pay less attention to the phonetic details of words that are syntactically predictable, and more attention to those which are syntactically unpredictable. In order to test this, target words will be drawn from function-content word homophone pairs in order to control for acoustic variables. In this experiment, subjects will perform an error detection task in order to assess whether they attended to the phonetic details of content (low syntactic predictability) words more faithfully.

4.2.1. Methodology

4.2.1.1. Stimuli

The target word list included eighteen pairs of function-content word homophones. Function words included determiners, auxiliary verbs, prepositions, and conjunctions. The category of these words is usually determined by the surrounding syntax, as demonstrated by 6a-d, such that only words of the target word category could occur in their location within the syntax. Pronouns, while members of a closed class of words, were omitted since often nouns could also occur in the same syntactic locations as pronouns. Content words in this experiment included nouns, verbs, and adjectives. Examples of homophone pairs are shown in table 4.2. While many of these pairs included semantically and historically unrelated words ('or' conj. vs. 'oar' n.) others are more closely related in form and meaning ('on' prep. vs. 'on' adj.)

6a) Katie picked ___ flowers from the garden (determiner)

6b) The man ___ see the bear in the forest. (auxiliary)

6c) The cat is ___ the brown box. (preposition)

6d) Jenny wants squash ___ zucchini in her garden. (conjunction)

6e) John saw ___ yesterday (pronoun OR noun)

Table 4.2: Examples of function-content homophone pairs

Function word	**Content word**
or (conj.)	oar (noun)
can (aux.)	can (noun)
like (prep.)	like (verb)
some (det.)	sum (noun)
on (prep.)	on (adj.)

The target words occurred in sentences which, as in 6a-6d, made their category unambiguous. The content-function pair sentences contained the same number of syllables to the left and right of the target word, as to not provide any greater or lesser amount of phonetic material for the listener to process. The sentences were also constructed in such a way as not to bias the listener toward particular members of their word categories. That is to say, semantic predictability was controlled for such that most members of a given lexical category would be suitable in place of the target word. This was particularly a concern with content words, because if semantic context biased the listener towards only a handful of grammatical *and* semantically reasonable content word choices in a given environment, the overall predictability of that word would be similar to that of function word. For example, in (7a) the word 'inn' could be replaced by thousands of possible choices that are both grammatical and semantically plausible: {tree, house, car, dog, snake, water, thing, apple, etc.), while the word 'inn' is not primed in this context. On the other hand, in (7b), only a few other words could replace 'in,' but only because the syntax limits a word in this position to prepositions. The semantic context does little to narrow the lexical candidates that could occur in place of the target word, as most locative prepositions could also fit in this position, though this set would be much smaller: {on, by, near, above, below, etc.}. The difference here is the small subset of allowable words in (7b) is driven by the syntax rather than the semantics. All target sentences can be found in Appendix C.

7a) Ted saw the inn at the lake.

7b) The cat is in the brown box.

The target stimuli always contained a phonemic level error in pronunciation. While a discrimination task, rather than error detection, would have been preferable for consistency with the experiments in chapters 2 and 3 as well as to investigate the details of exemplars as opposed to what could potentially be abstracted phonological forms of words, this was not deemed possible considering the limitations of the target word list. Because of the small number of content-function homophones, there was no ideal phonetic cue that could be manipulated easily for all tokens, such as VOT. Additionally, given the complications that arise with the reduction of function words, it seemed necessary to avoid manipulations of pitch, duration, and vowel quality. That is to say, since function words tend to have more imprecise phonetic targets, with more vowel reduction, shorter durations, and lower pitch, modulations of one of these might evade the notice of the listener because of they are more likely to match pre-existing stored exemplars of function words, whereas phonemic level errors would not. In any case, a bias in phonemic-level error detection for function and content words should still indicate different levels of acoustic attention.

The target word errors were realized as single phoneme, single feature aberrations from the intended phoneme. For example, an /n/ might be realized as /m/ (change in place of articulation), /t/ as /d/ (change in voicing) or /w/ as /l/ (change in place); these small changes were not notably glaring. Furthermore, as the error occurred in both the function and content word target sentence, errors that might be "natural" based on the surrounding context of one sentence but not the other were avoided. For example, assimilatory changes such as /n/ to /m/ before a bilabial consonant would be natural and expected and would likely be less noticeable than /n/ becoming /m/ before /g/. Thus, examples 7a and 7b were pronounced as shown in 8a and 8b.

8a) Ted saw the [ɪm] at the lake.

8b) The cat is [ɪm] the brown box.

Additional complications due to the reductive affinity of function words required further adjustments in the experimental design. The intention was to record the full target sentences and then copy and paste the content words into their function word sentences (similar to what was done in the discrimination and imitation experiments of chapters 2 and 3) such that the recording of the word being heard was identical in both sentences with only the surrounding context being different (which caused the same auditory signal to be reinterpreted as a different word of a different category). However, as previously mentioned, function words naturally avoid stress and are often pronounced with shorter duration and lacking pitch accent as opposed to their content word counterparts, despite often being phonemically identical. Because of this, even when instructing our linguistically-informed model to read the target sentences with little to no reduction, copying sentences from either context and pasting them into the other almost always yielded unnatural sounding stimuli that resulted in the pasted target words standing out in such a way as to attract the listener's attention.

Instead of this method, the individual words of the stimuli were recorded in isolation in their citation forms and pieced together to form the sentences. Next, all intonation was removed by applying monotone pitch to the entire utterance. The resulting stimuli were, unsurprisingly, very artificial sounding, and subjects were told they were listening to synthesized speech. In this way, every word had equal prominence. If for some reason the function words sounded more salient than their content counterparts due to not being reduced as is normally expected, this could yield the opposite effect as was expected, being that error detection would be better for content words; in other words, this would only lead to a false-negative result for our hypothesis. If despite this, listeners still display better accuracy in detecting errors for content words, this would be strong evidence in support of the hypothesis.

4.2.1.2. Procedure

The experiment consisted of a single block of 40 questions. Of these 40, 18 were target error detection stimuli, while the remaining 22 were filler content type questions meant to induce a more natural mode of listening focused on meaning rather than sound, following the observations of Lindblom et al. (1995) and Hickok & Poeppel (2004, 2007) suggesting two modes of listening. This method proved effective in chapters 2 and 3 for the discrimination and imitation tasks. For each question, the subject would hear the audio stimulus but, since the questions were randomized, would not know whether an error detection or content question was about to follow.

After being presented with one of the target error detection audio stimuli while the screen was blank, the screen then showed each word of the sentence just heard, along with checkboxes. Subjects were instructed to indicate the mispronounced words by checking the box, and that there might be no errors or occasionally more than one (while in reality each target sentence contained only one error). Subjects thus most commonly indicated only a single error. "Mispronounced" was defined as indicating that "one of the sounds in the word is wrong, replaced by another (English) sound / letter." While it would have been ideal to avoid referring to orthography, it was

the best way to convey the concept of phonemic errors to linguistically-naïve subjects, and all errors in the target words would result in different spellings following regular rules of English orthography.

A total of 20 subjects was divided into two counterbalanced groups of 10 subjects each. Whereas subjects in group A might be presented with a sentence containing a content counterpart in a homophone pair (e.g., 'inn'), subjects in group B were presented with a sentence containing the function word counterpart (e.g., 'in'). This was necessary to avoid having subjects hear both words in each pair, which could have biased them towards paying attention to repeated words when heard a second time (despite being different words, their identical phonological forms could have resulted in a priming effect).

4.2.1.3. Equipment and Subject Recruitment

As with the discrimination tasks in chapter 2, this experiment was conducted via the online crowdsourcing application Amazon Mechanical Turk, which has been shown to be a reliable tool for recruiting subjects in speech perception experiments (Yu & Lee 2014). The experiment was run using the online survey writing software site SurveyGizmo. Participants were asked to wear headphones and take the experiment in an environment free from distractions and noise. Before officially beginning, the subjects took a short demonstration to ensure that they understood the nature of the error detection task.

Subjects were recruited at random using Mechanical Turk, although they were required to be located in the United States and had been granted "masters" status (in the hopes of recruiting a reliable pool of subjects). Subjects provided informed consent and upon completion of the task they were compensated $1.50 for the approximately 10-15 minute experiment.

4.2.2. Results

4.2.2.1. Statistical Analysis

A mixed-effects regression model was used to consider the effect of word class (function vs. content), and in turn syntactic predictability, on the listener's attention to errors in speech. The model was fitted to the data in R using the *lmer()* function in the lme4 package. The response variable DETECTION was assigned either "Yes" or "No" to indicate whether the subjects detected the speech error for a particular target word or had failed to notice it. The fixed predictor variables included CLASS, which referred to the word class of the target word, function, or content, which also corresponded to whether the word was syntactically predictable or not. Additionally, the variable GROUP was included to control for any differences in the stimuli for the two counterbalanced groups (where each group was presented with only one member of the homophone pair).

Additionally, the fixed variable FREQUENCY was included in the model in order to consider the effects of frequency. Following the findings of Goldinger (1998) concerning phonetic accommodation, lower frequency words showed more accommodation. This was explained to be the result of higher frequency words having more stored exemplars, so a new exemplar would show a smaller effect on shifting the aggregate phonetic characteristics of a particular word. In

terms of speech perception then, Goldinger's observation could also mean that the phonetic details of high frequency words are more likely to be abstracted in some way due to the stronger pull of the large exemplar cloud. Additionally, studies of sound change (Bybee 1994, Phillips 2006) have pointed to frequency as a conditioner of phonetic erosion, particular that which is common in function words. While these studies view the phenomenon from the point of view of speech production, it could imply that phonetic attention correlates only with word frequency and not with word class and syntactic predictability. This is a particularly relevant concern since most function words are far more frequent than content words, but if both variables are included in the model the effect of each can be considered.

Frequency values were obtained from the Corpus of Contemporary American English (COCA) (Davies, 2008) which separated entries into lemmas (whereas homophonous words having different parts of speech have separate entries). Only the spoken frequency values were considered, which would most closely reflect the frequency of having heard these words in speech. Due to limitations of the available corpora, including COCA, three of the 18 target homophone pairs had to be removed from the data: been/bin, have/have, and had/had. The frequency of the word 'bin' was too low to appear in the corpus, and the search tool did not distinguish between the auxiliary verbs 'have' and 'had' and their main verb counterparts.

Lastly, the random intercepts for WORD and SUBJECT were incorporated into the model in order to establish that any bias in behavior based on word class is regular across subjects and words.

4.2.2.2. Results

Overall, subjects detected far more of the errors made for content words than their function word homophone counterparts. Out of a total of 150 target content words containing errors (excluding the 30 removed from the data due to lack of frequency data), subjects detected 108 of these errors (72%), whereas for the 150 function words containing errors, subjects detected only 74 (49.3%), a 22.7% decline in error detection. For the individual content-function word homophone pairs, subjects regularly showed better detection of errors for content words, with a better detection rate in 13/15 of the homophone pairs (as well as in the three pairs removed from the statistical analysis, thus 16/18).

The results of the mix-effects regression model show significant effects of both CLASS ($\beta = 0.1422$, $t = 2.561$, $p = 0.011$) and FREQUENCY ($\beta = 1.001e\text{-}07$, $t = 2.489$, $p = 0.014$). This result aligns with previous findings for word frequency on speech perception and production, suggesting that speakers show less attention to the phonetic details of words with higher frequencies. More importantly for the focus of this chapter, the results show greater phonetic attention to content words, and in turn, to words of greater syntactic predictability.

4.2.3. Discussion

The results show that syntactic predictability significantly influences the perception of phonetic details in speech. In the experiment conducted in this chapter (#4.1), listeners were presented with the same recordings of the same phonological word forms that differed only in how

their word class was assigned relative to the surrounding context. All intonational cues were removed to avoid acoustically highlighting any particular words by making them sound more out of place than other words. Despite controlling for these factors, and including word frequency in the statistical model, word class, and thus syntactic predictability was shown to have a significant effect on speech perception, in which listeners notice fewer details of words that are syntactically predictable in their given context.

4.2.3.1. Syntactic Predictability and the role of subsequent context

This result may be somewhat surprising based on the findings in chapters 2 and 3 concerning subsequent contextual predictability, although those results remain ambiguous and inconclusive. Both in the discrimination task in chapter 2, and in the imitation task in chapter 3, subjects showed no significant effect of subsequent context on the attention to phonetic details in the way which was found for preceding context. In many, though not all cases concerning syntactic predictability, the category of the target word would not be predictable until hearing at least a few words following it. Consider again the example stimuli reprinted below in 9a and b. In sentence 9a, if hearing this sentence spoken, the word category of the word following the determiner 'the' would not be predictable at that point in time, as the remainder of the utterance might be "… old inn," (adjective following 'the'), "very old inn" (adverb following 'the') or "inn" (noun following 'the'), although only a content word would be permissible in this case. In (9b), the potential word following 'is' would not even be limited to only a function or content word, as potential utterances such as "fluffy," (adjective following 'is') or "very fluffy," (adverb following 'is'), "eating" (participle verb following 'is'), "my pet," (determiner following 'is') or "in the brown box" (preposition following 'is') are all permissible. However, in both cases, the syntactic category of the word is determined by the syntax by the end of the utterance, but this means the syntactic predictability is as much based on subsequent context as preceding context.

9a) Ted saw the ___ (at the lake).

9b) The cat is __ the (brown box).

The results found for this experiment align with much of the literature concerning subsequent context and speech perception. Warren & Sherman (1974), Kawashima et al. (1998), Connine et al. (1991), and Szostak & Pitt (2013) found experimental evidence showing that a perceptual phenomenon such as phoneme restoration or misidentification could be triggered by subsequent context, suggesting that the interaction between the attention to the auditory signal and context is a post-perceptual phenomenon, rather than perceptual. Following this account, phonetic details would be stored in a raw, unfiltered form, but would effectively decay or become altered after contextual information biases a previous word in some way. This would mean, as represented in (10), that listeners first perceive the phonetic details accurately up to the point of the target word (stage 1), but once the following words determine the target word must be a preposition (stage 2), the auditory signal is overridden in some way (stage 3), likely prior to any conscious realization of a mispronunciation.

10)

	Det	N	V	N/Adj/Adv/Det/Prep ?				
stage 1	[ðə	kæt	ɪz	ɪm]				
	Det	N	V	**Prep**	Det	Adj/N?		
stage 2	[ðə	kæt	ɪz	**ɪm** ?	ðə	bɹaʊn….]		
	Det	N	V	**Prep**	Det	Adj	N	
stage 3	[ðə	kæt	ɪz	~~ɪm~~ **ɪn**	ðə	bɹaʊn	baks]	

Nevertheless, the findings from this experiment (#4.1) may indicate the flaws in both subsequent contextual predictability experiments in chapters 2 and 3 may in fact be obscuring these results. One possibility, which might be supported in the findings of this experiment, is that for some of the stimuli in the discrimination (chapter 2) and imitation tasks (chapter 3), the biasing word, that which would confirm the identity of a previous word based on semantic congruity, occurred too far after the target word to induce an effect. As suggested in Connine et al. (1991), the window of time for subsequent context to affect the perception of prior speech may be about one second. The conclusive results of the current experiment show the effect of syntactic predictability on speech perception, and as demonstrated in 9a and 9b, a word's category often cannot be determined by preceding context alone. However, in the 36 sentence stimuli used in the present experiment, in about 31/36 the target word was limited to a single word category by the immediately following word, with the remainder determined within two, or in one case three words (often some very improbable sentences make this number inexact). In any case, the subsequent context determining a word's category, which results in syntactic predictability, regularly comes shortly after a target word, while the context confirming a word's identity through semantic congruity may not come within this limited window of time. More discussion of these results with respect to the perceptual or post-perceptual nature of this phenomenon can be found in chapter 5.

4.2.3.2. Syntactic Predictability and Sound Change

The results of this chapter suggest a novel approach to understanding the process of phonetic erosion that commonly occurs in function words. Traditional accounts have considered this phenomenon primarily in terms of the role of the speaker rather than the listener. For example, Lindblom's (1990, et al. 1995) Hyper & Hypo (or H & H) theory suggests that speakers hyper- or hypo-articulate portions of their utterances in order to meet the needs of the listener. That is, speakers keep "a running estimate of the listener's needs for explicit signal information on a moment-to-moment basis and then adapts the production of the utterance elements… to those

needs" (pg. 5). Likewise, Alyett & Turk (2004) propose that "prosodic prominence is a linguistic means of achieving smooth signal redundancy. Prosodic prominence increases… with unpredictable sections of speech" (pg. 31). Essentially, the claim being made is that the speaker is the source of the types of changes that result in prominence in less predictable speech, and with a bias towards reduction and stress-avoidance in more predictable speech. Applying this model to the issue of the reduction that occurs in syntactically predictable words such as function words, it would be the speaker who is, on some level of consciousness, purposely reducing function words due to their predictability to the listener.

Other researchers have also suggested speaker-oriented models of function word reduction, however in terms of word frequency as opposed to contextual predictability, showing that words produced more frequently undergo phonetic erosion (Bybee 1994, Phillips 2006). Bybee & Hopper (2001) provide a difference approach to this reduction, stating that "the origins of reduction are in the automatization of neuro-motor sequences which comes about with repetition. This automatization involves the reduction of the magnitude of articulatory gestures and the increased overlap of these gestures" (pg. 11). This would also account for function word reduction but with less intention on the part of the speaker, who would not be considering the needs of the listener in this case.

But what if, instead, the trajectory of sound change, in particular that of function word reduction, was determined not by the speaker biases but by those of the listener? The findings from the current chapter, as well as those of chapters 2 and 3, lay the experimental groundwork which makes this account plausible. The results from all three chapters show that phonetic attention is modulated by contextual predictability, which is in agreement with much of the previous literature. In the current chapter we found that listeners pay less attention to the phonetic details of function words, which are often syntactically predictable. Along with the findings from chapter 3 in the phonetic accommodation experiments, it becomes clear that these details are stored in memory and in turn affect the production of speech. Thus, if listeners are storing details of function and content words differently, word classes which regularly display different degrees of syntactic predictability, we would expect the spread and adoption of sound changes to affect them differently. The exemplar clouds of function words should, following these conclusions, resist change, as listeners ignore their phonetic details in favor of activating abstracted forms of these words. This is in contrast to the observations that generally speaking, function words change more rapidly. Chapter 5 will consider this paradox in light of the experimental findings of chapters 2, 3, and 4, and will propose a model which will account for both these findings and the linguistic evidence demonstrating how function and content words change.

Chapter 5
Discussion

5.1. Introduction

The results of the three sets of experiments in this dissertation corroborate the literature showing that context modulates speech perception and that this bias in perception in turn affects speech production. These findings largely show that listeners rely on contextual knowledge whenever available, and that the perception of subphonemic as well as featural and phonemic information is influenced by higher level linguistic knowledge and expectations based on abstracted forms of words whenever possible. In this chapter I will address aspects of (1) speech perception and (2) sound change models which are supported or suggested by the results of the experiments of this dissertation. Concerning speech perception, these results suggest the existence of both abstract and more veridical representations of exemplars which are modulated by the listening mode. Additionally, in accordance with the findings of previous literature, the results support a model in which the raw auditory signal is held in working memory for a short period where top-down and bottom-up processing work together in word recognition. In my model of sound change accounting for content and function word differences, I propose a strong role of speech perception and the indirect consequences of 'what' mode listening in shaping the exemplar cloud. Additionally, I propose that first language acquisition provides the necessary requirement--- the lack of 'what' mode listening--- for the swift selection of new variants.

5.2. Relevance of Findings to Speech Perception

5.2.1. Dual Modes of Listening

The results align well with a dual model of speech perception, in which listeners either focus on recognizing words and extracting meaning ('what' mode) as opposed to processing and analyzing individual sounds and gestures ('how' mode). In particular, chapters two and three show an inverse relationship between the amount of context that is available and the attention to phonetic detail. The first imitation experiment in chapter three (#3.1), which considered attention to phonetic detail in isolated words versus those occurring in sentence context, was inconclusive

and did not yield strong evidence that the 'what' mode is used exclusively for processing speech in sentences whereas the 'how' mode is used for processing words in isolation. That is to say, the mere presence of linguistic structure and context was not enough to "flip a switch" from one mode to the other. Rather, the results from the imitation studies considering word predictability, rather than the presence of absence of sentential context, were much more conclusive. When given no instruction to imitate, word predictability showed a small but significant effect on how closely listeners attended to phonetic details and reproduced them in their own speech. Thus, the mode of listening is shown to be most influenced by word predictability or expectancy, such that surrounding context is able to aid the listener in speech recognition. Thus, expanding on Samuel's (1981:481) observation that "restoration is a function of context; the greater the context, the greater the expectation, the greater the restoration," we might say that *phonetic attention*, and along with it the mode of listening, is a function of context. This phenomenon is shown to work online as speech is processed, as listeners make predictions about the words they expect will occur, and process words with either more or less emphasis on fine phonetic details depending on the degree to which they need to rely on the auditory signal.

A second point of discussion concerning the dual modes of speech perception is that these two modes likely indicate two extremes along a continuum rather than two categorically distinct modules. Neurologically speaking (Hickok and Poeppel 2004, 2007), it may be the case that two categorically distinct and independent streams of processing exist in the brain (dorsal and ventral), however it could be that, depending on the nature of the listening task and the context available for top-down processing, listeners can utilize both modes to varying degrees for identifying particular words. The experimental results, especially those from the word predictability and imitation studies in chapter three, suggest at least three patterns, not two. Subjects showed the greatest degree of phonetic attention *when told to imitate*. In these cases, the subjects most closely imitated the model and contextual information such as word predictability became irrelevant. Subjects may have been paying little attention to the meaning of the sentences in order to focus on the pronunciation of the words (perhaps much like a parrot imitates human speech). Likely, a fairly extreme 'how' mode of listening was induced for this task. On the other hand, when given *no instruction to imitate*, subjects did not imitate the phonetic characteristics of the model as closely, probably processing sentences for meaning, with many words being abstracted (with only previously stored exemplars or phonological forms being activated) as a means of aiding memory and storage of word forms. However, when words were unpredictable in this context, subjects were forced to rely on the auditory signal to identify these words, and the 'how' mode may have been activated to some intermediate degree. We could imagine a model then, in which unpredictable words are "shipped off" to the dorsal stream for whatever amount of processing is needed in order to identify them (e.g., identifying all the segments in a particular word, identifying just the first few segments, identifying only the place or manner of particular segments, full subphonemic processing of the entire word, etc.).

Lastly, the very different behavior that subjects exhibited when given different instructions (told to imitate or not), and the fact that contextual effects on speech perception largely disappeared when listeners were encouraged to focus on pronunciation, underscores the methodological importance of encouraging subjects to focus on meaning when trying to analyze such contextual

effects. This may account for some of the unexpected results in the literature, for example Samuel's (1981) finding that subjects actually performed better in determining whether a sound was added or replaced when the words were predictable, or when Marslen-Wilson and Welsh (1978) found no difference in error detection for predictable and unpredictable words. Samuel reasoned this occurred because word recognition is easier for predictable words, so more processing of the auditory signal was available. However, in ordinary conversation, listeners are not typically concerned with identifying low-level phonetic information like they would be in a speech perception experiment, and it is possible in these cases that the opposite effect (or no effect) could be observed. Thus, in order to assess the effect of contextual information on speech perception, subjects should be given tasks that encourage them to focus on meaning, which was achieved in the discrimination (#2.1 and #2.2) and error detection experiments (#4.1) in chapters two and four by randomly mixing in content questions with questions asking subjects to consider the auditory signal.

5.2.2. Exemplar Theory

In addition to new observations concerning the dual model of speech perception, the results of this dissertation also yield related observations concerning the exemplar model of lexical activation and organization. Aligning with the notion of a gradient continuum between the 'what' and the 'how' mode of listening may be a continuum of exemplars from *abstraction* to the *raw auditory signal* (while the listener does not actually have direct access to the acoustic signal itself, as she is only able to perceive what the auditory system captures). This is supported by the varying degrees of acoustic accuracy that subjects showed in the experimental results of this dissertation when imitating speech--- particularly the fact that more veridical detail was stored for unpredictable words, which in turn had a greater effect on future productions of those words. As shown in figure 5.1, the most extreme abstraction may not even be a word, but a concept or idea, while the most veridical (and accessible) is the raw auditory signal. In many cases, the activation of a more veridical form will result in the activation of a more abstract (such that a phonological form might activate a word, and then a concept), however the opposite (activating a more veridical form from an abstract one) could only be done by the listener drawing from his own memory, in which case the activated form would not be faithful to the auditory signal. Concepts or ideas would not normally be the only forms activated in speech recognition (with no more veridical form activated even momentarily, unless a listener was paying so little attention he did not notice the specific word uttered but got the 'gist'; however function words are probably more likely to be activated as concepts), however even more veridical forms of words may eventually decay to mere concepts over time (for example, we rarely remember the exact words in conversation, but recall the general message). Words themselves are somewhat less abstract, and could be the forms that are activated during the process of word recognition given enough context (as opposed to activating sublexical information such as segments). Subjects in our imitation studies may have recognized more predictable words based on whole word comparisons between the auditory signal and stored form (similar to Klatt's 1979 LAFS module). If a subject in a phonetic accommodation study were to "imitate" a word that was originally activated as a whole word rather than a sequence of sounds, this subject may produce an articulatory plan based on her knowledge of the sounds that should occur in this word, rather than imitating the raw auditory signal.

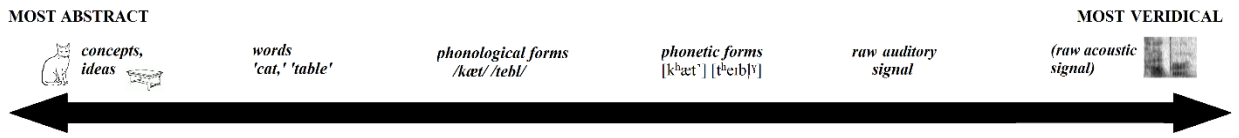

Figure 5.1: Abstractness scale

The activation of phonological forms themselves requires some level of segmental processing but still results in diminished attention to subphonemic characteristics. If recognizing speech based on phonological forms of words, a listener may notice missing phonemes (such as in a phoneme restoration task) but may fail to accurately imitate subphonemic, allophonic detail such as the type of stop release in a phonetic accommodation experiment. Similarly, recognizing speech by processing the auditory signal even to a point of attending to subphonemic detail, activating a phonetic form of a word, might still mean failing to notice a slight lengthening of VOT or minor deviations in vowel formant frequencies. The most veridical representation of the speech signal, that is, the form closest to the actual production in the physical world is the raw auditory signal, where potentially any acoustic detail that is captured by one's auditory faculties may be stored, processed, or activated (the raw acoustic signal is technically not directly accessible to humans, and must be minimally filtered through the auditory faculties). Of course, as previously suggested, instances of words that are activated or stored in memory, that is to say, exemplars, may represent a blend of any of these forms, whereby different parts of words (such as the onset vs. coda) might be stored in a form more or less faithful to the auditory signal.

Additionally, it is as of yet unclear the difference between what forms are initially activated in the course of recognizing speech, as opposed to the exemplars that persist in memory by the time one is asked to recall the pronunciation of a word or produce that word. Thus we might refer to two different types of exemplars. One is the actual "activated" or "processed" form of a word which is what is initially extracted from the raw auditory signal. This is the form that is used in word recognition, compared to the stored forms of words in the lexicon. The second is the "stored" form of a word itself which may be used in production or as a reference of comparison within the lexicon of new speech percepts in order to recognize them. Generally we can assume that stored forms may decay to some extent from their original activated or processed form such that they become more abstract (moving to the left on the scale) over time (which can happen through repetition of episodic traces as described in Hintzman 1986), since new phonetic details cannot be accurately restored once lost. It is also not clear whether the difference in what details are stored (depending on a word's predictability) is a matter of what is processed during perception or in what details are maintained in memory over a given period of time. It is possible that all speech is processed at some initial level of detail, followed by additional processing if needed in the absence of context, and also followed by differing degrees of memory decay depending on the needs of the listener in storing the speech signal. The perception vs. post-perception issue will be further considered in section 5.2.3.

What is clear from our results is that listeners do not attend to all the available acoustic detail and this detail is effectively filtered by context, and in the task of recalling (discrimination or error detection tasks) or reproducing this information (imitation tasks), such detail may be restored based on prior experience of the language. As Goldinger (2007) states, "each stored exemplar is actually a product of perceptual input combined with prior knowledge…" (p. 50). Thus, the actual form of an exemplar, and how abstract or faithful it is to the raw auditory signal, is modulated by the available context. Thus, listening in the 'what' mode, which relies on more context for top-down processing, we would expect the activation and storage of more abstracted word forms. If listening in the 'how' mode, due to absence of the context needed for word recognition or if one is purposely attending to details of pronunciation, exemplars will more closely reflect the details of the raw auditory signal. In any case, the weight of higher versus lower level information in the composition of the exemplar is dependent on these factors. Figure 5.2. shows the relative weight of abstracted vs. raw acoustic information in the composition of the exemplar, based on observations from the imitation experiments.

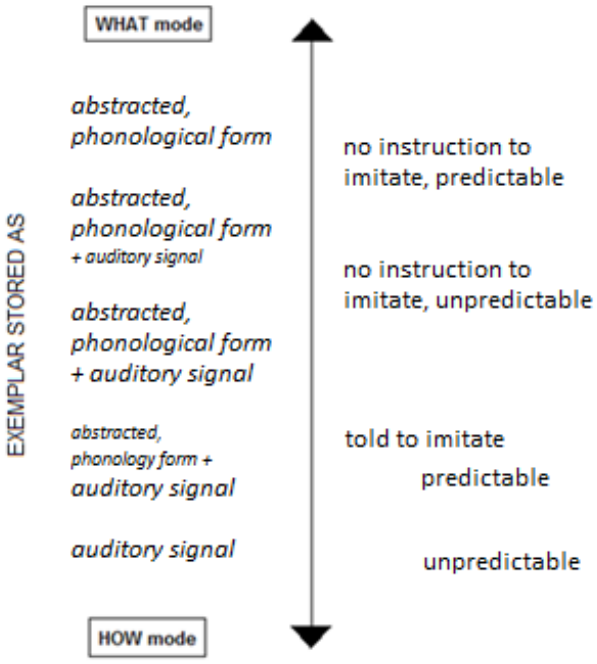

Figure 5.2: 'What'/ 'how' continuum and the weight of the abstraction vs. auditory signal

5.2.3. Speech Processing and Word Recognition

Some studies of contextual effects on speech perception, such as phoneme restoration (Samuel 1981) have explored whether these effects are *perceptual* or *postperceptual*. A perceptual phenomenon refers to the original processing of information sent from the sensory organs to the

brain. In contrast, postperceptual phenomena involve some sort of change from the original percepts by means of decisions based on new information or perhaps decay of the original percept in memory. For example, in speech, we might consider whether something like phoneme restoration is a perceptual or postperceptual effect. If perceptual, this would suggest the raw auditory signal was never processed to the extent that the listener noticed the absence of a sound--- the goal of the listener, perhaps to identify a particular word, was achieved without needing to analyze the speech signal on a phonemic level. If this effect is postperceptual, it would mean that the listener, on some level, conscious or not, perceived the absence of a sound, but upon lexical retrieval, determined the sound was in fact there and restored it in memory by essentially overwriting the original percept. Since this could occur subconsciously, similar behavior could occur (lack of consciously noticing the absence of a sound) regardless of whether the effect is perceptual or postperceptual.

The results from the current experiments concerning the effect of preceding and subsequent context may shed light on this particular question, although the ambivalence of the results may only obscure any clear explanation without future research. If effects are only found for preceding context, this might suggest that the cognitive system in some way primes itself to attend more or less to the phonetic details of upcoming speech, suggesting purely perceptual processes are at work. If effects are found for both preceding and subsequent context, it minimally suggests some processing of the phonetic details takes place in some window of time following the initial auditory input. This allows for the possibility of the effect to be post-perceptual, however still could mean the phenomenon results either from differences in *processing* (arguably still perceptual) as opposed to differences in memory decay (postperceptual).

Two versions of both the discrimination experiment in chapter two and the imitation experiment in chapter three were conducted, one considering the effect of preceding context and the other considering the effect of subsequent context. In both of these, reliably significant results were found pointing to the effect of preceding context on the attention to the phonetic details of upcoming words. However, also in both cases, the same effect was not found for subsequent context. This was in contrast to a number of studies (Warren & Sherman 1974, Kawashima et al. 1998, Connine et al. 1991, Szostak & Pitt 2013) which found perceptual effects based on subsequent context. In contrast, the results of the error detection task in chapter 4 seem to indicate that subsequent context played a role in attending to more phonetic detail for content (low syntactic predictability) words. As discussed in chapter four, often the class or category of a target word could not be determined from preceding context alone, however in a majority of the cases the category could be determined once the immediately following word is identified. Thus, the assessment of a word being "syntactically predictable" is determined at least in part by subsequent context. This result agrees with the findings of the literature but is at odds with the results of the discrimination (#2.2) and imitation experiments (#3.2b).

In comparing the three sets of experiments we find two major differences that might effect this disparate outcome. First of all, following the observations of Connine et al. (1991) the window of time for subsequent context to affect the perception of a previous word may be only about one second. As determined in chapter two, the biasing words in the stimuli for subsequent context--- the ones essentially confirming a prior word, making it "predictable," were often several words after the target word, even occurring in some cases at the end of the sentence. Thus, these words may have occurred outside this eligible window. In contrast, in the error detection experiment in chapter 4 (#4.1), most of the words that ultimately determined the target word's category, making

them syntactically predictable or not, occurred immediately after the target word, and thus may have occurred well within this perceptual window.

Secondly, both subsequent context experiments (#2.2 and #3.2b) involved more phonetic material between the target word and either its repetition (as in the discrimination experiment, #2.2) or its production by the speaker (in the case of the imitation experiment, #3.2b), due to the fact that the target words in the subsequent context sentences were usually near the beginning of the sentence as opposed to being at the end. In the discrimination experiment in particular there was also more time between initially hearing the target word and its repetition (the delay between hearing the target word and its production would be similar for both the preceding and subsequent context imitation experiments since subjects repeated the entire carrier sentences). Either additional time or additional phonetic material could act to further decay the stored form of the exemplar. It is possible that while perceiving unpredictable speech might result in more accurate activation and storage of phonetic details, that these details fade faster over time if they fail to be incorporated into a more abstracted representation. Remez (2003) refers to the "ephemeral nature of the auditory trace" (295) and contends that after a window of about 400 ms, such an auditory trace fades and that any form surviving longer in memory "must at least be lightly abstracted" (295). This implies that abstraction is a mechanism used to aid in storing details over longer periods of time; if listeners transform the raw auditory signal into something related to past experience and generalized concepts, it is easier to store in memory for longer durations (at the cost of less faithfully storing the unfiltered physical characteristics of a particular event). Thus, given enough time, it is possible that phonetic details of unpredictable speech would fade from memory to a similar extent as they would for predictable speech, or even more so in the sense that those details might be restored better from the more abstracted stored forms of predictable speech.

Thus, as a whole, the results of all three sets of experiments do not conflict with the existence of a small window of time during which subsequent context may influence speech perception. However, it is not clear that this implies a postperceptual phenomenon, and ultimately may challenge any categorical distinction between the two. Remez's (2003) 400 ms window may indicate the short time during which the listener has full access to the raw auditory signal stored in a sort of working memory. This may also reflect the relatively short intervals between articulatory rehearsal proposed in Baddeley & Hitch's (1974) phonological loop model of working model, such that rehearsing a word stored in memory refreshes its memory trace which would otherwise rapidly decay. During this short window, a listener may have the ability to revisit the trace and continue processing it at increasingly precise levels of phonetic detail. We could imagine that for any speech--- which may be either predictable or unpredictable--- the listener may begin processing the signal in as broad and abstracted a means as possible. This may first begin by comparing an entire chunk of speech, potentially a word, to other full word form exemplars stored in memory. In the case of words which are predictable based on preceding context, a number of likely candidates are already primed. If the superficially processed speech chunk is a close enough match to a primed candidate word, no further processing is needed. On the other hand, more thorough processing of the signal may be needed for unpredictable speech in the absence of the guidance of top-down processing--- if the candidate pool includes thousands of words, segmental or subsegmental level processing becomes more necessary in identifying a closely matching word.

These insights align with those of Craik & Lockhart (1972) who propose two types of processing. Any stimuli may pass through levels of processing, passing from the senses to short term memory and to a deeper level of abstraction where a stimulus is compared to past experience and knowledge. A more familiar stimulus will pass to this stage more rapidly, where it is retained

in memory better but is more abstracted (Type II processing). On the other hand, a less familiar stimulus, which in the current experiments might be something like an unpredictable word in context, may be held longer in primary memory where it can be revisited and further analyzed (Type I processing). While it is more difficult to hold such a stimulus is memory since it has yet to be connected to past experience and abstracted, it likely retains detail that is more faithful to the auditory signal. Over enough time, however, details that fail to be abstracted will be lost entirely.

It is not clear whether Connine et al.'s (1991) window during which subsequent context can affect speech perception is merely a more liberal version of Remez's 400 ms window or denotes a different period in the speech perception process. In any case, there is probably not a strict cut off point, but instead, the details of the raw auditory signal held in working memory will gradually fade from memory during this time. This window of time would seem to indicate a period of time during which the activated form of the exemplar--- that used in the process of comparing and recognizing a word--- can still be modified or refined. Such a phenomenon could be considered postperceptual if this involves overwriting the original perceived sound (being essentially a different window than that proposed by Remez). On the contrary, this could merely be the period during which bottom-up, signal-oriented information races against top-down, contextual information to achieve word recognition (and thus simply a longer version of Remez's window during which the raw auditory signal can be accessed). If no new biasing or confirming information comes within some range between 400 ms or one second, the low-level acoustic information wins out in being the primary force in determining the activated form of the exemplar used in speech recognition. This account of the phenomenon defies easy categorization as either a perceptual or postperceptual event.

To better illustrate this chain of events, consider how listeners may have processed a mispronounced function word such as [ɪm] for 'in,' such that the error went entirely unnoticed. As demonstrated in chapter four, upon hearing the syllable or potential word [ɪm], the listener would not be able to determine the part of speech based entirely on preceding context. With only having heard the preceding portion "The cat is [ɪm]," assuming successful identification of these preceding words, [ɪm] could be a noun, adjective, adverb, determiner, or preposition. At this point in time, [ɪm] is temporarily held in working memory, complete with all the acoustic details that the ear was capable in receiving and sending to the brain. However, only through processing this signal will the listener notice, or be able to store, any of these details. The earliest form of the processed or activated exemplar would not necessarily be [ɪm] at all, but perhaps something like a blurred, detail-impoverished spectrogram. The brain would not rule out [n] final words, because the brain has not processed the signal at the segmental level. Thus, possible nouns, adjectives, adverbs, determiners, or prepositions that, as a whole, are similar to this word, or at least if this were the first syllable, are assessed as possible word candidates. This could include words like 'him,' 'empty,' 'impossible,' 'thin,' 'indispensable," or, of course just 'in,' that could plausibly fit this first syllable (though many words may also be ruled out based on semantic implausibility). Once the next word, 'the,' ('The cat is in the,') is spoken, the target word [ɪm], as a single syllable word could only be a preposition. Still within a possible 400-1000 ms window, this piece of higher level linguistic knowledge would drastically reduce the pool of candidate words, and 'in,' as a semantically reasonable word in this position, by far the best acoustic match to the "blurred spectrogram" exemplar, would be selected as the intended word for the signal held in working memory. At the point of 400-1000 ms, the raw signal would fade from memory and could not be accessed again (the memory fading event could be instigated upon word recognition, as there would be no need to review or further process this portion of the auditory signal). The form of the

activated exemplar was only the word 'in,' and/or the "blurred spectrogram" which only contained broad acoustic details. This activated word instance would then be committed to storage in memory, joining the 'in' exemplar cloud, but would do little to change its pronunciation as only a broad, abstracted word form was ever activated. Thus, the [m] would go unnoticed by the listener entirely, as no segmental processing would have been needed for the linguistic goal of understanding the content of the sentence. Had the biasing information in subsequent context--- either syntactic information confirming word category or semantic information ruling out implausible candidate words--- come much later, after the 400-1000 ms window, the listener would be forced to continue with bottom-up processing, and the activated form of the exemplar, derived from the raw auditory signal held in working memory, would include more detail and likely more specific segmental or subsegmental information--- in other words, would be closer to the form of the raw auditory signal. In this case, [ɪm] would likely be perceived and the final [m] would be noticed; however, a purely postperceptual decision process may act to rewrite what was originally heard when the listener realizes 'in' was the intended word, however he is more likely to be conscious of the speech error.

Perhaps the model best suited for handling these aspects of the speech perception architecture that are illuminated by the findings of this dissertation is that of Adaptive Resonance Theory (ART) as proposed in Grossberg (1980) and in the context of linguistics in Goldinger and Azuma (2003). At the core of ART is the rejection of a fixed unit of speech perception. Goldinger and Azuma review a wide body of literature that considers whether segments, syllables, or words are the units of speech extracted from the speech signal used in speech recognition, however there is wide disagreement in the findings with no firm conclusion about a fixed unit of speech perception. ART proposes that there is no fixed unit, but that listeners are flexible in activating whatever units of speech are best in a given situation. In general, the activation of more functional units, such as words, mask lower units like syllables or phonemes. This model works particularly well with the results of the current experiments as well as a continuum version of dual speech perception model, such that the availability of contextual information (as well as the listening goal or experimental task) will influence the units of speech perception that are activated and in turn stored in exemplar memory. Top-down and bottom up processing essentially compete in processing the signal over a short amount of time until word recognition is achieved. Goldinger and Azuma (2003:310) state the following:

> "Processing in ART is self-optimizing, allowing rapid coherence across variable situations (Grossberg 1980). Strong bottom-up information can support resonance with minimal top-down matching, as when clearly spoken non-words are readily understood. Conversely, distorted bottom-up signals are readily identified with top-down support, as in phoneme restoration (Samuel 2001)."

Thus, ART readily handles the observations from the current experiments showing that stored exemplar forms are filtered and poor in acoustic detail in the presence of predictive or confirming contextual information. The results also suggest a short period of processing during which higher and lower level information is taken into consideration in the process of word recognition, which determines the contents of the activated exemplars that are in turn stored in memory. This can account for the effects of preceding context as well as immediately following subsequent context. These contextual effects, which are more likely to transform the raw auditory signal into abstracted forms such as entire words, can be overridden in phonetic listening tasks, such as when subjects were asked to try to imitate the model's speech in chapter three. In these

cases, phonetically-rich exemplars were activated and stored after processing the auditory signal for lower level phonemic and subphonemic detail.

5.3. How the perceptual effect of word predictability guides sound change

5.3.1. Predictability and the function word reduction paradox

The results of the current experiments show that top-down processing of the auditory signal affects the details perceived and stored in memory, and that this difference even resulted in a measurable change of the phonetic details of these words produced by listeners turned speakers. This by itself strongly suggests that contextual clues effectively shape exemplars in some way. It is important then to consider if there are sound changes which might logically result from this perceptual bias.

In the discussions of chapters three and four I considered a few such sound changes. Unlike with word frequency, it is not immediately apparent that words might have their own unique predictability values since most words can occur in an almost infinite number of sentential contexts in which those words may be either predictable or unpredictable. In chapter three, however, I suggested that obscured compounds, such as 'day's eye' > 'daisy' or 'God sibb' > 'gossip' could be one such case where the second word in the compound would be highly predictable due to its frequent occurrence after the first word. This could cause lower attention to these forms in particular, which actually results in something of a lexical split between the free word forms and those occurring in the obscured compounds. In chapter four, I suggested function words, as a class, were much more predictable than content words primarily due to *syntactic* (rather than semantic) considerations. That is, the surrounding syntax can determine a particular word's class or category, and in the case of closed-class function words, which tend to have very few members, these words are relatively more predictable due to having fewer competing candidate words. The historical record in fact shows unique phonological developments that occur to function words to the exclusion of content words, particularly in the form of reduction, lenition, and contrast neutralization.

However, it is at this point where a paradox emerges. Several studies (Phillips 1984, Bybee 2000, Pierrehumbert 2002) show that sound changes affect highly frequent words first, such that these are leaders in adopting new sound changes. Function words, most of which are among the most frequent words in languages such as English, thus will undergo sound change most rapidly. The reasoning here is usually that new innovations in these words will be heard more frequently with the perception of a higher number of exemplars resulting in more rapid phonetic change (though recent work such as Hay et al. 2015 suggests non-lenitive sound changes such as vowel shifts may actually affect low frequency words first). Bybee (2000) considers the effect of production as well, though specifically for the phenomenon of lenition, such that frequent production (rather than perception) of these words leads to reduction, and that these production exemplars are incorporated into the cloud. In any case, function words, and other more predictable words, would seem to adopt new changes more easily, while content words and other more unpredictable words should resist change. However, the imitation experiments in chapter three actually appeared to reveal the opposite, as a sound change was induced in *unpredictable* words. This experimental evidence would seem to indicate that predictable words would actually resist change over time, since the changes that occur would be more likely to go unnoticed by listeners.

By extension then, we might expect syntactically predictable function words to resist change whereas content words might adopt new changes more readily.

Results from past literature suggest a different effect of predictability on function and content words, which might in turn allude to differences in lexical access, retrieval, and as a result, how they may be influenced by perceiving them. For example, Bell et al. (2009) found that word frequency only significantly affected the duration of content words (shorter when more frequent), whereas predictability based on subsequent context primarily only affected function words (shorter when more predictable, and the effect was stronger for highly frequent ones). Predictability based on preceding context affected only less frequent function words and content words, while repeating a word multiple times only resulted in shorter duration of content words. Bell et al. propose that this is the result of different means of access and retrieval for content and function words, such that function words have privileged access. In light of the findings of this dissertation, privileged access could suggest that function words are more likely to be accessed from abstracted forms stored in memory, despite whether they were recently heard or not, while content words may be more likely to be accessed from recent, veridical exemplars including those recently perceived in conversation. This could account for the differing effects of frequency. Thus, it could be possible that variation in function words is primarily intrapersonal, occurring frequently but is ignored and not spread to other fluent speakers. Meanwhile, this different means of access could result in the spread of new variation for content words among fluent speakers, which is in line with the experimental results of this dissertation.

5.3.2. A model of predictability-modulated sound change

The results of the imitation experiments induced a sound change with a perceptual bias, such that newly perceived variants of unpredictable words showed a stronger effect of spreading and affecting the listener-turned-speaker's own productions. A similar perceptual bias was shown for function and content words. Thus, can a model of sound change be proposed from which the long term phonological changes that have been observed (e.g., function word erosion) naturally result, and does this model better account for the data than production-based accounts? The proposed model must first of all be able to account for these distinctive patterns of individual and language-wide change, such that function words show stability throughout the life of the fluent speaker but relatively rapid change over the course of the life of a language. Additionally, to review the criteria established in chapter one, the current model will avoid teleological explanations of sound change if possible. That is, there is no reason to assume that speakers are purposely shaping their language in order to meet the needs of the listener. Thirdly, the model will not assume biases in the production errors or pronunciation drift in one group of words or another (nor will the model discredit such a proposal, but it does not rely on this phenomenon). I will only assume an equal amount of variation will occur for function and content words (from a starting point at which they follow the same phonological patterns), but with a function-content word selection bias rooted in the mechanisms of speech perception, not production.

First of all, the model I propose follows a basic two-step architecture of *variation* and *selection* (Ohala 1989, Lindblom et al. 1995, Kiparsky 1995, Blevins 2004, Garrett & Johnson 2012, Yu et al. 2013). Variation occurs primarily in speech production, when speakers produce forms that have in some way deviated from the forms they intended to produce, due to articulatory or aerodynamic limitations (we could say variants of forms result from speech perception, as in the case of hyper- or hypocorrection (Ohala 1989), however, in some sense this is the selection of

forms in the state in which they were perceived). Selection, on the other hand, occurs when new variants (including misperceived variants) become stored in memory and become the actual linguistic targets of speech. Without selection, sound change could only occur intrapersonally, with a certain amount of variation occurring for particular individuals, but with this variation never spreading throughout the language and never even becoming intentional (i.e., phonologized) for any particular speakers.

The development of distinct subphonologies for function and content words suggests some sort of bias such that initially homophonous function-content words might eventually diverge, with function words adopting reduction. Such a bias could be either in the actual variation that occurs, such that speakers are prone to reductive type errors more often in function words, or in selection, such that listeners are more likely to notice and/or adopt reductive type errors occurring in function words. Without a doubt, there is bias in the types of errors which occur in speech (e.g., assimilation due to articulatory limitations), but there is no need to assume a speech production bias in the variation occurring for function and content words. While some (Bybee & Hopper 2001) have suggested frequent words reduce due to the repetition of articulatory gestures resulting in simplification, this does not account for class or category-wide sound changes (such as initial eth-voicing in function words that does not simply target the most frequent words in the language). Lindblom's (1990) H&H theory, or Alyett & Turk's (2004) model of achieving smooth signal redundancy could result in production biases resulting in reduction of function words, which, in addition to being fairly teleological in nature, is also challenged by the findings of Gahl, Yao, & Johnson (2012) that contend that the hypoarticulation of predictable speech is rooted in the ease of lexical retrieval. While Gahl, Yao, & Johnson's (2012) model could also yield production-oriented bias, again with more easily retrieved function words becoming reduced, their finding is rooted in the fact that words of high neighborhood density are also reduced and are more easily retrieved, yet there are no attested sound changes that are shown to target words of high neighborhood density (along with more predictable words), suggesting this phenomenon of production may not be relevant in language-wide sound change. Thus, we should consider the role of perception-based biases in the process of selection in a model of sound change, rather than biases in the production of variation. A perceptual bias leading to a bias in the selection of new variants is precisely the phenomenon that occurred in the imitation experiments (#3.2a, #3.2b) which I will show better accounts for the observed behavior and linguistic patterns.

In beginning to forge the structure of a model which addresses all of these issues and which predicts the distinct sound changes that target function and content words, let us assume an equal amount of variation of production of content and function words, perhaps for a content-function homophone pair like 'can'-'can.' This would be at a point in history at which the phonological targets are identical for both pairs, and perhaps at a point of pre-grammaticalization of the auxiliary 'can' such that both words had similar syntactic predictabilities, such that the only intended pronunciation is /kæn/, initially only ever surfacing as [kæn]. The eventual variation which might occur would include either (a) fully faithful and accurate productions which perfectly match (or only imperceptibly err from) stored representations (b) variants which only err minorly from their intended targets and (c) variants which miss their mark entirely. The difference in (b) and (c) might be that (b) word variants could still be recognized in isolation, whereas (c) may not.

In the case of content words, as a whole, listeners will pay attention to more phonetic details of these words due to their lower syntactic predictability. The phonetic details of type (b) variation will be attended to by listeners (though perhaps not consciously noticed) and as these forms are

still recognizably variants of the intended words, they will be adopted into the exemplar clouds for those words. On the other hand, the phonetic details of type (b) variation will not even be noticed when processing the auditory signal of function words (again, generally speaking). Function words are probably rarely if ever granted phonemic or any other sort of sublexical processing for

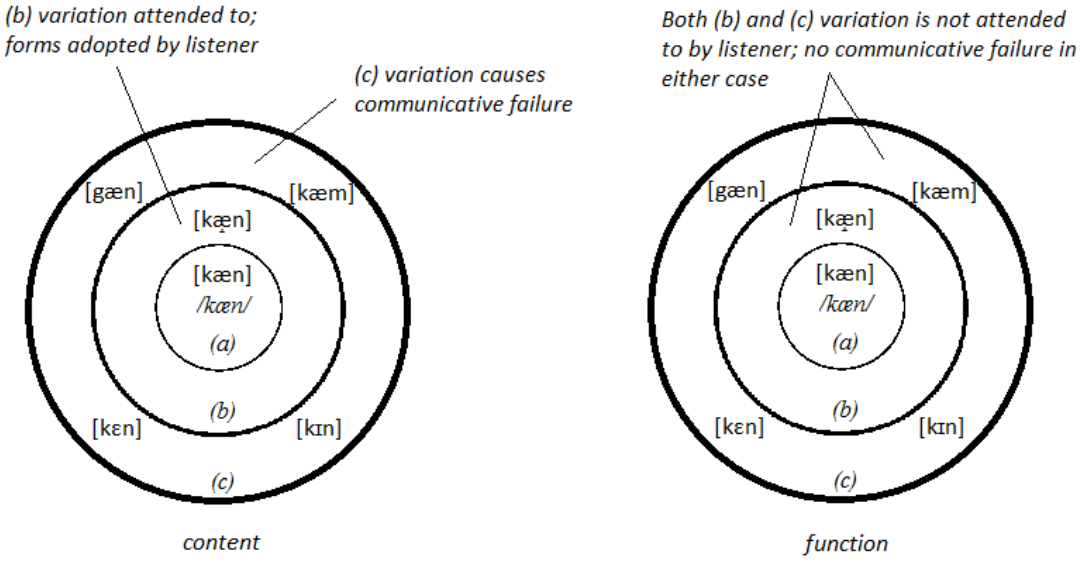

Figure 5.3: Content and function word articulatory target inputs and production outputs

the purpose of word recognition. The distinction in (b) type variation is what yields the results of the phonetic accommodation experiments in chapter three (#3.2a) which showed that unpredictable words underwent greater phonetic drift as opposed to predictable words (where VOT lengthened and pitch raising were fairly minor forms of variations). Additionally, the perceptual bias in (b) type variation would result in slow phonetic drift of content words (being syntactically unpredictable) and the relative stasis of function (syntactically predictable) words.

Type (c) variation includes aberrations of the magnitude that could result in the failure to identify a word under the right conditions and is more likely to drive the larger scale, generation to generation change that may occur in function words. In the case of the more unpredictable content words, type (c) variation would more likely result in communicative failure. In these cases, the listener would not be able to identify the word at all, and could not adopt the new exemplar into the cloud for that particular word. The speaker may be prompted to repeat the word, in which case she would be more likely to produce a clearer form of the word (this would be a case in which purposeful hyperarticulation would occur but only with the intent of successful communication, and not for the means of altering the language to anticipate the listener's needs). In the case of function words, variation of type (c) would still go unnoticed, and such variation away from the intended representation would likely be ignored and/or perceptually restored by the listener. One

could imagine that perhaps type (c) variation for function words could still be correctly associated with the intended word, and that at least *some* of the phonetic detail could make its way into memory storage, resulting in the more profound changes that can occur to function words. However, this would predict more rapid changes to function words over the course of individuals' lives, which would seem not to occur.

While the experiments in this dissertation did not directly consider the effects of type (c) variation, nor did the results show sound changes mimicking function (or predictable) word reduction, the finding that more attention is allotted to unpredictable and content words can be extended to hypothesize how this phenomenon could influence such changes. To account for how type (c) variants could manage to infiltrate the exemplar clouds of function words and eventually result in reduction to the exclusion of content words, we need to consider the role of first language acquisition, particularly with respect to the current findings regarding speech perception and top-down processing. Children who are learning their first language begin the process with little to no higher level knowledge of the language and are unable to use contextual information to aid speech perception (such as that seen in the experiments in this dissertation). When they hear type (a), (b), or (c) variants they may assume that these are the intended (phonological) targets of words, and most of these will be equally eligible to be included as part of the formative exemplar clouds for words they are learning. In other words, if a child hears [kæ̱n], [kɛn], or [kɪn] for the word 'can'--- as either a function or a content word--- he likely assumes that these are equally acceptable variants of the word 'can.' That is to say, at the earliest stages of first language acquisition, children can only conceivably process the auditory signal in the 'how' mode and should be unable to perceptually restore the intended forms of words based on pre-existing knowledge of the language.

This still leaves us with some question as to how any sort of perceptual bias would manifest in phonological distinctiveness between function and content words. Babies and young children, having little to no higher level linguistic knowledge, would show no difference in type (c) errors whether occurring for function or content words, and thus would not show the same behavior as adults who would be able to perceptually restore aberrant function word forms. Here I appeal to some role of speech production in a bias in propagation of variants in function vs. content words though crucially resulting indirectly from a perceptual bias similar to that observed in this dissertation. In this case, at some point in the course of speech development, likely during childhood where language is still being learned, the consequences of type (c) variants in content word production, in which communicative failure occurs, would train these developing speakers to rein in their productions of content words, while more variable productions of function words, going unnoticed by listeners, would never be discouraged. Since the results of the current dissertation indicate that predictable words, and likely by extension function words, are less flexible in adopting new variation, it seems likely that such articulatory training would occur relatively early in speech development, until a point at which the speaker finds a point of the maximum allowance of variation for function and content words that does not affect intelligibility. While this account suggests bias in the variants produced for function and content words, this bias results from the unintended consequences of the bias in speech perception, resulting in essentially three steps: (1) The listener's bias in attending to phonetic details of content words leads to (2) a bias in production learned through articulatory training, which results in (3) this bias being learned by the next generation due to that fact that language acquirers cannot make use of higher level linguistic information. This cycle would repeat until a point at which further reduction of function words *would* result in communicative failure.

Traditionally, many linguists (Meillet 1951, Halle 1962, Kiparsky 1965) have invoked first language acquisition as a source of sound change, such that children are imperfect learners who have to construct their language from the examples they encounter, without direct access to grammatical rules. However, more recent studies caution against arguments in favor of imperfect learning during acquisition as the primary motivator of sound change (particularly in the spread of sound change, rather than in the actuation of it). Foulkes & Vihman (in press) point out two counterarguments to the imperfect acquisition hypothesis in particular. First of all, many of the common patterns encountered in child phonology at best only vaguely resemble examples of actual sound change. Furthermore, commonly cited patterns are not displayed by all children and each type of pattern actually occurs in only a small percentage of children. Secondly, they point out that in cases of known sound changes in progress, that young children (up to two or three years) do not show evidence of extending the pre-existing pattern, but rather follow similar patterns of use to their parents. Older children, however, at a period past about three years, show higher occurrence rates of certain sound changes in progresses. This points to the role of sociolinguistic factors, rather than imperfect learning.

In my model the process of acquisition is an important part of transmitting variation to the next generation. In this case, children are not "imperfect learners" but in fact learn the surface or output pattern of the previous generation with potentially perfect accuracy. Fluent, adult speakers produce a wider range of variation than they may intend to or even notice, due to the restorative effect of using high level linguistic information to aid in word recognition; thus variation occurring in predictable words goes unnoticed and is not phonologized (though it is not clear whether child-directed speech might minimize some of the usual phonetic reduction). The variation produced by this model is represented in figure 5.4 over two generations. Here, individual instances of function and content words (such as /kæn/) are represented by dots which aim for a certain target. Again, we consider (a) variation to be unnoticeably different from the intended forms, (b) variation representing small, noticeable deviation, and (c) forms to represent aberrant forms which would not be recognized if they were words in isolation with no context. The first generation's target forms--- the exemplars stored in memory (figure 5.4, 1a)--- all fall within (a) variation. However, in developmental years in particular (perhaps ages 3-10), random variation occurs for both function and content words and these articulatory exemplars are shown in figure (5.4, 1b). Type (c) variation is penalized in words which are generally speaking less predictable, and figure (5.4, 1c) shows the actual output forms resulting from this training.

Here I assume that the same cloud of exemplars is used in speech perception as in production, and these forms in (5.4, 1a) and (5.4, 2a) serve as the input for production. Some random factor of target *accuracy* determines how closely the output forms (those in fig. 5.4, 1b/2b and 1c/2c) resemble the stored forms in (5.4, 1a) and (5.4, 2a) used as the input for production. Here I used the term *accuracy* following Byun, Inkelas, & Rose's (2016) definition of accuracy as how closely the child matches the adult's target, as opposed to *precision*, defined as the ability to maintain stable representations. Again, I assume that children are capable of learning the adult's target, but will lag for some time in the ability to match that target due to lack of articulatory development, or because they have not yet learned how accurate they need to be for successful communication (or eventually, to speak in a socially acceptable manner). To emphasize this point,

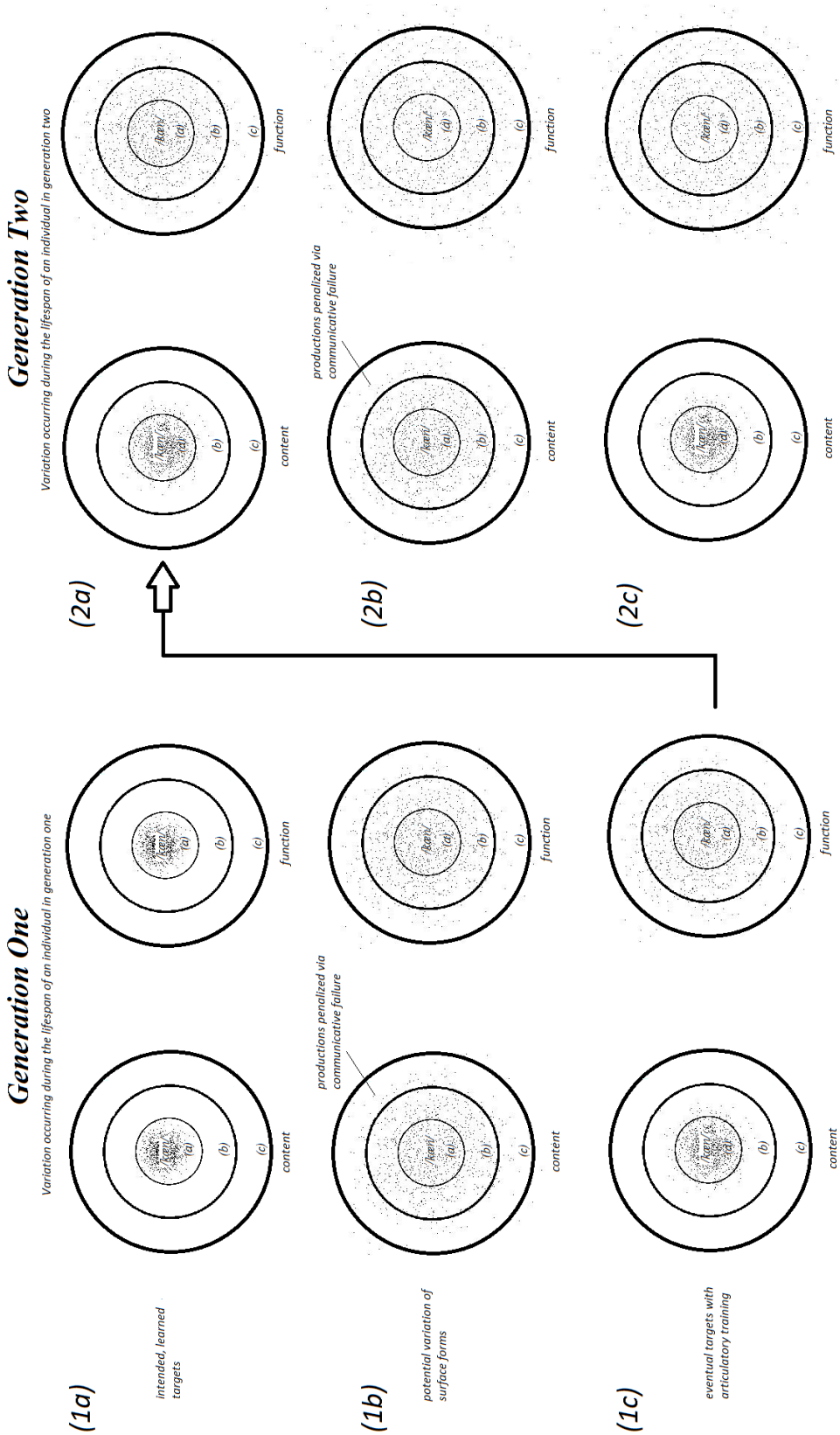

Figure 5.4: Function-content exemplar cloud divergence over two generations

the surface variants produced by the speaker (1b)/(2b)/(1c)/(2c) need not become part of the speaker's exemplar cloud, however whether these variants are accepted by the listener or result in communicative failure will determine whether target *accuracy* increases or decreases; in other words, failure of the listener to recognize a variant results in the speaker "trying harder" to accurately output an intended target. In any case, the output forms of the first generation in (1c) will become the forms faithfully learned by the second generation as their own stored exemplar forms used as the input targets of their own production. The process then repeats itself until such wide variation occurs even for function words such that these would result in communicative failure as well.

Other linguists have also suggested selection biases of word forms that result from differences in auditory processing among two different groups of people. In the present case, I am proposing that variants are propagated and tolerated in predictable words due to that fact that adult speakers make use of 'what' mode listening that involves focusing on meaning rather than segmental or subsegmental processing, whereas children acquiring language would be unable to do so from a starting point of having no high level linguistic knowledge. This results in children noticing and selecting this variation. On the other hand, Yu et al. (2013) proposes that individual differences in speech perception (among fluent, adult speakers) can account for the selection of new variants. His findings show that individuals who rank low on the *Autism Quotient* (AQ) spectrum (showing few if any autistic traits) show weaker effects of compensating for coarticulation and stronger lexical effects on speech perception. On the other hand, high-AQ individuals attend more to the speech signal, showing stronger compensation for coarticulation and weaker lexical effects on speech perception. In other words, high-AQ individuals listen in the 'how' mode more often than low-AQ individuals. Yu claims that low-AQ individuals listening in the 'what' mode will be more likely to select new variants due to inattention to the phonetic signal (although this is the opposite analysis proposed by Lindblom et. al 1995 who suggest that listening in the 'how' mode would result in selecting new variants).

Additionally, Dimov, Katseff, & Johnson (2012) suggest that social factors such as an individual's power within society may affect his or her ability to notice phonetic detail. That is, lacking power "sharpens one's attention to linguistic variation" (Johnson & Garrett:60-61). It could be the case then that people of lower power may pick up on phonetic variation, allotting more processing resources towards 'how' mode listening even when not necessary for speech recognition. Whether or not these new details are adopted would certainly depend on sociolinguistic factors, such as the social identity that may become attached to the particular variants. In any case, higher power individuals would not be able to adopt new variants that escape their attention entirely.

In contrast, the current proposal considers the role of language acquirers (infants and young children) as opposed to fluent adult speakers. Much of the variation produced in predictable speech would likely go unnoticed by fluent, adult speakers who can tap into contextual knowledge without relying on minute variation in the auditory signal. The model I propose only addresses the difference in the development of sound changes in predictable versus unpredictable speech. This leaves open many questions concerning the mechanisms of sound change as it may occur for unpredictable, open class word categories. While generally speaking, listeners give more attention to unpredictable and content words, and less attention to predictable and function words, it does not mean that no phonetic attention is given to predictable words as certain individuals may in

some cases devote more processing to the auditory signal than is necessary purely for the goal of speech recognition (there are clearly other goals in speech, such as conveying identity, understanding emotional cues, etc., which could require more processing of the auditory signal). Additionally, the act of perceiving certain phonetic variation does not necessarily lead to the adoption of it, where the literature seems to point to sociolinguistic factors as either inhibiting or facilitating the spread of variation (Babel 2010, 2012). In any case, the current model seeks to explain the differences in sound change that affect function and content words, and this model does not rely on perceptual or social differences among adult speakers, but also does not in any way challenge their possible roles in other aspects of sound change models. Future research in language acquisition (including second language acquisition), individual differences, and social factors in speech perception will help understand the relationships between these factors and how they interact in a more comprehensive model of sound change.

5.4. Conclusion and Future Research

The results of this dissertation have shown that (1) word predictability modulates the listener's attention of phonetic details, that (2) this perceptual bias influences the production of these words and (3) in observing listeners' perceptual behavior with respect to function and content words, predictability-modulated phonetic attention is a relevant factor in shaping the exemplar clouds of words and ultimately guiding sound change. Despite these results, a few important questions remain concerning some elements of the speech perception and sound change models proposed. Additionally, extensions of this research into other domains of linguistics may corroborate the findings and lead to a broader theory of the interaction between the expectations of past experience and perception.

5.4.1. The effect of subsequent context

One central question left unresolved from this research is whether subsequent context affects speech perception, and if so, how this might work differently from the effect of preceding context. The discrimination experiments of chapter two and the imitation experiments of chapter three both revealed a significant effect of preceding context whereas no effect was found for subsequent context, which was at odds with findings in the literature. At the same time, chapter four showed that listeners devoted more phonetic attention to content words. In this case, listeners would not be able to anticipate the part of speech of the next word, thus this perceptual bias would have to occur after the word was heard, and possibly not until one or two words later when part of speech could be determined. The most likely reason for the failure to find an effect of subsequent context in chapters two and three may be that the "confirming" words--- the ones following the target word that make it predictable--- came outside the perceptual window of time during which this effect would occur. A few possible ways of overcoming this issue in the discrimination and imitation experiments would be to make the target word predictable or confirmed by the *immediately* following word. This might only be possible in the case of compound words. Secondly, this research could be expanded to consider morphemes as well as words. A possible line of experiments could involve determining whether listeners allot equal attention to predictable prefixes as they would for predictable suffixes, in which both would be predictable based on the word root which may occur within the perceptual window necessary to induce this effect.

5.4.2. The effect of unexpected context

A related question is how listeners' attention would be modulated by *unexpected* context, that is, contrary to what was expected. In my experiments, I considered only target words that were predictable ('Kings and queens live in *castles*') as opposed to ones in which the listener would form no expectation at all ('The woman suddenly saw the *castles*'). However, it is not clear how listeners would react when hearing something like 'Kings and queens live in *canteloupes*.' In this case, the listener would be expecting to hear 'castles,' and because it would thus be predicted, we would expect less attention to the phonetic details perhaps regardless of what the word ends up being. However, if listeners showed an equal or greater amount of phonetic attention to unexpected words, it would suggest that they could, during some short period of time, revisit the auditory signal while it is held in working memory. If this is the case, it would provide more evidence that all details of the auditory signal are held in working memory (though not processed or stored) for a short window of time for further processing if necessary.

5.4.3. The role of language acquisition

One possibly controversial aspect of the model proposed for sound change is the role of first language acquisition in the spread of new sound changes. I contend that my model does not rely on "imperfect learning," such that infants and young children simply have a "close enough" approach to learning linguistic patterns, opting to give up on fine tuning their language once reaching a point of communicability. On the contrary, my model relies on language acquirers being rather exceptional at learning and imitating the pronunciation of words, but merely proposes that they will be unable to access the underlying intentions of other speakers and will only be able to learn the surface outputs of their productions. In other words, upon beginning to learn their language their processing of the speech signal is entirely bottom-up and they cannot use top-down processing and contextual clues to restore the intended acoustic signal. Nevertheless, lab research of this question is needed to corroborate this hypothesis. While the study of infant speech perception is possible (Eimas 1985, etc.), research in second language acquisition may also serve to address this question. While second language learners' perception of the sounds of a new language has been shown to be influenced by their first language (Maye 2007), it would still be of value to understand how listeners develop a 'what' mode of listening, incorporating and developing elements of top-down processing as their fluency increases. Such research may also provide insight into how such learners' early bottom-up only perception affects the interpretation of phonetic and phonological patterns.

5.4.4. Individual differences

In addition to the role of first language learners in adopting sound changes, it is also not clear how individuals may differ either cognitively or socially and how this affects speech perception and sound change. Following Yu et al. (2013), it is possible that cognitive differences, such as those relating to the autism quotient, may be relevant in whether speakers attend more or less to the raw auditory signal and that this may be an additional factor in the spread of sound change. All the analysis of the data in this dissertation focused on group behavior, and while subject variation was included in the statistical models, it is not clear whether listeners may differ, even greatly, in the relatively weighting given for top-down vs. bottom-up processing. Likewise, this weighting could also be influenced by an individual's social position and their level of power within society (Garrett & Johnson 2012). Future research will look more closely at variation

among listeners in how the contextual effects of word predictability interact with phonetic attention.

5.4.5. Syntax-Phonetics Perceptual Interface and the Development of Prosody

Other questions left unaddressed include how the listener makes calculations about predictability in an online system of modulating the trade-off between top-down and bottom-up processing, how regular is this process and how mentally taxing would it be? However, some aspects of my findings suggest some portion of listening modulation could be achieved via a syntax-phonetics perceptual interface mapping which is the default plan used in the perception of fluent speech in a particular language. For example, the results of the error detection task in chapter four reveal a sharply distinct pattern of behavior for listening to function and content words (as opposed to smaller differences in listening to low and high predictability words), paralleling the often categorical phonological differences that emerge in function and content words, which hints at the possibility that listeners may utilize categorically different and pre-determined modes of listening to process different parts of the morphosyntactic structure (as opposed to calculating syntactic predictability online, which seems to be the case for semantic predictability). Such a perceptual mapping could facilitate the process of speech perception, providing listeners with a perceptual plan that targets high and low predictability morphosyntactic regions based on knowledge of the grammar of one's language.

Fitting in with the observations proposed for the models of speech perception and sound change in this chapter, we could imagine that the listener first processes chunks of speech at a very broad level. Prosodic cues and contextual information may first aid the listener in determining word boundaries and at least basic part of speech categorization, followed by building a syntactic structure of the speech-chunk (which could occur alongside--- before or after--- recognition of particular words within the speech-chunk). Once the listener has constructed some portion of a syntactic structure fitted onto the speech-chunk, this may guide the speech perception faculties, directing more attention to the phonetic details of content words. A similar process could occur for word recognition, such that more processing of the phonetic signal may automatically occur for word onsets than for word codas (or thinking non-segmentally, the beginning vs. the end of a particular word).

Such a perceptual mapping would facilitate the speech perception process such that listeners would not have to make as many online calculations concerning word predictability and other semantic considerations. This in turn could drive the development of intonational systems, whereby lower phonetic attention to certain regions of the syntax would result in more freedom of articulatory variation, while this variation would in turn be selected and interpreted as an intended part of the intonational system by first language acquirers. Intonation could, in turn, influence segmental changes, such as the development of onset /ð/ in function words in English. Future research will continue to investigate this hypothesis.

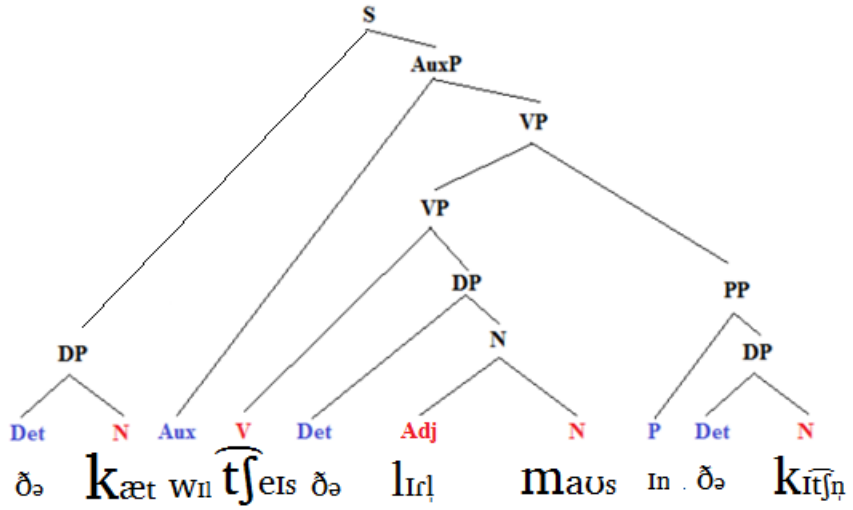

Figure 5.5: Syntax to phonetics attentional mapping interface. Size of the phoneme relates to the amount of phonetic processing that may be given to that portion of the speech signal (with more attention given to content words and onsets) for the sentence 'The cat will chase the little mouse in the kitchen.'

5.4.6. Applications to other related linguistic subdisciplines

Future research may also apply these findings to other disciplines within linguistics. For instance, "predictability" could be considered more broadly and in terms of sociolinguistic information not directly relevant to linguistic structure. For example, current research in sociolinguistics suggests that listeners use expectations about the speakers to prime certain aspects of speech perception (Drager 2010, McGowan 2015). For example, when recognizing an individual as being of a certain ethnicity, gender, or other social identity, a listener might form expectations about the speaker's accent or vocal characteristics which in turn causes a shift in perceptual boundaries between sounds (such as expecting certain vowel formant ratios from men as opposed to women, relating to vocal tract length). Thus, my current model of how predictability influences speech perception could be extended to social predictability or expectation, to investigate how non-linguistic factors influence attention. For example, if a speaker's voice is somehow unexpected (accent or voice not matching expectations or stereotypes of a person's face) would the listener attend more closely to the auditory signal, such as is the case with unpredictable speech?

Additionally, this research is well-suited for applications in neurolinguistics and the study of the structure and functioning of the brain. Observing the electrocortical responses and event-related potentials (ERPs) output by the brain would give neurological insight to how predictability-modulated phonetic attention occurs. Kutas (1997:385) states that the N400 component of the ERP indicates "semantic expectancy and its consequences for online processing." An EEG experiment measuring this electrocortical activity during discriminability or imitation tasks might find a correlation between the magnitude of this component and the accuracy of acoustic discrimination or the degree of phonetic accommodation. This may feed into research which incorporates these findings with Hickok and Poeppel's (2004, 2007) dual stream model of speech perception.

The finding that contextual predictability modulates phonetic attention can also be applied to other research which connects speech perception and sound change. For example, Ohala's (1981, 1983) types of misperception--- hypocorrection and hypercorrection--- involve the listener failing to undo natural coarticulatory effects, or incorrectly determining that a surface form of a sound has resulted from coarticulation. Future research should consider how these processes of misperception might interact with contextual predictability. Here the hypothesis would be that listeners would be more likely to ignore coarticulation with abundant context and that misperception of the intended signal would be more likely in the lack of context.

Another line of research should consider whether listeners attend to and adopt different types of phonetic innovations differently. In the imitation experiments in this dissertation, subjects were shown to attend and adopt variation perceived in unpredictable words, but these innovations involved fortition, in lengthening VOT and raising pitch. For the purposes of understanding the spread of lenition, it would be interesting to see if subjects behave similarly with lenitive innovations. Following Hay et al.'s (2015) recent research showing that non-lenitive sound changes spread to less frequent words first, it may then be the case that non-lenitive changes would be more likely to affect unpredictable speech, while lenition may actually be in some way blocked from spreading to unpredictable or low frequency words.

Finally, more research into the effects of visual context and its relationship to semantic context might reveal how these different high-level perceptual effects interact. For example, does the presence of visual cues (e.g., seeing a picture of a castle before hearing 'castles' in an unpredictable context) or reading a sentence before hearing it, have a similar effect as does word predictability? Also, how do visual linguistic cues such as seeing the movement of the articulators interact with linguistic context? In other words, would a McGurk type of perceptual effect (McGurk and McDonald 1976), such that seeing the articulators overrides the perception of the auditory signal, override or be subordinate to an opposing effect of phonemic restoration induced via word predictability?

5.5. Conclusion

This dissertation presents the results of several experiments providing additional support for the theory that contextual knowledge influences speech perception, providing insight into many details of this complex connection. Extending this hypothesis I have shown the relevance of this phenomenon to speech production and sound change. A wide range of promising future research programs shows a far-reaching applicability of these findings to many subdisciplines of linguistics and other related fields where this research may be relevant in broad investigations into our scientific understanding of memory, experience, perception, and change.

Bibliography

Andreewsky, E. and Seron, X. (1975). Implicit processing of grammatical rules in a classical case of agrammatism. Cortex XI, 379-390.

Altenberg, B. (1987). *Prosodic patterns in spoken English*. (*Studies in the correlation between prosody and grammar for text-to-speech conversation* 76). Lund: Lund University Press.

Alyett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31-56.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(04), 437-456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177–189.

Babel, M., and Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech* 55(2): 231-248.

Baddeley, A. D., and Hitch, G. (1974). Working memory. In G. H. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Bates, D., Hess, T., and Hilbert, V.. (1994). Lushootseed Dictionary. Seattle: University of Washington Press.

Bates, E. and Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. In G. Altmann (Ed.), Special issue on the lexicon, *Language and Cognitive Processes, 12(5/6)*, 507-586.

Beckman, J. (1997). Positional faithfulness, positional neutralization and Shona vowel harmony. Phonology 14(1), 1-46.

Beckman, J. (1998). *Positional Faithfulness.* Doctoral dissertation. Amherst: University of Massachusetts.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, 113:1001-1024.

Bell, A., Brenier, J., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language* 60:1, 92-111.

Biassou, N., Obler, L. K., Nespoulous, J.-L., Dordain, M., and Haris, K. S. (1997). Dual processing of open- and closed-class words. Brain and Language 57:360-373.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Bloomfield, L. (1933). *Language*. New York: Holt.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 6.0.14.

Bradley, D. (1978). Computational distinctions of vocabulary types. Unpublished Ph.D. dissertation, MIT, Cambridge.

Brown, E. L., (2013). Word classes in phonological variation: conditioning factors or epiphenomena? In *Chad Howe, Sarah Blackwell and Margaret Lubbers Quesada (eds.) Selected Proceedings of the 15th Hispanic Linguistics Symposium*. Somerville, MA: Cascadila Proceedings Project, pp. 179–186.

Brugmann, K. 1892. *A Comparative Grammar of the Indo-Germanic Languages – A Concise Exposition of the History:* vol. III – Morphology, part 2. London: Kegan Paul, Trench, Trübner and Co.

Bybee, J.. Perkin, R. D. and Pagliuca, W.. (1994). *The Evolution of Grammar*. Chicago: University of Chicago Press.

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (eds.), *Usage-based models of language*. Stanford: CSLI. 65-85.

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

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14, 261-290.

Bybee, J. (2014). 'Does grammar impact sound change?' Presentation given at the 3rd Biennel Workshop on Sound Change: Sound Change in Interacting Human Systems. University of California, Berkeley. May 29, 2014.

Bybee, J. & Hopper, P. (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Byun, T., Inkelas, S., & Rose, Y. (2016). The a-map model: Articulatory reliability in child-specific phonology. Language, 92(1), 141-178.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Cole, R. and Jakimik, J. (1980). A model of speech perception. In Cole, R. (ed.) *Perception and Production of Fluent Speech*, (pp. 133-163). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cole, R., Jakimik, J., and Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America,* 64, 44-56.

Collins, A. & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review* 82(6):407-428.

Connine, C.M., Blasko, D.G., and Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraint. *Journal of Memory and Language* 30(2), 234-250.

Craats, I. van de (2000). *Conversation in the acquisition of possessive constructions*. Doctoral Dissertation, Tilburg University.

Craik, F., and Lockhart, R. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11, 671–684.

Cutler, A., and Foss, D. (1977). On the role of sentence stress in sentence processing. *Language and Speech* 20, 1-10.

Cutler, A., and Norris, D. (1979). Monitoring sentence comprehension. In *Psycholinguistic studies presented to Merrill Garrett* (ed. W. E. Cooper and E. C. T. Walker), pp. 113-134. Erlbaum.

Davies, M. (2008-) *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

Dimov, S., Katseff, S. & Johnson, K. (2012). Social and personality variables in compensation for altered auditory feedback. In M.-J. Solé & D. Recasens (Eds.), The initiation of sound change: Perception, production, and social factors (pp. 185–210). Amsterdam, the Netherlands: John Benjamins.

Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass* 4(7): 473-480.

Eimas, P. (1985). The perception of speech in early infancy. *Scientific American* 252: 46–52.

Foulkes, P., & Vihman, M. M. (in press). Language acquisition and phonological change. In P. Honeybone & J.C. Salmons (eds.) *The Handbook of Historical Phonology.* Oxford: OUP.

Foss, D.J., Harwood, D.A., Blank, M.A. (1980) Deciphering decoding decisions: Data and devices. In: Cole, R.A., editor. *Perception and production of fluent speech*. Hillsdale, N.J: Erlbaum; 1980.

Gahl, S., Yao., Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66, 789-806.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, Vol 6(1), Feb 1980, 110-125.

Gardner, H. and Zurif, E. 1976. Critical reading of words and phrases in aphasia. Brain and Language, 3, 173-190.

Garrett, A. & Johnson, K. (2012). Phonetic bias in sound change. In Alan Yu (ed.), *Origins of sound change: Approaches to phonologization* (pp. 51-97). Oxford: Oxford University Press.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 17th International Congress of Phonetic Sciences*, 49-54.

Goldinger, S. D. & Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics* 31, 305-320.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review* 87, 1-51.

Halle, M. (1962). Phonology in generative grammar. *Word* 18:54-72.

Halle, M. & Stevens, K. N. (1959). Analysis by synthesis. In W. Wathen-Dunn & L. E. Woods (eds.) *Proceedings of the seminar on speech compression and processing*. USAF Camb. Res. Ctr. 2: Paper D7.

Halle, M. & Stevens, K. N. (1962). Speech recognition: A model and a program for research. *IRE Transactions of the PGIT, IT-8*, 155-159.

Hay, J., Pierrehumbert, J., Walker, A., LaShell, P. (2015) 'Tracking word frequency effects through 130 years of sound change.' *Cognition* 139, 83-91.

Hautus, M. J. (1995) Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments, & Computers* 27:46.

Heath, J. S. (2014). Accommodation can lead to innovated variation. *UC Berkeley Phonology Lab Annual Report*, 119-145.

Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.

Hickok, G. and Poeppel, D. (2007). Opinion – The cortical organization of speech processing. *Nature Reviews and Neuroscience*, 8(5), 393-402.

Hillenbrand, J. M. and Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America 105*, 3509-3523.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93(4), 411-428.

Honorof, D. N., Weihing, J., & Fowler, C.A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, 391(1), 18-38.

Hooper, J. B. (1976). *An Introduction to Natural Generative Phonology*. New York: Academic Press.

Jarvella, R. J. and Meijers, G. (1983). Recognizing morphemes in spoken words: Some evidence for a stem-organized mental lexicon. In Giovanni B. Flores d'Arcais and Robert J. Jarvella (eds.), *The process of language understanding.* J. Wiley and Sons, Ltd., New York. 81-112.

Jetté, Jules and Eliza Jones (authors) and James Kari (ed.). (2000). *Koyukon Athabaskan Dictionary*. Fairbanks: Alaska Native Language Center.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America* 88:642-654.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In Johnson & Mullennix (eds.) *Talker Variability in Speech Processing* (pp. 145-165). San Diego: Academic Press.

Jurafsky, D., Bell, A., Gregory, M., Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J., and Hopper, P., (eds.) *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins. 229-254.

Kabak, B. and Schiering, R. (2006). The phonology and morphology of function word contractions in German. *The Journal of Comparative Germanic Linguistics* Vol. 9, No. 1, pp. 53-99.

Kawashima, T. (1998). The enhancing effect of subsequent context in the perception of the sentence-initial word. *Journal of the Acoustical Society of America* 103(5), 2771.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. *Phonetic Linguistics,* ed. V. Fromkin, Academic Press, 115-132.

Kiparsky, P. (1965). *Phonological Change.* PhD. dissertation, MIT.

Kiparsky, P. (1995). The phonological basis of sound change. In *The handbook of phonology theory* (ed. J. A. Goldsmith), pp. 640-670. Blackwell, Oxford.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. In *Perception and production of fluent speech* (ed. R. A. Cole), pp. 243-288. Erlbaum, Hillsdale, N.J.

Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the functions of various language structures. Psychophysiology, 34(4), 383-398

Labov, W. (1981) Resolving the Neogrammarian controversy. *Language*, 57:267-308.

Lehiste, I., & Peterson, G. E. (1961). Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America 33*, 268-277.

Lewandowski, N. (2012). Automaticity and consciousness in phonetic convergence (abstract). *Proceedings of the Listening Talker Workshop,* 71.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review 74*, 431-461.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. J. and Marchal, A. (eds.) *Speech Production and Speech Modeling* (pp. 403-439). Dordrecht: Kluwer.

Lindblom, B., Guion, S., Hura, S., Moon, S-J., and Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica,* 7(1), 5-37.

Manker, J. (2012). *An Acoustic Analysis of Stem Prominence in Hän Athabascan*. Unpublished Master's Thesis, University of Alaska Fairbanks.

Manker, J. (2015). Phonetic Sources of Morphological Patterns in Sound Change: Fricative Voicing in Athabascan. Unpublished manuscript.

Mann, V. A. & Repp, B. H. (1980). Influence of vocalic context on the [s]-[ʃ] distinction. *Perception and Psychophysics,* 28, 213-228.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology,* 10, 29-63.

Marslen-Wilson, W. D. (1987) Functional parallelism in spoken word-recognition. *Cognition*25(1), 71-102.

Martinet, A. (1960). Éléments de linguistique Générale. Paris: Armand Colin.

Maye, J. (2007). Learning to overcome L1 phonological biases. *Proceedings of the 17th International Congress of Phonetic Sciences*, 63-66. Retrieved from: http://www.icphs2007.de/conference/Papers/1783/1783.pdf

McGowan, K. B. (2015) Social Expectation Improves Speech Perception in Noise. *Language and Speech* 58(4), 502-521.

McGurk, H. & McDonald, J. (1976). Hearing lips and seeing voices. *Nature (London), 264*, 746-748.

May, J. (1976). Vocal tract normalization for /s/ and /ʃ/. *Haskins Laboratories: Status Report on Speech Research, SR-48*, 67-73.

Maye, J. (2007). Learning to overcome L1 phonological biases. *Proceedings of the 17th International Congress of Phonetic Sciences*, 63-66. McCarthy, J. J. (2005). Optimal paradigms. In Laura Downing, Tracy Alan Hall, and Renate Raffelsiefen, eds., Paradigms in Phonological Theory. Oxford: Oxford University Press. Pp. 170-210.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology* 18, 1-86.

Meillet, A. (1951). *Linguistique Historique et Linguistique Générale*, vol. 2. Paris: Klincksieck.

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America 25*, 114-121.

Miller, George A.; Heise, George A.; Lichten, William. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, Vol 41(5), May 1951, 329-335.

Minkova, D., and Stockwell, R. (2005). Clash avoidance in morphologically derived words in Middle English. (Why [-hʊd] but [-dəm])? In: Nikolaus Ritt and Herbert Schendl (eds.) *Rethinking Middle English. Linguistic and Literary Approaches,* 263-280. Bern: Peter Lang.

Mitterer, H. & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition,* 109(1), 168-173.

Mixdorff, H., Cole, J., & Shattuck-Hufnagel, S. (2012). Prosodic similarity: Evidence from an imitation study. *Proceedings of the 6th International Conference on Speech Prosody, Sp 2012, 2, 571-574.*

Morton, J., & Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior* 15, 43-51.

Namy, L., Nygaard, L., Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21(4): 422-432.

Neville, H. J., Mills, D. L., and Lawson, D. S. 1992. Fractioning language: different neural subsystems with different sensitivity periods. Cerebral Cortex 2, 244-258.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics,* 39(2), 132–142.

Nye, P. & Fowler, C. (2003). Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics,* 31(1), 63–79.

Ohala, J. (1981). The listener as the source of sound change. In Masek, C., Hendrick, R. & Miller, M. (eds.) *Papers from the parasession on language and behavior* (pp. 178-203). Chicago: Chicago Linguistics Society.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P. (ed.), *The Production of Speech* (pp. 189-216). New York: Springer-Verlag.

Ohala, John (1989). Sound change is drawn from a pool of synchronic variation. In Leiv E. Breivik and Ernst H. Jahr, (eds.) Language Change: Contributions to the study of its causes. [Series: Trends in Linguistics, Studies and Monographs No. 43]. Berlin: Mouton de Gruyter.

Pardo, J. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America,* 119(4), 2382-2393.

Penke, M. 2012. The acquisition of inflectional morphology. University of Cologne.

Peterson, G. E. (1961) Parameters of vowel quality. *Journal of Speech and Hearing Research 4*, 10-29.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language* 60: 320-342.

Phillips, B. S. (2006). *Word Frequency and Lexical Diffusion*. New York: Pargrave MacMillan.

Pickering, M. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences,* 27(2),169-189.

Pierrehumbert, J. (2002). Word-specific phonetics. In Gussenhoven, C., and Warner, N. (eds.), *Laboratory Phonology VII* (pp. 101-139). Berlin: Mouton de Gruyter. Mouton de Gruyter, Berlin.

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics* 34, 516-531.

Pierrehumbert, J.B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2, 33-52.

Polišenká, D. 2010. Dutch children's acquisition of verbal and adjectival information. Ph.D. dissertation, University of Amsterdam.

Pollack, I. and Pickett, J. (1963). The intelligibility of excerpts from conversation. *Language and Speech*, 6, 165-171.

Remez, R. E. (2003). Establishing and maintaining perceptual coherence: Unimodal and multimodal evidence. *Journal of Phonetics* 31: 293-304.

Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology*, 110(4), 474-494.

Samuel, A. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science* 12, 348-351.

Sancier, M. and Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421-436.

Schnoebelen, T. & Kuperman, V. (2010). Using Amazon Mechnical Turk for linguistic research. *Psihologija* 43(4), 419-428.

Shi, R., Gick, B., Kanwischer, D., Wilson, I. (2005). Frequency and category factors in the reduction and assimilation of function words: EPG and acoustic measures. *Journal of Psycholinguistic Research* 34:341.

Shockley, K., Sabadini, L., and Fowler C. (2004). Imitation in shadowing words. *Perception and Psychophysics*, 66, 422–429.

Sprouse, R. and Johnson, K. Forthcoming. Berkeley Phonetics Machine. To be presented at Interspeech 2016.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2), 73-74.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computer* 31*:*137.

Strand, E. A. (2000). *Gender Stereotype Effects in Speech Processing*. PhD Dissertation, Ohio State University.

Sussman, H. M. (1986). A neuronal model of vowel normalization and representation. *Brain and Language 28*, 12-23.

Szostak, C. M., and Pitt, M. A. (2013). The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics 75*, 1533-1546.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly 30,* 415–433.

Talkin, D. and Lin, D. (1996). Get_f0 Online Documentation. ESPS/Waves release 5.31. Entropic Research Laboratory.

Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, 37(3), 276-296.

Urbanczyk, S. 2011. Root-affix asymmetries. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice (eds.), *Blackwell companion to phonology, volume IV: phonological interfaces.* Malden, MA: Blackwell, 2490-2515.

Wang, William S.-Y. (1969). 'Competing changes as a cause of residue', *Language* 45: 9-25.

Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392-393.

Warren, R. & Sherman, G. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics 16*: 150.

Willerman, R. (1994). The phonetics of pronouns: Articulatory bases of markedness. Doctoral dissertation, University of Texas at Austin.

Yu, A., Abrego-Collier, C., and Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and 'autistic' traits. *PLOS ONE 8*(9):e74746.

Yu., A., and Lee, H. (2014). The stability of perceptual compensation for coarticulation within and across individuals. A cross-validation study. *Journal of the Acoustical Society of America 136*(1):382-388.

Yuan, J. and Liberman, M. (2008). Speaker identification on the SCOTUS corpus. Proceedings of Acoustics '08.

Zellou, G., Scarborough, R., & Nielsen, K. (2013). Imitability of contextual vowel nasalization and interactions with lexical neighborhood density. *Proceedings of Meetings on Acoustics* 19(1).

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. New York: Addison-Wesley.

Appendix A

20 target NP sentences:

1) The pail is full.
2) The old man smokes a pipe.
3) The dogs licked their paws.
4) The porch is white.
5) The nations made a pact.
6) The fisherman sees the pond.
7) The boy watched the pandas.
8) The pelican flew over the beach.
9) The pope led the service.
10) A pentagon has five sides.
11) The man found the portal.
12) Eric bought a pillow.
13) The pagans worshipped many gods.
14) John is a patron of the arts.
15) The chef chopped the parsley.
16) The palace is magnificent.
17) Your brother is a pest.
18) Tom saw a panther in the mountains.
19) Susan is a poet.
20) The pantry is empty.

20 novel NP sentences:

1) The pulp is thick.
2) The girl broke the pane.
3) The pears are delicious.

4) The peas are green.

5) The farmer harvests the peaches.

6) They climbed the peak.

7) William is a peasant.

8) Singing is her passion.

9) Emily is wearing a parka.

10) The pasture is very large**.**

11) The woman sees the path.

12) The pauper stole bread to eat.

13) Mary wants to see the pageant this year.

14) The soldier needs the powder for his gun.

15) The policy is strict.

16) The porcupine is eating twigs.

17) The pigeon is walking through the city.

18) Each student has a partner for the game.

19) The children used the paste.

20) Anna moved the pawn.

20 novel VP sentences:

1) Bill published three articles this month.

2) Kathy poached the egg.

3) Steve polished his shoes.

4) Sally purchased a new car.

5) The boy panicked about the test.

6) The governor pardoned the criminal.

7) Jane pondered the meaning of life.

8) The Vikings pillaged a monastery.

9) The woman pampers her dog.

10) Many citizens perished from the famine.
11) The referee penalized the team.
12) The boy poked the dog.
13) Martha paid the cashier.
14) The children popped the bubbles.
15) Steven paused the movie.
16) Sally pinned the banner up.
17) The woman pumped the tire up.
18) The workers paved the street.
19) Susan patted the dog on the head.
20) Rita punished her children.

Appendix B

Predictable stimuli for preceding context:

1) Pennies are made out of copper. (25/34, 73.5%)
2) The printer needs a new ink cartridge. (28/34, 82.4%)
3) When there is a blackout we light candles. (32/34, 94.1%)
4) A spreadsheet has rows and columns. (30/34, 88.2%)
5) The pioneers made log cabins. (28/34, 82.4%)
6) Harvard University is a prestigious college. (17/34, 50%)
7) The largest library in the U.S. is the Library of Congress. (27/34, 79.4%)
8) Root vegetables include parsnips and carrots. (9/34, 26.5%)
9) John stacked the plates and put them away in the cupboard. (16/34, 47.1%)
10) Sauerkraut is made from fermented cabbage. (29/34, 85.3%)
11) Kings and queens live in castles. (25/34, 73.5%)
12) Nobody puts baby in the corner. (17/34, 50%)
13) The dog is wearing a collar. (22/34, 64.7%)
14) The vampires are sleeping in coffins. (26/34, 76.5%)
15) Mary prefers milk and sugar in her coffee. (23/34, 67.6%)
16) The witch is brewing a potion in her cauldron. (15/34, 44.1%)
17) Butterfingers are my favorite type of candy. (26/34, 76.4%)
18) Bob spilled wine on the white carpet. (15/34, 44.1%)
19) Debbie drinks milk straight from the carton. (19/34, 55.9%)
20) The windows are covered with curtains. (10/34, 29.4%)
21) In the desert, we rode on camels. (25/34, 73.5%)
22) The great barrier reef has beautiful coral. (14/34, 41.1%)
23) A young cat is called a kitten. (33/34, 97.1%)
24) T-shirts are made out of cotton. (27/34, 79.4%)
25) Five nickels equal one quarter. (26/34, 76.5%)

26) In India, the most deadly snake is the cobra. (25/34, 73.5%)

27) The refrigerator and stove are in the kitchen. (32/34, 94.1%)

28) The needle always points north on a compass. (33/34, 97.1%)

29) We looked up at the night sky and saw Halley's comet. (31/34, 91.2%)

30) In Arizona, tourists come to the Grand Canyon. (30/34, 88.2%)

31) The girl dressed as a ghost for her Halloween costume. (18/34, 52.9%)

32) A one-person canoe is sometimes called a kayak. (19/34, 55.9%)

33) France, Spain, and Germany are European countries. (31/34, 91.2%)

34) In the desert, Anna saw a spiky Saguaro cactus. (18/34, 52.9%)

35) Dave likes hotdogs with mustard and ketchup. (17/34, 50%)

36) The photographer put film in her camera. (32/34, 94.1%)

37) Cats are felines while dogs are canines. (28/34, 82.4%)

38) The historian dug up a time capsule. (19/34, 55.9%)

39) At the circus they shot a man out of a cannon. (29/34, 85.3%)

40) A prism breaks light into separate colors. (34/34, 100%)

41) Bill woke up after twenty years in a coma. (26/34, 74.5%)

42) Six people were elected to city council. (29/34, 85.3%)

43) The president's words were taken out of context. (28/34, 82.4%)

44) Amanda won the pie eating contest. (28/34, 82.4%)

45) The horror movie is about a serial killer. (32/34, 94.1%)

46) Kathy wears glasses instead of contacts. (30/34, 88.2%)

47) The pitcher threw the ball to the catcher. (20/34, 58.8%)

48) The women are singing Christmas carols. (30/34, 88.2%)

49) Matt doesn't like riding big roller coasters. (27/34, 79.4%)

50) The Wizard of Oz is set in Kansas. (18/34, 52.9%)

51) Boxes are made out of cardboard. (27/34, 79.4%)

52) Separate words in a list with commas. (20/34, 58.8%)

53) The lion asked the Wizard of Oz for courage. (24/34, 70.6%)

54) A popular spice in Indian cuisine is curry. (14/34, 41.2%)

55) Sarah's favorite video game is Mortal Kombat. (34/34, 100%)

56) The team was down but made a huge comeback. (20/34, 58.8%)

57) Painters paint on a material called canvas. (29/34, 85.3%)

58) Jim baked chocolate chip cookies. (29/34, 85.3%)

59) Panthers or mountain lions are also called cougars. (14/34, 41.2%)

60) Art and music are part of a nations' culture. (22/34, 64.7%)

Stimuli with contextually *unpredictable* target words, preceding context:

1) The next word is 'copper.'

2) What I need is a cartridge.

3) The man is looking at the candles.

4) Chris said he saw the columns.

5) Joe turned and saw the cabins.

6) Linda is thinking about college.

7) Albert is always thinking about congress.

8) Molly's favorite thing is the world is carrots.

9) Ron decided to look at the cupboard.

10) Mary wishes she had more cabbage.

11) The woman suddenly saw the castles.

12) The dog stopped and looked at the corner.

13) The first word on the page is 'collar.'

14) The first thing Mary saw was the coffins.

15) My grandmother needs more coffee.

16) Matilda said that she wants a cauldron.

17) I have been thinking a lot about candy.

18) Bob is always thinking a lot about carpet.

19) Debbie sat and stared at the carton.

20) Wanda says she wants curtains.

21) Nick turned around and saw the camels.

22) Barbara turned the page and saw the coral.

23) The first thing I saw was a kitten.

24) Ashley really needs more cotton.

25) Paul opened his eyes and saw a quarter.

26) The woman kept thinking about the cobra.

27) Jack sat thinking about his kitchen.

28) The first thing the man thought of was a compass.

29) The people looked and saw a comet.

30) Don knows there is a canyon.

31) The first thing Julie saw was her costume.

32) Everybody began to stare at the kayaks.

33) John thinks a lot about different countries.

34) Anna looked to the right and saw a cactus.

35) Dave's favorite thing in the world is ketchup.

36) Lisa decided that she wants a camera.

37) The word at the bottom of the page is 'canines.'

38) The last word of the book is 'capsule.'

39) I know that the man has a cannon.

40) The boy said he saw lots of colors.

41) The last word bill said was 'coma.'

42) The man looked directly at the council.

43) Lately I've been thinking a lot about context.

44) Amanda wants to have a contest.

45) Frank has been thinking a lot about the killer.

46) Cindy wants to get more contacts.

47) The man turned and looked at the catcher.

48) Zack says he really likes carols.

49) The first word Matt spelled was 'coasters.'

50) Courtney likes to think about Kansas.

51) I really need to get more cardboard.

52) The women were talking about commas.

53) Ben wishes that he had more courage.

54) Tom's favorite thing in the world is curry.

55) They have been thinking a lot about combat.

56) What I really need now is a comeback.

57) Eric decided to get more canvas.

58) All day long Jim thought about cookies.

59) The boy decided to look at the cougars.

60) Lately everyone has been discussing culture.

Predictable sentences, subsequent context:

1) A layer of copper covers pennies.

2) The cartridge in the printer is out of ink.

3) The candles melted wax on the birthday cake.

4) There are more columns than rows in this spreadsheet.

5) The cabin in the woods is haunted.

6) The best college in the east is Harvard.

7) The U.S. congress is composed of the House and Senate.

8) The carrots in my garden are long, crunchy, and orange.

9) The cupboard is where we store our clean dishes.

10) A cabbage must be fermented to make sauerkraut.

11) This castle was built for kings and queens to live in.

12) Out of the corner of my eye I saw a shadow.

13) The collar of his shirt was loose around his neck.

14) The coffin held Dracula's sleeping corpse.

15) The coffee beans were ground to make a cappuccino.

16) In a cauldron the witch's potion brewed.

17) I think that candy apples are my favorite Halloween treat.

18) The carpet cleaners can remove the stain from the rug.

19) A carton of eggs is only two dollars.

20) The curtains block all the light from the windows.

21) This camel has one hump but others have two.

22) A coral reef has lots of colorful ocean life.

23) A kitten is a young cat.

24) The cotton gin separates the seeds from the fibers.

25) The quarter-pounder is Tom's favorite hamburger.

26) The cobra is a deadly snake from India.

27) Our kitchen has a stove and a refrigerator.

28) Use a compass to tell where north is.

29) The comet's tail left a long streak in the night sky.

30) This deep canyon was slowly carved by a river.

31) For Sue's costume party you must dress in disguise.

32) A kayak is like a small one-person canoe.

33) Our country has fifty states.

34) A cactus is a spiky desert plant.

35) Dave likes ketchup and mustard on his hotdogs.

36) Lisa's camera needs film to take pictures.

37) Sam knows canines include dogs and wolves.

38) Two capsules or tablets have one dose of medicine.

39) The cannon fired iron balls at the pirate ship.

40) The colors on the flag are red, white, and blue.

41) Bill's coma lasted twenty years before he awoke.

42) Six council members and a mayor were elected.

43) Out of context his words make no sense.

44) This contest's winner ate the most pies in an hour.

45) The number one killer in America is heart disease.

46) Her contact lens fell out of her eye.

47) The catcher receives the ball from the pitcher.

48) Zack knows that carols are traditional Christmas songs.

49) Using coasters keeps cups from staining rings on tables.

50) The Kansas City Royals are Sharon's favorite team.

51) These cardboard boxes come in all sizes.

52) The comma is used to separate words in a list.

53) Katie lacked the courage to stand up to the bully.

54) Tom knows that curry is a popular Indian spice.

55) The combat zone is a deadly place during battle.

56) I didn't have a comeback witty enough for John's insult.

57) The canvas was blank before she started to paint.

58) Her cookie jar was full of snickerdoodles.

59) The cougar can also be called a puma or panther.

60) The culture of Europe is rich in art and music.

Unpredictable target sentence, subsequent context:

1) The first word is 'copper' followed by 'cup.'

2) The cartridge I laid on the table is gone.

3) The candles are lying on the kitchen floor.

4) Chris saw four columns on the hill by the lake.

5) The cabin was seen by the woman.

6) By the college is a small grocery store.

7) The second congress was so much better than the first one.

8) Some carrots were shown on the second page of the book.

9) The cupboard was never completed by John.

10) A cabbage is all that Mary needs to buy today.

11) A castle was seen seven miles to the east.

12) In the first corner is a picture of my mother.

13) A collar was sitting on top of the boxes.

14) A coffin was lying next to the road.

15) The coffee on the table was ignored by the guests.

16) By the cauldron is a bag of carrots.

17) I think that candy is the greatest thing in the whole world.

18) This carpet has been here since 1975.

19) This carton was missing since last Saturday.

20) These curtains are yellow but those are purple.

21) That camel is the one I saw yesterday.

22) Some coral is larger and prettier than others.

23) The kitten is very sad.

24) Some cotton is piled up in a yellow basket.

25) This quarter is heavier than the other one.

26) A cobra was in the center of the picture.

27) My kitchen is off-white but my mother's is yellow.

28) Grab the compass on the tabletop.

29) The comets were discussed by everyone in town.

30) This canyon is my favorite place to visit.

31) Lisa's costume will be discussed by everyone.

32) My kayak is the most expensive thing I own.

33) This country is my favorite.

34) A cactus is on the side of the road.

35) That ketchup was some of the best I've had.

36) This camera was very expensive.

37) Bob knows 'canines' is the first answer.

38) These capsules are too big to fit in my minivan.

39) The cannon was the very first picture on the page.

40) These colors are better than the other ones.

41) A coma is a horrible thing to have happen.

42) This council is the worst one that I have ever known.

43) Lots of context would be helpful here.

44) These contests were the greatest ones in the whole country.

45) I know that the killer is still something Mary thinks about.

46) The contacts didn't come in the mail.

47) The catcher did not know what to do at all.

48) Zack said these carols are some of the best that he knows.

49) Besides coasters we had to buy thirteen other things.

50) In Kansas we spent several days looking for antiques.

51) The cardboard we left there is missing now.

52) The comma is what we studied in class yesterday.

53) Katie's lack of courage was going to be discussed.

54) Tom saw the curry was put at the end of the table.

55) In combat many things happened very quickly.

56) Lisa says that a comeback is what she really needs right now.

57) The canvas they sell is not very expensive.

58) That cookie was left outside overnight.

59) The cougar was what she was the most excited to see.

60) Their culture is just as good as anyone else's.

Appendix C

1a) Ted saw the inn at the lake.

1b) The cat is in the brown box.

2a) The big can is sitting on the table.

2b) The man can see the bear in the forest.

3a) John thinks Sam is two right now.

3b) Jonathan ran to the park.

4a) There is some dew in the bathroom.

4b) The women do not see the books.

5a) Jenny wants the oar that she put on the table.

5b) Jenny wants squash or zucchini in her garden.

6a) The wood is sitting next to the doorway.

6b) Bob would read the newspaper every day.

7a) Tom saw the will lying on the table.

7b) The children will open presents today.

8a) Matt saw the witch by the yellow house.

8b) Matt told us which color he wanted.

9a) I think the weather is what he thinks about.

9b) Jack doesn't know whether Sue is at work right now.

10a) Katie picked some flowers from the garden.

10b) Ted thinks the sum is hard to figure out.

11a) The teacher might give us homework today.

11b) It is his might that has led him through this.

12a) The cookie has been eaten already.

12b) Mike saw the bin in the living room.

13a) James has eaten breakfast already.

13b) James has something but I don't know what.

14a) Mary had slept for slept hours.

14b) Mary had something in her bedroom.

15a) Jacob had no idea for the next project.

15b) The women know something but they didn't say what.

16a) Tom read the book like his friend.

16b) We know that they like something.

17a) Chris doesn't know where that thing is.

17b) I know that they wear something today.

18a) Pam hit the ball for her friend.

18b) Pam knows Sam is four right now.