SPOKEN WORD RECOGNITION OF THE REDUCED AMERICAN ENGLISH FLAP

by

Benjamin V. Tucker

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF LINGUISTICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2007

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Benjamin V. Tucker
entitled "Spoken Word Recognition of the Reduced American English Flap"
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

____ Date: 05/25/07
Natasha Warner

____ Date: 05/25/07
Brad Story

____ Date: 05/25/07
Michael Hammond

____ Date: 05/25/07
Merrill Garrett

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College. I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 05/25/07
Dissertation Director: Natasha Warner

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Benjamin V. Tucker

ACKNOWLEDGEMENTS

I would like to thank my advisor, Natasha Warner, for her instruction and guidance in helping me to complete this degree, especially for always believing that I could succeed. Her belief in my potential as an academic has helped me reach beyond what I thought possible. I would also like to thank the rest of my dissertation committee: Brad H. Story, Michael Hammond, and Merrill Garrett for their contribution to and help in writing my dissertation. Writing a dissertation was the part of getting a Ph.D. that always made me wonder if this is what I really wanted to do. I would like to thank Michael Hammond for allowing me to be a member of his lab and giving me the opportunity to find a love for psycholinguistic research. I would like to thank Brad H. Story for introducing me to speech perception and production outside of the Department of Linguistics.

As a student I have been fortunate to have had multiple mentors other than my academic advisor. While these people may not have participated as members of my dissertation committee they have helped me develop my many interests in the field of linguistics. I would like to thank D. Terence Langendoen for being a mentor and informal advisor and for taking a chance and hiring an undergraduate student who was new to the field of linguistics as an assistant. I will always value his friendship and his vast knowledge of the field of linguistics. I am also immensely grateful to Susan Penfield who has also acted as a mentor, advisor and teacher. Susan has helped me develop an interest in working with endangered languages and has given me the opportunity to work with people on the front lines of documentation and revitalization. I am also grateful for the friendship that has developed. Four-hour drives across the Arizona desert to Parker, Arizona provide a wonderful opportunity to create a friendship. I would like to thank Diane Ohala for her kindness, support and friendship. I would like to thank Diana Archangeli for her advice and support all through out graduate school. Special thanks to Dick Demers for showing me that language could be studied scientifically.

To the excellent faculty at the Department of Linguistics who have all helped and supported me along the way, thank you for all you do to create a wonderful learning and working atmosphere. Without the support of the Department of Linguistics Staff both past and present, Yolanda, Rosemary, and Gwen and Marian, Jennie, Kimberly, and Jennifer, graduate school would have been a nightmare.

I would like to thank my cohort of graduate students, who all began this journey with me: Jordan Brewer, Jaime Parchment, Dave Meideros, Peter Richtsmeier, Polly O'Rourke, Sumayya Racy, Stacey Oberly, Adam Baker. I am glad to be able to call you friends and I will miss our "end of the semester feedings".

Thank you to all of the graduate students that have helped me along the way and have helped make graduate school "fun": Robert Kennedy, Rachel Hayes-Harb, Cathy Hicks, Peter Norquest, Jason Haugen, Luis Barragan, Tania Granadillo, Scott Jackson, Xu Xu, Ruby Basham, Debbie Cole, Shannon Bischoff, Lindsay Butler, Phill Cash Cash, Jerid Francom, Selene Gardner, Andrew Glines, Erin Good, Keisha Josephs, Amy LaCross, Naomi Ogasawara, Sylvia Reed, Jeffrey Berry, Angelina Serratos, Azita Taleghani, Jianyuan Wang, Mercedes Tubino, Jeffrey Witzel, Kang Li, and Brian Monson.

Special thanks to Mary Dungan and Karen Morian without whose help it would have taken much longer to finish. Thanks also to Raiza Betters and Chris Crawford. I would also like to thank the members of the Douglass Phonetics Laboratory and Speech Laboratory who have listened to me present and talk about various portions of my dissertation.

I would like to thank my parents, siblings and grandparents for all of their love and support. Without you I would never have achieved my goal.

Thank you Dallin, for helping me smile at the end of a long day by exclaiming “dada!” as I walk through the door. Lastly, I would like to thank my wife Camille. Her love and support have been the steadying influence in my life.

TABLE OF CONTENTS

TABLE OF CONTENTS-Continued

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Phonetic variation as found in various speech styles is a rich area for research on spoken word recognition. Research on spoken word recognition has focused on careful, easily controlled speech styles. This dissertation investigates the processing of the American English Flap. Specifically, it focuses on the effect of reduction on processing. The main question asked in this dissertation is whether listeners adjust their expectations for how segments are realized based on speech style. Even more broadly, how do listeners process or recognize reduced speech? Two specific questions are asked that address individual parts of the broad question. First, how does reduction affect listeners' recognition of words? Is it more difficult for listeners to recognize words pronounced in reduced forms, or is it perhaps easier for listeners to recognize reduced forms? Second, do listeners adjust their expectations about reduction based on preceding speech style (context)?

Four experiments were designed using the auditory lexical decision and cross-modal identity priming tasks. Listeners' responses to reduced and unreduced flaps (e.g. unreduced [pʌɾl̩] as opposed to reduced [pʌɾ̞l̩]) were recorded. The results of this work show that the phonetic variation found in speech styles containing reduction causes differences in processing. Processing of reduced speech is inhibited by weakened acoustic information or mismatch to the underlying phonemic representation in the American English flap. Listeners use information about speech style to process the widely varying acoustic reflections of a segment in connected speech. The implications of these findings for models of spoken word recognition are discussed.

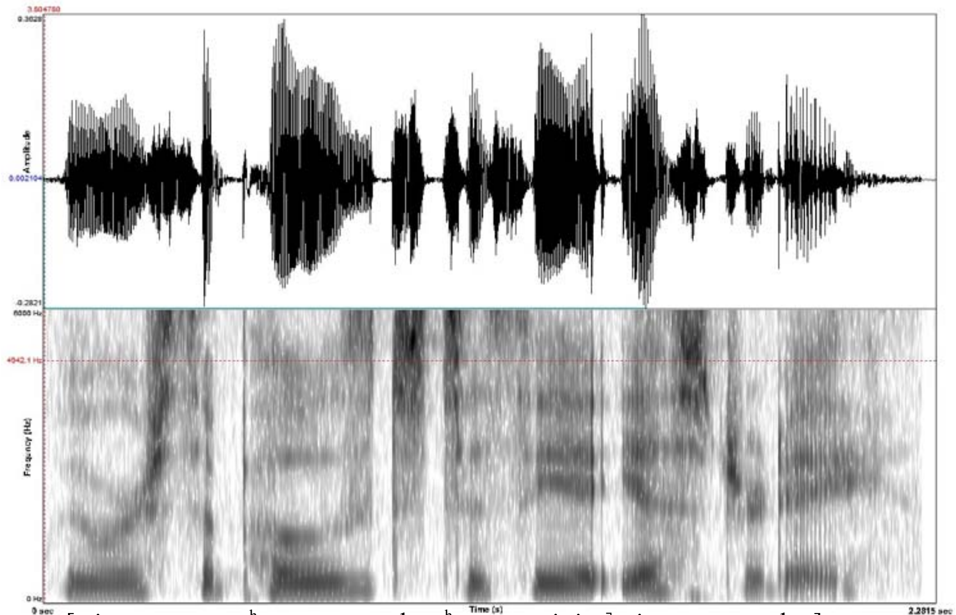
CHAPTER 1

INTRODUCTION

1.0 Introduction

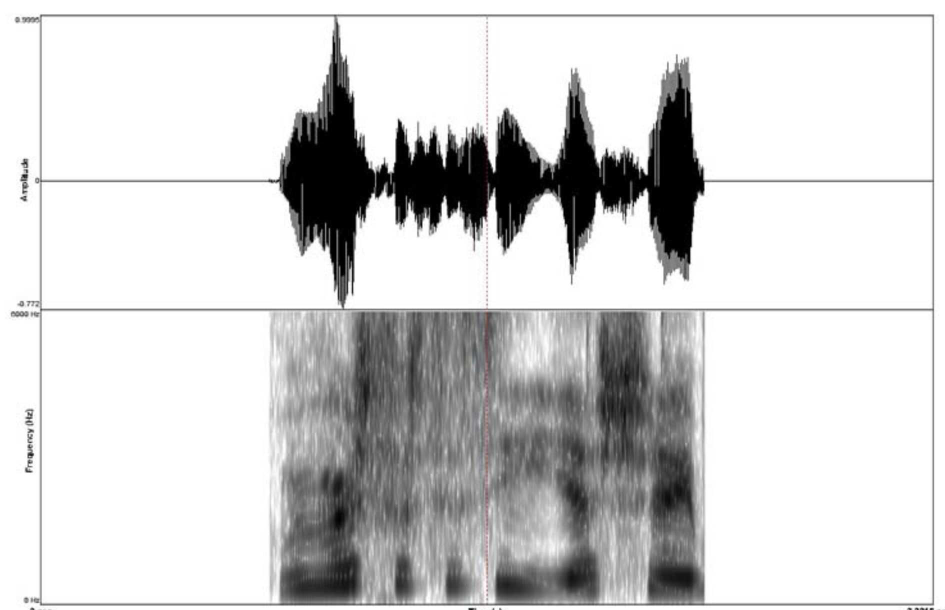
Humans use speech to communicate in their daily interactions. The spontaneous speech used in these interactions is highly erratic and contains large amounts of variation. However, most speech research has focused on laboratory (also defined as read or careful) speech (Cutler, 1998; Johnson, 2004), which is different from the spontaneous (also defined as natural, conversational, or casual) speech that humans encounter in their normal, daily communicative exchanges.

Extreme laboratory speech differs from spontaneous speech in that attention is given to the enunciation of every segment contained in that speech phrase or that the speech produced has been carefully controlled to test a specific effect on speech recognition (as might be produced by a phonetician for experimental stimuli). For example, the phrase “We were supposed to see it yesterday.” recorded in a laboratory setting might be produced as [wi wɚ səpʰoʊzd tʰə si ɨt̚ jɛstɚdeɪ] as seen in Figure 1.1 below. However, in a recent recording of a spontaneous conversation from Warner and Tucker (2007), that same sentence was pronounced as [wiɚ sɸʊsə si jɛʃe]. A waveform and spectrogram is provided in Figure 1.2 below. Silence is added to both sides of the speech in Figure 1.2 so that the time domain is equivalent to Figure 1.1.



[wi wɚ s ə pʰ oʊ z d tʰ ə s i ɨ t̚ j ɛ s t ɚ d eɪ]

Figure 1.1 A spectrogram and waveform of a careful elicitation (laboratory speech) of the phrase “We were supposed to see it yesterday.”



[w i ɚ s ɸ ʊ s ə s i j ɛ ʃ e]

Figure 1.2 A spectrogram and waveform of a spontaneous production of “We were supposed to see it yesterday.” from Warner & Tucker (2006)

The difference between the two speech samples is immediately apparent by comparing the mean duration of the utterances. The difference is also apparent when the number of IPA symbols used to transcribe the utterances is used as a metric for comparison. In the laboratory speech a total of 25 segments are used, while in the spontaneous speech a total of 15 segments are used. Of the 15 segments used in the spontaneous speech, two do not occur in the original transcription ([ɸ] and [ʃ]). Furthermore, examination of additional alternations, in the spontaneous speech reveals the deletion of segments (e.g., [w] from 'were'), syllables (e.g., three syllable [səpoʊzd tə] becomes two syllable [sɸʊsə]), and whole words (e.g., 'it') as well as segmental changes (e.g., /s t r/ in 'yesterday' form [ʃ]).

Spontaneous speech often contains instances of articulatory undershoot. For example, a production of a stop may not reach full closure, thus, acoustically what formerly was a stop can become more approximant-like (Ernestus, 2000). These variations in production are called speech reduction. This type of reduction, where segments, syllables, and words can be deleted and/or changed, is prolific throughout spontaneous speech (Johnson, 2004; Jurafsky et al., 2000; Greenberg, 1998). Despite this variation in naturally occurring speech, human listeners understand it with little apparent effort. Surprisingly, relatively little research has been performed on the production and recognition of spontaneous or reduced speech (Johnson, 2004; Ernestus et al., 2002; Cutler, 1998) even though listeners encounter spontaneous speech more frequently than laboratory speech. For example, an undergraduate student might have several conversations with friends, listen to several lectures where several quotes are read, and participate in a speech perception experiment all in one day. In this example the listener participates in several activities where spontaneous and less spontaneous speech is processed. The vast majority of the listeners' interactions involve perceiving and processing spontaneous speech.

Because spontaneous speech is so highly variable, it has not frequently been used in spoken word recognition research. This variability makes spontaneous speech difficult to control and specific speech-related questions difficult to examine. However, it is possible to create items that provide sufficient control over relevant variables if first the acoustic aspects of production of reduced and unreduced speech are investigated (Arai, 1999; Ernestus, 2000; Ernestus et al., 2002; Mitterer and Ernestus, 2006). For example, Mitterer and Ernestus (2006) first conducted a corpus study that investigated the acoustic and distributional characteristics of /t/-lenition in Dutch. In their investigation they identified two residual acoustic cues from /t/-lenition in the speech sample. Based on the information from the corpus study, they designed perception experiments which allowed them to test the perception of reduced speech (i.e. /t/-lenition). They found that the presence of residual cues and the segmental context both play important roles in listeners' recovery of lenited /t/.

One type of reduction is the variation found in productions of the American English flap. Specifically, recent work by Warner and Tucker (2007) investigating the acoustic properties of American English flaps and other stop consonants has found that the degree of flap reduction is highly dependent on speech style (aspects of speech style are discussed in greater detail in section 1.3 below). As a result, a large body of descriptive information regarding the acoustics of flaps and the other stop consonants was generated (Warner and Tucker, 2007). Using the knowledge about the reduction of flaps based on speech style and the acoustic characteristics of flaps from Warner and Tucker (2007), it is possible to generate reduced and unreduced items which mimic the acoustic characteristics found but that are still controlled. This allows for a comparison of the recognition of reduced and unreduced speech while carefully controlling the items.

1.1 Research questions

There is variability and reduction in naturally produced speech (Johnson, 2004; Jurafsky et al., 2000; Warner and Tucker, 2007; among many others). This is especially true in spontaneous conversation, and even more so if one compares spontaneous speech to a careful speech style as seen in the example above (Johnson, 2004; Jurafsky et al., 2000). Human listeners, at least when listening in their native language and dialect, can understand speech, with all of its variability and reduction, almost effortlessly. The broad research question investigated here is whether listeners adjust their expectations for how segments are realized based on speech style. Even more broadly, how do listeners process or recognize reduced speech?

To investigate this difference it is necessary to examine the effect of reduction on listeners' recognition. First, how does reduction affect listeners' recognition of words? Is it more difficult for listeners to recognize words pronounced in reduced forms, implying that weaker acoustic cues cause a conversion of such forms into their abstract, underlying representation? Or is it perhaps easier for listeners to recognize reduced forms, because reduced forms are more commonly produced and encountered, thus more quickly accessed (frequency of occurrence effects)? This assumes that the more frequently occurring stimuli are easier to process. Second, do listeners adjust their expectations about reduction based on preceding speech style or context?

This research investigates the question of whether a processing difference exists between laboratory (or unreduced) and spontaneous (or reduced) speech using the reduction found in American English flaps. Unlike previous research (e.g. Charles–Luce, 1997; McLennan, Luce, and Charles-Luce, 2003; Connine, 2004) which investigate the relationship between flapping and /t/ or /d/, this research focuses on reduction of flaps. For example, the unreduced form is not [t] but [ɾ] and the reduced form is [ɾ̞]. Four experiments are outlined that explore the effect of reduction of flaps on spoken word recognition. The first two experiments explore how listeners respond to reduced and unreduced words in isolation. Specifically, these experiments investigate whether reduction of flaps inhibits or facilitates the processing of words produced in isolation. The third and fourth experiments test whether leading listeners to expect different speech styles (reduced and unreduced preceding context) affects listeners' responses to reduced versus unreduced words.

1.2 Organization of the dissertation

This first chapter discusses the gradient nature of speech style and addresses the notion of reduction and summarizes current reduction research. This summary of current research explores reduction in speech production and spoken word recognition, with a focus on recognition. This chapter also discusses research on the American English flap and describes relevant acoustic characteristics in the reduction of flaps. The second chapter presents Experiments 1 and 2. These experiments investigate the processing of reduced words in isolation. Chapter Three discusses research on the effect of context in spoken word recognition. The main body of this chapter describes and reports the results of Experiments 3 and 4, which investigate the role of context/speech style in the recognition of reduced words. In the fourth chapter the results of Experiments 1–4 are discussed with respect to their implications for spoken word recognition, models of lexical access, and lexical representation.

1.3 Speech style

Listeners generally have a sharp ear for variation of speech style (Zwicky, 1972a). A listener can tell whether the speech was produced quickly or carefully and many other aspects of the speech. In general, research has historically not investigated more natural, casual speech styles (Zwicky, 1972a; Culter, 1998). Speech style thus far in this work has been characterized as either laboratory speech or spontaneous speech. While *laboratory* and *spontaneous* speech are convenient cover terms they can make the notion of speech style sound categorical, which is misleading. An entire continuum of speech style might be constructed as in Figure 1.3 below.

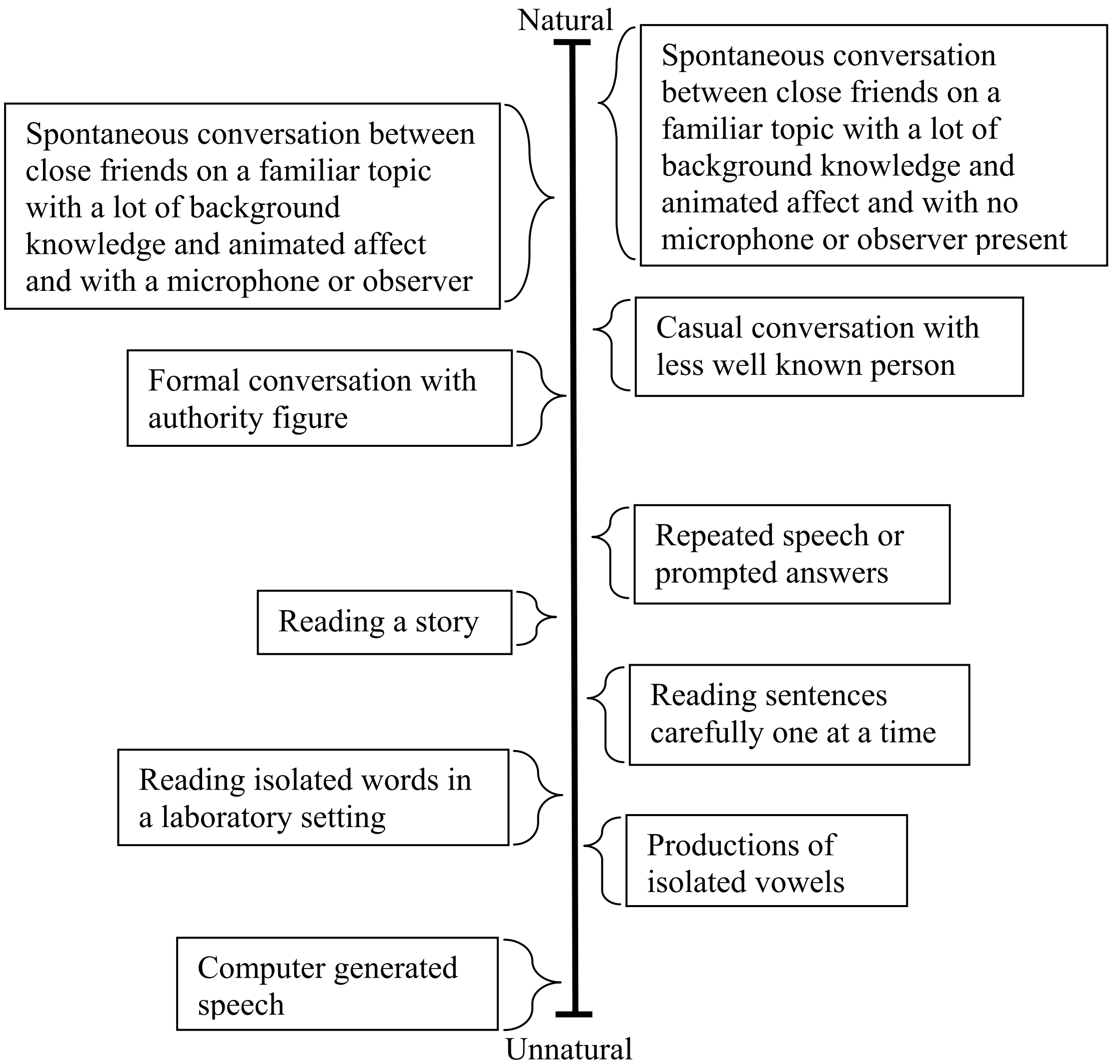


Figure 1.3 A one dimensional diagram of possible speech types, illustrating the gradient nature of speech style.

While Figure 1.3 is a one dimensional illustration, the variation contained within speech style isn't necessarily one dimensional. Speech style can contain any number of relevant dimensions. For example the type of speech might be one dimension (e.g. read speech) and a second dimension might include the situation the speech is produced in (comfortable vs. uncomfortable). Many researchers have investigated speech style and have divided speech style into many discrete groups as opposed to a gradient continuum.

Zwicky (1972a) uses the term *casual speech* to identify running connected speech and the processes that occur in running speech. He discusses causal speech and its potential contribution to synchronic and diachronic linguistics. Zwicky argues for the need to account for the processes that occur in casual speech. Zwicky points out that many linguists have identified the need to make a distinction between speech styles. He cites several examples of researchers who have posited speech styles from careful to more casual in various languages. For example, Dressler (1972) posits three speech styles in Breton, Harris gives four for Spanish (Harris, 1969), Cheng (1968) posits five for Chinese, and Wescott (1965) seven for Bini (from Zwicky, 1972a). Dressler (1972) discusses *fast speech rules* (*allegro rules*) which he indicates are phonological rules that apply specifically in fast speech. Dressler and Zwicky group speech rate and speech style together as the same phenomenon.

Labov (1972) proposed a classification of five different levels of speech style affecting phonetic studies: casual speech, careful speech, reading, word lists, and minimal pairs. Labov in fact proposes a continuum similar to the one presented in Figure 1.3. Labov's continuum "ranges along a single dimension of attention paid to speech, with casual speech at one end of the continuum and minimal pairs at the other" (1972: 991). In Lindblom's H&H Theory (1990) speakers vary their output along a hyper- and hypospeech continuum. On one end of the continuum is hyperspeech or careful "output-oriented speech" and on the other end is hypospeech or casual "system-oriented speech". H&H Theory is discussed in greater detail in section 4.3 below.

Other researchers have investigated *clear speech*. *Clear speech*, as defined by this body of literature, is produced with extreme care given to the production of segmental and suprasegmental units. Clear speech is often used to increase speech intelligibility (Picheny et al., 1985). The effects of clear versus conversational speech on production are shown in English and Croatian to be decreased speaking rate, expansion of pitch range, and increased F1 X F2 space (Smiljanić and Bradlow, 2005). Clear speech versus conversational speech has also been shown to be different in speech prosody (Smiljanić and Bradlow, 2006; Smiljanić and Bradlow, in press).

Llisterri (1992) provides a summary of speech styles in speech production research based on presentations from the workshop on the "Phonetics and Phonology of Speaking Styles" at ESCA in Barcelona (1991). In this summary Llisterri shows that that there is a large variety of terms used by researchers investigating speech style. For example, "spontaneous speech" has been used to describe recordings of interviews, free conversation with a friend, political debates and sports commentary. He also describes a set of segmental, suprasegmental, and phonological processes that are used as correlates of speech style. Based on his survey, Llisterri (1992) calls for standardization of terminology and collection procedures.

Many researchers make a two way distinction of speech style. Ernestus (2000), for example makes a distinction between casual and careful speech. Cutler (1998) uses the terms "read-speech" and "spontaneous" to distinguish between planned speech and less planned speech respectively. Fosler-Lussier, Greenberg, and Morgan (1999), similar to the Cutler (1998), also distinguish between read and spontaneous speech. In the current work this two-way distinction is adopted.

1.4 Reduction in speech

As already discussed, there is a small but growing body of research on phonetic reduction in spontaneous speech (Johnson, 2004; Ernestus, 2000; Greenberg, 1998; Cutler, 1998; Smiljanic and Bradlow, 2006; Kilanski and Wright, 2006; Smith, 2006). Reduction can occur in many different forms: vowels can become more centralized, segment durations can be decreased, and segments may be affected by many other qualitative changes (e.g., an unstressed vowel /i/ can be produced as /ɨ/). The terms “reduced” and “reduction” are used to describe phonetic change (perhaps due to target undershoot) and deletion of speech. Reduced speech encompasses any kind of speech or speech style that contains reduction, generally characteristic of spontaneous speech. The term “unreduced speech” is used in this work in opposition to the term “reduced speech”. Unreduced speech is used to refer to careful speech which contains little or no apparent reduction.

This section discusses speech reduction in greater depth and provides a summary and discussion of previous literature in speech reduction. The discussion of speech reduction is divided into the reduction that occurs in speech production (section 1.3.1) and how reduction affects spoken word recognition (section 1.3.2).

1.4.1 Reduction in speech production

Spontaneous speech is highly variable. This variability is a major source of difficulty for the designers of machines that would perform automatic speech recognition comparable to human speech recognition (Greenberg, 1998, 1999; Fosler-Lussier et al., 1999). As a result of this difficulty, much of the reduction literature focuses on or has implications for automatic speech recognition systems. The body of research on reduction in speech production is much larger than that of spoken word recognition, though many of the production studies discuss processing implications along with implications for automatic speech recognition. This section discusses a sample of recent literature on reduction in speech production (e.g., Johnson, 2004; Jurafsky et al., 2000; Byrd, 1994; Ernestus, 2000), specifically types of reduction that occur and the reported causes of reduction in spontaneous speech.

Most research investigating reduction in speech production relies on large spontaneous speech corpora. Work by Greenberg (1998, 1999) and colleagues (Fosler-Lussier et al., 1999) show that Automatic Speech Recognition systems, which have been trained on corpora of read and careful speech, recognize well on these styles of speech (85-98%, Greenberg, 1998). However, when the Automatic Speech Recognition system is applied to a spontaneous speech corpus (recorded telephone conversations) like the Switchboard Corpus (Godfrey et al., 1992) the recognition level falls to 50-70% accuracy (Greenberg, 1998). This is due to the high degree of variation found in spontaneous speech along with pauses, hesitations, repetitions, corrections and speech errors common in spontaneous speech. Greenberg and colleagues argue for the use of higher level units (the syllable) and their potential to help account for some of the variation in spontaneous speech.

Patterson, LoCasto, and Connine (2003) investigated the frequency of schwa deletion in English using two corpora. The first corpus was the Switchboard Corpus (Godfrey et al., 1992), a spontaneous speech style corpus, and the second was a narrative corpus created by the authors. Schwa deletion or syncope in the phonology literature has long been a topic of research (e.g. Zwicky, 1972b; Hooper, 1978; Pérez, 1992; Hammond, 1999). Patterson et al. (2003) found that when schwa occurred in a pre-stress environment (*support* /səpoɹt/ -> *s'port* /spoɹt/) deletion was lees likely then when schwa occurred post-stress (*corporate* /koɹpəɹɪt/ -> *corp'rate* /koɹprɪt/). They also found that in the pre-stress environment three syllable words were much more likely than two syllable words to contain a deleted schwa. In comparing the results of this study to results of a study by Dalby (1986) they found an indication that schwa deletion is more frequent in a spontaneous setting (Patterson et al., 2003) than when recorded in a laboratory setting as a monologue (Dalby, 1986).

In an examination of the TIMIT Corpus (Seneff and Zue, 1988; Zue et al., 1990), Byrd (1994) investigated the factors of sex and dialect on reduction. One disadvantage of using the TIMIT Corpus for this type of research is that it consists of laboratory style speech, with subjects reading sentences in a controlled setting. Byrd's study examined reduction in this corpus by investigating speaking rate, vowel reduction (mean occurrence of central vowels), frequency distribution of stop releases (e.g. /ðæt̚/ as compared to

/ðæt^{h}/), flaps (e.g. occurrence of flapping in *suit in* /sut^{h}ɪn/ as compared to /suɾɪn/), glottal stops (e.g. distribution of /ʔ/), breathy vowels, and syllabic alveolar nasals. The results of the study indicate that gender and dialect influence reduction. More specifically, male speech is more likely than female speech to contain features of reduced speech. The effects of dialect were difficult to interpret due to lack of statistical power and poor dialect classification in the database, dialect nonetheless played a role.

Jurafsky, Bell, Gregory, and Raymond (2000) investigated a sub-portion of the Switchboard Corpus (Godfrey et al., 1992). In their investigation they showed that reduction of both content and function words was dependent on the lexical frequency of the word and the conditional probability of the neighboring words. Their investigation compared high to low frequency words and high to low probability words. This comparison showed that for high frequency and high probability words the mean duration of content and function words was shorter, that vowels were more reduced (for example a full vowel [i] becoming [ɨ]), and that deletion of word final /t/s and /d/s was more frequent.

Also utilizing production data from the Switchboard Corpus (Godfrey et al., 1992), Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, and Gildea (2003) expanded the work of Jurafsky, Bell, Gregory, and Raymond (2000). They investigated the effects on reduction of contextual predictability, disfluencies (e.g., *uhh* and *umm*), and the location of the function word in a phrase. They found that all of these factors played a strong role in the reduction of words. Specifically, they found that contextual predictability plays a large role in the reduction of function words, i.e., words with high predictability based on context are more likely to be reduced, reinforcing the findings by Jurafsky, Bell, Gregory, and Raymond (2000). The also found that when function words follow disfluencies or occur at the beginning or end of a phrase they were less likely to be reduced.

Work by Ernestus (2000) investigated spontaneous speech in Dutch using a corpus of face-to-face spontaneous speech. This work studied the variations found in productions of casual speech as compared to what would be produced in careful speech, or what is proposed as the underlying representation. Generalizations were made as to what consonants could be deleted in what contexts and when a full vowel pronounced carefully might occur as a schwa in casual speech or not occur at all. Specifically, this research showed that segments are frequently deleted at the end of words (e.g., /t/ and /r/) and words and fixed expressions may lose unstressed syllables. For example, the Dutch word *eigenlijk* [‘ɛixələk] “actually” can be produced as [‘ɛik].

Pluymaekers, Ernestus, and Baayen (2005) used the Corpus of Spoken Dutch (Oostdijk, 2000), a corpus of face-to-face spontaneous style speech, to investigate the effect of lexical frequency on the duration of four affixes (e.g., *lijk* in *natuulijk*) and the durations of their segments in Dutch. They showed a strong effect of lexical frequency on affix duration, which showed shorter affix durations for words with high frequency stems.

Ernestus, Lahey, Verhees, and Baayen (2006) investigated regressive voice assimilation in Dutch (e.g. [vɛt] and [buk] ‘law book’ would be [vɛdbuk]) also using the Corpus of Spoken Dutch (Oostdijk, 2000). In their investigation they found that in correlation to increased lexical frequency, acoustic indicators such as cluster duration and period of glottal vibration decreased while the duration of release noise increased. They interpreted this result as an indication of reduced articulatory effort for higher frequency words.

Johnson (2004) investigated what he calls "massive" reduction using the Variation in Conversation Corpus (Pitt et al., 2003). He investigated reduction in terms of the frequency of segment deletion and segmental change and found that more than 22% of four- to six-syllable words contain two or more syllable deletions (e.g. *hilarious* pronounced as [hlɛrɛs]). The investigation of segmental change found that over 60% of the analyzed corpus deviates from its citation (unreduced) form by at least one phoneme. Based on the "massive" reduction results in the corpus study, Johnson discusses the implications for models of auditory word recognition, arguing that models using abstract representation are at a disadvantage due to the massive reduction contained in spontaneous, reduced speech.

The studies summarized thus far have shown that reduction is common throughout spontaneous speech and even occurs in laboratory speech. It was also shown that reduction is more likely in less careful speech as compared to careful speech. In fact it might be posited that the more "natural" the speech, as in the continuum in Figure 1.3, the more reduction that would occur. Because reduction is so common in speech it is important to investigate the nature of reduction. For example, computer speech recognition systems are generally good at recognizing laboratory speech. However, reduced speech poses a problem (Greenberg, 1998, 2001) for these systems. The accuracy of these systems falls dramatically when presented with spontaneous speech.

The research above shows that there are differences between spontaneous and laboratory speech. Investigation of spontaneous speech provides insight into speech that might have been missed if research continued to focus on laboratory speech. The variation caused by reduction in speech provides a rich area of research in understanding how speech is produced and how it is recognized.

1.4.2 Reduction in spoken word recognition

Prior to the arguments made by Johnson (2004) about auditory word recognition, Cutler (1998) said: "Most of the speech we listeners hear is spoken spontaneously, […] Psycholinguistics has compiled an extensive body of research on human speech recognition. […But] the speech mode most frequently encountered in psycholinguistics is *not* the speech mode most frequently encountered outside of the laboratory" (as quoted by Cutler (1998) from Mehta and Cutler (1988): 135-136). This quote summarizes the state of speech recognition in 1998 and a decade earlier in1988 when the quote originated. In 1998 there was a small, but growing, body of research investigating reduced speech. Since 1998 the body of research has continued to grow, but is still small relative to the existing body of research on speech recognition using laboratory speech (Johnson, 2004; Ernestus et al., 2002). This section summarizes existing research on the recognition of reduced speech and its contribution to the understanding of speech recognition processes.

Pollack and Pickett (1963, and Pickett and Pollack, 1963) investigated the effect of reduced speech on intelligibility. They found that fast speech, both laboratory and spontaneous, is more difficult to recognize than slow speech. Bard, Shillcock and Altmann (1988) used recorded speech from spontaneous conversation in a gating task to investigate the effect of following context on recognition of connected speech. They found in support of Pollack and Pickett (1963) that spontaneous speech was difficult to recognize. They also found that the addition of following context (approximately two words) was helpful in recognition.

Mehta and Cutler (1988) compared spontaneous (reduced) and laboratory (unreduced) speech using a phoneme detection task. In this study the authors recorded a spontaneous conversation and selected stimuli from that conversation. The speaker was asked to return and read the same items from the recorded conversation. A phoneme detection task was then performed with the reduced and unreduced stimuli. The effects differed dependent on the type of speech to which listeners responded (i.e., reduced versus unreduced). The results of the unreduced speech were similar to previous results in the phoneme detection task. Listeners responded to later targets and targets preceded by longer words more rapidly than earlier targets and targets preceded by short words. In reduced speech, listeners responded to targets in accented words and strong syllables more quickly than in unaccented and weak syllables. Mehta and Cutler (1988) go on to argue that this experiment illustrates the usefulness of exploring reduced speech in spoken word recognition because it indicates the possibility that different strategies are used when processing reduced versus unreduced speech.

McAllister (1991) investigated the effect of spontaneous speech on the processing of syllables. Gating tasks were used to investigate the processing load of laboratory (unreduced) and spontaneous (reduced) speech. Differences in stressed and unstressed syllables were used to investigate processing. McAllister reported that in unreduced speech, listeners had more difficulty processing unstressed syllables than stressed syllables. However, in reduced speech no difference in processing time was found between stressed and unstressed syllables.

Cutler (1998) builds on the argument made in Mehta and Cutler (1988) arguing for more research into the investigation of spontaneous speech, presenting a summary of research on the recognition of reduced speech as part of the argument. Cutler (1998) discusses four phenomena that contribute to the knowledge of recognition in spontaneous speech: resyllabification, assimilation, deletion, and epenthesis. A summary of each of these phenomena are provided below with the addition of some more recent research in these areas. Each of these phenomena is an example of a type of change seen in spontaneous speech.

Research investigating resyllabification refers to the phenomenon wherein connected speech segments like [k] in *make*, which are usually in the coda, (Kahn, 1976; Harris, 1983; Hammond, 1999) resyllabify in the phrase *make up* to become the onset of *up*. The question posed by researchers is whether this resyllabified [k] activates other [k] allophones to cause inhibition of lexical access. Studies investigating resyllabification (Whalen et al., 1997; Matter, 1986; Dejean de la Batie and Bradley, 1995) have found that it has very little effect on processing. However, Vroomen and de Gelder (1999) found longer response latencies for phoneme identification in resyllabified targets than in non-resyllabified targets. They explain this as a possible result of the phonotactics of Dutch, where the resyllabified targets are phonotactically possible sequences while the non-resyllabified targets are not.

Assimilation is another phenomenon which commonly occurs in reduced or spontaneous speech. Most of the work in this area has investigated phonological assimilation. For example, English has nasal-obstruent place assimilation (i.e., *ran back* becomes /ræm bæk/) where the nasal matches the place of articulation of the following obstruent. The results of many studies of assimilation have been generally consistent: items where legal or optional assimilation occur pose no processing difficulties, but items containing illegal assimilations impede processing (Cutler and Otake, 1998; Koster, 1987; Otake et al., 1996; Gaskell and Marslen-Wilson, 1996).

Recent work by Coenen, Zwitserlood, and Bölte (2001; Zwitserlood and Coenen, 2000) on assimilation in German using a cross-modal form-priming experiment shows priming effects for assimilated forms. Specifically, they investigate the phonological pattern, which occurs in fast speech, of regressive (a segment adopts characteristics of the preceding segment) and progressive (a segment adopts characteristics of a following segment) assimilation and how assimilation affects lexical access and representation. The results suggest that regressive place assimilation has a larger priming effect than progressive voice assimilation. However, they find a smaller effect in the amount of priming for assimilated primes than for unchanged primes. This effect was interpreted as indicating that there is some processing cost, even if minor, for assimilated forms when compared to unchanged forms. The effect was only found for assimilation occurring in context and not when assimilated words occurred in isolation.

Research investigating deletion refers to the phenomenon wherein segments are frequently deleted from words. The example given at the beginning of this chapter, *supposed to* becoming [sɸʊsə] where the *u* is completely deleted (along with other changes), is common in the languages of the world. Matter (1986) found that deletions at the beginning of words inhibited processing in French. Racine and Grosjean (1997) found similar inhibition of processing in French when the vowel in the initial syllable was deleted (e.g. *semaine* "week" was shortened to *s'maine*). Kuijpers, Van Donselaar, and Cutler (1996) found in Dutch that deletion of a vowel in medial position (e.g. the second vowel in *referaat* "scholarly report" spoken as *ref'raat*) inhibited processing. These three studies suggest that deletion of a segment has an inhibitory effect on processing.

LoCasto and Connine (2002) investigated English schwa deletion for two- and three-syllable words (e.g. *police* spoken as *p'lice*). Using an auditory form priming task (Zwitserlood, 1996), they presented listeners with vowel-reduced (schwa) and vowel-deleted forms. They found larger priming effects when the prime and probe were identical as opposed to when they were mismatched. They also found that the two types of syllable-deleted targets showed little difference in matched and mismatched pairs. This experiment generally supports the hypothesis that deletion of a syllable inhibits processing, though no effect was found in one of the conditions.

The last phenomenon discussed by Cutler (1998) is epenthesis, which involves the insertion of a segment into a string of sounds where it did not originally exist. Dupoux et al. (1999) investigated Japanese and French listeners' perception of targets like *ubzo* and *ubuzo* (which are not words in French or Japanese). Japanese listeners had difficulty distinguishing the two words, while French listeners had no problem making the distinction. When Japanese listeners were asked to transcribe the *ubzo* items they were written as 'ubuzo'. They argue that because Japanese does not allow consonant clusters that Japanese listeners 'hear' an epenthesized 'u', even though acoustically it is not physically there. To verify that Japanese listeners were paying attention to the middle vowel they performed further experiments investigating targets like *ubuzo* and *ubuuzo*. Japanese has a vowel length distinction making this a natural distinction for Japanese listeners and Japanese listeners were very good at distinguishing these items. French listeners on the other hand, had difficulty distinguishing between the pairs.

Dupoux et al. (2001) in further work on this phenomenon investigate targets with close lexical neighbors. For example, there were two types of targets: the u-set like *sokdo* /sokudo/ 'speed' which were targets that became lexical items with 'u' epenthesis) and the non-u-set *mikdo* /mikudo/ which did not become lexical items with 'u' epenthesis but had close lexical neighbors, like *mikado* 'emperor'. Japanese listeners transcribed both types of items with 'u' even though the non-u-set contained close lexical neighbors. In a lexical decision task no difference in response time was found between 'yes' responses for u-set words and 'no' responses for non-u-set words. These findings taken with the previous findings indicate that phonological knowledge is necessary at a prelexical level and that listeners can sometimes 'hear' of repair sequences which do not conform to their linguistic knowledge.

Cutler (1998) discusses data indicating that epenthesis in some cases can facilitate processing of speech. For example, Dutch listeners responded to epenthesized /tuləp/ faster than to /tulp/ 'tulip'. This suggests that processing can actually be facilitated by the insertion of segments in certain locations.

Arai (1999), similar to Johnson (2004), investigated the effects of "massive" reduction on processing in Japanese. Arai first investigated and described pronunciation variations (reduction) from the OGI Multi-Language Telephone Speech Corpus for Japanese. Unlike Johnson (2004), Arai performed a perception experiment with one target selected from the corpus. Listeners were asked to identify the target with varying degrees of context (full context: /hayaku ieba hokani/ 'in short, other...'; no context: /kebo/). Targets identified within phonetic context showed a substantial increase in accuracy of identification over those identified within the isolation context. Identification accuracy continued to increase as the amount of context increased, but the slope of increase is much closer to one.

Ernestus, Baayen, and Schreuder (2002) investigated the perception of highly reduced word forms. In this study, reduced word forms taken from Ernestus' (2000) corpus like ['ɛik] for *eigenlijk* ['ɛixələk] "actually" were presented to listeners, who were asked to identify the items. Listeners heard items divided into high, medium, and low reduction words. Items were presented in one of three contexts: Full, Limited, and Isolation. The contexts were manipulated so that Full context provided semantic and syntactic context, while Limited context provided only phonetic context (the target word and its neighboring vowels) and Isolation context provided only the target item. The results showed that listeners performed very well in the Full context environments and that their identifications were increasingly less accurate for Limited and Isolation contexts. Furthermore, the results showed that in the Limited and Isolation contexts low and medium reduction forms generally corresponded, while identifications of highly reduced forms were less accurate. They also claim, contrary to Johnson (2004) that since the degree of reduction inhibits recognition it supports claims of abstract representation.

A growing body of research suggests that understanding listeners’ responses to reduced speech is essential to understanding spoken word recognition. Many aspects of how speakers and listeners actually communicate would be left undiscovered if reduced speech were not investigated. The remainder of this chapter describes aspects of the reduced and unreduced American English flap. It also introduces some of the previous work investigating the production and perception of the flap.

1.5 American English flap

Many of the experiments described in previous sections on the processing of reduced speech investigated the production of phonetic and phonological variability (e.g. Ernestus, 2000; Ernestus et al., 2002; among others). This focus allowed researchers to gain an understanding of the production of specific reduction phenomena, thus allowing them to perform recognition experiments. This section discusses research on the production and perception of the American English flap. A flap is used here to describe what has been identified in the phonetics literature as a tap or a flap (Ladefoged, 2006). Articulatorily, a flap is a voiced stop that is produced by a rapid gesture of the tongue toward the alveolar ridge. It taps the alveolar ridge, creating a brief closure, and then releases the closure. A flap occurs in American English if it is intervocalic and precedes an unstressed vowel (e.g. ˈVɾV *butter* /bʌɾɚ/or VɾV *ability* /ʌbɪlɪɾi/) (Kahn, 1976; Kiparsky, 1979). Flapping can also occur in these same environments across word boundaries (e.g. *He said it.*

/hisɛɾɪt/) (Zue and Lafierre, 1979; Fukaya and Byrd, 2005). There is an abundance of research on the phonological distribution of the American English flap (Kahn, 1976; Kiparsky, 1979; Selkirk, 1982; Steriade, 2000; Riehl, 2003). This section focuses discussion of previous research, instead, on the articulatory and acoustic aspects of the American English flap and then describes experiments investigating the perception of the American English flap.

1.5.1 Articulatory investigation of the American English flap

Research investigating the articulatory aspects of flaps in American English has shown that production of flaps and [d] fall on a gradient scale. de Jong (1998) showed, using the University of Wisconsin X-Ray Microbeam Database, that articulatory productions of flap and [d] are dependent on gradient changes in lingual positioning. This study also suggested that while there was a gradient effect articulatorily, listeners were consistent in categorizing flaps and non–flaps (other allophonic occurrences of /t/ or /d/) acoustically, such that they reflected a categorical judgment.

Fukaya and Byrd (2005) also investigated the articulation of flaps, specifically “voiceless and voiced (flap) realization of underlying word–final /t/” (Fukaya and Byrd, 2005: 47). They investigated flaps occurring phrase-internally (e.g. *Pat said that* *<u>hot asparagus</u>* *is the tastiest dish.*) and phrase-final (e.g. *When it’s served* *<u>hot</u>*, *<u>asparagus</u>* *is the tastiest dish.*) in falling and level stress contours. Their investigation of flaps used the EMMA (Electromagnetic Midsagittal Articulometer) magnetometer system (Perkell et al., 1992) to acquire kinematic measurements of the tongue tip. In addition to gestural movement, they also reported articulatory velocity and acoustic duration of flaps. The analysis of three speakers shows that speakers' articulations are highly speaker dependent. Individuals produce acoustic shortness by accelerating the articulators or by decreasing the closure duration. The results support the results of de Jong (1998) in that they support a gradient production in the articulatory domain between the flap and voiceless stop, which does not support a categorical explanation. In the acoustic domain the duration of the flaps, however, suggests a "systematic" explanation (Fukaya and Byrd, 2005: 57).

1.5.2 Acoustic investigation of the American English flap

The articulatory studies suggest that flapping is a gradient phenomenon; categorical effects have been reported in acoustic studies. In the studies discussed below the research focuses on the distinction between the flap and the non-flapped oral stop. Zue and Lafierre (1977), in a seminal study on the acoustics of flaps, showed that word medial /t/ and /d/ following a stressed vowel were flapped 99% of the time. Turk (1992) found that stops occurring in a flapping environment (post-stress or between unstressed vowels) were shorter than stops preceding a stressed vowel.

Work by Horna (1998) also studied the acoustics of American English flaps. This research compared the results of elicited (unreduced) speech, like that in Zue and Lafierre (1977), to fluent (reduced) speech. It was suggested in this study, similar to the Cutler (1998) study of spoken word recognition, that effects found in researching unreduced speech are relevant to reduced speech, but that there are additional effects found in fluent speech that do not occur in investigations of unreduced speech. Specifically, Horna (1998) found that stress environment (ˈVɾV or VɾV) was less important in reduced speech than in unreduced speech, meaning, for example, that *butter* and *ability* are more likely to flap in spontaneous speech than in laboratory speech, regardless of the stress pattern. Horna (1998) also found that tongue position in the preceding vowel affected the occurrence of flapping in reduced speech.

In an investigation on word-medial stops of American English, focusing on flapping, Warner and Tucker (2007) compared reduced and unreduced flaps. While most of the previous studies described above focused on whether a phoneme is realized as a flap in a given environment, this study focused on the acoustic effects of the production of flaps in different speech styles. This study investigated the acoustic characteristics of flaps in a word-list production, read story, and spontaneous conversation. A corpus of speech from 22 speakers was created for the study; the results of seven of those speakers are reported in Warner and Tucker (2007). The acoustic characteristics typical of reduced and unreduced flaps are described below. An example of the canonical flap is illustrated below in Figure 1.4.

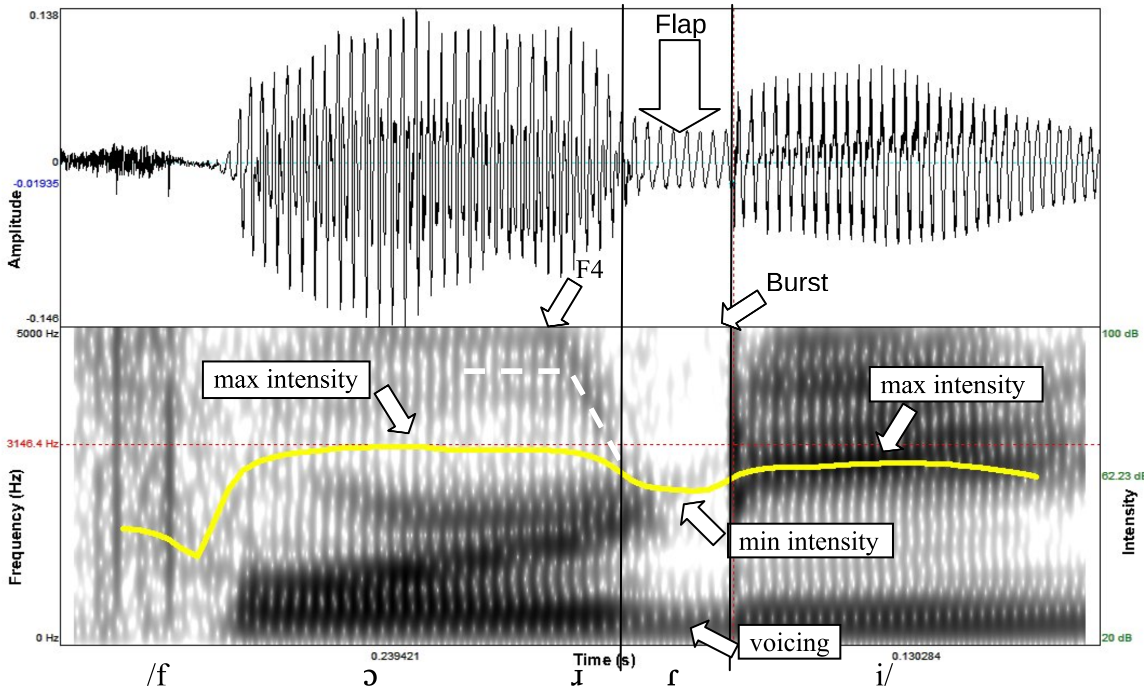


Figure 1.4 A spectrogram and wave form of the canonical flap from the word 'forty' with intensity superimposed. The flap, duration, intensity change, F4, and burst are indicated.

Acoustically, there are several canonical characteristics indicating a flap. First, there is voiced occlusion between two vowels, i.e., voicing occurs during an alveolar closure. Second, there is a general decrease in intensity relative to the surrounding vowels. Third, there is frequently a high intensity burst-like release at the end of the flap. It has been noted based on data by Dungan, Morian, Tucker, and Warner (2007) that in many instances there is a large drop in the F4 around a flap. This generally occurs when the flap is produced before or after an /ɹ/ though it has also been found to occur occasionally when no /ɹ/ is present (F4 is indicated by the dashed line in Figures 1.4 and 1.5).

Articulatorily, as a speaker produces a reduced flap the tongue gesture might not reach full closure; the gesture may be smaller so that the tongue moves only a little (more like a gesture for an approximant). Figure 1.5, below, provides a contrasting spectrogram of a reduced flap.

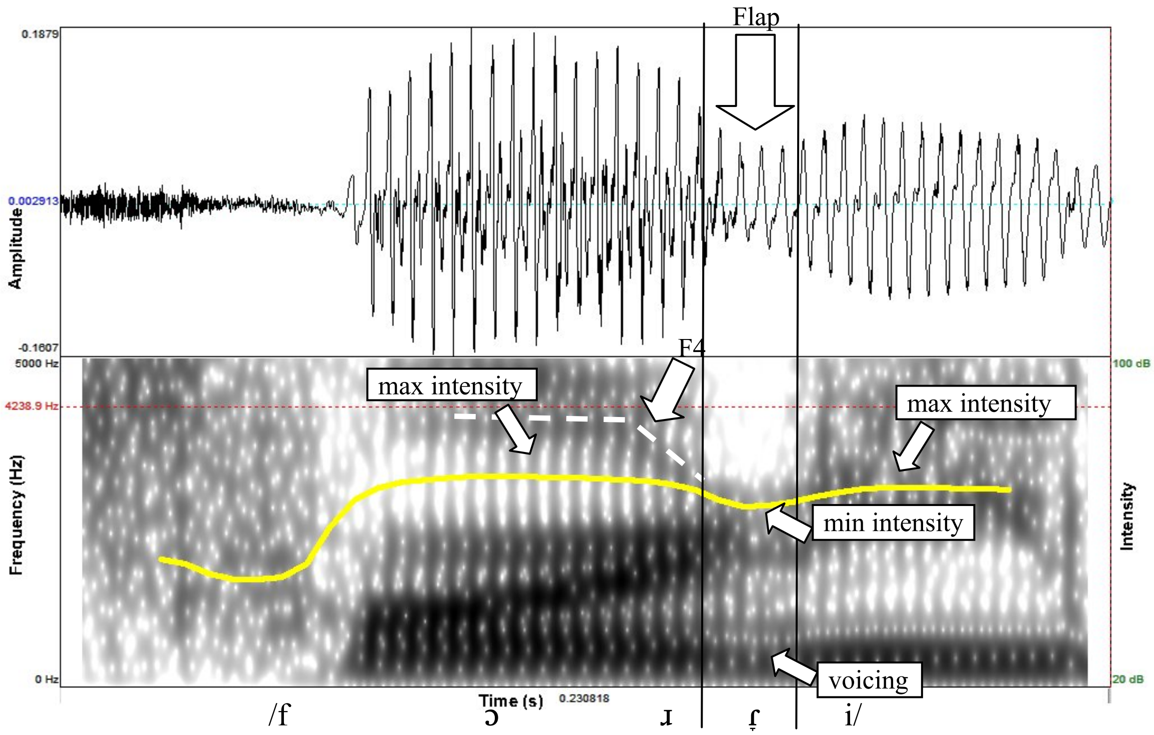


Figure 1.5 A spectrogram and wave form of the reduced flap from the word 'forty' as described by Warner & Tucker (2006) with intensity superimposed. Duration, maximum intensity, minimum intensity, F4, and voicing are indicated.

Acoustically, there are several characteristics indicative of reduced flaps. First, the formants from the surrounding vowels carry through the flap. Second, while there is still a decrease in intensity, that decrease is generally flatter with a smaller amount of change. In some cases the flap can become so reduced that there is very little acoustic realization of the flap. However, the drop in intensity and the dip in F4 generally persist regardless to the amount of reduction.

The results of the Warner and Tucker (2007) study suggest a phonological, categorical effect in the production of flaps (as opposed to [t] and [d] in the same environment) regardless of the speech style and also suggest both systematic and random phonetic variability. The results crucially show that reduction of the flap ([ɾ] to [ɾ̞])is more likely to occur in the spontaneous speech style as opposed to the more careful speech styles (read story and word list). Based on the results of this production study, American English flaps provide an important case in which to test spoken word recognition of reduced speech.

The main criteria used for the selection of the American English flap are: (1) flap reduction is more representative of spontaneous speech, thus will provide a good comparison for unreduced speech, and (2) a categorical, phonological abstract effect occurs in the acoustics of flap production, which may affect processing as compared to other intervocalic stops. In this dissertation a distinction is made that is different than previous work on flaps. Most work investigating flaps consider flapping a form of reduction and the non-flap oral stop [t] or [d] the unreduced form. In this dissertation flap reduction is investigated within the domain of the production of flaps. In other words, an unreduced target or flap refers to a canonical flap (e.g. Figure 1.4) and a reduced target or flap refers to an approximant-like or even nearly deleted flap (e.g. Figure 1.5).

1.5.3 Perception and processing of the American English flap

Several researchers have used the American English flap as a test case in perception experiments. Charles-Luce (1997) performed a production experiment followed by a perception experiment, which together investigated the influence of semantics and pragmatics on the phonemic contrast between /t/ and /d/ (e.g. *writer* versus *rider*). The voicing contrast between /t/ and /d/ is incompletely neutralized. The contrast is maintained by one of the following: “(1) closure duration is longer for /t/ but shorter for /d/ and/or (2) vowel duration preceding a stop is longer preceding a /d/ and shorter preceding a /t/” (Charles–Luce, 1997: 232). Subjects in the production experiment produced the targets by either reading a passage or a list of sentences which were either semantically neutral or biased. The production experiment showed a contrast when semantic content was biased and when a listener was present during the reading of target phrases. A contrast did not occur when semantic content was neutral and when there was no listener present. The perception experiment sought to determine whether listeners used these durational differences to distinguish between phonemic /t/ and /d/. The results showed that listeners significantly identified correct targets like *writer*/*rider* in an auditory forced-choice task. The results of this study showed that semantic and pragmatic context significantly affected both production and perception of incomplete neutralization in American English flaps.

Isardi and Sung (2004) performed a cross-linguistic study of the acoustics and perception of American English and Korean flaps. While American English flaps are underlyingly /t/ and /d/ Korean flaps are underlyingly /r/. They found that acoustically, American English and Korean flaps are very similar and have no statistically significant differences in closure duration, voicing, or occurrence of bursts. In the perception studies they suggested that the results indicate that listeners perceive flaps based on their first language background. When American English speakers hear both Korean and American English flaps they consistently categorize them as /d/. However, Korean speakers consistently categorize flaps from both languages as a liquid /r/ or /l/.

McLennan, Luce, and Charles-Luce (2003) investigated whether during lexical access flaps activate their phonemic counterparts, /t/ and /d/. In six repetition priming experiments combinations of two tasks were used: shadowing and lexical decision. Stimuli were presented in two blocks: the prime block and the target block. An item like *Adam* was produced carefully [ædəm] and casually [æɾəm], so that the carefully produced stimuli were non-reduced oral stops (not flaps). This is different from the distinction made above for reduced and unreduced flaps (as in Warner and Tucker, 2007), where an item like *Adam* would be produced as reduced [æɾ̥əm] and unreduced [æɾəm].

McLennan, Luce, and Charles-Luce (2003) found, as a result of these experiments, that flaps activated their underlying phonemic counterparts. In their Experiments 1 and 2, using a shadowing task in both the prime and target blocks, matches (identical stimuli, e.g., listener heard [ædəm] in the prime block and heard [ædəm] in the target block) and mismatches (e.g. [ædəm] with [æɾəm]) both showed a priming effect indicating activation of the underlying counterparts. These experiments also contained a non-alveolar comparison group. In the non-alveolar group, only the match condition primed significantly. Their Experiments 3 and 4 used lexical decision tasks, which varied in degree of difficulty in non-word stimuli. Experiment 3, which had non-words labeled as "easy" (e.g. *thushshug*) to discriminate, showed no priming effect, while Experiment 4, which had non-words labeled as "difficult" (e.g. *bacov*) to discriminate, showed activation of the underlying phonemes. Their Experiments 5 and 6 combined the shadowing and lexical decision tasks so that in Experiment 5 the priming block was a shadowing task and the target block was an "easy" lexical decision task and in Experiment 6 the task pairing was reversed, i.e. the priming block was the "easy" lexical decision task and the target block was the shadowing task. For both experiments, activation of the underlying phonemes was found, which they argued supported models of mediated lexical access.

Connine (2004) investigated the American English flap and the nature of its representation in the lexicon. A phoneme identification experiment was used to investigate the nature of the representation. Listeners identified the initial segment of words like *pretty* and *bretty*, which typically contain flaps. Similar to McLennan et al. (2003), Connine contrasted flapped items with [t], a continuum was made containing 6 steps from [t] to [ɾ]. Connine found that listener responses forming real words (e.g. *pretty* as opposed to *bretty*) were more common for targets in a flap context than a [t] context. She claims that these results illustrate that the highly frequent flap is contained in the lexical representation of words containing flaps.

1.6 Summary

In daily communication a wide variety of speech styles are used. Research on reduction indicates that production of speech varies depending on the particular speech style a speaker uses. The literature on reduction also indicates that in spoken word recognition different strategies can be employed by listeners when recognizing speech produced in different speech styles. While the reduction literature indicates differences which are dependent on speech style, there remains only a small amount of research on the topic.

Research by Warner and Tucker (2007) indicates that the American English flap can provide an excellent test case for investigation. Previous research on American English flaps has shown that there are articulatory and acoustic differences in flap production which are dependent on speech style. Perception research on flaps has not investigated the effect of reduction, per se, but has investigated the effects of semantics and pragmatics, first language, and lexical form on the recognition of flaps.

The research presented thus far has shown that a need exists for further investigation of reduced speech and that flap reduction in American English can provide a domain in which this can be studied. For convenience, the main research questions are recapitulated: How do listeners process reduced speech? Do listeners adjust their expectations for how segments are realized based on speech style? Further breaking down these questions:

1. How does reduction affect listeners' recognition of words?
    a. Is it more difficult for listeners to recognize words pronounced in reduced forms, implying that weaker acoustic cues cause a conversion of such forms into their abstract underlying representation?
    b. Or is it perhaps easier for listeners to recognize reduced forms, implying that reduced forms are more commonly produced and thus more quickly accessed (frequency of occurrence effects)?
2. Do listeners adjust their expectations about reduction based on preceding speech style (context)?

This research investigates these questions using four experiments, outlined above and summarized here. Experiments 1 and 2 examine listeners' responses to reduced and unreduced words in isolation. Experiments 3 and 4 probe whether the introduction of different speech styles (preceding context) affects how listeners respond to reduced versus unreduced words. The findings of these experiments designed to address the above questions have implication to representational issues in spoken word recognition models. The findings also may show that processing of surrounding context, containing speech style information, can influence processing of other speech.

CHAPTER 2

PROCESSING OF REDUCED FLAPS IN ISOLATED WORDS

2.0 Introduction

Conversational, or reduced speech, while more common than careful speech in daily communication, is relatively unstudied (Cutler, 1998; Ernestus et al., 2002). This chapter investigates the processing of reduced and unreduced flaps in words occurring in isolation. Two types of experimental tasks are used to investigate the effect of reduction on processing: auditory lexical decision and cross–modal identity priming. Both tasks require that the subjects access the lexicon in order to respond.

This chapter is organized as follows: Section 2.1 introduces and reports the results of Experiment 1, the auditory lexical decision task. Section 2.2 introduces and reports the results of Experiment 2, the cross–modal identity priming task. Section 2.3 provides a general discussion of the results of Experiments 1 and 2 as well as conclusions drawn from these results.

2.1 Experiment 1: Lexical decision

The first task, an auditory lexical decision task (Marslen–Wilson, 1980; Goldinger, 1996a), is used to determine the effect of reduction of a particular segment, the American English flap on processing. The auditory lexical decision task is a speeded response task in which subjects are presented with an auditory stimulus and asked to identify it as either a word or nonword of the language in question. In order for subjects to make judgments based on the auditory input they must access information in the lexicon; as the lexicon is accessed, it becomes possible to indirectly investigate its structure and contents (Goldinger, 1996a).

The auditory lexical decision task has been used to show the effect of lexical variables on access time as well as the nature of lexical representation (Goldinger, 1996a). Effects of word frequency and neighborhood density as well as phonological, inhibitory phonetic, and semantic priming have all been shown using auditory lexical decision (see Goldinger, 1996a for a summary). For example, word frequency effects (Luce, 1986; Marslen–Wilson, 1990), (i.e. listeners respond to high frequency words faster than low frequency words) have been shown to influence lexical access. Neighborhood density, where words with dense neighborhoods are more difficult to process than words in less dense neighborhoods, has also been shown to influence lexical access (Luce, 1986; Luce and Pisoni, 1998). The demonstrated influence of these variables on lexical access has contributed to the understanding of lexical representation.

The lexical decision task is used here to establish whether a difference exists between the access time required by a listener to respond to a word containing a reduced flap and the access time required to respond to a word containing an unreduced flap, as described in Chapter 1. A response time that is longer or shorter for a reduced word than an unreduced word indicates an effect of reduction on processing. For example, a listener might process reduced items more slowly than unreduced items due to decreased acoustic information in the speech signal. Conversely, a listener might process reduced items more quickly than unreduced items, meaning that lexical access is facilitated by frequency of occurrence. This result would be similar to the result as described in the lexical frequency work mentioned previously. However, this result differs in that it refers to the frequency of occurrence of a particular acoustic pattern as the realization of a phoneme as opposed to the lexical frequency, or the number of times that a particular item occurs. In this sense, the reduced acoustic patterns are more common; it is this acoustic pattern that listeners encounter in their day–to–day interaction as opposed to careful, or "laboratory" acoustic patterns.

2.1.1 Participants

Listeners were 64 native speakers of American English, recruited at the University of Arizona. The majority of listeners were recruited from the linguistics undergraduate pool, and consisted mainly of students enrolled in Introduction to Linguistics courses. All received extra credit for participating and none reported any known hearing loss. Listeners' ages ranged from 18-60 years. All listeners were native speakers of American English, over 50% were either studying a foreign language (generally Spanish or French), or were late bilinguals. Nineteen of the listeners were self-reported monolinguals.

2.1.2 Materials

Eighty words were selected as target items, of which 40 contained flaps (represented by 'd' orthographically) and 40 contained word–medial /g/ (represented by 'g' orthographically). As described in Chapter 1, flapping in American English can occur for both 't' and 'd', when they occur in the appropriate environment. Flapped items were selected so that they were represented orthographically by 'd' so as to reduce any potential effects of orthography. Flapped items are referred to as phonemic /d/ items which reflects the orthographic distinction already established. All /d/ and /g/ items were bisyllabic (CVCV, CVCVC, VCV, VCVC) and occurred in flapping environments, i.e., post stress. Two instances of each item were selected (recording procedures described below), one in a reduced form and the other in an unreduced form. Unreduced targets were canonical flaps (not a stop [d], [t] or [t^h], see Figure 2.1 below) while reduced targets were reduced (approximant-like) flaps (see Figure 2.2. below).

In addition to the 80 real–word target items there were 220 fillers (200 non–word fillers and 20 real–word fillers) and 10 additional practice items. All non–words were bisyllabic (CVCV, CVCVC, VCV, VCVC) and phonotactically possible English words. Table 2.1 below contains samples and distribution of the items used. A complete list of the target items and all fillers is provided in Appendix A.

Table 2.1 Example lexical decision items used in Experiment 1 with the total number of items occurring.

	Flap /d/	**/g/**	**Real–word Filler**	**Non–word Filler**
Example Item	puddle	baggy	copy	kooga /kuga/
Number	40	40	20	200

Word–medial /g/ items were selected to act as a comparison group to the flap items (/d/ phonemes). Word–medial /g/ items were selected because they do not undergo a phonological process similar to the process that /t/ and /d/ undergo to become a flap. Word–medial /g/ does undergo reduction similar to flaps. Having both word–medial /g/ and the flap allows for a comparison that may indicate if there are phonological effects on the processing of flaps as opposed to word–medial /g/. Frequency and neighborhood density of target items was controlled by using a within-items design for the factor Target Reduction (reduced versus unreduced). As a within–items factor, the comparison involves the same word, which has the same frequency and neighborhood density.

Items were recorded over several recording sessions by a female native speaker of American English who is trained in phonetics and was aware of the purpose of the experiment. Recordings were made in a sound attenuated booth using identical equipment for each session. The speaker was instructed to produce multiple tokens (at least three) of both reduced and unreduced versions of the words. The speaker is involved in the flap production project (Warner and Tucker, 2007) and was therefore familiar with what reduced and unreduced targets are like. Thirty–seven non–word fillers containing a word–medial /d/ or /g/ were produced several times each with various degrees of reduction. From these non–word fillers twenty–four were selected containing a reduced production similar to the target items, and 13 non–word fillers were selected as unreduced productions.

Item selection was based on a visual comparison of the targets to the data from Warner and Tucker (2007). Specifically, acoustic cues such as existence of a burst release, change in intensity, and strength or existence of formants during the flap were used as visual criteria. Experimental items were then compared with the data from Warner and Tucker (2007, discussed in Chapter 1) to confirm that the acoustics of reduced and unreduced versions are comparable to production data. Items were measured using the same methods from Warner and Tucker (2007) and the results of these measurements are discussed below in section 2.1.3.

Figures 2.1 and 2.2 below show examples of the item *puddle*. These examples show the waveform on the top and the spectrogram on the bottom, with intensity superimposed on the spectrogram. These images specifically illustrate the difference between the reduced and unreduced target items in the experiment.

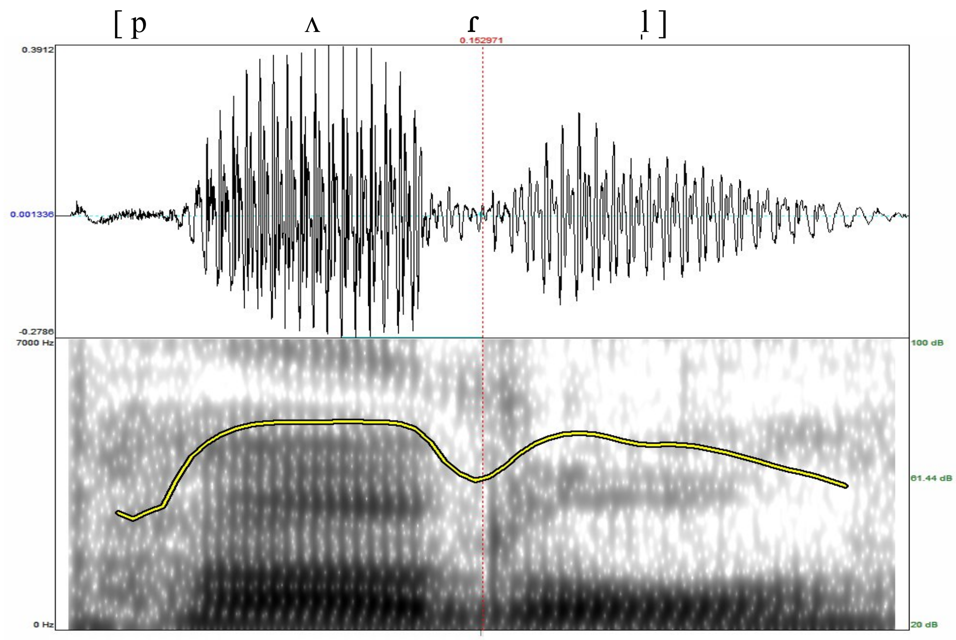


Figure 2.1 Waveform and spectrogram of the unreduced production of the item *puddle*. The line represents the intensity of the sound overlaid on the spectrogram.

In Figure 2.1 the unreduced flap has a noticeable burst release, stop closure, voicing occurring throughout the closure duration, and a large fall in intensity during the flap duration closure. Figure 2.2 below contrasts the reduced token with the unreduced token from above.

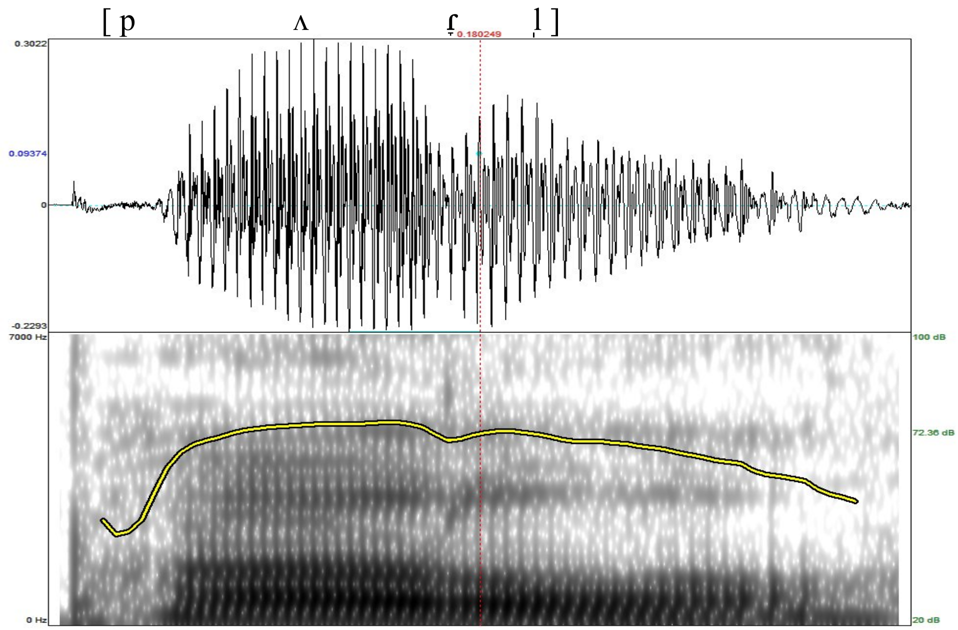


Figure 2.2 Waveform and spectrogram of reduced item *puddle*. The line represents the intensity superimposed on the spectrogram.

In Figure 2.2 the reduced flap has no noticeable burst release, a less obvious stop closure, voicing occurring throughout the flap duration, and a much smaller fall in intensity. Importantly, the formants continue through the flap, causing it to resemble an approximant, while the formants in the unreduced token do not continue through the flap, which is more stop–like. Similar images are provided for the word *baggy* in Figures 2.3 and 2.4, below.

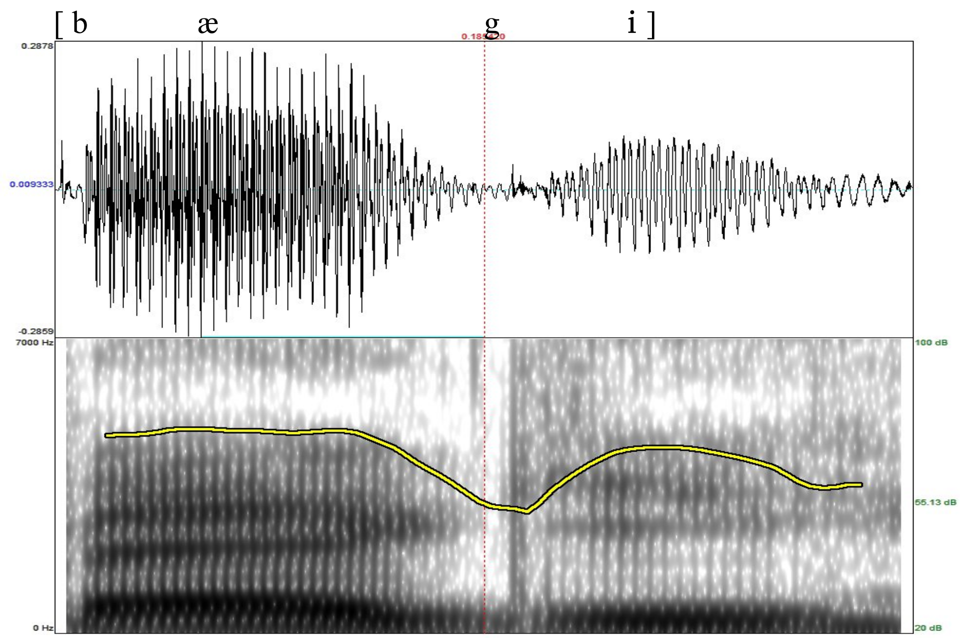


Figure 2.3 Waveform and spectrogram of unreduced item *baggy*. The line represents the intensity of the sound overlaid on the spectrogram.

In Figure 2.3 the word–medial /g/, like a flap, has a noticeable burst release, stop closure, voicing occurring throughout the closure duration, and a large fall in intensity during the closure. Figure 2.4 below contrasts the reduced /g/ with the unreduced /g/ from above.

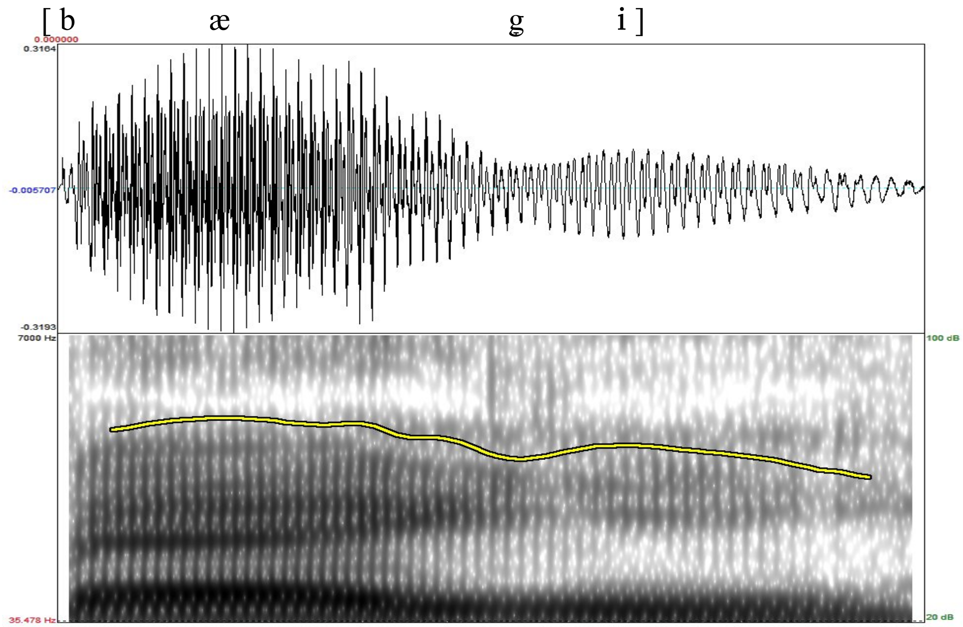


Figure 2.4 Waveform and spectrogram of reduced item *baggy*. The yellow line represents the intensity of the sound overlaid on the spectrogram.

In Figure 2.4 the reduced /g/ has no noticeable burst release, a less obvious stop closure, voicing occurring throughout the consonant duration, and a much smaller fall in intensity. Importantly, the formants continue through the consonant, causing it to resemble an approximant, while in the unreduced token the formants do not continue through the consonant, which is more stop–like.

2.1.3 Item measurements and comparison to reduction

As previously stated, after items were selected they were analyzed using the same methodology as Warner and Tucker (2007) so that reduced and unreduced target items could be compared to the production data. The purpose here is to show that the items in this experiment, Experiment 1, are comparable to reduced and unreduced forms (from Warner and Tucker, 2007) produced in a natural setting and to show that the reduced and unreduced tokens are acoustically different. Of the 160 total target items, three items were too reduced to obtain accurate measurements and were excluded from results described. Four measures were used to examine the target data. All of the measures were compared directly to the production data. All of the measures were shown by Warner and Tucker (2007) to be strong acoustic indicators of reduction.

The intensity difference was created by measuring the intensity of the preceding and following vowels and finding their mean intensity value. The minimum intensity was subtracted from the mean intensity value, giving the amount of change in intensity. This difference can be used as a measure of reduction. The closer the difference is to 0 the more reduced the item is. For example, in Figures 2.1 and 2.3 above, the difference for the unreduced item *puddle* (14.44, Figure 2.2) is larger because of the large drop in intensity during the flap closure (as compared to 3.57 for reduced *puddle*). However, in the reduced item *baggy* (7.55, Figure 2.3) the intensity drops only a very little during the flap (more of a flat line, as compared to 19.82 for unreduced *baggy*). Mean difference values of reduced and unreduced targets were then compared to a percentile distribution of the production data. The mean intensity difference for both reduced /d/ (3.6331) and for /g/ (5.5696) targets fell below the 10th percentile (comparable to the bottom 10% of reduced stimuli, i.e. the most reduced stimuli) in the production data. For the /d/ unreduced targets (16.4332) the difference fell above the 88th percentile and for the /g/ targets (21.0464) the difference fell above the 70th percentile. A table of the complete distribution is provided in Appendix B. Figures 2.5 and 2.6 below are histogram plots illustrating the distribution of the differences for both /d/ (Figure 2.5) and /g/ (Figure 2.6).

The histograms compare the distribution of the difference data for the reduced and unreduced tokens to the distribution of the production data from Warner and Tucker (2007) for all speech styles (conversational, story, word list).

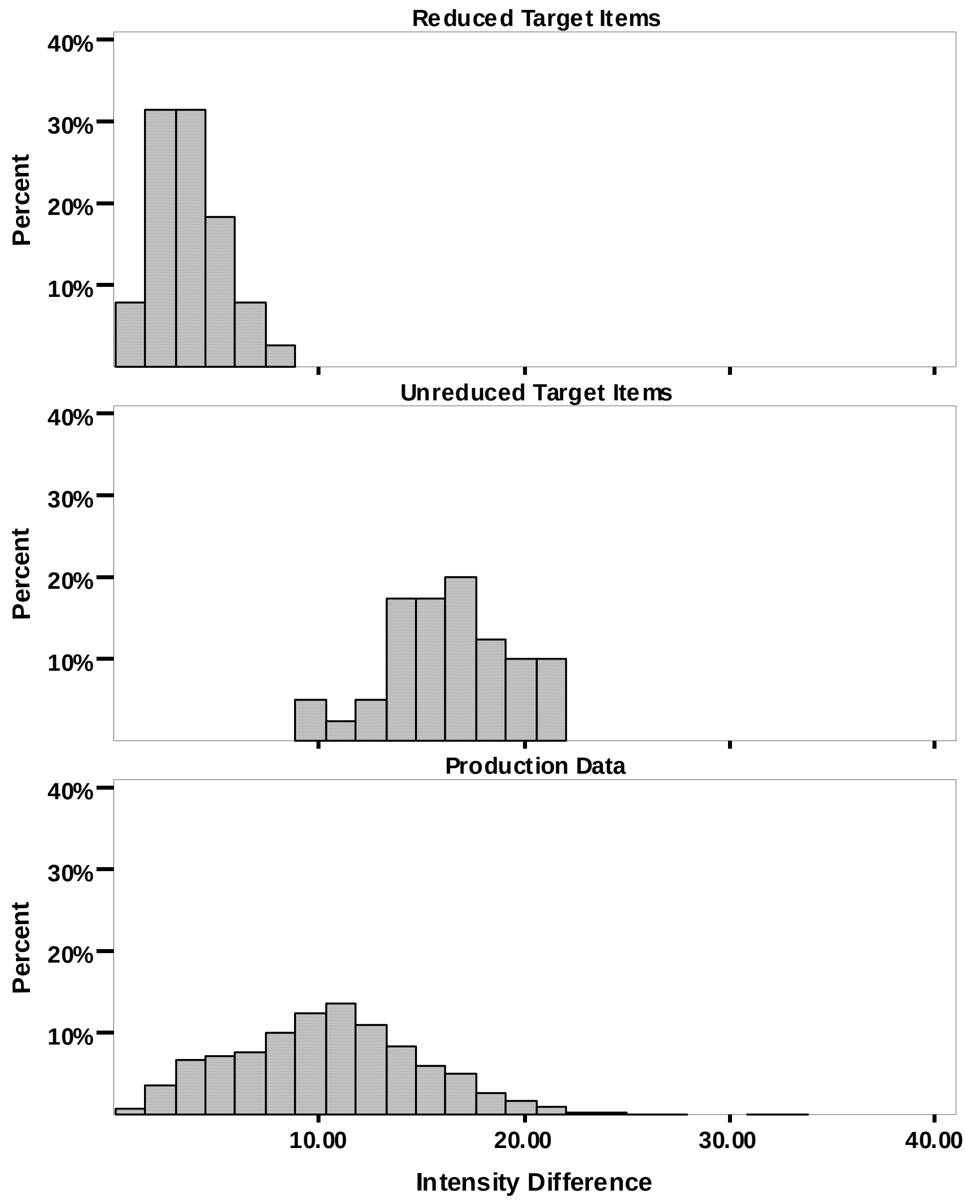


Figure 2.5 Histograms of intensity difference for /d/ targets comparing unreduced and reduced stimuli from Experiments 1 and 2 to production data from previous experiment (different speakers, Warner and Tucker 2007).

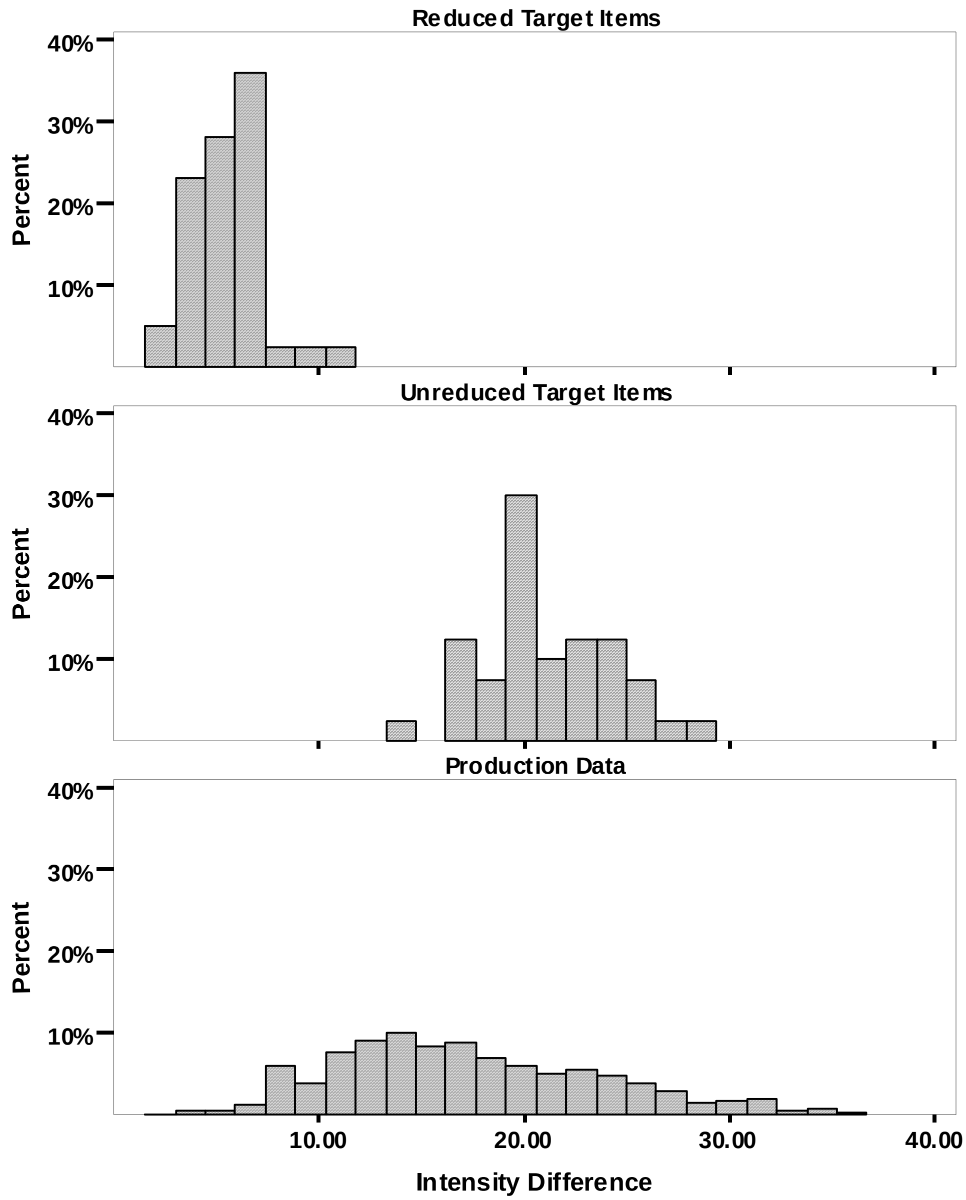


Figure 2.6 Histograms of intensity difference for /g/ targets comparing unreduced and reduced stimuli from Experiments 1 and 2 to production data from previous experiment (different speakers, Warner and Tucker 2007).

The segment duration of the target items was also measured. Consonant duration was much more difficult to determine because of the need for various criteria for onset and offset of the consonant, because the consonant can be an approximant or a flap (or /g/). In the unreduced tokens the consonant onset was selected at the end of the second formant (F2) in the preceding vowel and the offset was selected at the onset of F2 in the following vowel. In the reduced tokens this selection criterion could not be applied because the formant generally carried through the consonant of the reduced items (see Figure 2.5 and accompanying discussion below). In these cases, duration was determined by a combination of "best–fit" judgment of onset and offset or at half–way points along the intensity curve between peak and minimum intensity. As such, the duration measure is less accurate. The mean duration of the reduced /d/ targets (12ms) fell above the 80th percentile and the duration of the reduced /g/ targets (15ms) above the 90th percentile. The unreduced targets for /d/ (34ms) and /g/ (55ms) fell below the 40th and 60th percentiles, respectively. A table of the complete distribution is provided in Appendix B. Figures 2.7 and 2.8 below are histogram plots illustrating the distribution of the durations for both /d/ (Figure 2.7) and /g/ (Figure 2.8). The histograms compare the distribution of the duration data for the reduced and unreduced tokens to the distribution of the production data from Warner and Tucker (2007) for all speech styles.

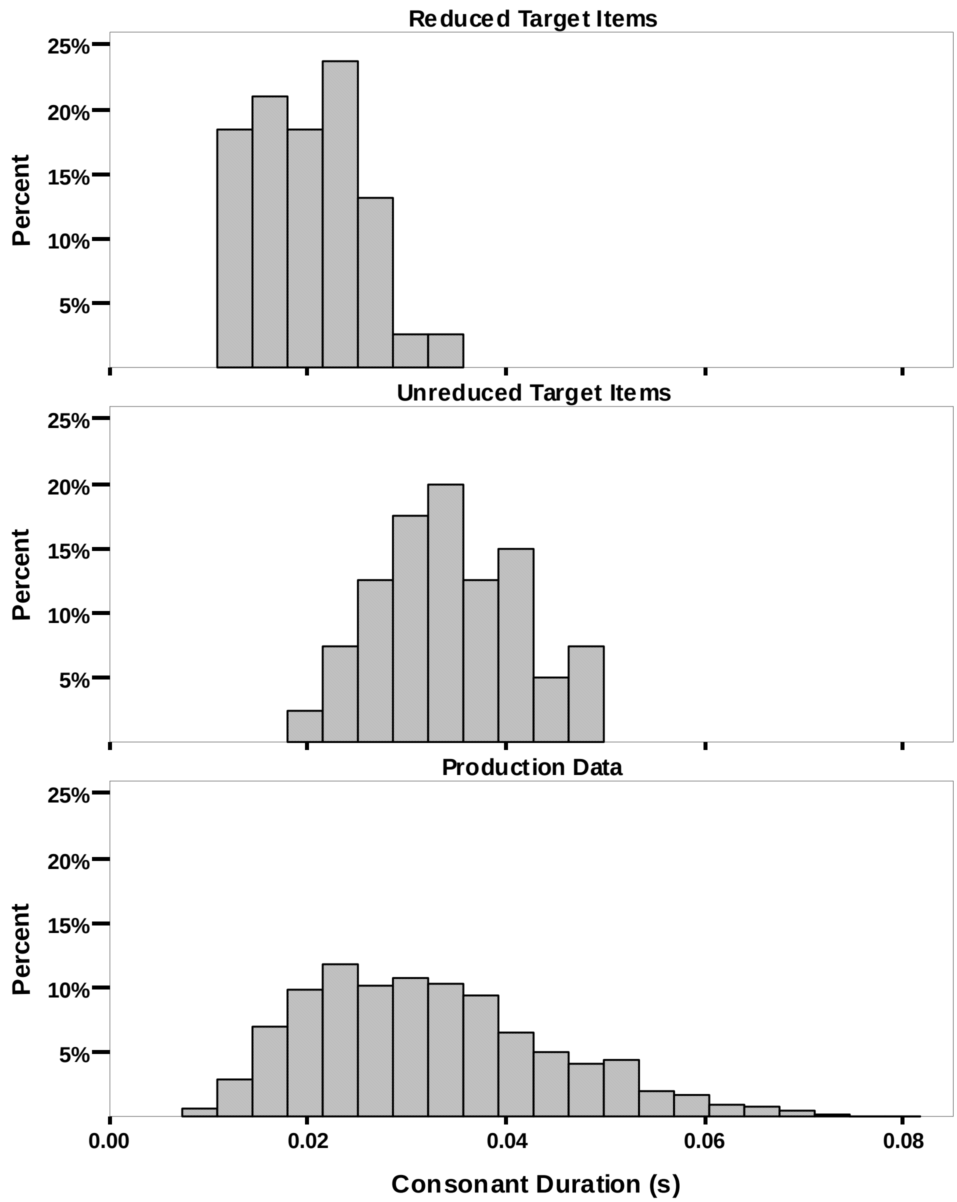


Figure 2.7 Histograms of consonant duration for /d/ targets comparing unreduced and reduced stimuli from Experiments 1 and 2 to production data from previous experiment (different speakers, Warner and Tucker 2007).

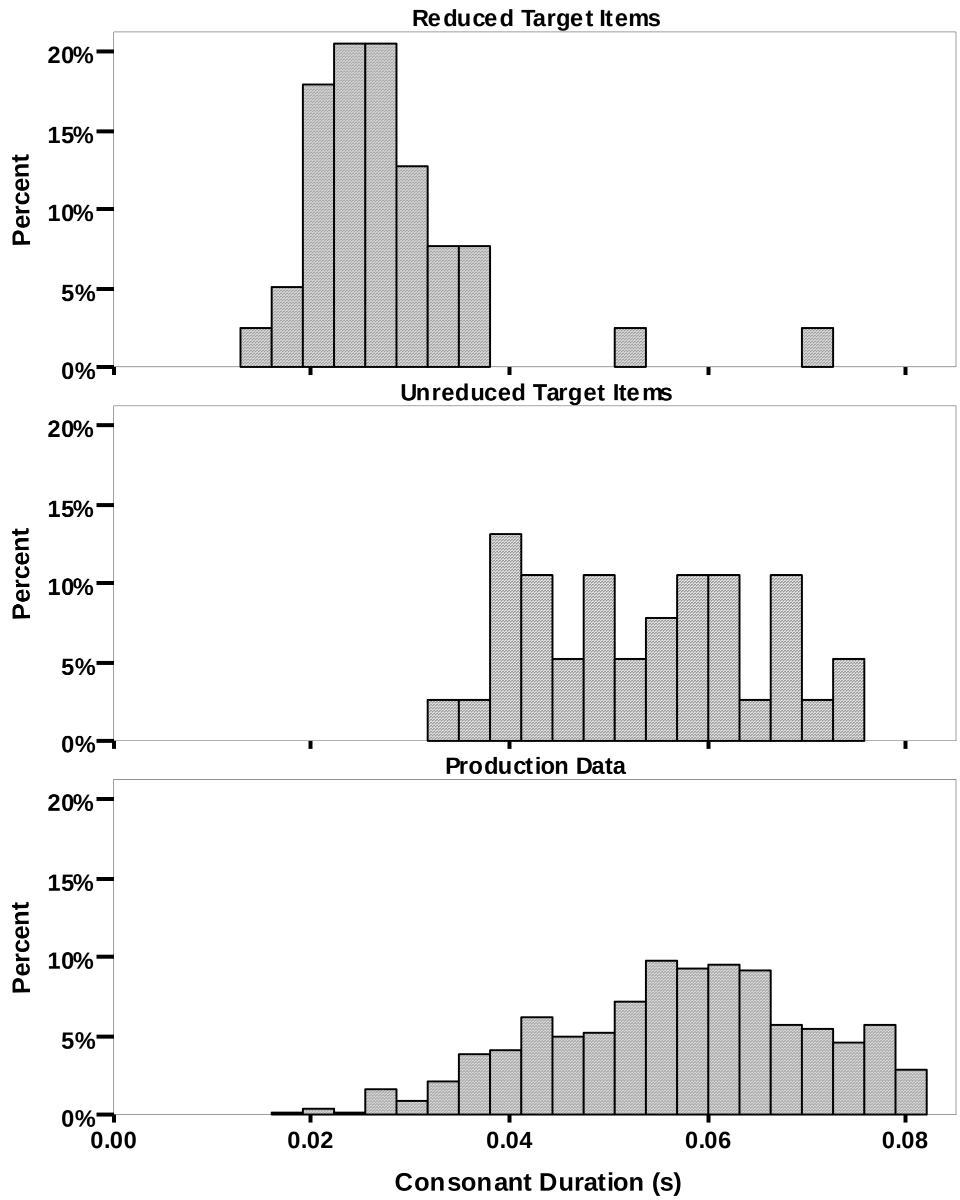


Figure 2.8 Histograms of consonant duration for /g/ targets comparing unreduced and reduced stimuli from Experiments 1 and 2 to production data from previous experiment (different speakers, Warner and Tucker 2007).

As mentioned previously, in the reduced items it is more common for formants to continue through an entire reduced target than the unreduced targets (as illustrated in Figures 2.1–2.4), above. Figures 2.9 and 2.10 below illustrate the mean percentage occurrence of formants in target segments as compared to individual speakers from Warner and Tucker (2007). For the experimental stimuli below the reduced stimuli were coded as “Conversational” and the unreduced stimuli as “Word List Reading”.

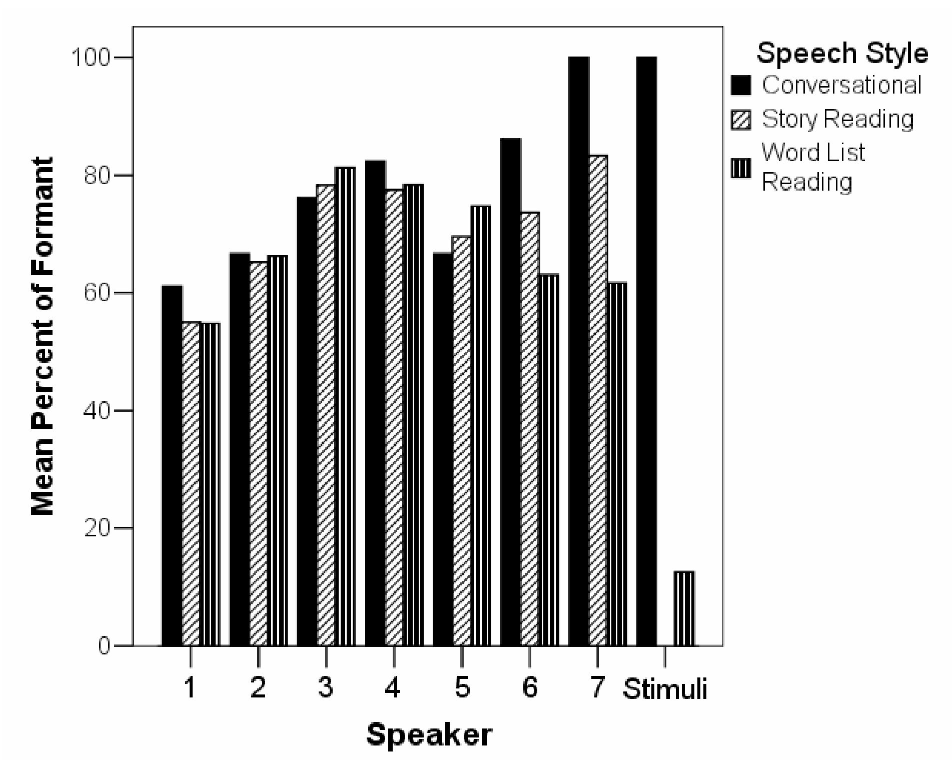


Figure 2.9 Mean percentage occurrence of formants through the stop closure for /d/ target items. Experimental stimuli are labeled as “Stimuli”, reduced targets are labeled as Conversational and unreduced targets are labeled as Word List Reading.

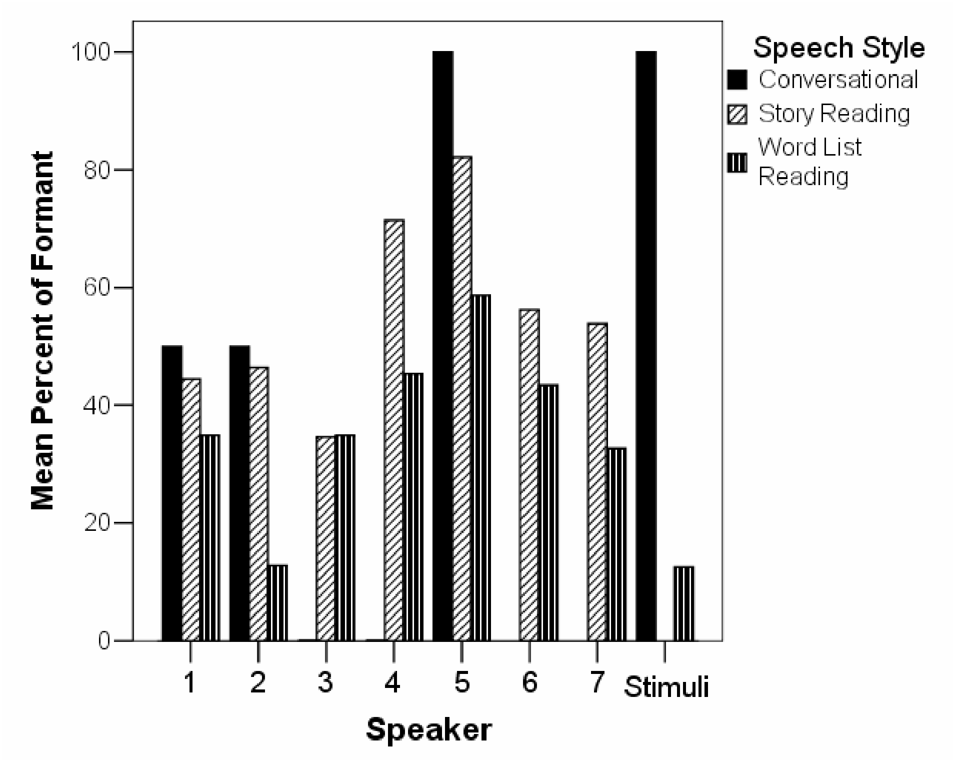


Figure 2.10 Mean percentage occurrence of formants through the stop closure for /g/ target items. Experimental stimuli are labeled as "Stimuli", reduced targets are labeled as Conversational and unreduced targets are labeled as Word List Reading.

In the reduced targets both /d/ and /g/ items occur at 100%, while in the unreduced targets both phonemes occur at around 15%, indicating that in about 6 of the 40 items strong formants carried through the closure duration. The production data show that for most of the speakers formants occur more frequently in the more casual speech style. For the experimental stimuli the reduced and unreduced tokens fall on the extreme ends of the distribution for individual speakers, showing that in this measure they are good examples of reduced and unreduced speech.

The last measure considered here is the occurrence of a burst release in the items (as seen in Figures 2.2 and 2.4). Following a stop closure, a sudden broadband noise is generally visible in the spectrogram. This noise results from the release of pressure built up in the oral cavity during the stop closure. When a burst does not occur following a closure, it may indicate that the tongue did not create a full closure with the alveolar ridge or velum during the consonant. Data from Warner and Tucker (2007) indicate that the occurrence of bursts correlates well with the speech style. Figures 2.11 and 2.12 below show the inverse percentage occurrence of bursts in the /d/ and /g/ target items as compared to individual speakers from Warner and Tucker (2007).

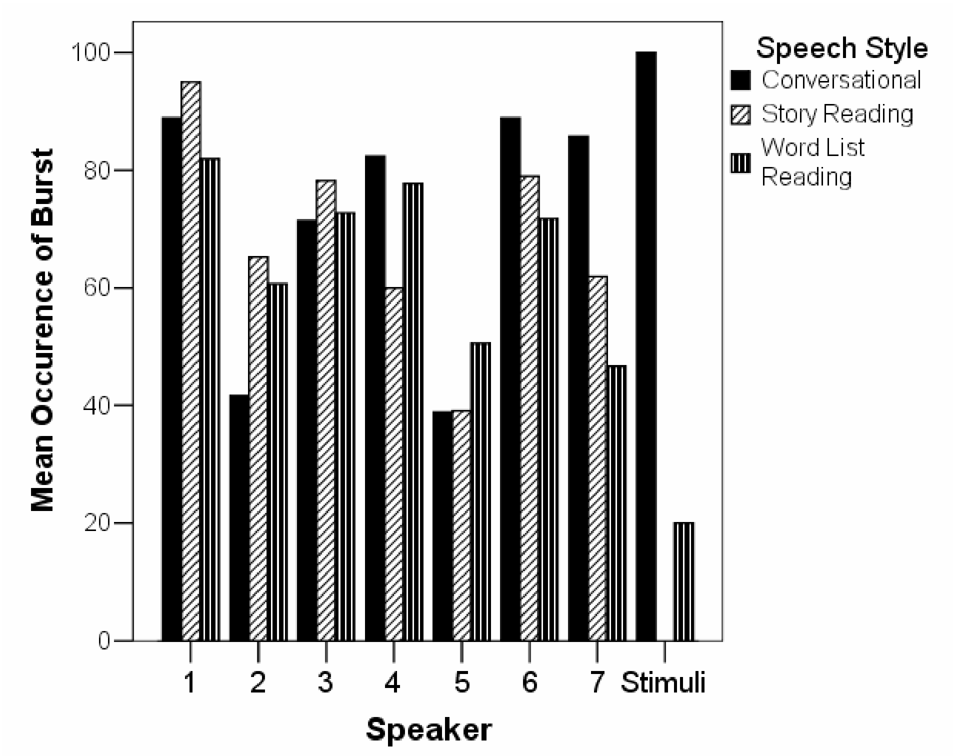


Figure 2.11 Mean inverse percentage occurrence of bursts for /d/ target items, where 100% indicates no bursts. Experimental stimuli are labeled as "Stimuli", reduced targets are labeled as Conversational and unreduced targets are labeled as Word List Reading.

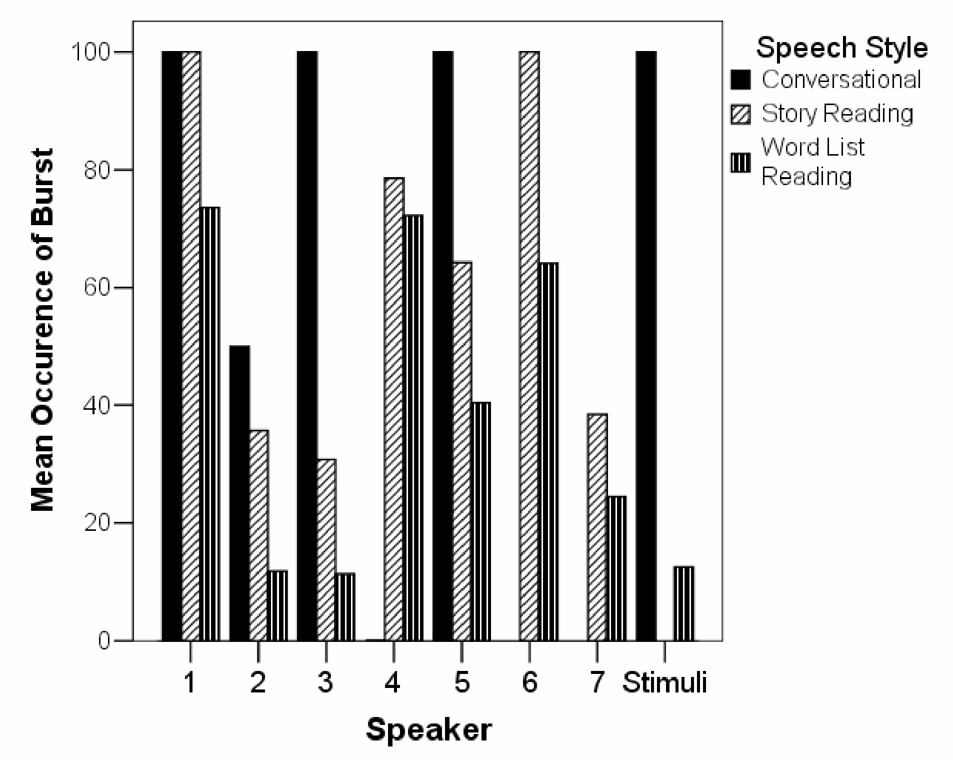


Figure 2.12 Mean inverse percentage occurrence of bursts for /g/ target items, where 100% indicates no burst occurred. Experimental stimuli are labeled as "Stimuli", reduced targets are labeled as Conversational and unreduced targets are labeled as Word List Reading.

In the reduced items bursts never occur (100%), while in the unreduced items bursts are much more likely to occur. The production data show that for most speakers, bursts occur more frequently in the more casual speech style. However, for the /d/ items there are some exceptions in the production data, which may indicate that the occurrence of bursts in flaps is less predictable. Again the experimental stimuli fall on the extreme edges of the production data for this measure. This is an additional indication that the experimental stimuli are good examples of typical reduced and unreduced speech segments.

In summary, the targets produced for the processing experiment are very similar to stops from the production study of Warner and Tucker (2007). The intensity differences generally fall in the appropriate areas on the spectrum of reduced versus unreduced. Both formant and burst indications show that there is a difference between the reduced and unreduced stops. The duration data are slightly less conclusive, in that the duration means fall around the middle of the spectrum of the production data, particularly the unreduced /g/ durations. While the duration distribution is not ideal, the other three measures show that there is a strong difference between the reduced and unreduced items.

2.1.4 Procedure

All items were excised from the recordings described above. Items were then counterbalanced across the independent variable Target Reduction (reduced vs. unreduced) creating two lists of targets, so that no one listener received the same item in both reduced and unreduced forms. Each list was pseudo-randomized with all of the 220 fillers. Item lists were presented in one of two orders (opposite orders from each other) to control for possible order effects, thus creating four possible lists of 300 items each. Each listener heard 80 target items and 220 fillers. As part of the pseudo–randomization of the lists, all target items were always preceded by a non–word.

Listeners were seated in separate sound–attenuated booths for the experimental procedure. This experiment and all other experiments in this dissertation were presented using the *E–prime* experimental presentation software package. Listeners received written directions instructing them to determine as quickly as possible whether the word presented over the headphones was a real word of English. Two orthographic examples were given to listeners with one a real word and the other a nonword, and each was identified for the listeners as part of the instructions. After the instructions, listeners proceeded to a 10–word practice session before moving on to the main experiment. Listeners were presented with the visual cue "Get ready" for 500ms and then the auditory stimulus was presented. Listeners were asked to perform a lexical decision task on the words presented by selecting a button labeled either "YES" or "NO". The "YES" button was always on the listener's left and the "NO" button was always on the listener's right. Each listener's response latency and percent error were recorded for analysis. Response latencies were measured from the offset of the auditory stimulus.

2.1.5 Results

The results from the response latencies are presented first, followed by the percent error results. Items with response latencies less then 220ms or greater then 1500ms were considered errors and excluded. Cutoffs for response latencies were chosen based on a histogram of the distribution of all response latencies. The cutoffs were selected at points were the distribution leveled out on either end. Additionally, data for five of the items was excluded due to high error rates (below 30% correct)[1]. Nine subjects' data was excluded from the analysis because they scored below 70% correct on the target items. All incorrect responses were excluded from the response latency analysis. Figure 2.13 below summarizes the response time data.

[1] The items exclusion cutoff for error rate was set low as a result of the difficulty of the auditory lexical decision experiment. Had a cutoff been selected at 70% (the same as for subjects) 32 items would have been excluded from the analysis, decreasing the statistical power and ability to generalize the results.

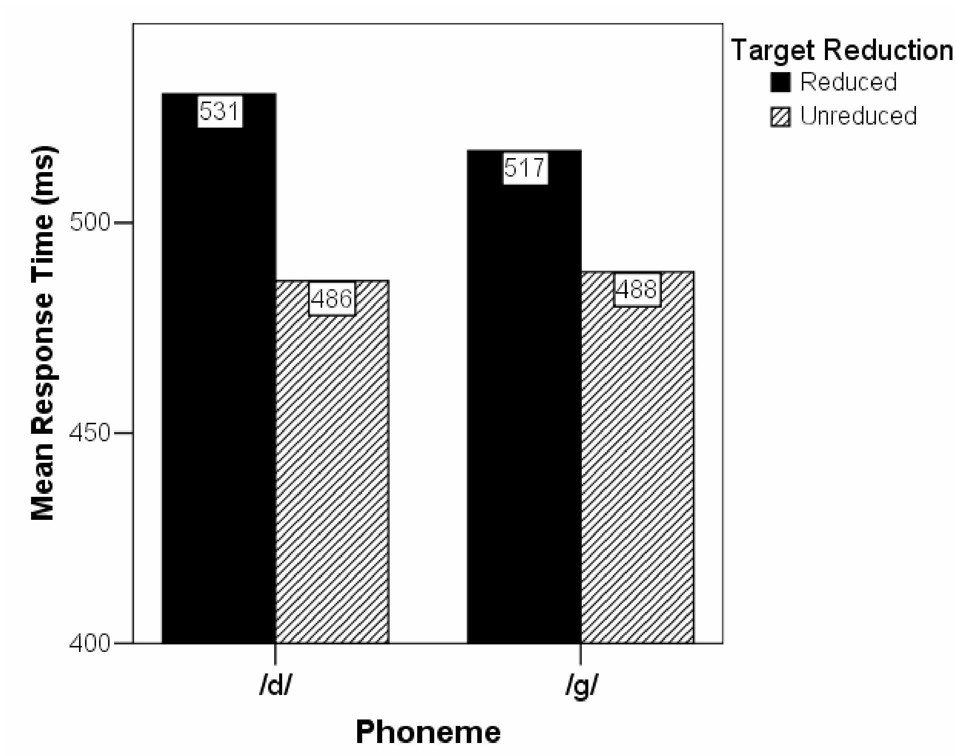


Figure 2.13 Mean response time for auditory lexical decision for reduced versus unreduced /d/ and /g/ word–medial phonemes.

An Analysis of Variance (ANOVA) was performed on the response latencies for listeners (*F1*) or items (*F2*) as repeated measures. For this analysis there are two independent variables: Target Reduction (reduced vs. unreduced) and Phoneme (/d/ vs. /g/)[2]. Both are repeated measures factors for the by–subjects analysis and only Target Reduction is a repeated measure in the by–items analysis. The main effect of Target Reduction showed that listeners' response latencies are significantly inhibited by reduction: $F1(1,43)=27.831$, $p<0.001$; $F2(1,73)=18.636$, $p<0.001$. The main effect of Phoneme was not significant: $F1(1,43)=1.756$, $p>0.05$; $F2<1$. The interaction of Target

[2] In this and all subsequent analyses an additional factor, Counter-Balanced List, is included in the analyses. If no significant effects are found then this factor is excluded from discussion.

Reduction and Phoneme was also not significant: *F1*(1,43)=1.207, p>0.05; *F2*<1. Percent error data were also analyzed; Figure 2.14 summarizes the percent error data for Experiment 1.

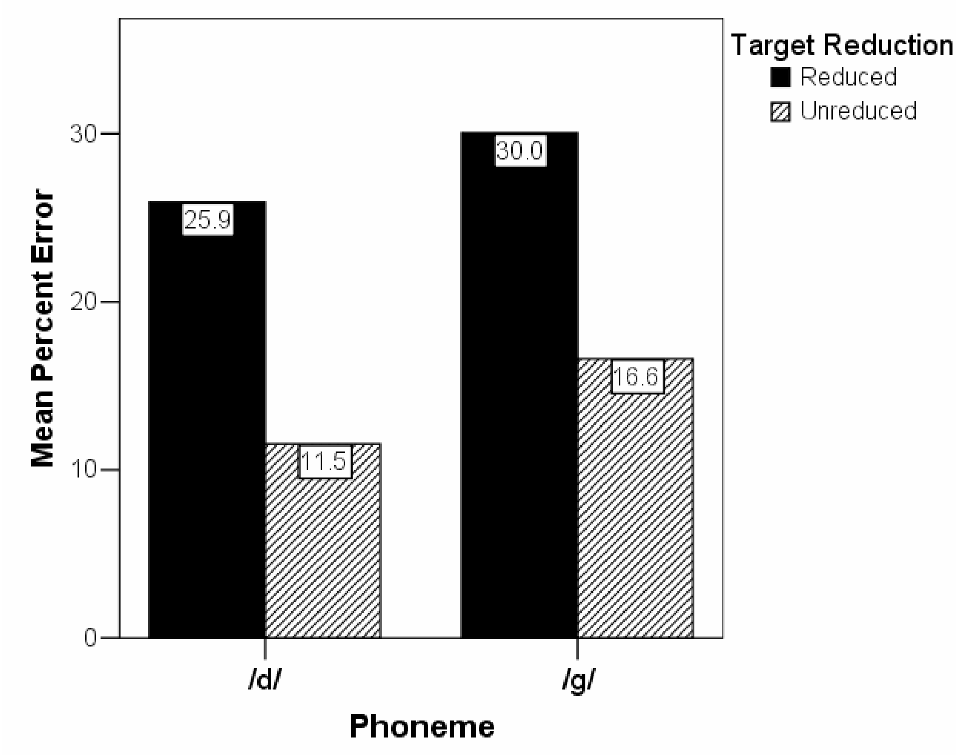


Figure 2.14 Mean percent error for auditory lexical decision for reduced versus unreduced /d/ and /g/ word–medial phonemes.

An ANOVA was performed on percent error of responses by–subjects (*F1*) or by–items (*F2*) as repeated measures, using the same statistical design as for the response latency data. Subjects made significantly more errors for reduced tokens than for unreduced tokens (main effect of Target Reduction: *F1*(1,51)=102.744, p<0.001; *F2*(1,73)=58.259, p<0.001). The main effect of phoneme was significant by–subjects but not by–items (*F1*(1,51)=14.561, p<0.001; *F2*(1,73)=1.98, p>0.05). The interaction of Target Reduction and Phoneme was not significant: *F1*<1; *F2*<1.[3]

---

[3] This experiment was designed specifically to investigate processing of reduced versus unreduced productions of words. Since this comparison compared different versions of the same word, frequency was not controlled. It is possible that Frequency interacts in some way with Target Reduction. A post hoc analysis was performed to investigate whether Frequency statistically interacts with Target Reduction (frequency counts are from the Brown Corpus (Kučera and Francis, 1967)). Items were split into high and low frequency tokens. This split was made by dividing the complete list of /d/ items and the complete list of /g/ items into high and low groups based on the frequency distribution of items in each list. For /d/ items low frequency words fell between 0 and 11 counts per million (18 items) and high frequency words were above 12 counts per million (18 items). For /g/ items 0-2 counts per million were low frequency items (22 items) and 3 and greater were high frequency words (17 items). The goal of having an even split between high and low groups created a different distinguishing frequency between high and low frequency groups in the /d/ and /g/ items. This follows based on the fact that the /g/ items are generally less common and lower frequency. Therefore the split for the /g/ items would necessarily be lower than the split for the /d/ items. One subject was excluded because data was not available for all cells. As a post-hoc test the results are considered tentative possibly suggesting further investigation.

A three factor ANOVA was performed with Target Reduction (reduced and unreduced), Phoneme (/d/ and /g/) and Frequency (high and low) as factors. In the by-subjects analysis all factors were within-subjects factors and in the by-items analysis Phoneme and Frequency were between-items factors while Target Reduction was a within-items factor. The results are shown in Figure 2.15 below.

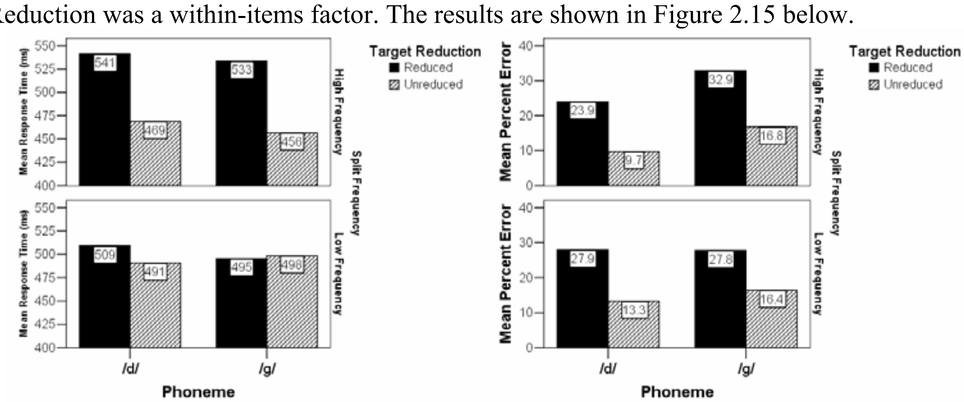


Figure 2.15 Mean response time and percent error for auditory lexical decision for reduced versus unreduced /d/ and /g/ word–medial phonemes split by frequency.

The statistical analysis of the response time data showed a significant interaction of Target Reduction with Frequency (*F1*(1,53)=9.967, p<0.005; *F2*(1,71)=11.920, p<0.005) and also a main effect of Target Reduction (*F1*(1,53)=36.931, p<0.001; *F2*(1,71)=20.359, p<0.001). All other interactions and main effects were not significant. The analysis of the Error Rate data showed only a significant main effect of Target Reduction (*F1*(1,53)=96.672, p<0.001; *F2*(1,71)=58.113, p<0.001) all other interactions and main effects were not significant both by-items and by-subjects.

The significant interaction in the response time data is the relevant comparison. Visually the result of this interaction is that the effect of Target Reduction is very strong in the high frequency and disappears or is very weak for the low frequency items. As a post-hoc test the statistical power is not very high, however this result indicates that it is possible that frequency is affecting the way in which listeners process reduced and unreduced targets. In order to further confirm this result additional experimentation is necessary which uses a new set of target words which are designed to specifically test the interaction of frequency and reduction.

Some of the reduced stimuli could be interpreted as one of two lexical items, thus increasing lexical competition for an item and potentially affecting the result. For example, the reduced form of *regal* might be interpreted as *regal* [rigl□], but if the flap reduction is very extreme the listener might perceive it as *real* [ril□]. A post–hoc analysis excluded these potentially ambiguous items (11 flap items and 12 word–medial /g/ items) to verify that the results are not caused by these items. Figures 2.16 and 2.17 show response latencies and percent error with the potentially homophonous items excluded.

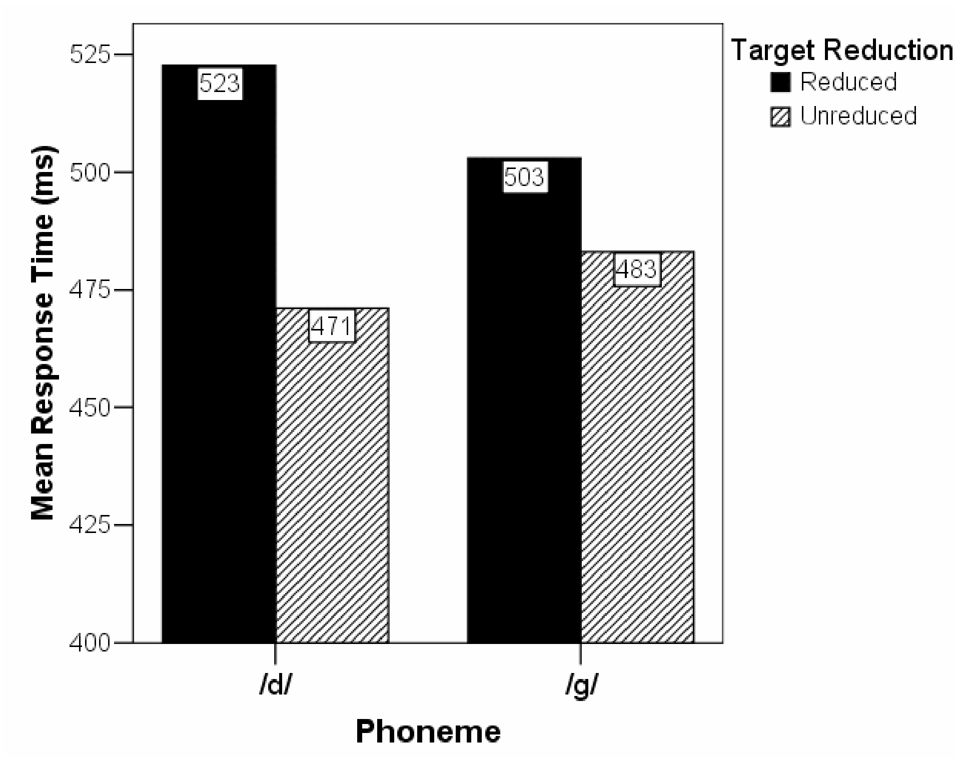


Figure 2.16 Mean response latency for auditory lexical decision for unambiguous reduced versus unreduced /d/ and /g/ word–medial phonemes.

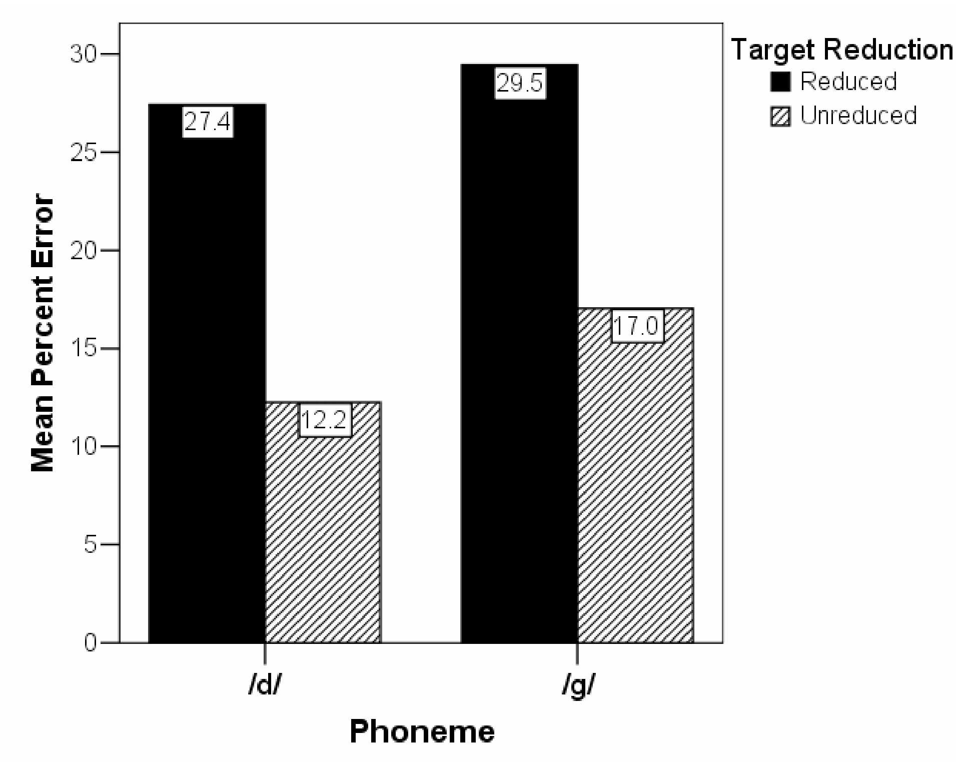


Figure 2.17 Mean percent error for auditory lexical decision for unambiguous reduced versus unreduced /d/ and /g/ word–medial phonemes.

ANOVAs performed on both the response latency and percent error rate data show the same results as reported above where ambiguous items were included. Since this experiment was not designed to test ambiguity, there are too few items (insufficient power) to do a full statistical analysis. It would thus seem that the ambiguous items do not have any substantial effect on the results reported above.

2.1.6 Item duration results

A possible interpretation of the results in Experiment 1 is that the item durations for the reduced and unreduced items are different and are the cause of the effect of Target Reduction. If this exploration holds it would mean that a strong correlation would exist between word duration and any other effect that might occur as a result of the experiments described above. If there is such a strong correlation with duration it could be difficult to determine if any observed effect is due to flap reduction or general duration of the individual words in Experiment 1.

Figure 2.18 below shows a summary of the duration data.

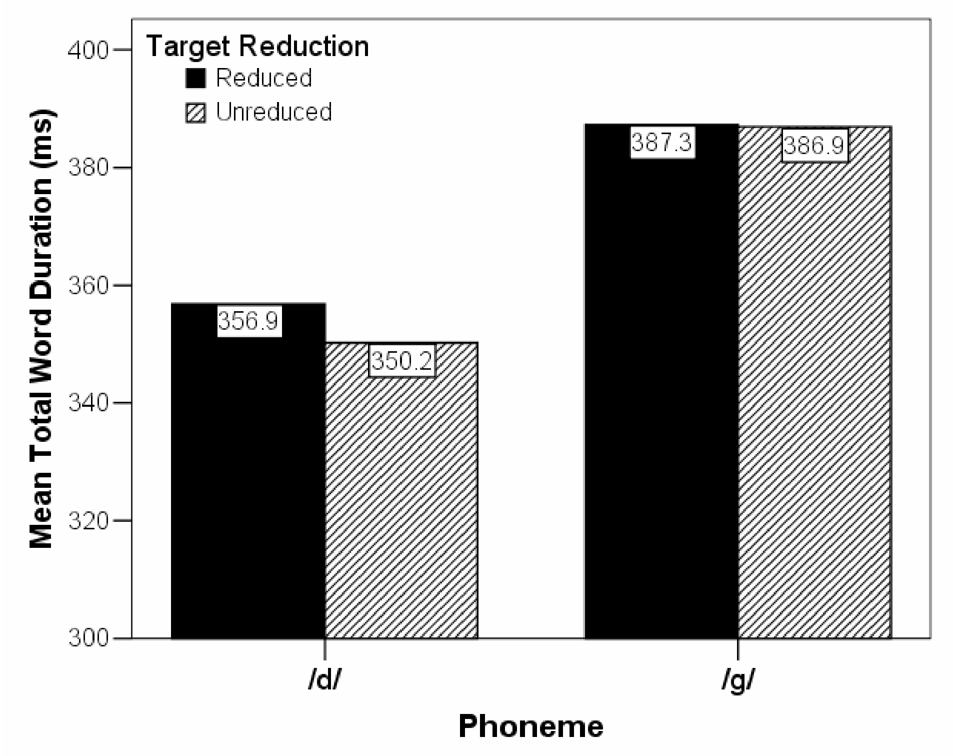


Figure 2.18 Mean duration of target stimuli from auditory lexical decision items for reduced versus unreduced /d/ and /g/ word–medial phonemes.

An ANOVA with items as a repeated measure was performed on the total duration of target words. Factors were Target Reduction (reduced vs. unreduced) and Phoneme (/d/ vs. /g/). The analysis shows that there is very little difference between reduced and unreduced stimuli for either flap or /g/ items (main effect of Target reduction:

$F(1,78)<1$). The analysis showed that /g/ items were significantly longer than the flap items (main effect of Phoneme: $F(1,78)=9.128$, $p<0.005$).

The effect of item duration is not surprising. Generally speaking, word–medial /d/ words are of higher frequency than word–medial /g/ words. Frequency counts of the target items show that the mean frequency for all /d/ items was 53.225 per million, in contrast to a mean frequency for all /g/ items of 6.55 per million (frequency counts are based on the Brown Corpus (Kučera and Francis, 1967)). Jurafsky et al. (2000) have shown in a corpus study of production that item frequency highly correlates with item durations (for opposing results on homophones cf. Whalen, 1991a; 1991b). Thus, higher frequency items are likely to be shorter than lower frequency items. Therefore, the fact that /d/ items are generally more frequent and shorter than the /g/ items is an expected result.

In summary, listeners respond to reduced words in isolation more slowly than to the unreduced pairs and there was no difference found between listener response to /d/ and /g/ phoneme items. Furthermore, the effect of reduction on processing was not influenced by the potentially homophonous items. The lack of an effect of reduction for stimulus duration indicates that duration cannot account for the effect. The result that reduced words inhibit processing of words is therefore reliable.

A second experiment was designed to further test the effect of reduction on speech processing. In Experiment 2, the cross–modal identity priming task (Norris, Cutler, McQueen and Butterfield, 2006) is used to investigate further the effect of reduction on lexical access.

2.2 Experiment 2: Cross–modal identity priming

A cross–modal identity priming task (Swinney, Onifer, Prather and Hirshkowitz, 1979; Marslen–Wilson, Nix and Gaskell, 1995; Norris, Cutler, McQueen and Butterfield, 2006) uses both auditory and visual information to investigate lexical access. In this task, listeners are presented with an auditory prime, and following the prime they receive a visual probe. In the priming instances, listeners hear and see identical stimuli, therefore, a listener would hear *puddle* and then see *puddle* spelled out on a computer screen. Listeners then perform a lexical decision on the visual stimulus. This task has shown significant priming effects when compared to a non–prime control condition (e.g., listener hears *puddle* and sees *bracket*).

This task has been used to show priming effects of words in both isolation and context. Experimenters change the phonetic or phonological identity of the prime in a controlled fashion to test how the change affects priming. For example, effects of phonological boundaries (Spinelli, McQueen and Cutler, 2003), segmentation of speech (McQueen and Cho, 2003), and detailed phonetic variability (Alphen and McQueen, 2006) have been shown using cross–modal identity priming.

Spinelli, McQueen and Cutler (2003) tested phonological boundaries by investigating items from French occurring in a liaison environment. They used, for example, ambiguous phrases due to liaison like d*ernier oignon* (last onion) and *dernier rognon* (last kidney) (Spinelli, McQueen and Cutler, 2003). These phrases contain the same sequence of phonemes because in the “last onion” case the /r/ becomes the onset of the following word. Listeners heard one of these productions and were presented a visual lexical decision with the words *oignon* or *rognon*. Visual targets were presented halfway through the last word of the auditory prime. The results showed facilitation for the matched cases (e.g. speaker intended *oignon* matched with visual *oignon*), and weaker facilitation for the mismatch cases (e.g. speaker intended *oignon* matched with visual *rognon*).

Alphen and McQueen (2006) tested the effect of detailed phonetic variability in spoken word recognition using voice onset time of plosives in Dutch. Auditory primes were created that varied in the amount of prevoicing they contained (12, 6 and 0 periods). The auditory primes were paired with their identical visual target for the lexical decision. Visual targets were presented 200ms following the onset of the prime, so that the visual stimulus appeared before the offset of the auditory stimulus. They showed that for the 12 and 6 period auditory primes patterned together while the 0 period primes had a different pattern. Alphen and McQueen (2006) show that phonetic detail (the presence of prevoicing) facilitates lexical access.

Previous research using the cross-modal identity priming task have sequenced the visual onset in varying places in the experimental sequence. Spinelli, McQueen and Cutler (2003) and Alphen and McQueen (2006) timed the visual target so that it occurred before the offset of the auditory sequence. van der Lugt (1999) found that presenting the visual target at the onset of the prime may be to early and that presenting it at the offset of the prime may be to late. However, other work (Bölte & Coenen, 2002; Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1996) has timed the occurrence of the visual target to the offset of the auditory prime and shown significant priming effects.

Experiment 2 tests the effect of reduction on listeners' recognition of words using the cross–modal identity priming task (Norris, Cutler, McQueen and Butterfield, 2006). The outcomes for this experiment are the same as those predicted in Experiment 1; however they are reformulated here to fit the cross–modal identity priming task. (1) Due to the decreased clarity of acoustic information, reduced primes might facilitate recognition of visual probes less effectively than unreduced stimuli, hence reduced primes will show a smaller priming effect relative to controls. (2) Reduced/spontaneous speech targets are more frequently heard by listeners, so priming might be greater for reduced items than for unreduced items. Based on the findings in Experiment 1, it is expected that priming in reduced items will be smaller than in unreduced items relative to controls.

2.2.1 Participants

Thirty–nine listeners from the University of Arizona community participated in the experiment. The majority of listeners were recruited from the linguistics undergraduate pool, students mainly in Introduction to Linguistics courses. Listeners' ages ranged from 18-60 years. All listeners were native speakers of American English either studying a foreign language (generally Spanish or French) or late bilinguals with no known hearing problems. Listeners received either extra credit or were paid for their participation. Listeners in Experiment 2 and had not participated in Experiment 1.

2.2.2 Materials

Target items consisted of 30 word–medial /d/ (flap items) and 30 word–medial /g/ items, and were a subset of items from Experiment 1. Target items were produced in both reduced and unreduced forms. Sixty additional real word items were recorded as auditory controls for the visual probes. Each real word auditory control item was selected by matching its frequency (Kučera and Francis, 1967) to the frequency of one of the target items. For example, the item *puddle*, which has a frequency of 1 per million, was matched with *bracket*, which also has a frequency of 1 per million. In addition matching frequency all auditory control items were bisyllabic (CVCV, CVCVC, VCV, VCVC). The auditory control items then served as the auditory stimulus when the item occurred as a control.

One–hundred–and–sixty filler and 9 practice items were constructed in addition to the targets. Of the 160 fillers, 20 contained real–word visual stimuli matched with different real–word auditory stimuli. Real–word items were also bisyllabic words, which did not contain a word–medial /d/ or /g/. The remaining 140 fillers contained phonotactically possible non–word visual stimuli matched with real–word auditory fillers. The visual non–word for 40 of these fillers contained partial phonological overlap with the auditory filler (e.g. auditory *level* with visual non–word *mavel*). These items, items with phonological overlap, were created so that they would be phonologically similar to real words preventing response strategies based on detecting overlap. Table 2.2 gives examples of auditory primes with their corresponding visual probes. A complete list of auditory primes with visual probes appears in Appendix C.

Table 2.2 Example items from Experiment 2 illustrating both auditory prime and visual probes

	Flap /d/	/g/	Non–word Filler		Phonological Overlap Filler
Auditory Stimulus for Prime or Overlap Condition	puddle (ɾ) puddle (ɾ̞)	beggar (g) beggar (g̞)		level	
Auditory Stimulus for Control or Non-Overlap Conditions	bracket	torture	offer		dozen
Visual Probe	puddle	beggar	prunshin	mavel	orphan

2.2.3 Procedures

Target and filler items were recorded in a sound–attenuated booth over several sessions by the same trained female phonetician, who is a native speaker of American English. Target items were taken directly from Experiment 1. Additional fillers were recorded in later sessions using the same speaker and recording set up. Filler items were produced in reduced and unreduced fashions where reduction of a word–medial stop was appropriate. Items were excised from the original recordings for stimulus presentation.

Each listener saw 30 /d/ target visual items and 30 /g/ target visual items. For each listener, 10 /d/ and 10 /g/ items were paired with an auditory prime consisting of the same word in reduced pronunciation, 10 /d/ and 10 /g/ items were paired with an auditory prime consisting of the same word in unreduced pronunciation, and 10 /d/ and 10 /g/ items were paired with an unrelated word as the auditory control stimulus. The two factors, visual item (/d/, /g/) and target reduction (reduced. unreduced, auditory control) were counterbalanced, creating three lists. For example, the visual stimulus *puddle* appeared in the first list with the reduced auditory stimulus, in the second list with the unreduced auditory stimulus, and in the third list with the auditory control. All fillers (160) appeared in each list. Lists were pseudo–randomized for presentation and were presented in one of two opposite orders, creating six total lists. Thus, any one listener heard 40 target items and 20 auditory control items (25% of the total number of items presented).

Listeners were tested individually in sound–attenuated booths. The auditory primes were presented over headphones and the visual probe appeared on a monitor outside of the booth in lower–case, 36–point Arial font. The visual probe appeared at the offset of the auditory stimulus (0ms interval). Visual probes remained on the monitor for 2 seconds, and if the listener did not respond in the time allotted, the experiment proceeded to the next stimulus. Listeners were instructed to respond as quickly and as accurately as possible to the visual probe using a button box with buttons labeled “YES” and “NO”. A “YES” response indicated that the listener decided the visual probe was a real word of English. There is a danger that listeners will stop paying attention to the auditory probe because the task is purely a visual lexical decision. If this were to occur then the effect of priming would diminish. A comprehension test was designed for the end of the experiment to verify that subjects were listening to the auditory stimuli.

The experimental procedure for each listener was: instruction, then practice session, then one of the six randomly assigned stimulus lists, and finally a comprehension test. Listeners were informed at the beginning of the procedure in the instructions that there would be a comprehension test at the end of the experiment. The test contained 10 visually presented items, half of which were presented auditorily in the main experiment and the other half of which were new. Listeners were asked if they had heard these visually presented words in the main part of the experiment. Listeners scored an average of 78% correct on the comprehension test. A single factor ANOVA of the number of 'yes' responses for new and presented targets showed that subjects responded 'yes' significantly more to the already presented stimuli ($F(1,308)=165.402$, $p<0.001$). This verifies that listeners did attend to the auditory stimulus despite the danger of ignoring the auditory stimulus and responding to the visual probe.[4]

2.2.4 Results

Response latencies from the lexical decision task were measured from the onset of the visual probe. Responses slower than 1100ms or faster than 350ms (less than 5% of total responses) were treated as errors and removed from the analysis. Five listeners scored below 85% correct and were also excluded. One was excluded after a self–report of not paying attention to the auditory stimulus. Two were excluded after a self–report, that they repeated the visual probes aloud. One subject was excluded due to a technical glitch in the software. Mean response times are shown in Figure 2.19.

---

[4] The same items were always used in the comprehension test, regardless of the counter-balanced version of the experiment presented. As a result it is possible that a listener would be presented with a target both auditorily and visually, which would facilitate responses in the comprehension test. Only two of the five "yes" responses could potentially have been presented both auditorily and visually depending on the experiment list. This confirms that listeners attended to the auditory stimulus and not just the visual stimulus.

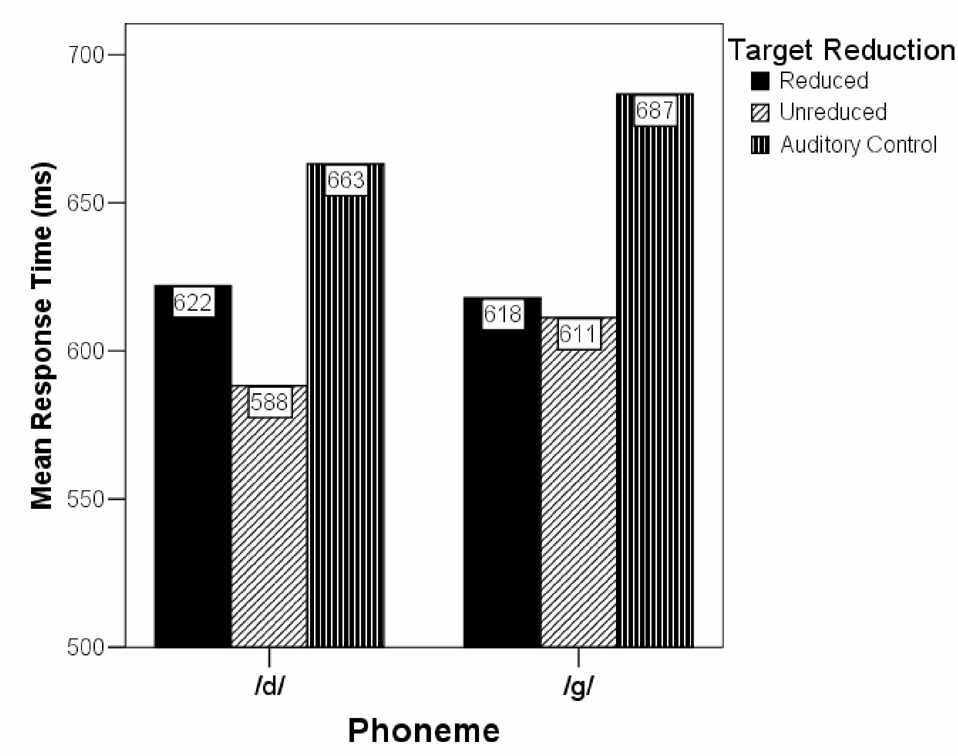


Figure 2.19 Mean response time for cross–modal identity priming of reduced versus unreduced /d/ and /g/ word–medial phonemes and control items.

An ANOVA of the response time data was performed by–subjects (*F1*) and by–items (*F2*). The interaction of Phoneme and Target Reduction was not significant (*F1*(2,48)=1.235, p>0.05; *F2*(2,116)=1.000, p>0.05). A significant main effect of Phoneme was found in the by–subjects analysis (*F1*(1,24)=11.397, p<0.005); however, the by–items analysis was not significant (*F2*(1,58)=2.153, p>0.05).

The analysis revealed an identity priming effect for Target Reduction: listeners responded to the identity primes more quickly than the non–primes (auditory controls) (*F1*(2,48)=38.292, p<0.005; *F2*(2,116)=33.402.753, p<0.001). Furthermore, planned pairwise comparisons of each priming condition compared to the control condition showed that listeners' responses for /d/ items were primed significantly for both reduced ($F1(1,24)=13.351$, $p<0.005$; $F2(1,29)=7.295$, $p<0.05$) and unreduced ($F1(1,24)=46.324$, $p<0.001$; $F2(1,29)=26.007$, $p<0.001$) targets. A pairwise comparison of unreduced to reduced /d/ items showed a significant difference ($F1(1,24)=13.843$, $p<0.005$; $F2(1,29)=6.073$, $p<0.05$) indicating that the degree of priming for both the reduced and unreduced stimuli is significantly different. The planned pairwise comparisons of the /g/ items were primed significantly for reduced ($F1(1,24)=33.123$, $p<0.001$; $F2(1,29)=21.049$, $p<0.001$) and unreduced ($F1(1,24)=27.072$, $p<0.001$; $F2(1,29)=44.022$, $p<0.001$) targets as compared to the auditory control condition. However, the comparison of reduced vs. unreduced did not show a significant effect ($F1<1$; $F2<1$) but it did show a non–significant trend in the same direction as the /d/ items. The results from the mean percent error data are shown in Figure 2.20, below.

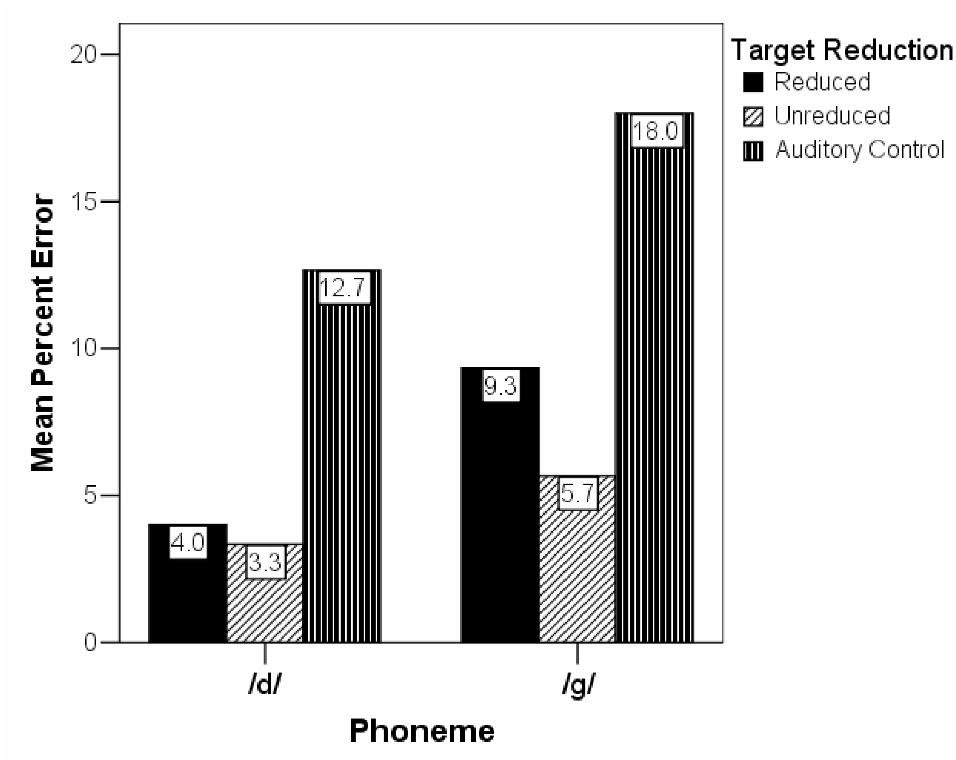


Figure 2.20 Mean percent error for cross–modal identity priming of reduced vs. unreduced /d/ and /g/ word–medial phonemes and control items.

An ANOVA of the error data was performed by–subjects (*F1*) and by–items (*F2*). The interaction of phoneme and target reduction was not significant (*F1*<1; *F2*<1). The main effect of phoneme was significant (*F1*(1,24)=14.139, p<0.005; *F2*(1,58)=4.378, p<0.05). Listeners made fewer mistakes in the identity primes than the non–primes in target reduction (*F1*(2,48)=19.137, p<0.001; *F2*(2,116)=18.558, p<0.001). Planned pairwise comparisons of each priming condition compared to the control condition showed that listeners' responses were significantly primed for reduced (*F1*(1,24)=13.331, p<0.005; *F2*(1,29)=11.25, p<0.005) and for unreduced targets (*F1*(1,24)=19.733, p<0.001; *F2*(1,29)=10.411, p<0.005). Targets showed a larger non–significant trend of priming for unreduced /d/ items (*F1*<1; *F2*<1) than for reduced /d/ items. The comparison of the /g/ items is similar to the /d/ items. A significant difference was found in priming for both the reduced (*F1*(1,24)=7.955, p<0.01; *F2*(1,29)=7.917, p<0.01) and unreduced (*F1*(1,24)=16.52, p<0.001; *F2*(1,29)=14.533, p<0.005) targets. The comparison of reduced vs. unreduced appeared to follow the same direction but was not significant (*F1*(1,24)=3.788, p>0.05; *F2*(1,29)=2.233, p>0.05). [5]

---

[5] To further investigate and confirm the results from the post hoc experiment in Experiment 1 on the effect of frequency on reduction (footnote 3, section 2.1.5), a similar post hoc analysis was performed on Experiment 2 (frequency counts are from the Brown Corpus (Kučera and Francis, 1967)). Items were split into high and low frequency groups following the same procedure as above. For the /d/ items low frequency words fell between 0 and 11 counts per million and high frequency words were 12 counts per million and above (15 items in low frequency and 15 items in high frequency). For the /g/ items 0-2 counts per million were low frequency items and 3 and greater were high frequency words (18 items in low frequency and 12 in high frequency). Again as a post hoc test the results are only considered tentative and may suggest further investigation.

A three factor ANOVA was performed with Target Reduction (reduced, unreduced, auditory control), Phoneme (/d/ and /g/) and Frequency (high and low) as factors. In the by-subjects analysis all factors were within-subjects factors and in the by-items analysis Phoneme and Frequency were between-items factors while Target Reduction was a within-items factor. The results are shown in Figure 2.21 below.

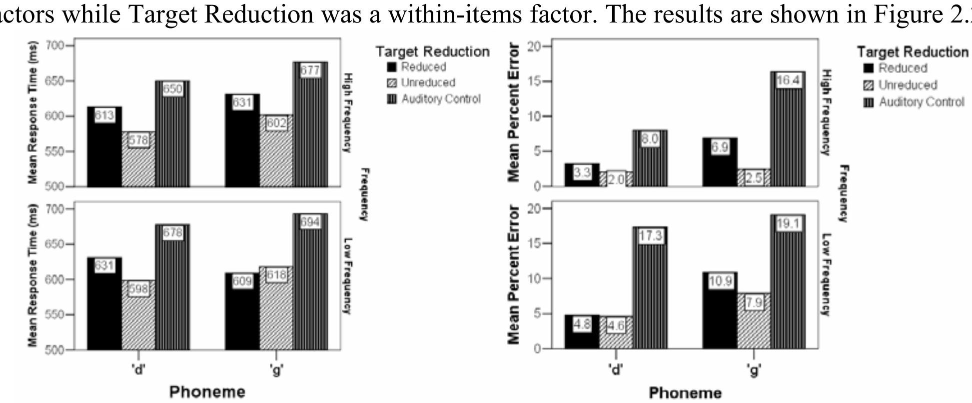


Figure 2.21 Mean response time and percent error for cross-modal identity priming for reduced versus unreduced /d/ and /g/ word–medial phonemes compared to auditory control stimuli split by frequency.

The statistical analysis of the response time data (with the auditory control level excluded) showed no significant interactions and only a main effect of Target Reduction (*F1*(1,28)=15.319, p<0.001; *F2*(1,56)=6.904, p<0.05). All other main effects were not significant. The analysis of the Error Rate data showed no significant interactions and significant results for all of the main effects in the by-subjects analysis but not in the by-items analysis (Target Reduction (*F1*(1,29)=12.336, p<0.001; *F2*(1,56)=2.27, p>0.05), Phoneme (*F1*(1,29)=6.021, p<0.05; *F2*(1,56)=3.261, p>0.05) and Frequency (*F1*(1,29)=9.922, p<0.005; *F2*(1,56)=2.939, p>0.05)).

While the interaction in the response time data was not significant in this comparison, visually, the pattern of results appears somewhat similar to the results found in Experiment 1. The results of Experiment 1 and the absence of a significant effect here warrant further investigation of the effect of frequency on reduction.

A post hoc analysis removing the ambiguous items, as in Experiment 1, was also performed. The results of removing the ambiguous items are shown in Figures 2.22 and 2.23, below.

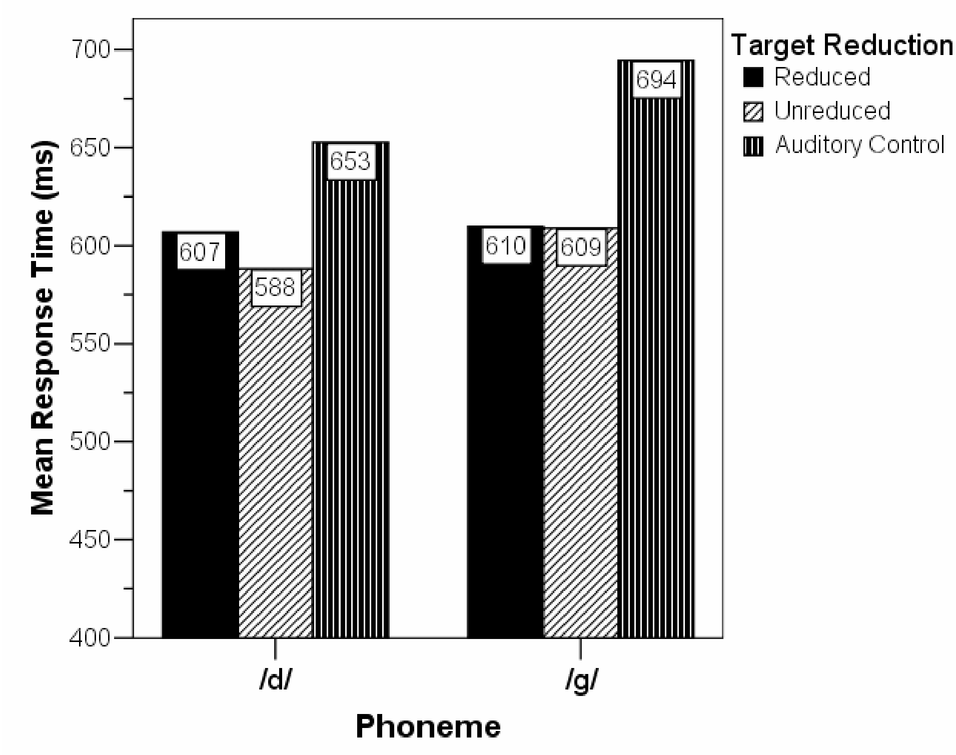


Figure 2.22 Mean response time for cross–modal identity priming task of words in isolation for /d/ and /g/ word–medial phonemes and control items of unambiguous reduced and unreduced targets.

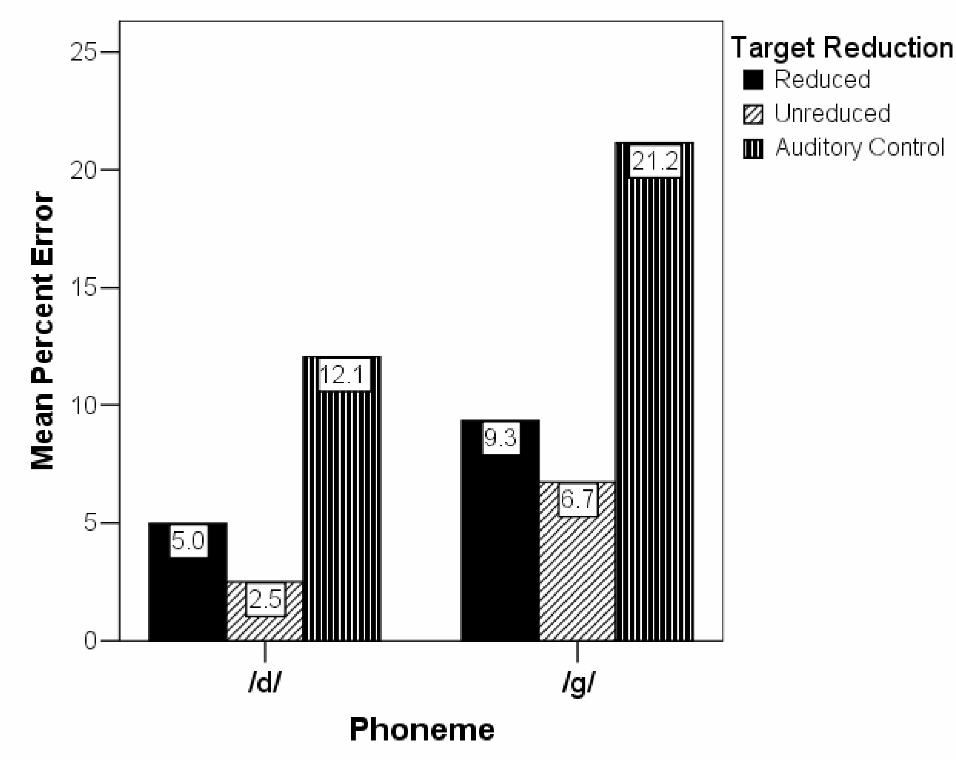


Figure 2.23 Mean percent error for cross–modal identity priming task of words in isolation for /d/ and /g/ word–medial phonemes and control items of unambiguous reduced and unreduced targets.

The result after removing the ambiguous items differed from the original result only in minor ways. The main difference is that the interaction of phoneme and target reduction for response latency is significant in the by–subjects analysis but not significant in the by–items analysis ($F1(2,48)=29.022$, $p>0.001$; $F2(2,70)=1.485$, $p>0.05$). The analysis of the percent error results also showed the same difference ($F1(2,48)=11.939$, $p>0.001$; $F2<1$).

In the cross–modal identity priming task, subjects showed a significant priming effect of reduced and unreduced targets for both word–medial stops. The pairwise comparisons showed that priming significantly occurred for both reduced and unreduced items. The pairwise comparisons also showed that the difference in priming between the reduced and unreduced primes was only significant for the /d/ items and only in the response time results, but overall, tended to show non-significant greater priming for unreduced targets as opposed to reduced targets. There was a main effect of phoneme in the percent error analysis and in the response latency results the by-subjects analysis was significant but the by-items analysis was not significant.

2.3 General discussion

This section returns to the main question of how reduction affects word recognition. The results from Experiments 1 and 2 show that reduction inhibits word recognition. Specifically, listeners from both experiments recognized unreduced words faster and more accurately than reduced words.

In the auditory lexical decision task, listeners' processing of reduced words was inhibited by the reduced items. Thus, their response times are slower and less accurate for the reduced items than for the unreduced items. In the cross–modal identity priming task, the response time and percent error results indicate that priming occurred in both the reduced and unreduced conditions. Listeners responded more quickly and accurately to the unreduced items than to the reduced items. For example, there was a greater difference in priming for unreduced *puddle* than for reduced *puddle*. The comparison, while only significant for response time data for the flapped items, also showed a non–significant trend in the same direction as the remaining data.

The reported difference between phonemes might be attributed to word frequency. Medial /d/ words are generally more frequent than medial /g/ words (mean frequency of

/d/ target items = 53.225 and /g/ = 6.55 (Kučera and Francis, 1967)). This could have led to faster responses to the more frequent /d/ items as opposed to the less frequent /g/ items. However, what is more important is the effect of reduction which is tested within–items and thus frequency does not play a role.

The results from Experiments 1 and 2 indicate that listeners rely heavily on acoustic information in processing reduced vs. unreduced speech. Furthermore, this indicates that even though reduced speech is a more common speech style, it does not facilitate recognition of words in isolation. The results suggest that speech reduction influences auditory word recognition.

This chapter has investigated the processing of reduced and unreduced words appearing in isolation. In spontaneous, conversational speech processing of words does not occur in isolation. Instead processing occurs with regard to contextual factors such as phonetic, phonological, speech style, syntax, and semantics. This raises the question: Do listeners adjust their expectations based on the degree of reduction in context?

CHAPTER 3

PROCESSING OF REDUCED FLAPS: THE EFFECT OF PRECEDING CONTEXT

3.0 Introduction

Reduced speech has thus far been investigated in words that occur in isolation. Cutler (1998) points out that one of the reasons for investigating speech recognition using laboratory speech is that it allows the researcher to carefully control contributing factors and separate possible effects. In Chapter 2, processing of reduced forms was investigated in isolation to isolate the effect of reduction. However, spontaneous connected speech generally occurs in large chunks of spoken phrases. Listeners' daily interaction with other listeners involves the processing of spontaneous, connected speech. In connected speech a listener rarely hears a word occurring in isolated unconnected speech as in laboratory speech. In this chapter, experiments that test the effect of reduction on the processing of connected speech are reported.

Past literature has shown that context affects speech recognition (Ladefoged and Broadbent, 1957; Miller and Volaitis, 1989; Miller and Liberman, 1979, Ernestus, Baayen, and Schreuder, 2002; Mitterer and Ernestus, 2006, Zwitserlood and Coenen, 2000; Coenen, Zwitserlood, and Bölte, 2001). The main goal of this dissertation as a whole is to investigate the effect of reduction on speech recognition. In Chapter 2, it was established that reduction inhibits processing of words in isolation. Specifically, words containing a reduced flap or word–medial /g/ are more difficult to process than their unreduced counterparts. This chapter builds on the results of Chapter 2, where, in a controlled way, the effect of reduced and unreduced context on speech recognition was investigated.

In section 3.1 some previous work on the effect of context on speech perception is introduced. Experiment 3, which is a lexical decision task using targets in a connected speech context is described and the results are reported in section 3.2. In section 3.3, Experiment 4 is described and the results of a cross-modal identity priming task using targets in a connected speech context are reported. Both experiments replicate Experiments 1 and 2, described in Chapter 2, while introducing context as a new factor. Both experiments investigate the role of context in the processing of reduced and unreduced speech. Finally, in section 3.4 all results are summarized and discussed.

3.1 Context effects

Context can be defined as surrounding information. This information can either precede or follow a particular target or both. Contextual information can provide phonetic, syntactic, semantic, and other types of information which can aid in the recognition of a particular target. Context has been shown in many instances to affect speech perception. Listeners have been shown to “normalize” their responses in phonetic categorization based on context. Examples of context effects in speech recognition are discussed in this section: vowel normalization (Ladefoged and Broadbent, 1957), syllable duration (Miller and Liberman, 1979), speaking rate (Summerfield, 1981; Miller and Volaitis, 1989) and some previously mentioned context effects (Section 1.3.2) are briefly summarized: syntax/semantics (Ernestus, Baayen, and Schreuder, 2002), phonological context

(Mitterer and Ernestus, 2006), and assimilation (Zwitserlood and Coenen, 2000; Coenen, Zwitserlood, and Bölte, 2001).

One of the first studies that showed an effect of context was a vowel normalization study by Ladefoged and Broadbent (1957). In this study they found that varying the mean frequency range of the preceding context affected the way listeners perceived a target word. They selected the preceding context "Please say what the word is _____" which was followed by one of four *bVt* target words. Listeners heard low and high ranges for F1 and F2 throughout the preceding context. Ladefoged and Broadbent created six versions of the frame sentence that manipulated the range of F1 and F2 frequencies presented in the context. The listeners first categorized the target stimuli in isolation and then categorized them in context. For example, if a listener heard the *bVt* (F1 = 375Hz, F2 =1700Hz) stimulus in isolation it was generally categorized as *bit* (containing the vowel /ɪ/). When presented with the same target following the neutral frame sentence (F1 range = 275–500Hz, F2 range = 600–2500Hz) 87% of listeners labeled it as *bit*. However, when the F1 range was lowered ((F1 range = 200–380Hz, F2 range = 600–2500Hz) 90% of listeners categorized the target as *bet* (vowel /ɛ/). By adjusting the frequency range of the formants in the preceding context, Ladefoged and Broadbent were able to influence the perception of the target vowel. This effect has since been shown to be highly robust (e.g. Holt, 2005; Mitterer, 2006; among others). Thus, listeners adjust their expectations for a vowel in a target word based on the mean formant frequencies of preceding context.

Syllable duration has been shown to play a role in the perception of [b] and [w] (Miller and Liberman, 1979). A major acoustic difference between [b] and [w] is the duration of the formant transitions: otherwise identical syllables with shorter formant transitions are perceived as [b]; with longer transitions they are perceived as [w] (Liberman, Delattre, Gerstman, and Cooper, 1956). Miller and Liberman (1979) found that by varying the duration of the following vowel, and thus the duration of the syllable, they could shift the perception of [b] versus [w] in a categorical perception experiment. For example, if the duration of the vowel was long then the perception of a [ba] – [wa] continuum would shift more toward [wa]. If the duration of the following vowel was short then the perception would shift more toward [ba]. The relative duration of the following vowel and the duration of the entire syllable affected the perception of [b] versus [w]. Figure 3.1 uses imaginary results to create a hypothetical graph which illustrates the results of this type of experiment.

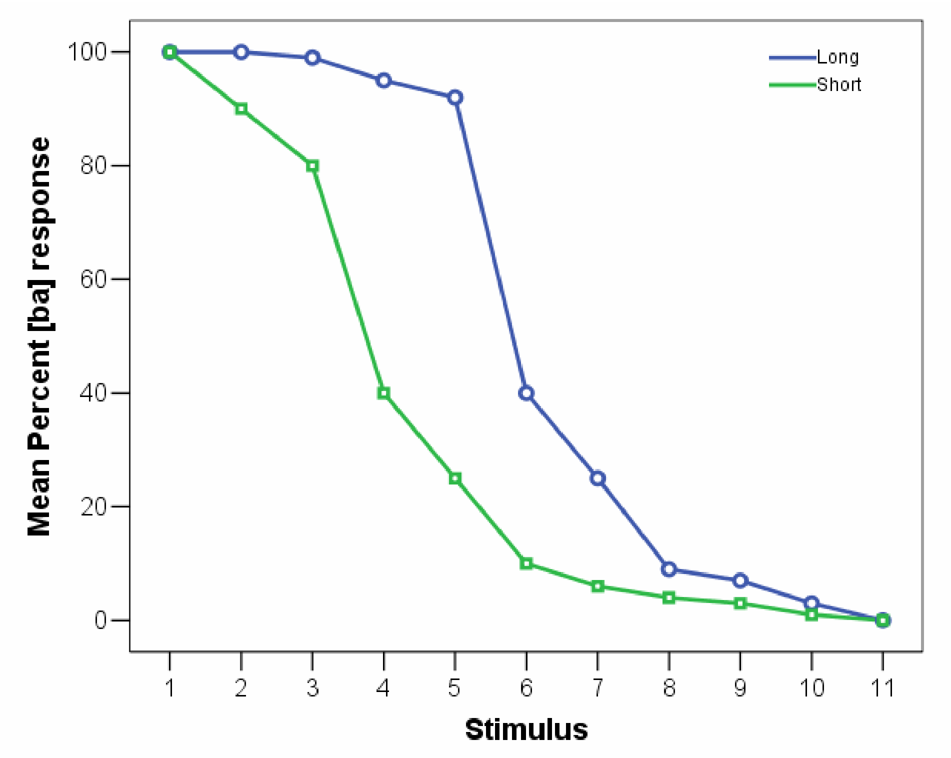


Figure 3.1 Sample identification function of a hypothetical [ba]–[wa] continuum from Miller and Liberman (1979). Circles correspond to long syllables and squares correspond to short syllables.

Miller and Liberman explained this effect as a type of normalization for speaking rate, the listener normalizes for fast speech and perceives targets relative to the context (See Diehl and Walsh (1989) for an alternative explanation). The effect of syllable duration is also a very robust effect (Diehl and Walsh, 1989; Pisoni, Carrell, and Gans, 1983; For infants Eimas and Miller, 1980; Miller and Eimas, 1983).

Speaking rate (Summerfield, 1981) has also been shown to affect perception. Summerfield (1981) showed that the preceding speaking rate influenced the perception of English voiced and voiceless obstruents (/p, t, k/ and /b, d, g/). The relevant acoustic factor in this study was voice onset time (VOT). For example, the obstruent /d/ has a very small VOT, while the obstruent /t/ has a much longer VOT. Summerfield (1981) showed that if the rate of the preceding speech was fast then listeners were more likely to perceive /t/ from a continuum of [t]-[d] stimuli. However, when the rate of the preceding speech was slow, listeners were more likely to perceive /d/ on the same continuum. In fast speech the relative syllable duration in the preceding speech is short, thus the VOT of the [t]-[d] stimuli is treated relative to the mean syllable duration of the preceding context and /t/ is perceived more often. The opposite is true for slow speech. In slow speech the mean duration of syllables is longer, thus the listener normalizes and expects longer VOTs for [t]; as a result, listeners select [d] more often on the same continuum.

In Section 1.3.2 several types of context effects were described in detail. The results of those studies and their relevance to context effects are briefly summarized here. Assimilation of phonemes across word boundaries was shown only to inhibit processing of assimilated targets in context and not when appearing in isolation (Zwitserlood and Coenen, 2000; Coenen, Zwitserlood, and Bölte, 2001). Contextual semantics and syntax have been shown to facilitate processing of reduced forms in Dutch (Ernestus, Baayen, and Schreuder, 2002) and the phonological context has also been shown to influence the perception of the presence or absence of a segment (Mitterer and Ernestus, 2006).

In the experiments described in this chapter, the effect of contextual speech style (reduced versus unreduced) is investigated. Two main questions are investigated: (1) Do differences exist in the way words containing reduced flaps and word-medial /g/s are processed in isolation as opposed to in context? (2) Does speech style (reduced versus unreduced) of the preceding context affect processing of reduced versus unreduced targets?

3.2 Experiment 3: Lexical decision in context

This experiment uses the auditory lexical decision task (Goldinger, 1996a) to explore the effects of reduced and unreduced context on processing. This experiment replicates Experiment 1 but adds a preceding phrase that provides context for the target stimulus (reduced versus unreduced speech style).

The possible expected outcomes in this experiment are similar to those in the experiments from Chapter 2. First, processing may be inhibited by the acoustically degraded (reduced) signal, thus response times will be slower for reduced items. Second, the reduced items may be processed faster because they have a higher frequency of occurrence in conversational speech than the unreduced items. Experiments 1 and 2 indicate that the predictions in the first outcome are maintained. However, it is possible that by introducing context, listeners will respond more quickly to the reduced target items as opposed to the unreduced items because they are occurring in an environment where reduction is expected. In addition to the outcomes discussed thus far, it is possible that the introduction of reduced and unreduced context might facilitate processing of targets matched for style. In other words the preference for unreduced might be smaller after hearing a reduced speech context. For example, a reduced context occurring with a reduced target may have less processing cost than a reduced context mismatched with an unreduced target.

3.2.1 Experiment 3a

3.2.1.1 Participants

Forty–two listeners from the University of Arizona community participated in the experiment. Listeners were recruited from the linguistics undergraduate pool, students mainly in Introduction to Linguistics courses, or they were recruited from the general University community. Listeners' ages ranged from 18-60 years. All listeners were native speakers of American English either studying a foreign language (generally Spanish or French) or late bilinguals with no known hearing problems. Listeners received either extra credit or were paid for their participation. Listeners in this experiment had not participated in Experiments 1 or 2.

3.2.1.2 Materials

In this experiment the target items are identical to the items from Experiment 1. Table 3.1 below shows the distribution of items.

Table 3.1 Example lexical decision items used in Experiments 1 and 3with the total number of items occurring.

	Flap /d/	**/g/**	**Real–word Filler**	**Pseudo–word Filler**
Example Item	puddle	baggy	copy	kooga /kuga/
Number	40	40	20	200

In order to introduce context, a phrase was created that preceded all items, both targets and fillers. The preceding context used was a quotative phrase, so that it did not contribute any semantic or phonetic context to the target stimulus. The frame sentence (context) was: “A lot of the time he says _____”. The frame sentence was produced in a reduced and unreduced form, to test the effect of reduction of context on processing of the target stimulus. The same trained phonetician that produced the stimuli in Experiments 1 and 2 created the frame sentence. Waveforms and spectrograms of the frame sentence are provided in Figures 3.2 and 3.3, below.

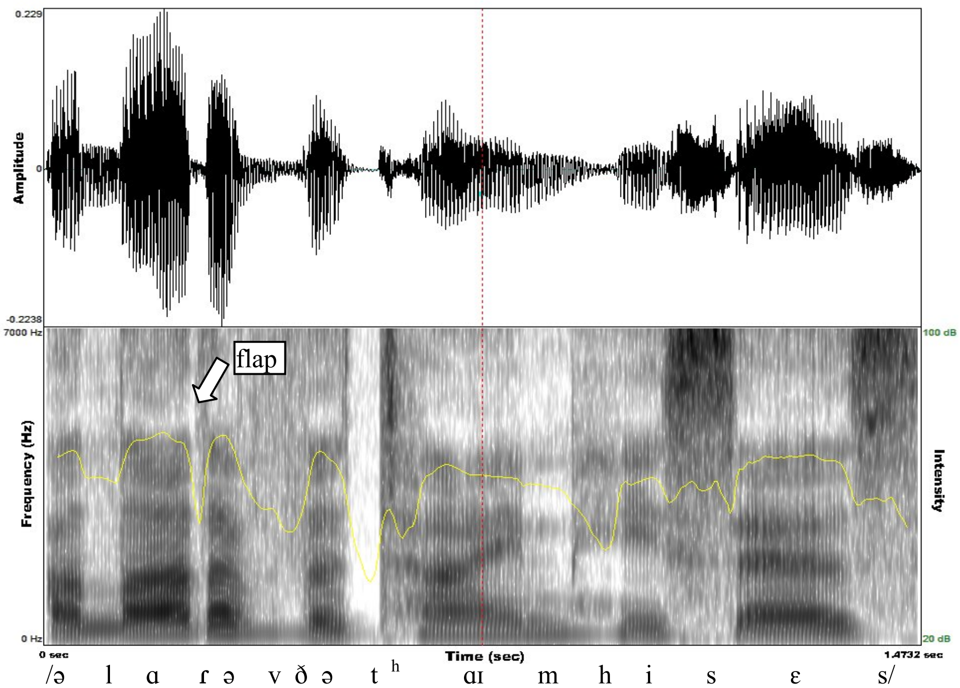


Figure 3.2 Waveform and spectrogram with superimposed intensity of unreduced frame sentence, “A lot of the time he says _____”, from Experiment 3. Total duration of frame is 1.47sec.

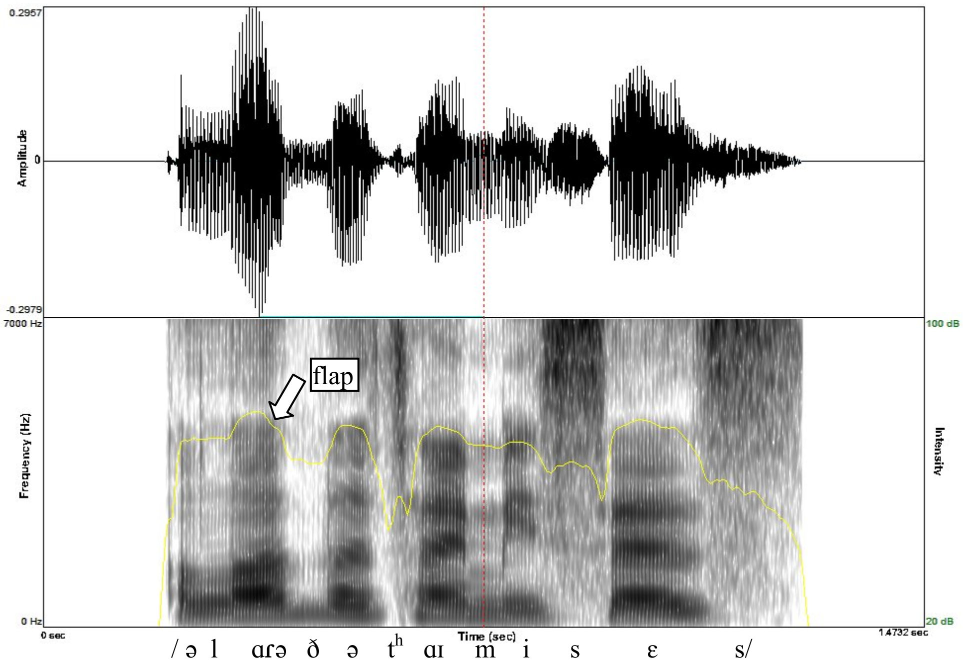


Figure 3.3 Waveform and spectrogram with superimposed intensity of reduced frame sentence, “A lot of the time he says _____”, from Experiment 3 with silence added. Total duration is matched to the duration of Figure 3.2, frame duration is 1.06sec.

The reduced frame is approximately 400ms shorter in duration than the unreduced frame. [6] Both frames also contained a flap, which occurred across a word boundary in “lot of.” In the unreduced frame a clear flap occurs, while in the reduced frame the flap is so reduced that it is nearly deleted. The only visual indicator for the flap is the small dip in intensity during the vowel (indicated in Figure 3.3). In the reduced frame the first vowel of the frame is essentially deleted, as well as the vowel in ‘of’. The frication for the /h/ in ‘he’ is also deleted in the reduced frame.

[6] In Figure 3.3 silence was added so that it appeared on the same time scale as Figure 3.2, making a visual comparison of the duration difference possible. The silence was not part of the original context so that the offset to the speech was the offset of the context.

3.2.1.3 Procedure

The procedure was identical to the procedure in Experiment 1 except that listeners heard the target stimulus immediately after the phrase sentence. For example, for each stimulus a subject would hear the frame sentence “A lot of the time he says ______” followed by the target stimulus, with an interval of 0ms between the end of the frame and the onset of the target. A 0ms interval was selected as sounding natural with the combination of frame and target. The frame sentence varied in whether it was the reduced or unreduced form of the sentence. Thus a listener might hear “A lot of the time he says puddle,” with the frame sentence reduced and the target reduced (matched reduced case), or the opposite may occur and the listener would hear the unreduced version of the frame sentence and the unreduced target (matched unreduced case). A listener might also hear mismatch cases where an unreduced frame would occur with a reduced target (mismatch unreduced–reduced case) or a reduced frame would occur with an unreduced target (mismatch reduced–unreduced case). The listeners’ task was to perform the lexical decision on the auditory probe *puddle*. The two factors Frame Sentence Reduction (reduced versus unreduced) and Target Reduction (reduced versus unreduced) were counterbalanced into four lists. Lists were pseudo-randomized with the 220 filler items and each list was presented in one of two opposite orders, creating a total of eight lists.

Listeners were tested individually in sound-attenuated booths. Listeners were instructed to respond as quickly and as accurately as possible to the visual probe using a button box with buttons labeled “YES” and “NO”. A “YES” response indicated that the listener decided the visual probe was a real word of English. Auditory probes were presented over headphones. The experimental procedure for each listener was: instruction, practice trial, and then one of the eight stimulus lists. Response latency and percent error were measured for each listener. Response latency was measured from the offset of the target stimulus.

3.2.1.4 Results

Items with response latencies less than 185ms or greater then 1500ms were considered errors and excluded (less than 5% of the total responses). Cutoffs for response latencies were chosen based on a histogram of the distribution of all response latencies. The cutoffs were selected at points where the distribution leveled out on either end. Additionally, responses to six of the items were excluded due to high error rates (below 30% correct) [7]. Accuracy scores for subjects below 70% were excluded from the analysis, which excluded a total of nine subjects. The results from the response latencies are presented first, followed by the percent error results. Figure 3.4, below, summarizes the response time data.

[7] As for Experiment 1 the items exclusion cutoff for error rate was set low as a result of the difficulty of the auditory lexical decision experiment. Many items would have been excluded, decreasing the statistical power and ability to generalize the results.

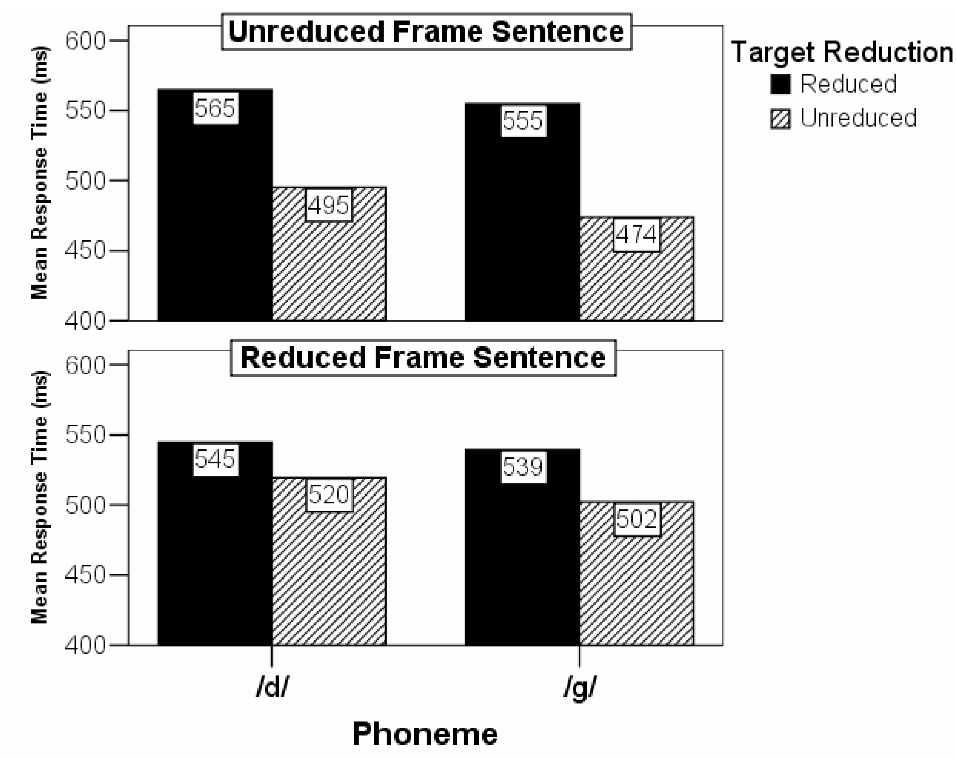


Figure 3.4 Mean response time for auditory lexical decision of words in reduced and unreduced contexts for /d/ and /g/ word–medial phonemes of reduced and unreduced targets.

A three-factor ANOVA was performed on the response latencies for listeners (*F1*) or items (*F2*) as repeated measures. There are three independent variables for this analysis: Frame Sentence Reduction (reduced versus unreduced), Target Reduction (reduced versus unreduced), and Phoneme (/d/ versus /g/). The three-way interaction of Frame Sentence Reduction, Target Reduction, and Phoneme was not significant ($F1<1$; $F2<1$). The two-way interaction of Target Reduction with Phoneme ($F1(1,25)=1.273$, $p>0.05$; $F2<1$) and Phoneme with Frame Sentence Reduction ($F1<1$; $F2<1$) was also not significant. The two-way interaction of Target Reduction and Frame Sentence Reduction, however, was significant in the by-items analysis and showed a trend in the by-subjects analysis: $F1(1,25)=3.383$, $p=0.078$; $F2(1,72)=5.951$, $p<0.05$. The interaction of Target Reduction and Frame Sentence Reduction is an important one: this interaction indicates that listeners react differently to the targets dependent on the frame sentence. The effect for unreduced targets is smaller after the reduced frame and the effect for reduced targets is smaller after the unreduced frame. When the type of reduction is inconsistent the effect size is smaller than when the reduction is consistent. The main effect of Target Reduction showed a significant effect: $F1(1,25)=29.659$, $p<0.001$; $F2(1,72)=14.49$, $p<0.001$. The main effects of Phoneme and Frame Sentence Reduction were not significant: $F1(1,25)=2.246$, $p>0.05$; $F2<1$ (Phoneme) and $F1<1$; $F2<1$ (Frame Sentence Reduction).

Due to the partial significance of the interaction, the factor Phoneme was collapsed and a two-factor ANOVA was performed with Target Reduction and Frame Sentence Reduction as factors. As with the three-factor analysis, the main effect of Frame Sentence Reduction was not significant ($F1<1$; $F2<1$) while the main effect of Target Reduction was significant ($F1(1,25)=27.125$, $p<0.001$; $F2(1,73)=16.674$, $p<0.001$). The interaction of Target Reduction with Frame Sentence Reduction showed a strong trend toward significance in the by-subjects and the by-items analysis was significant ($F1(1,25)=4.192$, $p=0.051$; $F2(1,73)=6.008$, $p<0.05$). Figure 3.5 shows the data from Figure 3.4 collapsed by phoneme. This graph specifically illustrates the interaction of Target Reduction with Frame Sentence Reduction.

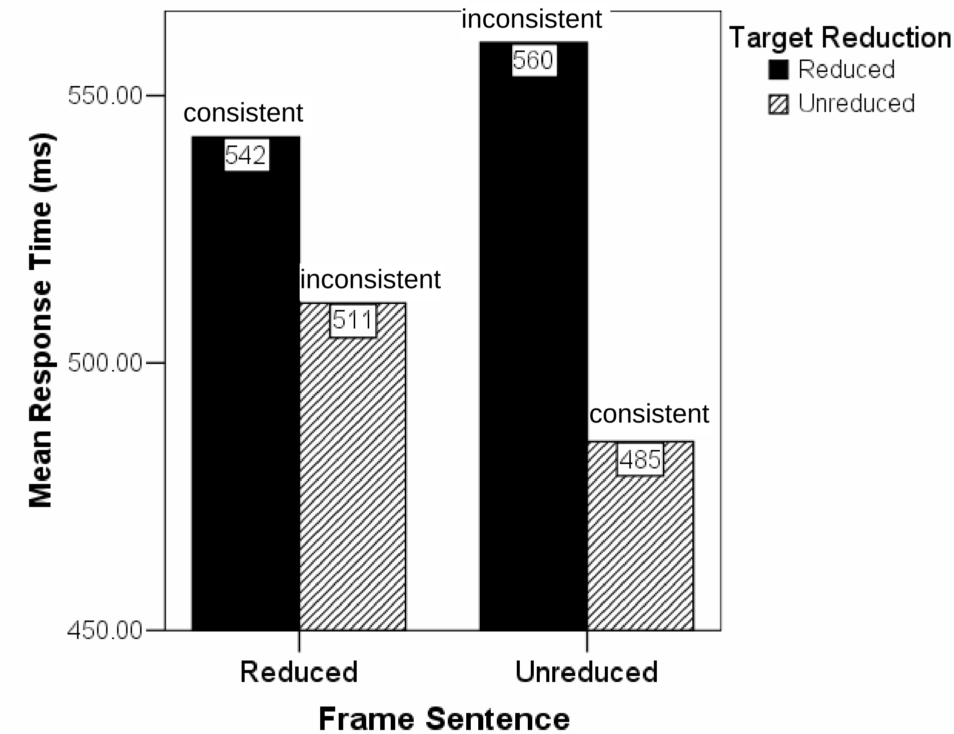


Figure 3.5 Mean response time for auditory lexical decision of words in reduced and unreduced contexts for reduced and unreduced targets to illustrate the interaction.

Figure 3.5 also illustrates the how the factor Frame Sentence Reduction affects response latencies for reduced and unreduced targets. When target and frame are consistent (where reduction is matched), response latencies are faster; when they are inconsistent (reduction is mismatched) response latencies are slower. Tests of simple effects showed that Target Reduction was significant for the unreduced Frame Sentence ($F1(1,25)=29.155$, $p<0.001$; $F2(1,73)=22.36$, $p<0.001$). The simple effect for the reduced Frame Sentence was significant for the by-subjects analysis and showed a non-significant trend in the by-items analysis ($F1(1,25)=8.475$, $p<0.01$; $F2(1,73)=3.313$, $p=0.073$). Effect size (by-subjects = $\eta_p^2 1$, by-items = $\eta_p^2 2$) of Target Reduction calculated for unreduced

Frames Sentences ($\eta_p^2 1$ = 0.538; $\eta_p^2 2$ = 0.234) showed a larger effect than for reduced Frame Sentences ($\eta_p^2 1$ = 0.253; $\eta_p^2 2$ = 0.043). Simple effects of Frame Sentence Reduction were not significant for reduced targets ($F1$<1; $F2(1,73)=1.203$, $p>0.05$) and simple effects for unreduced targets ($F1(1,25)=4.465$, $p<0.05$; $F2(1,73)=5.825$, $p<0.05$) were significant. The significant interaction and the difference in effect size demonstrate that Frame Sentence Reduction does influence the response times for targets.

Follow-up questions asked after listeners completed the experimental procedure noted that a large number of listeners found the frame sentence highly repetitive and some even reported that they stopped paying attention to the frame sentence and only listened for the target word. It is possible that the small effect in the tests of simple effects of Frame Sentence Reduction and Target Reduction is due to this effect reported by listeners. Percent error data were also analyzed; Figure 3.6 summarizes the percent error data for Experiment 3.

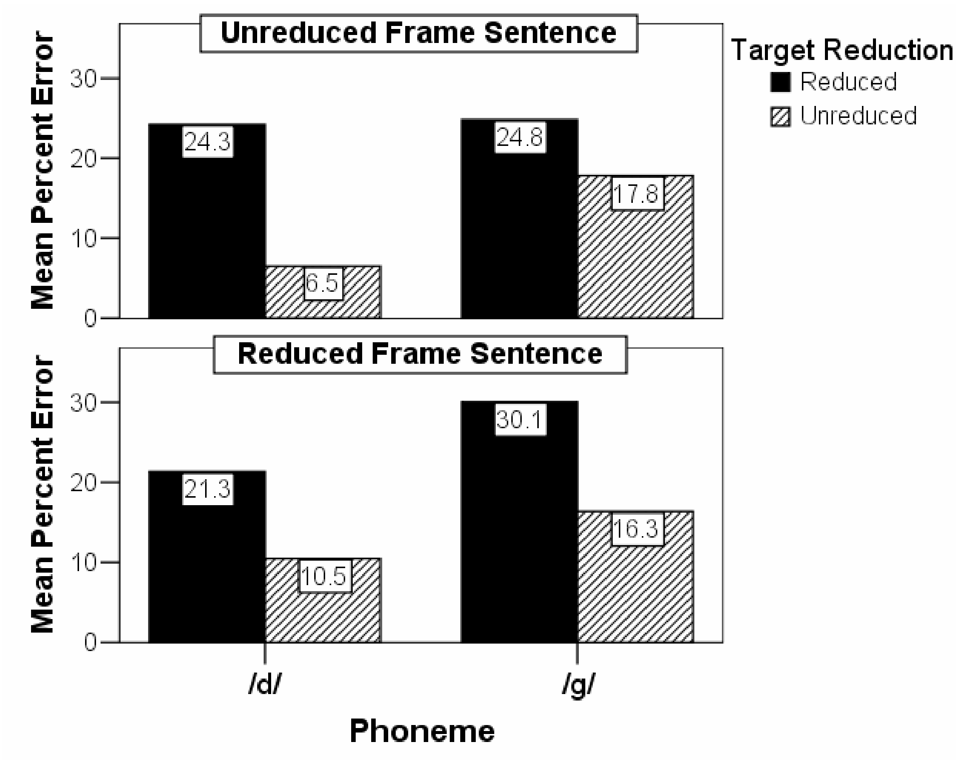


Figure 3.6 Mean percent error for auditory lexical decision of words in reduced and unreduced contexts for /d/ and /g/ word–medial phonemes of reduced and unreduced targets.

A three-factor ANOVA was performed on percent error of responses for listeners (*F1*) or items (*F2*) as repeated measures. The three-way interaction of Frame Sentence Reduction, Target Reduction, and Phoneme was significant ($F1(1,25)=8.266$, $p<0.01$; $F2(1,72)=9.91$, $p<0.005$). No two-way interactions were significant: Phoneme with Target Reduction ($F1(1,25)=1.004$,$p>0.05$; $F2<(1,72)=1.515$, $p>0.05$), Phoneme with Frame Sentence Reduction ($F1<1$; $F2(1,72)=1.969$, $p>0.05$), Target Reduction with Frame Sentence Reduction ($F1<1$; $F2<1$). The main effect of Target Reduction was significant: $F1(1,25)=70.164$, $p<0.001$; $F2(1,72)=39.356$, $p<0.001$. The main effect of Phoneme was significant in the by-subjects analysis but was not significant for by-items

($F1(1,25)=23.078$, $p<0.001$; $F2(1,72)=2.468$, $p>0.05$). The main effect of Frame Sentence Reduction was not significant: $F1(1,25)=1.226$, $p>0.05$; $F2<1$. In the response latency results the interaction between Target Reduction and Frame Sentence Reduction illustrated an important relationship between the two factors. The interaction between Target Reduction and Frame Sentence Reduction was further investigated by splitting the data by phoneme.

The two-way ANOVA of the flap items for the Target Reduction and Frame Sentence Reduction factors showed only a main effect of Target Reduction ($F1(1,25)=39.915$, $p<0.001$; $F2(1,36)=25.04$, $p<0.001$). The main effect of Frame Sentence Reduction ($F1<1$; $F2<1$) was not significant. The interaction was not significant by-subject but was significant in the by-items ($F1(1,25)=1.785$, $p>0.05$; $F2(1,36)=5.0427$, $p<0.05$). The two-way ANOVA for the /g/ items also shows a significant interaction in the by-items but not in the by-subjects ($F1(1,25)=1.823$, $p>0.05$; $F2(1,36)=5.184$, $p<0.05$). The main effect of Frame Sentence Reduction was not significant ($F1(1,25)=1.167$, $p>0.05$; $F2(1,36)=1.481$, $p>0.05$). The main effect of Target Reduction was significant ($F1(1,25)=18.998$, $p<0.001$; $F2(1,36)=14.521$, $p<0.005$). As with Experiments 1 and 2, possible ambiguous items were removed and the analysis[8] was

[8] Statistical analysis of the non ambiguous response latency data showed only one minor difference to the main analysis otherwise all other comparisons were identical. The three-way interaction was not significant ($F1(1,25)=2.507$, $p>0.05$; $F2<1$). The main effect of Target Reduction ($F1(1,25)=14.801$, $p<0.005$; $F2(1,50)=9.176$, $p<0.005$) persisted. The main effects of Phoneme ($F1<1$; $F2<1$) and Frame ($F1<1$; $F2<1$) were not significant. The interaction of Target Reduction with Frame Sentence Reduction was not significant by-subjects but was significant for items ($F1<1$; $F2(1,50)=4.849$, $p<0.05$). It is possible that the lack of significance in the by-subjects analysis is due to the decrease in power of this post-hoc test of the original 74 items only 51 remain for this analysis. The remaining two-way interactions were not significant (Target Reduction with Phoneme ($F1<1$; $F2<1$) and Frame Sentence Reduction with Phoneme ($F1<1$; $F2<1$)). The percent error analysis showed no major difference.

again performed. No major differences were found between the main analysis and the analysis of non-ambiguous items.

The results of Experiment 3a have shown an effect of Target Reduction such that reduced targets are slower to process than unreduced targets, supporting the finding in Experiments 1 and 2. Context was shown to facilitate processing of targets when the preceding context matched the reduction style of the target for the response latency data. This is supported by the significant interaction and the simple effect of items preceded by an unreduced frame sentence. The simple effect of items preceded by the reduced frame sentence was not significant, but showed a trend in the direction expected. No interaction of Target Reduction with Frame sentence was found in the percent error data. In fact a visual inspection of the /g/ percent error data shows a non-significant trend in the opposite direction as might be predicted by the response latency results.

3.2.2 Experiment 3b

In Experiment 3a, the almost significant interaction of Target Reduction and Frame Sentence Reduction indicated that in the inconsistent cases the priming effect was smaller than in the consistent cases. The interaction indicates that there may be an effect of context, albeit a small effect. A large number of listeners reported that the repetition of the same frame sentence became monotonous, causing them to stop attending to the frame sentence. In Experiment 3b, 12 new frame sentences were introduced, in addition to the frame from Experiment 3a, for a total of 13 frames. The new frames like the previous frame contributed no semantic content. It was expected that the greater diversity of frame sentences would help listeners attend more to the frames and would increase the effect of frame sentence on the processing of the reduced and unreduced targets.

3.2.2.1 Participants

Forty–six listeners from the University of Arizona community participated in the experiment. Listeners were recruited from the linguistics undergraduate pool, students mainly in Introduction to Linguistics courses or they were recruited from the general University community. Listeners' ages ranged from 18-60 years. All listeners were native speakers of American English either studying a foreign language (generally Spanish or French) or late bilinguals with no known hearing problems. Listeners received either extra credit or were paid for their participation. Listeners in this experiment had not participated in Experiments 1, 2 and 3a.

3.2.2.2 Materials

In this experiment the target items are identical to the items from Experiment 1 and 3a. In this experiment, 13 frame sentences were used. These frame sentences were produced by the same phonetician using the same recording setup as for Experiments 1, 2, and 3a. Frame sentences were created so that they were all approximately quotative and ended in a variety of sounds. This allowed any frame sentence to be matched with any target, without providing semantic or phonetic cues to the target. Each frame sentence was produced in reduced and unreduced forms. Multiple repetitions of each frame sentence, including a variety of levels of reduction, were recorded. Frame sentences were chosen based on an auditory comparison of reduced to unreduced forms and on clarity of the sentence. One reduced and one unreduced version of each frame sentence was selected, totaling 26 total sentences. Frame sentences were normalized using Adobe Audacity's "Group Waveform Normalize" function which normalized all of the files to the average level of the source files and applied an equal loudness contour which emphasized the mid frequencies. The 13 frame sentences and their total durations are provided in Table 3.2 below.

Table 3.2 Frame sentences used in Experiment 3b with the total duration of reduced and unreduced frames chosen for use in the experiment.

Frame Sentence	Reduced (sec)	Unreduced (sec)
"Yesterday I heard him say ____"	1.069	1.577
"And then yesterday he says ____"	0.898	1.734
"A lot of the time he says ____"	1.064	1.473
"I thought what I heard was ____"	0.970	1.477
"Last week she was telling me ____"	1.128	1.747
"Today I told her ____"	0.649	1.353
"I can't believe she said ____"	0.907	1.287
"I don't think she would mean ____"	0.812	1.230
"Over the phone I heard ____"	0.686	1.262
"In class the teacher told me ____"	1.086	1.726
"On the radio they recorded ____"	1.066	1.434
"I downloaded this off the web ____"	1.235	1.535
"What they always mean is ____"	0.976	1.600

3.2.2.3 Procedure

The procedures were identical to the procedure in Experiment 3a except that listeners heard the target stimulus immediately after one of 13 frame sentences. Response latency and percent error were measured for each listener. Response latency was measured from the offset of the target stimulus.

3.2.2.4 Results

Items with response latencies less than 150ms or greater then 1300ms were considered errors and excluded (approximately 5% of the total responses). Cutoffs for response latencies were chosen based on a histogram of the distribution and were selected at points where the distribution leveled out on either end. Additionally, responses to eight of the items were excluded due to high error rates (below 32% correct) [9]. Subjects who scored below 70% correct were excluded from the analysis, excluding a total of eleven subjects. The results from the response latencies are presented first, followed by the percent error results. Figure 3.7, below, summarizes the response time data.

[9] As for Experiment 1 and 3a the items exclusion cutoff for error rate was set low as a result of the difficulty of the auditory lexical decision experiment.

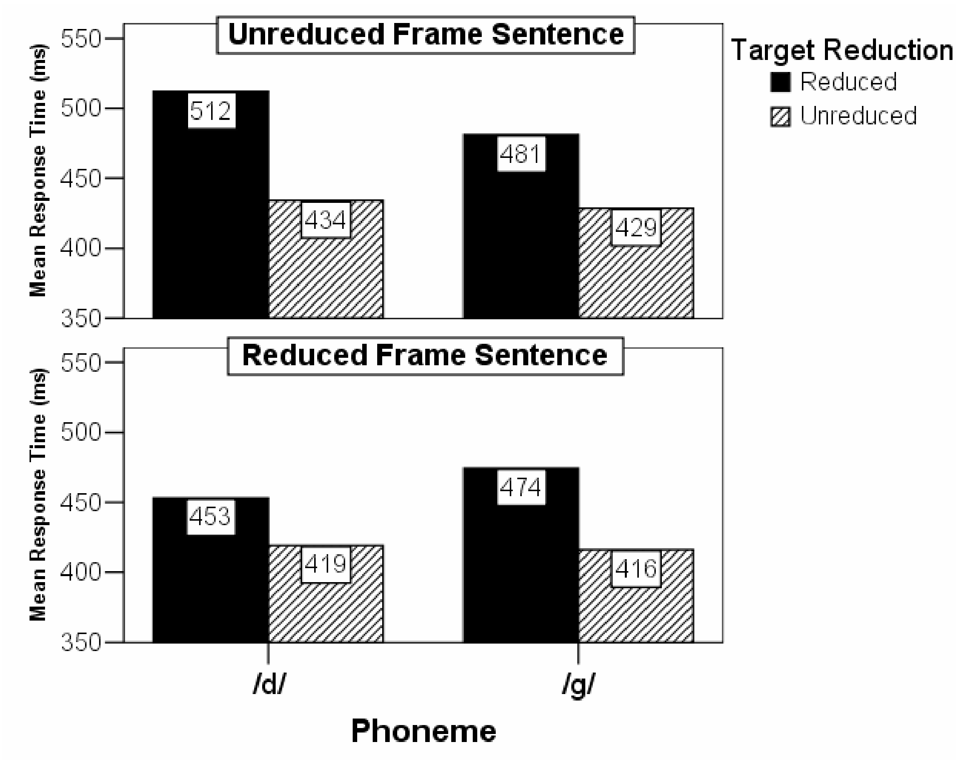


Figure 3.7 Mean response time for auditory lexical decision of words in reduced and unreduced contexts for /d/ and /g/ word–medial phonemes of reduced and unreduced targets.

A three-factor ANOVA was performed on the response latencies for listeners (*F1*) or items (*F2*) as repeated measures. As in Experiment 3a there are three independent variables: Frame Sentence Reduction (reduced versus unreduced), Target Reduction (reduced versus unreduced), and Phoneme (/d/ versus /g/). The three-way interaction of Target Reduction, Frame Sentence Reduction, with Phoneme was not significant ($F1(1,27)=1.696$, $p>0.05$; $F2(1,70)=1.34$, $p>0.05$). The two-way interaction of Target Reduction with Phoneme ($F1<1$; $F2<1$) and Phoneme with Frame Sentence Reduction ($F1(1,27)=3.278$, $p<0.05$; $F2(1,70=1.537$, $p<0.05$) was also not significant. The two-way interaction of Target Reduction and Frame Sentence Reduction was significant in the by-subjects analysis but the by-items analysis was not significant: $F1(1,27)=5.214$, $p<0.05$; $F2<1$. Indicating again that there may be a weak interaction and thus that responses to targets are influenced by the preceding context. The main effect of Target Reduction showed a significant effect: $F1(1,27)=62.025$, $p<0.001$; $F2(1,70)=28.449$, $p<0.001$. The main effect of Phoneme was not significant: $F1(1,27)=2.629$, $p>0.05$; $F2<1$, and main effect of and Frame Sentence Reduction was significant in the by-subjects analysis but not significant in the by-items analysis: $F1(1,27)=14.244$; $p<0.005$; $F2(1,70)=2.784$, $p>0.05$.

Due to the partial significance of the interaction in this experiment and in Experiment 3a, the factor Phoneme was collapsed and a two-factor ANOVA was performed with Target Reduction and Frame Sentence Reduction as factors. As with the three-factor analysis, the main effect of Frame Sentence Reduction was significant in the by-items analysis but not the by-subjects ($F1(1,27)=11.909$, $p<0.005$; $F2(1.71)=2.763$, $p>0.05$) while the main effect of Target Reduction was significant in both by-subjects and items ($F1(1,27)=65.518$, $p<0.001$; $F2(1,71)=28.808$, $p<0.001$). The interaction of Target Reduction with Frame Sentence Reduction like the three-factor analysis showed a significant effect in the by-subjects but the by-items was not significant ($F1(1,27)=4.431$, $p<0.05$; $F2<1$). Figure 3.8 shows the data from Figure 3.7 collapsed by phoneme, as in Experiment 3a. This graph specifically illustrates the interaction of Target Reduction with Frame Sentence Reduction.

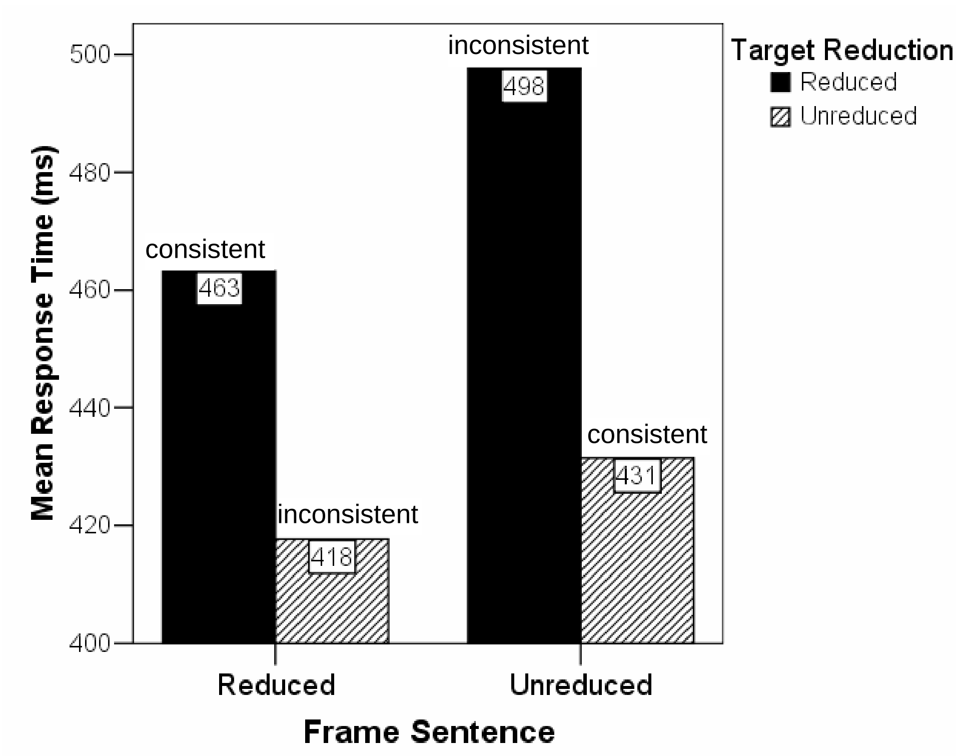


Figure 3.8 Mean response time for auditory lexical decision of words in reduced and unreduced contexts for reduced and unreduced targets to illustrate the interaction.

Figure 3.8 also illustrates how the factor Frame Sentence Reduction affects response latencies for reduced and unreduced targets. In this experiment the results are different because for the unreduced targets the subjects are faster to respond to the inconsistent stimuli as opposed to the consistent stimuli. The results for reduced targets in this experiment are the same as Experiment 3a. When target and frame are consistent, response latencies are faster; when they are inconsistent response latencies are slower. Tests of simple effects showed that Target Reduction was significant for the unreduced Frame Sentence ($F1(1,27)=66.214$, $p<0.001$; $F2(1,71)=17.594$, $p<0.001$). The simple effect for the reduced Frame Sentence was also significant ($F1(1,27)=19.682$, $p<0.001$;

$F2(1,71)=10.458$, $p=0.005$). Effect size (by-subjects = $\eta_p^2 1$, by-items = $\eta_p^2 2$) of Target Reduction calculated for unreduced Frames Sentences ($\eta_p^2 1 = 0.71$; $\eta_p^2 2 = 0.199$) showed larger effect than for reduced Frame Sentences ($\eta_p^2 1 = 0.422$; $\eta_p^2 2 = 0.128$). Simple effects of Frame Sentence Reduction were significant for reduced targets in the by-subjects analysis but not in the by-items analysis ($F1(1,27)=10.573$, $p<0.005$; $F2(1,71)=2.339$, $p>0.05$). Simple effects for unreduced targets showed an almost significant trend for the by-subjects analysis and was not significant in the by-items analysis ($F1(1,27)=4.194$, $p=0.05$; $F2<1$). The significant interaction and the difference in effect size demonstrate that Frame Sentence Reduction does influence the response times for targets. Figure 3.9 summarizes the percent error data for Experiment 3b.

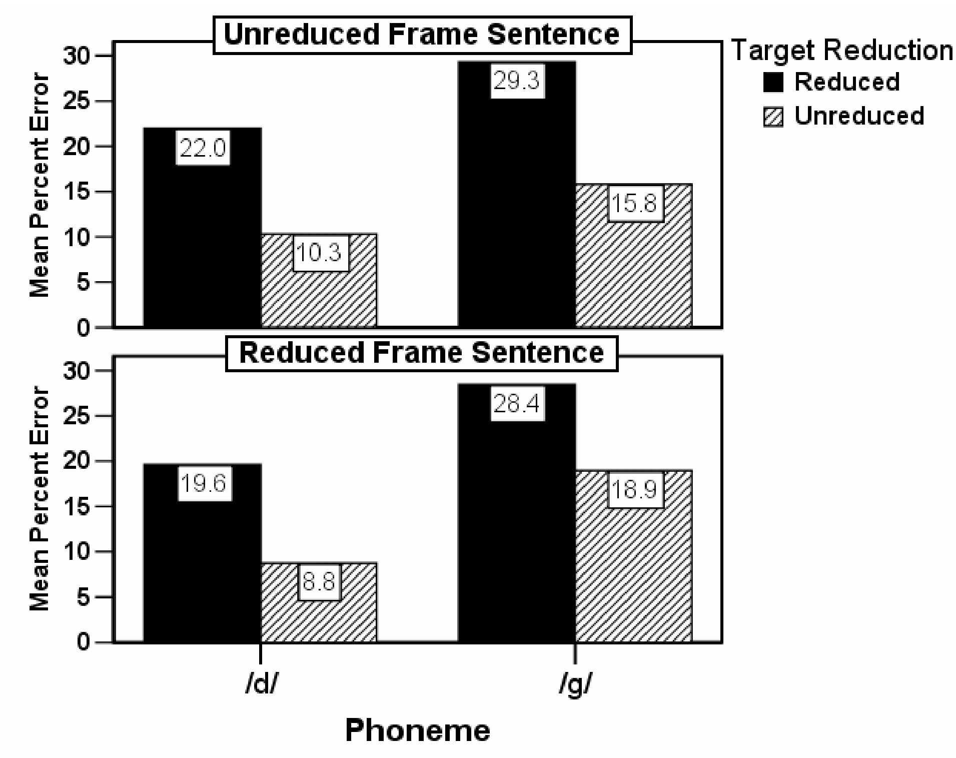


Figure 3.9 Mean percent error for auditory lexical decision of words in reduced and unreduced contexts for /d/ and /g/ word–medial phonemes of reduced and unreduced targets.

A three-factor ANOVA was performed on percent error of responses for listeners (*F1*) or items (*F2*) as repeated measures. The three-way interaction of Frame Sentence Reduction, Target Reduction, with Phoneme was not significant (*F1*<1; *F2*<1). No two-way interactions were significant: Phoneme with Target Reduction (*F1*<1; *F2*<1), Phoneme with Frame Sentence Reduction (*F1*(1,27)=2.649, p>0.05; *F2*(1,70)=2.842, p>0.05), Target Reduction with Frame Sentence Reduction (*F1*<1; *F2*<1). The main effect of Target Reduction was significant: *F1*(1,27)=80.825, p<0.001; *F2*(1,70)=19.907, p<0.001. The main effect of Phoneme was significant in the by-items analysis but was not significant for by-subjects (*F1*<1; *F2*(1,70)=8.111, p<0.01). The main effect of Frame

Sentence Reduction was not significant by-items but was significant in the by-subjects analysis: $F1(1,27)=20.819$, $p<0.05$; $F2<1$.

The results of Experiment 3b showed weak statistical evidence of an interaction of Frames Sentence Reduction with Target Reduction in the response latency analysis. As with all previous experiment there was strong support for the effect of target reduction in both the response latency and the percent error analyses.

3.2.3 Summary of Experiment 3

Considering the results from both Experiment 3a and 3b together, they both show weak indications that context affects listeners' responses, particularly in the response latency data, to reduced and unreduced targets. As has been consistent through out this current work there has been strong evidence for an effect of Target Reduction, such that processing of reduced targets is inhibited as compared to unreduced targets. It is possible that by replicating Experiment 3b, but using a cross-modal identity priming task, the task may provide better access to how speech style affects processing.

3.3 Experiment 4: Cross–modal identity priming in context

The cross-modal identity priming task (Norris, Cutler, McQueen, and Butterfield, 2006) was used in this experiment. The cross-modal identity priming task was selected to increase the strength of the effect based on its previous success investigating context (Coenen, Zwitserlood, and Bölte, 2001), as described in section 3.2. Cross-modal identity priming was used by Coenen, Zwitserlood, and Bölte (2001), who showed a significant effect of context in regressive and progressive place assimilation.

3.3.1 Participants

Sixty listeners from the University of Arizona community participated in the experiment. Listeners were recruited from the linguistics undergraduate pool, students mainly in Introduction to Linguistics courses or they were recruited from the general University community. Listeners' ages ranged from 18-60 years. All listeners were native speakers of American English either studying a foreign language (generally Spanish or French) or late bilinguals with no known hearing problems. Listeners received either extra credit or were paid for their participation. Listeners in this experiment had not participated in Experiments 1, 2, or 3.

3.3.2 Materials

Target items in this experiment were identical to target items in Experiment 2. Table 3.3 summarizes the item types. The frame sentences used were the same as the frame sentences used in Experiment 3b.

Table 3.3 Example items from Experiments 2 and 4 illustrating both auditory prime and visual probe

	Flap /d/	**/g/**	**Non–word Filler**		**Phonological Overlap Filler**
Auditory Stimulus for Prime or Overlap Condition	puddle (ɾ) puddle (ɾ̞)	beggar (g) beggar (g̞)		level	
Auditory Stimulus for Control or Non-Overlap Conditions	bracket	torture	offer		dozen
Visual Probe	puddle	beggar	prunshin	mavel	orphan

3.3.3 Procedures

As with Experiment 2, each listener saw 30 /d/ visual probe items and 30 /g/ visual probe items. For each listener, 10 /d/ and 10 /g/ items were paired with an auditory prime consisting of the same word in reduced pronunciation, 10 /d/ and 10 /g/ items were paired with an auditory prime consisting of the same word in unreduced pronunciation, and 10 /d/ and 10 /g/ items were paired with an unrelated word as the auditory control stimulus. Auditory targets were paired with one of the 13 frame sentences in either reduced or unreduced frame sentence. This was followed by the visual target (an interval of 0ms). Six lists were created by counterbalancing the three factors: Visual Item (/d/, /g/), Target Reduction (reduced, unreduced, auditory control) and Frame Sentence Reduction (reduced versus unreduced). For example, the visual probe *puddle* appeared in the first two lists with the reduced auditory prime preceded by either the reduced (matched reduced case) or unreduced (mismatch unreduced-reduced case) frame sentences, in the third and fourth lists visual *puddle* appeared with the unreduced auditory probe preceded by either the reduced (mismatch reduced-unreduced case) or unreduced (match unreduced case) frame sentence, and in the fifth and sixth lists visual *puddle* appeared with the auditory control preceded by either the reduced and unreduced frame sentences. All fillers (160) appeared in each list and were matched randomly with a frame sentence[10]. Thus, any one listener would hear 40 target items and 20 auditory control items (25% of

[10] The fillers in Experiment 3 always appeared with the same frame sentence. The individual fillers in Experiment 4 appeared with either the reduced or unreduced version of the same frame sentence depending on the list it occurred in. This variation was discovered after over half of the subjects had participated in the experiment. While having individual fillers matched with the same frame sentence is ideal, it was determined that the addition of this variation would most likely not affect the outcome of the experiment.

the total number of items presented) in any one session. Lists were pseudo-randomized for presentation and were presented in opposite orders, creating 12 total lists.

Listeners were tested individually in sound-attenuated booths. The auditory primes were presented over headphones and the visual probe appeared on the monitor viewable though a window outside the booth in lower-case, 36-point Arial font. The visual probe appeared at the offset of the auditory stimulus (0ms interval). Visual probes remained on the monitor for 2 seconds; if the listener did not respond in the time allotted the experiment proceeded to the next stimulus. Listeners were instructed to respond as quickly and as accurately as possible to the visual probe using a button box with buttons labeled "YES" and "NO". A "YES" response indicated that the listener decided the visual probe was a real word of English. The experimental procedure for each listener was: instruction, practice trial, followed by one of the six stimulus lists, and then a comprehension test. Listeners were informed at the beginning of the procedure in the instructions that there would be a comprehension test at the end of the experiment. The test contained 10 visually presented items, half of which were presented auditorily in the main experiment and the other half of which were new. Listeners were asked if they heard these visually presented words in the main part of the experiment. Listeners scored an average of 68.3% correct on the comprehension test. A single factor ANOVA of the number of 'yes' responses for new and presented targets showed that subjects responded 'yes' significantly more to the already presented stimuli ($F(1,456)=73.581$, $p<0.001$).

3.3.4 Results

Items with response latencies less than 350ms or greater then 1500ms were considered errors and excluded (less than 3% of the total responses). Cutoffs for response latencies were chosen based on a histogram of the distribution of all response latencies. The cutoffs were selected at points where the distribution leveled out on either end. Accuracy scores for subjects below 85% were excluded from the analysis, which excluded a total of twelve subjects. Upon analysis of Experiment 4 results an error was detected in the counterbalancing of item lists. Stimulus lists were not properly counterbalanced across the Frame Sentence Reduction factor, and as a result three additional subjects were excluded from the analysis[11]. The results from the response latencies are presented first, followed by the percent error results. Figure 3.10, below, summarizes the response latency data.

[11] These subjects were excluded because there was no data in some of the analysis cells due to the error in counter-balancing.

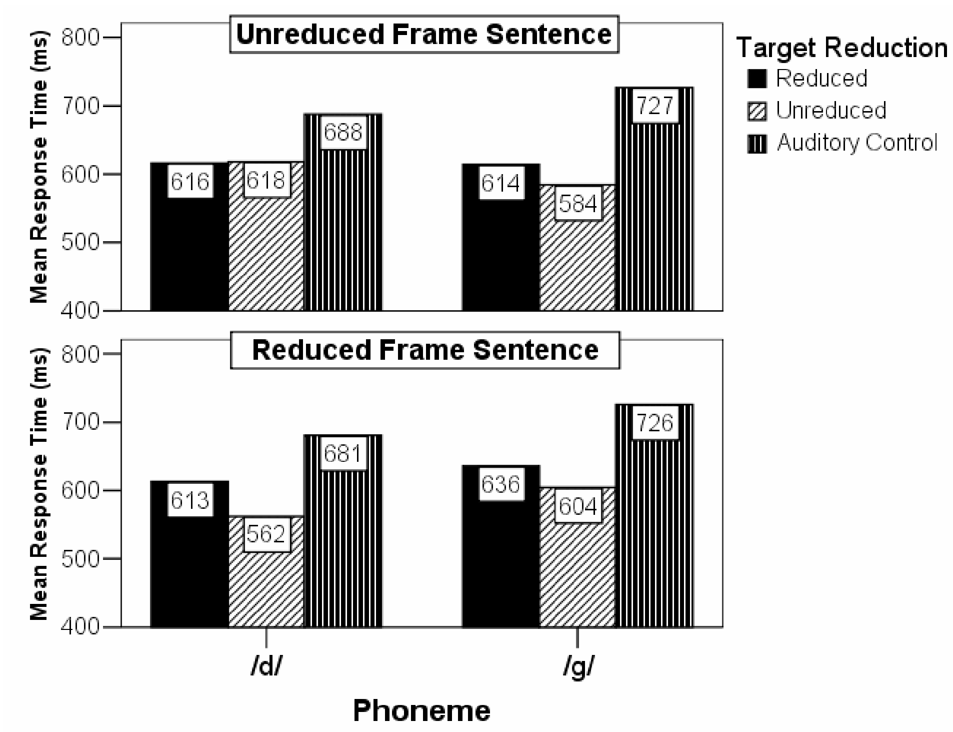


Figure 3.10 Mean response time of cross-modal identity priming of words in reduced and unreduced contexts for reduced and unreduced /d/ and /g/ word–medial targets.

By–subjects (*F1*) and by–items (*F2*) ANOVAs of the response latency data were performed. The following three factors were analyzed: Phoneme (/d/, /g/), Target Reduction (reduced, unreduced, auditory control) and Frame Sentence Reduction (reduced versus unreduced). All factors in the by-subjects analysis were within-subjects factors and in the by-items analysis Target Reduction and Frame Sentence were within-items factors and Phoneme is a between-items factor. The three-way interaction of Visual Item, Target Reduction, with Frame Sentence reduction was not significant ($F1<1$; $F2(2,116)=1.196$, $p>0.05$). The two-way interactions of Target Reduction with Phoneme ($F1(2,66)=1.808$, $p>0.05$; $F2(2,116)=1.5529$, $p>0.05$) and Target Reduction with Frame Sentence Reduction ($F1<1$; $F2(2,116)=1.012$, $p>0.05$) were also not significant. The interaction of Phoneme with Frame Sentence Reduction was significant ($F1(1,33)=8.751$, $p<0.01$; $F2(1,58)=9.018$, $p<0.005$). The main effects of Phoneme ($F1<1$; $F2(1,58)=2.45$, $p>0.05$) and Context ($F1(1,33)=1.081$, $p>0.05$; $F2<1$) were not significant. The main effect of Target Reduction was significant ($F1(2,66)=71.026$, $p<0.001$; $F2(2,116)=56.784$, $p<0.001$) showing that identity priming of the target items was occurring as compared to the Auditory Control items.

A second three-way ANOVA was performed removing the Auditory Control level from Target Reduction. The three-way interaction of Target Reduction, Phoneme with Frame Sentence Reduction was not significant ($F1(1,33)=1.039$, $p>0.05$; $F2<1$). The interactions of Target Reduction with Phoneme ($F1<1$; $F2<1$) and Target Reduction with Frame Sentence Reduction ($F1<1$; $F2(1,58)=1.93$, $p>0.05$) were also not significant. The interaction of Phoneme with Frame Sentence Reduction was significant ($F1(1,33)=9.075$, $p<0.01$; $F2(1,58)=11.196$, $p<0.005$), showing that the significant interaction was not due to differences in the Auditory Control items. Since the interaction of the Target Reduction with Frame Sentence Reduction is the main focus of this experiment, follow-up tests of interaction of Phoneme with Frame Sentence Reduction are not pursued. The main effects of Phoneme ($F1<1$; $F2<1$) and Frame Sentence Reduction ($F1(1,33)=1.532$, $p>0.05$; $F2<1$) were also not significant. The main effect of Target Reduction was significant ($F1(1,33)=22.415$, $p<0.001$; $F2(1,58)=14.088$, $p<0.001$), showing that listeners generally respond more slowly to reduced targets as compared to unreduced targets.

Planned pairwise comparisons of each priming condition compared to the control condition for both the reduced and unreduced Frame Sentence factors of the /d/ items showed that listeners' responses were primed significantly for both reduced targets (reduced frame $F1(1,44)=15.475$, $p<0.001$; $F2(1,29)=10.713$, $p<0.005$; unreduced frame $F1(1,44)=5.138$, $p<0.05$; $F2(1,29)=8.076$, $p<0.01$) and unreduced targets (reduced frame $F1(1,44)=22.693$, $p<0.001$; $F2(1,29)=42.231$, $p<0.001$; unreduced frame $F1(1,44)=16.614$, $p<0.001$; $F2(1,29)=10.336$, $p<0.005$). However, the pairwise comparisons of unreduced to reduced /d/ items in unreduced frame sentence showed a non-significant difference ($F1(1,44)=2.517$, $p>0.05$; $F2<1$) indicating that the degree of priming was not significantly different. The comparison of reduced and unreduced /d/ targets in the reduced frame sentence showed a non-significant difference in the by-subjects analysis and a significant difference in the by-items analysis ($F1(1,44)=2.331$, $p>0.05$; $F2(1,29)=11.401$, $p<0.005$) The general trend of the Target Reduction items follows the same trend as in the previous experiments, except in the case of the /d/ items in the unreduced frame sentence. Response latencies for reduced and unreduced targets in the unreduced frames are practically identical. The planned pairwise comparisons of the /g/ items were primed significantly when comparing the auditory control targets with the reduced targets (reduced frame $F1(1,44)=18.443$, $p<0.001$; $F2(1,29)=13.359$, $p<0.005$; unreduced frame $F1(1,44)=23.408$, $p<0.001$; $F2(1,29)=29.648$, $p<0.001$) and with the unreduced targets (reduced frame $F1(1,44)=37.724$, $p<0.001$; $F2(1,29)=32.066$, $p<0.001$; unreduced frame $F1(1,44)=45.519$, $p<0.001$; $F2(1,29)=33.598$, $p<0.001$), showing that both the reduced and unreduced auditory primes affected responses to the visual target regardless of the type of frame sentence. However, the comparison of reduced vs. unreduced for the unreduced frame sentence did not show a significant effect ($F1(1,44)=2.3$, $p>0.05$; $F2(1,29)=1.813$, $p>0.05$) but it did show a significant effect for the reduced frame sentence targets ($F1(1,44)=4.414$, $p<0.05$; $F2(1,29)=4.261$, $p<0.05$). In the non-significant case the response latencies again showed a trend of more priming for the unreduced targets in the unreduced frames sentences and a significant effect for the targets following the reduced frame sentence. The results from the mean percent error data are shown in Figure 3.11, below.

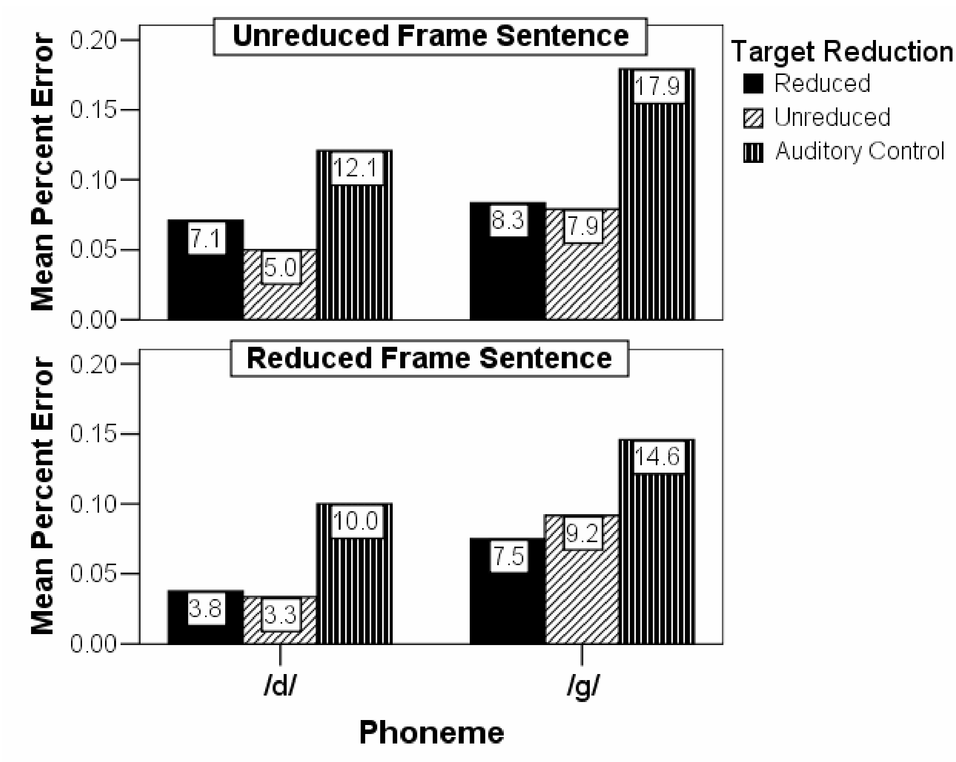


Figure 3.11 Mean percent error of cross modal identity priming of words in reduced and unreduced contexts for reduced/unreduced /d/ and /g/ word–medial targets.

A three-way ANOVA of the error data was performed by–subjects ($F1$) and by–items ($F2$). The three-way interaction and the two-way interactions were not significant (Frame Sentence Reduction, Phoneme with Target Reduction ($F1(2,66)=1.1$, $p>0.05$; $F2<1$); Target Reduction with Phoneme ($F1<1$; $F2<1$); Target Reduction with Frame Sentence Reduction ($F1<1$; $F2<1$); Phoneme with Frame Sentence Reduction ($F1<1$; $F2<1$). Listeners made fewer mistakes in the identity primes than the non–identity primes in Target Reduction ($F1(2,66)=19.262$, $p<0.001$; $F2(2,116)=20.058$, $p<0.001$). The main effect of Phoneme showed that /g/ targets had more errors than /d/ targets ($F1(1,33)=9.317$, $p<0.005$; $F2(1,58)=6.282$, $p<0.05$). The main effect of Frame Sentence Reduction ($F1(1,33)=1.922$, $p>0.05$; $F2(1,58)=2.982$, $p>0.05$) was not found to be significant.

A second three-way ANOVA was performed removing the Auditory Control level from the Target Reduction factor. None of the interactions showed a significant effect (all interactions $F1<1$; $F2<1$, except Phoneme with Frame Sentence Reduction: $F1(1,33)=1.219$, $p>0.05$; $F2(1,58)=1.773$, $p>0.05$). The main effects of Target Reduction ($F1<1$; $F2<1$) and Frame Sentence Reduction ($F1(1,33)=1.763$, $p>0.05$; $F2(1,58)=1.269$, $p>0.05$) were also not significant. No effect of Target Reduction indicates that listeners are making similar numbers of mistakes for both the reduced and unreduced targets. The main effect of Phoneme was significant ($F1(1,33)=7.091$, $p<0.05$; $F2(1,54)=3.969$, $p=0.051$), showing that the difference between /d/ and /g/ is not due to differences in the Auditory Control items.

Planned pairwise comparisons of each priming condition compared to the control condition for both the reduced and unreduced Frame Sentence factors of the /d/ items showed that listeners' responses showed significant priming for reduced targets in the reduced frame ($F1(1,33)=8.972$, $p<0.01$; $F2(1,29)=5.8$, $p<0.05$) but not in the unreduced frame ($F1<1$; $F2(1,29)=3.551$, $p>0.05$). The unreduced targets showed priming in the reduced frame ($F1(1,33)=13.01$, $p<0.005$; $F2(1,29)=6.605$, $p<0.05$) and the unreduced frame showed priming in the by-items analysis but not the by-subjects ($F1(1,33)=2.619$, $p>0.05$; $F2(1,29)=5.898$, $p<0.05$). The pairwise comparisons of unreduced to reduced /d/ items in both frame sentences showed non-significant differences (reduced frame $F1<1$; $F2<1$; unreduced frame $F1<1$; $F2<1$) indicating that the degree of priming was not significantly different. The general trend of the Target Reduction items follows the same trend as in the previous experiments. The planned pairwise comparisons of the /g/ items were primed significantly for reduced targets (reduced frame $F1(1,33)=6.928$, $p<0.05$; $F2(1,29)=8.906$, $p<0.01$; unreduced frame $F1(1,33)=8.504$, $p<0.01$; $F2(1,29)=10.796$, $p<0.005$). The unreduced targets showed significant priming in the unreduced frame ($F1(1,44)=18.799$, $p<0.001$; $F2(1,29)=19.503$, $p<0.001$) as compared to the auditory control condition but not in the reduced frame ($F1(1,33)=2.723$, $p>0.05$; $F2(1,29)=4.15$, $p=0.051$). The comparison of reduced vs. unreduced for the unreduced frame sentence did not show a significant effect ($F1<1$; $F2<1$) nor did targets in the reduced frame sentences ($F1<1$; $F2<1$). It is possible that the non-significant trend for the /d/ and the /g/ items is due to the counter-balancing error.

The results of Experiment 4 are somewhat different than those in Experiment 3, the auditory lexical decision with context. There was a significant effect of identity priming for both the response latency and the percent error data, which shows the subjects were primed in the way expected. There was also a significant interaction of Phoneme and Frame Sentence Reduction in the response latency analysis. Importantly, Frame Sentence Reduction did not significantly affect Target Reduction. Target Reduction as with previous experiments showed a trend that listeners' responses to reduced target items were slower than unreduced target items. However, the /d/ items in the unreduced frame sentences showed very little difference.

It is possible that the error made in the counter-balancing of the Frame Sentences is the source of the inconsistent results. The mistake in counter-balancing caused subjects to be presented with an unequal number of groups. For example, subject A, who received list A, could have been presented with 32 reduced /d/ targets preceded by an unreduced frame in one cell and only 8 unreduced /d/ targets preceded by an unreduced frame in another cell. Assuming that none of these items were discarded this particular comparison is unbalanced which could cause a bias toward one cell affecting the results of the analysis. It possible that the cell with 32 responses happens to consist of targets that are easy to respond too and the other cell with 8 responses consists of targets that happen to be difficult. When in fact, the cell with more responses in all of the other lists contains hard items and the cell with fewer responses contains easy items. The averaged response latencies would be skewed due to the large number of responses in this particular cell. It is also possible that the error in counter-balancing did not affect the results of this experiment and that results of the experiment have different implications than in Experiment 3. In order to verify these results this experiment will need to be replicated with the items properly counterbalanced.[12]

3.4 General discussion

The results of Experiments 3 and 4 are summarized in this section and the results are briefly discussed in light of the main questions of the dissertation. A brief explanation of the results is also provided. A more detailed discussion of all the results follows in Chapter 4.

In Experiment 3, the auditory lexical decision with context task, it was found in support of Experiments 1 (auditory lexical decision) and 2 (cross-modal identity priming) that reduced targets are more difficult to process as compared to unreduced targets, regardless of the preceding context. Experiment 4 also showed that reduced targets were

---

[12] Partial results of a replication of this experiment with the counter-balancing fixed show that the counter-balancing did influence the results (results are based on 33 subjects). For example, both analyses show a significant interaction of Phoneme with Frame Sentence Reduction. However, the targets interact in opposite ways. Figure 3.12 illustrates the difference visually.

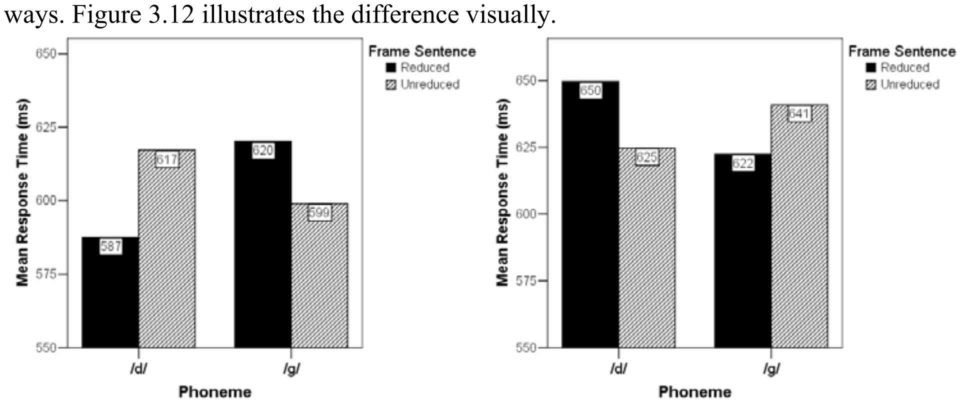


Figure 3.12 Mean response latency for /d/ and /g/ phoneme plotted with Frame Sentence Reduction illustrating the interaction. The left figure illustrates the data from Experiment 4 and the right figure shows the data for the experiment in progress.

generally more difficult to process than unreduced targets, though this effect was much smaller than in previous experiments and not found in the percent error analysis. Experiment 3 also showed an interaction of Target Reduction and Frame Sentence Reduction and differences in effect size in the comparisons of simple effects. Therefore, listeners' responses were facilitated when the Frame Sentence and Target were consistent as compared to the inconsistent cases. Another way of looking at this would be that when the effect size of the response latencies is smaller for the comparison of unreduced and reduced targets in the reduced frame sentence as compared to the same comparison in the unreduced frame sentence, then the frame sentence is facilitating responses for the reduced targets. The percent error results for the /g/ items did not support this generalization. In fact these results are the opposite of what would have been expected based on the response latency data. As already discussed in 2.3, it is possible that the source of this result is a frequency effect or the fact that the flap and /g/ items have different lexical representations.

Since the effect of Frame Sentence Reduction was not particularly strong in Experiment 3a, it was hoped that Experiment 3b would find a stronger effect of Frame Sentence Reduction by increasing the number of frame sentences and using the cross-modal identity priming task. Experiment 3b also found weak effects of the interaction of Frame Sentence Reduction with Target Reduction. These results taken together provide stronger evidence that Frame Sentence Reduction, i.e. speech style, affects how listeners process speech. Since the results from both Experiment 3a and 3b were still relatively weak, a cross-modal identity priming task was used to further explore the effect of speech style on processing. Experiment 4 however did not show an interaction between Target Reduction and Frame Sentence Reduction. Frame Sentence Reduction showed only an interaction with Phoneme in Experiment 4.

The disagreement between the results of Experiment 3 and 4 indicate that further investigation is necessary to understand the effect of speech style on the processing of reduced and unreduced targets. As already discussed it is possible that the error in counter-balancing is the source of the disagreement between Experiments 3 and 4. Further investigation is necessary to identify this error as the source of these differences.

Chapters 2 and 3 have shown that processing of reduced forms is different from the processing of unreduced forms. They have also shown that context can change a listener's expectations and thus affect processing of target items. The results from Chapters 2 and 3 (Experiments 1-4) are discussed in greater detail in Chapter 4, where the results are discussed in light of theories of spoken word recognition and general speech perception.

CHAPTER 4

DISCUSSION AND CONCLUSIONS

4.0 Introduction

Spontaneous speech styles are the types of speech that confront listeners most often in their daily interactions. Researchers, who seek to understand spoken word recognition, and more generally, how language is processed, must be aware of the processing effects of multiple speech styles. While investigation of the processing of a broad diversity of speech styles is becoming more common, much more research is currently performed on careful and controlled speech (Cutler, 1998). This dissertation contributes to the understanding of spoken word recognition in more spontaneous speech styles. One of the characteristics of spontaneous speech is the variation caused by reduction. The American English flap is an interesting test case because it can vary from a canonical flap, which has a short-voiced closure duration with a possible burst release, to something that more closely resembles an approximant (as seen in Chapter 1).

The experiments in this dissertation were designed to address two specific questions: (1) how does reduction affect listeners' recognition of words and (2) do listeners adjust their expectations about reduction based on preceding speech style (the style of the context)? The overarching question driving this research is how do listeners recognize and process reduced speech? Furthermore, do they base their expectations of segments on speech style? In this last chapter, findings of the experiments are summarized, and their implications for theories and models of spoken word recognition are discussed.

4.1 Summary of results

4.1.1 How does reduction affect listeners' recognition of words?

The results of Experiments 1 through 4 found that recognition of reduced word forms was inhibited during lexical access. Experiment 1 showed that when presented with reduced targets in an auditory lexical decision task, listeners made more mistakes and responded more slowly (showed an inhibited response) than when presented with unreduced targets in the same lexical decision task. In Experiment 2, a cross-modal identity priming task was used to further confirm the result of inhibition due to reduction. The results of Experiment 2 showed significant priming of both the reduced and unreduced stimuli. The amount of priming for unreduced stimuli was greater than the amount of priming for the reduced targets. Thus, listeners presented with reduced targets showed inhibited response latencies and increased percent error, confirming the results of Experiment 1. Experiments 3 and 4, while designed to test research question 2 (Do listeners adjust their expectations about reduction based on preceding speech style?), also contributed findings to research question 1 (How does reduction affect listeners' recognition of words?). Experiment 3, auditory lexical decision with context, and Experiment 4, cross-modal identity priming with context, both showed that listeners' responses to reduced targets were slower and less accurate than their responses to unreduced targets. In Experiment 2, listeners' response latencies showed a non-significant difference for reduced and unreduced /g/ items, but showed a significant difference for flaps. In both Experiments 1 and 2, subjects made more errors for the /g/ targets compared to the flap targets.

An additional set of questions was posed to further clarify the finding that listeners presented with reduced targets showed inhibited response latencies and increased percent error. To investigate the source of the findings, the following secondary questions were posed: Is it more difficult for listeners to recognize words pronounced in reduced forms (implying that weaker acoustic cues cause a conversion of such forms into their abstract, underlying representations)? Or is it easier for listeners to recognize reduced forms because reduced forms are more commonly produced and encountered, and are thus more quickly accessed (frequency of occurrence effects)?

As already discussed in the preceding chapters, weaker acoustic cues in the reduced forms of the target words inhibited processing of the words, showing a dependence on acoustic information in speech processing. The cross-modal identity priming task showed that both reduced and unreduced targets were significantly primed for the response latency data (Experiments 2 and 4) and the error rate date (Experiment 2 only). This indicates that even in the reduced form of the word-medial stop, where it could be argued that the manner of the articulation had changed from plosive to approximant, listeners accessed the lexical form of the target. Additionally, the percent error data in the lexical decision experiments show that both reduced and unreduced targets activated the lexical entry. Correct identifications of lexical items indicate lexical access. Thus, if listeners show a percent error of about 50%, then they are performing at a chance level. In the case of the reduced items, subjects averaged a percent error rate between 20% and 30%, which is well above chance accuracy.

4.1.2 Do listeners adjust their expectations about reduction based on preceding speech style (context)?

The investigations in Experiments 3 and 4, while also contributing to the understanding of reduction on spoken word recognition, focused on the effect of contextual speech style in spoken word recognition. In Experiment 3, an auditory lexical decision task with one preceding frame sentence, processing was facilitated when speech style was consistent across the frame sentence and target, i.e., the difference between reduced and unreduced targets was smaller. Processing was inhibited when speech style was inconsistent across frame sentence and target, i.e., the difference was greater. This was shown by a significant interaction of the factors: Frame Sentence and Target Reduction. However, the results of this experiment were statistically weak. Subjects often reported at the end of the testing phase that during the experiment they had stopped attending to the frame sentence and only listened to the auditory prime. The weak statistics were attributed to the use of a single frame sentence. It was hypothesized that by including additional frame sentences, subjects might attend longer to the frame sentence and not lose interest as quickly. Experiment 4 sought to find stronger statistical support for this generalization by including 13 frame sentences and using the cross-modal identity priming task. Unfortunately, the results did not support the generalization. In fact, as pointed out earlier, the response time data was the opposite of the expected effect for the /g/ data in Experiment 4.

The results of Experiment 3 indicate that listeners do adjust their expectations for speech style on the preceding context. When listeners are presented with a careful or spontaneous speech style, they base their expectations for the following speech on the preceding speech style. This result is similar to the results in other context literature (discussed in section 3.1). For example, Ladefoged and Broadbent (1957) showed that listeners adjusted their expectations of the vowel in a target word based on the formant frequency information in the preceding speech.

In Experiment 3, the error rate data for the /g/ targets showed the opposite effect as that seen in the response time data. Listeners made more errors when the speech styles of the frame sentence and target were consistent and fewer errors when they were inconsistent. In Experiment 4, response latencies were dependent on the speech style. Responses to /g/ targets were opposite those for flap targets (this may be due to the counterbalancing error). These differences, along with the results from Experiment 2, indicate that the processing of word-medial flap and /g/ may differ. This is, however, very weak support for differences in the processing of the flap and /g/. The occurrence of these differences was discussed as possibly being explained by different segmental representations for /d/ and /g/, and possibly being an effect of the frequency difference between word-medial flap and /g/. Additional sources for these differences are discussed in greater detail in section 4.3.1, below.

The remainder of this chapter considers the results summarized above in light of their implications for theories and models of phonetic variation and lexical access. It also discusses the implications of this work for a broader linguistic audience. Following the discussion of these implications is an overview of planned future research. The chapter finishes with some general conclusions.

4.2 Recognition of spontaneous speech

A broad goal of this research was to investigate the recognition of reduced or spontaneous speech. As discussed in Chapter 1, reduced speech has been shown to be very different from careful speech in production (e.g., recent work on production and careful speech: Smiljanić and Bradlow, 2006; Kilanski and Wright, 2006; Smith, 2006 among others) and spoken word recognition (e.g. Cutler 1998; Mehta and Cutler, 1988; Johnson, 2004; Ernestus et al., 2002, among others). This section compares the current results to previous results and discusses the effects of speech style on word recognition in isolation and in context.

4.2.1 Comparison to previous results

Results from previous research (as described in section 1.4.2) have shown that various types of reduction tend to inhibit lexical access. Ernestus et al. (2002) showed that reduced items presented to listeners out of context were more difficult to process than those presented with phonetic context or full context (i.e., syntactic or semantic information). It has also been shown that segment deletion inhibits lexical access (Matter, 1986; Racine and Grosjean, 1997; Kuijpers, Van Donselaar, and Cutler, 1998; LoCasto and Connine 2002; Zwitserlood, 1996). Others have found that speech intelligibility decreases in fast speech when compared to intelligibility in slow speech (Pollack and Pickett 1963; Bard, Shillcock and Altmann, 1988). Bard et al. (1988) also found that the addition of following context was helpful in recognition. The findings of the experiments in the current work show a difference between reduced and unreduced speech, or by extension, spontaneous and laboratory speech. Specifically, listeners' responses are inhibited when presented with a reduced flap and /g/ similar to the experiments cited above (e.g. Mehta and Cutler, 1988; Ernestus et al., 2002).

4.2.2 Speech style

Speech style varies in many ways producing a multi-dimensional gradient scale of speech style (see discussion in section 1.3). *Spontaneous* and *laboratory* speech were used as generic terms to distinguish between faster, reduced, less careful speech styles and planned, less reduced, and more careful speech styles. In particular, this work investigated the effect of highly controlled laboratory speech, which was generated to replicate the reduction found to occur in more spontaneous speech styles.

While this research generalizes to a large portion of the speech style continuum, there are also large portions to which it does not generalize. Most of the entire upper portion of the speech continuum, which contains purely spontaneous speech between family members or close friends, is uninvestigated by this research; future studies planned for this portion are described in section 4.6, below. This investigation compares speech styles that contain phonetic reduction of a particular segment. It is limited to the small portion of the speech style continuum that involves the phonetic variation of two segments ([ɾ] and [g]). Reduction of these segments inhibits processing, particularly if they occur in isolation. Listeners adjust their expectations based on preceding speech, or contextual information. The current work investigates only one small phenomenon in the domain of phonetic variation that occurs in reduced or spontaneous speech. The effect of phonetic variation within speech styles is an area of exploration that can potentially provide an abundance of knowledge about spoken word recognition.

4.2.3 Context effects: Perception of **X** *affects perception of* **Y**

Previous research has shown that listeners adjust their expectations based on contextual information (as mentioned in 4.1.2). In other words, listeners perceive the contextual information *X* and adjust their expectations for *Y* based on *X*. For example, Ladefoged and Broadbent (1957) showed that the formant frequencies of vowels in preceding speech affected the perception of an ambiguous vowel in a target word. Miller and Lieberman (1979, Eimas and Miller, 1980) showed that syllable duration affected the perception of a /b/ versus /w/. Summerfield (1981) and Miller and Volaitis (1989) showed that speaking rate affected the perception of VOT, and thus, the identification of voiced and voiceless obstruents in American English.

Results from Experiment 3 show that the perception of preceding information (speech style) affects the perception of the following speech (a target word). Unlike previous findings, this finding indicates that speech style should be included with the contextual factors that affect perception. Furthermore, this effect is shown in the domain of spoken word recognition, such that lexical access of the target word is facilitated when the speech style information is consistent and lexical access is inhibited when the speech style information is inconsistent. Listeners take speech style into consideration as they process incoming speech. Contextual speech style is used by listeners to focus the particular processing strategies (Mehta and Culter, 1988; Cutler 1988) that are more successful in various speech styles. In other words, a listener may give greater relevance to acoustic information when no semantic or syntactic context is provided, but when context is provided semantic or syntactic information may play a greater role (Ernestus et al., 2002). In the current research, a listener expects targets with reduction when the preceding context contains elements of a reduced speech style.

4.3 H&H theory (Lindblom, 1990)

Phonetic variation has been a challenge to speech researchers attempting to understand speech perception and spoken word recognition. Many models attempt to exclude as much variation as possible, which leads to a search for phonetically invariant structure. In an attempt to explain phonetic variation without searching for invariant structure, Lindblom (1990) developed the H&H theory, where H&H stands for hyper- and hypospeech. This theory is situated within the domain of speech production and speech perception. As discussed in section 1.3, a continuum of speech is created where speech can vary from hyper- to hypospeech. Lindblom argues that the search for phonetically invariant structure is replaced with a new search to explain "the notion of sufficient discriminability and defining the class of speech signals that meet that criterion" (Lindblom, 1990: 403).

In the development of H&H theory Lindblom defines hyperspeech as "output-oriented control" and hypospeech as "system-oriented control."[13] In output-oriented control, the speech production system is creating speech at a high degree of discriminability. Simultaneously, the system-oriented control (hypospeech) is constantly

---

[13] Based on this idea, Optimality Theoretic constraints e.g. LAZY (Kirchner, 1998) were developed to account for occurrences of consonant lenition.

attempting to create speech using the least amount of energy. Reduction in spontaneous speech is a result of the competition between hyper- and hypospeech. The production system allows target undershoot to occur as the result of the drive toward hypospeech.

Lindblom observes that lexical access is “a function of the distinctiveness (rather than invariance) of the acoustic stimulus” (1990:404). If lexical access is a function of the distinctiveness of the input, then Lindblom predicts that in cases where the input is less distinct there should be some sort of lexical inhibition. Reduction, or articulatory undershoot, creates a situation of decreased distinctiveness. H&H theory would predict that because reduced flaps are more distant from stops, more like approximants, they would be more difficult to access, as shown in Experiments 1-4.

Lindblom states that “the process of discrimination is facilitated by processes not in the signal and whose contributions show short-term variations. Accordingly, lexical access is assumed to be driven by ‘knowledge’, that is, by signal-complementary processes” (1990: 404). Experiment 3, auditory lexical decision in context, shows that listeners have speech style knowledge, which they use to adjust their processing of speech. This means that listeners store information about speech style, aiding their processing. For instance, other types of information stored include word frequency or neighborhood density (Luce, 1986, Luce and Pisoni, 1998) and speaker-specific information (Ladefoged and Broadbent, 1957), among the other types described in section 3.1).

H&H theory argues that in speech production and perception the speaker and listener adapt to one another. The results of Experiment 3 support this claim. Adaptation occurs for listeners as they are presented with different frame sentences representing various speech styles. This adaptation is further observed when listeners' responses are facilitated by the receipt of consistent speech style information, as opposed to inconsistent speech style information. Investigation of speaker and listener adaptation to situational contexts complements investigation of speech style and reduced speech. For Lindblom (1990), reduction is driven by adaptation. A speaker produces reduced speech as the result of an adaptation to a situational context, meeting the speaker's drive to economize effort. If a speaker receives negative feedback, i.e., the listener misunderstands the speech production, the speaker adapts to a less economic speech style to preserve clarity. Investigation of reduction processing can help researchers understand the perceptual and processing motivations that lead speakers and listeners to make speech style adaptations.

4.4 Models of spoken word recognition

H&H theory is an attempt to explain phonetic variation in production and perception. Many models of spoken word recognition have also been developed over the last 30 years, for example: Cohort (Marslen-Wilson and Welsh, 1978), Trace (McClelland and Elman, 1986), Shortlist (Norris, 1994), NAM (Luce and Pisoni, 1998), PARSYN (Luce et al., 2000) and ART (Grossberg, 1986; Grossberg and Myers, 2000). Psycholinguists are interested in spoken word recognition because it plays a central role in the communication process (McQueen, 2007). The goal of these models is to understand the processes and representation involved in the recognition of spoken words (Jusczyk and Luce, 2002). This section focuses on two issues under the domain of representation and addressed by all the above-mentioned models; levels of representation (direct versus mediated access) and nature of representation (Exemplars) (Jusczyk and Luce, 2002).

4.4.1 Levels of representation: Direct versus mediated access

Models of spoken word recognition can be divided into two groups based on whether they posit the existence of sublexical levels of the representation. Direct access models (e.g., Cohort, LAFS [Klatt, 1979]) contain no intermediate levels—the featural or spectral information from the speech stream is mapped directly to the lexical representation without interruption (Marslen-Wilson and Welsh, 1978; Marslen-Wilson and Warren, 1994; Klatt, 1979; Jusczyk and Luce, 2002). Mediated access models (Trace, Shortlist, NAM, PARSYN, and ART) contain prelexical representations (e.g., segmental information such as phonotactic probability [Vitevitch et al., 1997; Vitevitch and Luce, 1998] or syllabic information). Prelexical representation "acts as the interface between auditory and lexical processing" (McQueen, 2007:2). This prelexical level may consist of several levels (e.g. NAM and PARSYN) or just one intermediate level (e.g. Shortlist). The number of existing prelexical levels is beyond the scope of this discussion, which will instead focus on whether an intermediate level(s) is necessary at all. Intermediate levels may contain allophonic, phonemic, and syllabic information, which simplifies the processing load at the lexical level.

The results of the current experiments indicate that the processing of /g/ and flap are somewhat different. It is possible that these variations are due to competition between prelexical and lexical levels resulting in different outcomes for /g/ and flap, as would be seen by a main effect of the factor Phoneme in the statistical analysis. The direction of the main effect has two possible interpretations. Lexical frequency (flap words are more common than /g/ words) predicts that listeners will respond to flap targets more quickly and accurately than /g/ targets. Prelexical competition (flap has more possible phonemic and orthographic representations than /g/) would predict slower response times and more errors for flap targets. Competition between prelexical and lexical information may have varying effects in different tasks and with different segments.

Both Experiments 1 and 2 showed a difference between flap and/g/ in the percent error data. Listeners made more mistakes with the /g/ targets than they did with the flap targets. This finding indicates that lexical information is playing a greater role in the percent error responses, meaning that subjects make more errors due to the relatively low frequency of /g/ targets. Experiment 3 also showed this same effect: listeners generally made more mistakes for /g/ targets than for flaps. This effect was not found to be significant in Experiment 4, but the trend persisted. It is possible that the percent error data is showing an effect at the lexical level.

A post-hoc investigation of the response latency data shows that while there was no significant effect of Phoneme for Experiments 1 or 3, response times were slightly slower for flap targets. Experiment 2 showed mixed results: flap target response time was faster for unreduced targets, while /g/ target response time was faster for reduced targets. In this case, the reduced and unreduced /g/ targets were not statistically different. Generally speaking, response times that are slower for flap targets than /g/ targets indicate prelexical effects, i.e., access is inhibited by greater competition at the prelexical level. The percent error data indicate that lexical frequency effects at the lexical level are surfacing, while the response latency data indicate that prelexical effects are surfacing. Evidence for both lexical and pre-lexical competition is shown in current work, supporting models of mediated lexical access.

In some sense, the current experiments are similar to the phonetic mismatch experiments, which investigate mismatches between an input and stored lexical knowledge (e.g., Marslen-Wilson and Zwitserlood, 1989). For example, in Dutch a mismatch of the initial phoneme by changing the place of articulation, *honing* ‘honey’ to *woning* ‘dwelling’, prevented lexical access in a cross-modal form priming experiment of visual *bij* ‘bee’. Experiments investigating mismatches have shown that when mismatches occur, lexical access is prevented or inhibited (Alphen and McQueen, 2006; Streeter & Nigro, 1979; Whalen, 1984, 1991a, 1991b; Marslen-Wilson & Warren, 1994; Gow & Gordon, 1995; McQueen, Norris, & Cutler, 1999; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Davis, Marslen-Wilson, & Gaskell, 2002; Gow, 2002; Spinelli, McQueen, & Cutler, 2003). In this case, the match would be the unreduced stimulus and the mismatch would be the reduced stimulus. The mismatch is different in manner of articulation due to target undershoot (unreduced has a stop closure and reduced has a gesture toward the same place of articulation without closure and is thus more approximant-like). The phonetic mismatch is a naturally occurring phonetic mismatch caused by reduction. Responses to this mismatch indicate that listeners pay attention to segmental information. Changes in the representation inhibit lexical access (Experiments 1 and 3).

This match/mismatch comparison lends support for a prelexical level in which segments are identified and that information is passed on to the lexical level. The cross-modal identity priming experiment shows that both the reduced (mismatch) and the unreduced (match) targets are primed. This indicates that lexical access of both match and mismatch items is occurring, but the mismatch items are inhibited as compared to the match items. This result is similar to results found by Alphen and McQueen (2006), which show that changes in voice onset time of initial segments in Dutch can affect lexical access, but do not prevent access. The evidence from the Phoneme effects and the target mismatch effects supports models of lexical access that use mediated lexical access.

4.4.2 Nature of representation: Exemplars

Another major question in models of spoken word recognition is how lexical information is stored. One type of representation that has been adopted in some models of lexical access is based on episodic memory (Nosofsky, 1985). Episodic representation in speech (Hintzman, 1986; Goldinger, 1996b, 1998; Johnson, 1997) posits that a listener stores all acoustic information (exemplars) received for every speaker heard throughout the listener's entire life. For example, every instance of a flap ever perceived is stored in the listener's lexicon.

The findings in this dissertation do not support episodic models of lexical representation (Hintzman, 1986; Goldinger, 1996b, 1998; Johnson, 1997). An episodic representation would predict that all forms heard by a listener are stored in memory and that access occurs through comparison of stored lexical forms to the incoming stimulus. Based on previous discussion, most of the speech encountered by the average listener occurs in a spontaneous context, meaning that the average listener is presented with reduced forms of speech more often than unreduced forms. Therefore, the average listener would have more exemplars of reduced forms stored in their lexical representation. The exemplar cluster containing flaps would then have more exemplars of reduced flaps than unreduced flaps.

Data from Warner and Tucker support the assumption that reduced forms are more common than unreduced forms. If the occurrence of formants is used as an indication of reduced or more approximant-like productions (formant structure during the flap indicates little if no closure duration), then a comparison can be made of the number of sequences occurring with formants to the number of flaps without formants. The data illustrated in Figure 4.1, below, show that productions of flap with formants are more common than productions without formants. If weak formants and strong formants are taken together, the occurrence of formants is far more common than the absence of formants, regardless of speech style. By extension, reduced flaps are far more common than unreduced flaps in production.

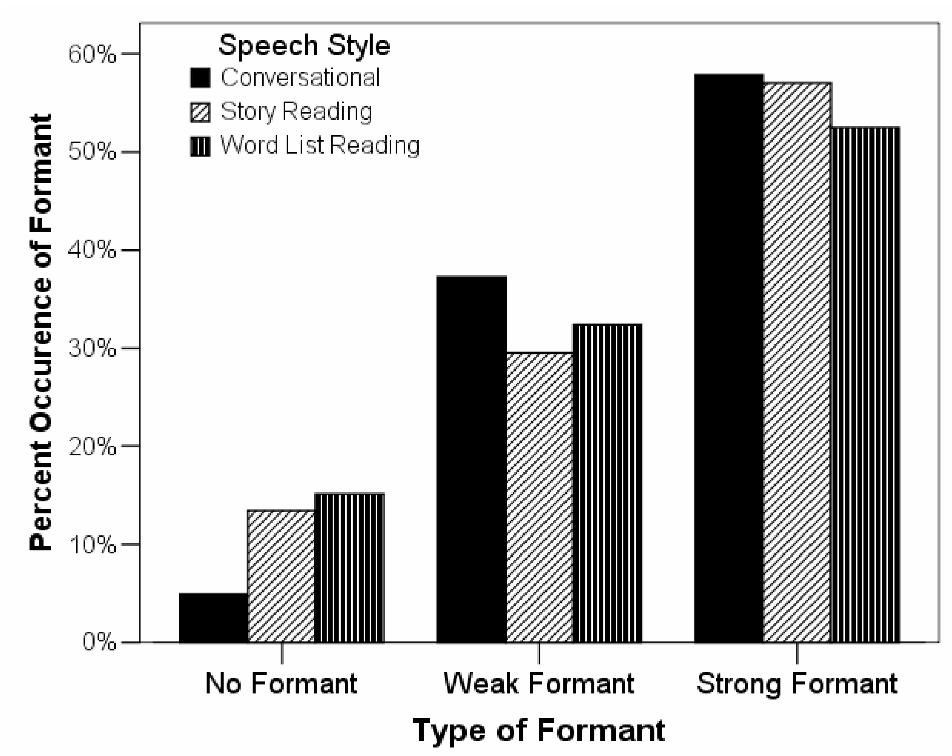


Figure 4.1 Percent occurrence of flaps containing formant structure in results from seven speakers (Warner and Tucker, 2007).

Similar to the pattern observed with lexical frequency effects, a more frequent type of exemplar will trigger lexical access more quickly than will a less frequent exemplar. As a result, and as described as one of the possible outcomes in section 1.1 above, listeners would respond to reduced word forms more quickly than unreduced word forms and with a higher degree of accuracy.

The results of Experiments 1-4 indicate that listeners respond more quickly to the unreduced items and that reduced items are more difficult to process. This finding does not support arguments for episodic representation. Instead, this finding indicates that acoustic information is more important than the presence or strength of episodic traces. The opposite prediction follows if it is assumed that some other form of representation is in use, that only the canonical form is stored (Gaskell and Marslen-Wilson, 1996, 1998), and that the unreduced target is of the canonical form. Under this prediction, reduced items would be more difficult to process than unreduced items because reduced items are acoustically further away from the canonical representation. The findings from the cross-modal identity priming experiment show that although reduced and unreduced targets both prime significantly, unreduced targets show more priming, indicating that they may be more representative of the canonical form.

4.5 Broader implications

The purpose of this section is to explore and introduce broader implications of the current research. These implications are discussed in four domains: speech pathology, second language acquisition and teaching, speech recognition and synthesis, and work with endangered languages.

4.5.1 Speech pathology

In the domain of speech pathology, the current research contributes to the understanding of how listeners process normal speech, which can then be compared to processing of disordered speech. This research also contributes to the body of literature on the processing of reduced speech, which can be explored in many aspects, including speech pathology. It is important for speech clinicians to be aware of the basics ways in which non-disordered human listeners perceive and process speech. As speech clinicians become more aware of these effects, they can use this knowledge to inform their interventions with patients. For example, Liss et al. (1998) argue that “speech intelligibility is central to the diagnosis, treatment and study of dysarthria.” Processing of reduced speech has particular relevance to dysarthria, a neurogenic disorder that causes patients to lose motor control of the speech apparatus, among many other things. Dysarthria can be caused by attacks on the central nervous system from such things as stroke, amyotrophic lateral sclerosis (ALS or Lou Gehrig's disease), Parkinson's disease, or Huntington's disease. ALS, for example, is a degenerative disease in which muscle tissue becomes severely weakened, leading eventually to increasingly reduced motor speech patterns (Shames and Anderson, 2002). As fine motor control skills are lost, patients have difficulty moving their articulators to speech targets and speech becomes more reduced.

Work by Liss et al. (1998, 2000, 2003; and Lansford et al. 2007) compare the effects of processing different types of dysarthric speech to that of normal speech. Dysarthia causes patients to produce speech that contains aspects similar to reduction due to target undershoot. Lansford et al.'s (2007) results are similar to the results found in this dissertation. Lansford et al. show that listeners' responses are slower and they make more mistakes for dysarthric speech as compared to normal speech. Further investigation of production and processing of reduced speech can contribute more information for comparison by researchers interested in patients experiencing dysarthria.

In the domain of audiology, listeners with hearing loss might show difficulty processing reduced speech. It has been demonstrated that use of clear speech improves speech intelligibility for hearing-impaired listeners (Picheney et al., 1985). As discussed in section 1.3, "clear speech" is a special speech style where particular emphasis is given to segmental and suprasegmental aspects to produce highly intelligible speech. It is possible that clear speech offers intelligibility benefits to listeners because one of the adjustments of clear speech is removal of reduction. Removing reduction would decrease the cognitive load for listeners and aid processing of the speech. Conversely, reduction may increase a listener's cognitive load, making it more difficult, if not impossible, to process the speech signal. Listeners experiencing hearing loss already have an increased cognitive load; addition of reduced speech would increase their cognitive load and the processing difficulty.

4.5.2 Second language acquisition and teaching

Nonnative contrasts pose a challenge for learners of a second language. For example, it is difficult for speakers of English to learn the Korean three-way distinction between /t/, /t*/ and /t^h/ (Kim, 1965; Kim and Duanmu, 2004). Research has shown second language learning is facilitated when learners are presented with a greater variability of productions of a non-native distinction (Kingston, 2003). For example, Japanese native speakers have difficulty learning English /r/ and /l/. Research has shown that stimulus variability facilitates the learning of nonnative distinctions (Logan et al., 1991; Lively et al., 1993, 1994). Kingston (2003) also showed that more variability aided English speakers in distinguishing German high and back distinctions not existent in English. The current work has shown that productions of the American English flap are highly variable and that reduced forms of the flap are more common than unreduced forms. Further, this work has shown that these reduced forms are more difficult to process. Learners of English as a second language who speak a language that does not have a [ɾ] or [d] in its inventory could have difficulty learning to distinguish between [ɾ], [d], and [t], particularly if the native language only contains [t] in its inventory. It is possible that learners of English as a second language would benefit from exposure to a large variety of flap productions. It is also possible that this variety would assist learners in acquiring flap as a phonological category.

Based on the results of this work, an investigation could be undertaken to discern how contextual speech style contributes to second language processing. Would contextual speech style facilitate processing in a similar manner as in Experiments 3? This leads to another question: how do second language listeners process reduction when it is frequent in daily conversation? Additionally, does reduction pose a problem for L2 listeners? Based on the results in the current work, a reasonable hypothesis might be that reduction inhibits processing by second language listeners. Furthermore, reduction is likely to prevent lexical access in L2 listeners, as opposed to simple inhibiting access experienced by L1 listeners.

4.5.3 Speech recognition and synthesis

The field of Automatic Speech Recognition (ASR) has already begun to investigate spontaneous speech. The results of the current research support the need for ASR systems to further investigate spontaneous speech and reduction. An understanding of the occurrences of reduction in particular situations may help ASR systems to better estimate the types of segments produced by a speaker by analyzing information in the speech stream to gauge the speech style. This approach may be similar to the way speech style cues aid human listeners in processing reduced speech. ASR systems generally require huge data sets to train the recognition system. It has historically been the case that these training data sets have contained careful read speech with very little variability (e.g., TIMIT, Wall Street Journal, Greenberg, 1998). After the training, the systems generally perform very well in recognition of this speech style. As discussed by Greenberg (1998), when these systems are then introduced to a corpus of spontaneous speech (Switchboard) accuracy levels fall to below 70%. The current work supports the need to include training data sets that include reduced speech.

Another application of this research might occur in designing speech synthesis systems. These results indicate that more careful speech will aid perception and processing of created speech. This means that the closer the speech created by the synthesis systems is to the abstract representation, the easier it will be to perceive and process by human listeners.

4.5.4 Endangered language work

The issues of speech style and speech reduction have indirect implications that endangered language workers should take into consideration. For example, a focus of this dissertation was the investigation of phonetic variation within the range of flap production and how it is processed. Phonetic variability is rampant throughout the languages of the world. However, most endangered language work, and in particular language documentation, rarely investigates phonetic variability. In fact, most documentation (i.e. language grammars) only briefly mentions the phonemic system before moving on to discussion and investigation of morphology, syntax, semantics, and pragmatics. This lapse in phonetic variability documentation decreases the general knowledge of phonetic variation worldwide. Amery (2000), in his work on Kaurna, discusses the presence of variation as a sign of the linguistic vitality of a language. Documentation and investigation of phonetic variation not only documents the vitality of the language, but also makes it possible for later learners of the language to observe the type of variation that existed. However, if phonetic investigation is generally missed in endangered language work, phonetic dialect differences will rarely surface. Even in phonetic documentation, the focus tends to be exclusively on careful speech or isolated tokens (cf. Ladefoged, 2003).

Along with phonetic variation, investigation and documentation of endangered language speech style must also be considered. In section 1.3 a discussion of speech style was given. There it was also shown that speech varied highly in production depending on the speech style that was produced. It is essential, not just for phonetic research, that workers involved in any kind of language documentation, consider documentation of as many speech styles as possible.[14] To do so will create, among other things, a more accurate account of the sound system of a language. This type of documentation can be helpful to advanced learners of a language or learners seeking to revitalize a language. Quite often, it is necessary for endangered language workers focusing on revitalization to specify a “right, pure” way to pronounce a word, which can benefit language acquisition as learners acquire the unreduced form. However, a healthy vital language contains a

[14] This also is highly dependent on the particular group the worker is working with and the number of speakers available. From one speaker it is possible to collect a range of styles, but conversational exchanges will be impossible.

large amount of phonetic variability that learners may eventually apply in more advanced stages of learning.

The vast majority of investigation of spoken word recognition has occurred on Indo-European languages with few examples of non Indo-European languages (e.g. Japanese, Chinese, Finnish, Korean, Sesotho). This paucity of diverse research is not limited to the field of spoken word recognition, but can be extended to psycholinguistic research generally. Many of these lesser-studied languages will exhibit processing strategies than have not been previously found (Cutler, 1985, 1996, 1997). Cutler (1997) has shown that by considering a large diversity of languages, a more complete understanding of spoken-word recognition develops. As endangered languages disappear, it becomes more difficult to investigate these lesser-studied languages.

4.6 Limitations of the current work

The difficulty facing all studies of spoken word recognition is that the researcher does not have direct access to the processing sub-system. Thus, researchers are limited to indirect procedures of investigating the perception and processing of speech. The current work collected response latency and percent error data from listeners as they responded to auditory or visual probes. Both experimental designs required listeners to respond by deciding whether the stimulus was a word or a non-word. Listeners rarely encounter non-words in their daily routine. When new words are encountered, listeners assume that it contains some semantic content. The current work is interested in the listeners' responses to real words; responses to the non-words did not play a role in the analysis. It is possible that because the task was unnatural in some regards that it affected the results.

As previously mentioned, the tasks used provide indirect data about processing. In an ideal situation with no limitations, access to the actual firing of individual neurons in particular locations of the brain would provide more direct processing data. A map of the brain and the time course of neuron activity during speech processing could be created. Comparisons could be made between this time course of neuronal firing and variations in the speech stream. Of particular interest to the current work would be to see whether the map and time course of neuron activity was the same for reduced and unreduced speech. It is currently possible gather this brain data, though it is very limited and of low resolution. In an ideal situation, the experiment could be masked so that the listener is unaware of any data collection, whether participating in a conversation with a friend or in more controlled conditions as auditory stimuli are presented.

Another limitation of the current work is that the main focus has been on reduction, which occurs in spontaneous speech. As above, it is recognized that the methodology does not replicate real conversational processing. Also, the stimuli produced are still carefully controlled, though with care to replicate the occurrence of these forms in spontaneous speech styles. Following is a description of an experiment currently underway that investigates processing of more naturally occurring speech.

4.7 Future research

The results of these experiments have shown a distinction in the way listeners process spontaneous versus laboratory speech. The results have also shown that listeners adjust their expectations based on preceding speech style information. The results of these experiments lead to many more questions.

A follow-up study is planned using reduced and unreduced flap stimuli in a phoneme monitoring task to supplement the current lexical decision and cross-modal identity priming tasks. This study will more directly test listeners' processing of sounds rather than words and will be designed to further investigate the current set of research questions posed in this dissertation. This paradigm might also provide interesting results supporting the distinction between the processing of flap and /g/. A phoneme monitoring task prevents lexical effects from influencing the responses because listeners are monitoring for a phoneme in a stream of speech sounds that do not contain any lexical information. However, segment duration of the targets could differ, possibly contributing to any effect found.

A second follow-up study will further investigate the effect of contextual speech styles. In the context experiments, listeners reported that the experiment was "boring" and that they had a hard time concentrating. In many cases, listeners reported that they stopped paying attention to the preceding frame sentence (even in the experiments containing 13 frame sentences) despite being verbally instructed to pay attention to what they heard. This follow-up study would shorten the experiment by leaving out the /g/ stimuli. It is then expected that the effect of speech style will be stronger.

A third follow-up study is designed to take advantage of the ambiguity of certain items when they are reduced, as noted in Chapter 2. In this study, target items will be ambiguous, i.e., 'regal' could become 'real' if the /g/ is sufficiently reduced. If a listener hears the reduced 'regal' in a cross-modal identity priming task, then would a difference in identity priming for visual stimuli 'regal' and 'real' occur? It is hypothesized that there would be more priming for the ‘regal’ target then for the ‘real’ target. This is due to minor phonetic cues indicating that the auditory stimulus is ‘regal’. The first half of this study would focus on targets in isolation; the second half would introduce speech style, and a listener would hear reduced and unreduced frame sentences. This might actually strengthen the priming effect, meaning that in the ‘regal’ visual target with reduced context there would be more priming. However, the ‘real’ visual target with the unreduced frame might show more priming. In this study, it is possible that competition between weak phonetic cues and speech style (reduced vs. unreduced) would occur to produce very interesting results.

The work by McLennan, Luce, and Charles-Luce (2003) discussed previously also shows effects of mediated lexical access. McLennan et al. (2003) use flaps, [t]s, and [d]s in their auditory stimuli for match and mismatch priming (where [t] and flap are mismatches). It is possible that the fact that they used [t] in a flap position unduly influenced their results. Patterson and Connine (2001) in their corpus study of flaps found that 96% of the items are flaps and the remaining 4% consists of articulated [t] and glottal stop. Connine (2004), as discussed in Chapter 1, claims that listeners do not recode /t/ to flap, but that the lexical representation must contain a flap. If the underlying representation for a flap variant is a flap, then it is possible that the use of [t] and [d] actually produced results not representative of the underlying representation.

A fourth follow-up study is planned, which will replicate the McLennan et al. (2003) paradigm, but will use stimuli that are either reduced or unreduced. It is possible that this study will replicate the McLennan et al. (2003) results, which would further support the need for mediated lexical access. It is also possible that it would contradict the McLennan et al. (2003) results. This would mean that priming would be facilitated for “careful,” unreduced tokens and inhibited for “casual” reduced tokens. This contradictory result would further support the results found in the current work. However, a contradiction of their results would still support a model which uses mediated lexical access.

In Warner & Tucker (2006) two acoustic characteristics, “F4 swoosh” (Dungan et al., 2007) and intensity dip (the change in intensity during a flap), were identified as characteristic of flaps. A fifth study would manipulate the F4 swoosh and intensity difference to identify how these acoustic cues contribute information in identifying flaps. A phoneme identification task will be used for this experiment. Items are VV sequences where, inserted between the vowels will be a C that contains variations of F4 and intensity. Variations of the flap will be synthesized using a Klatt synthesizer. Both F4 (from flat [VV] to large dip [VCV]) and degree of intensity dip or intensity difference (from none [VV] to a level defined by pilot data [VCV]) will be manipulated. Two continua will be created: one that has ambiguous, constant intensity dip and manipulates F4, and another that has ambiguous, constant F4 dip and manipulates intensity.

A fifth set of experiments are planned which investigate the effect of word frequency on reduction. Footnotes 3 (section 2.1.5) and 5 (section 2.2.4) indicate that frequency may play a role in the processing of reduced speech. While the results of the analyses in footnotes 3 and 5 are inconclusive additional research investigating the relationship between reduction and frequency is necessary. This experiment will use the auditory lexical decision and cross-modal identity priming tasks as in the experiment described in Chapters 2 and 4. This experiment will introduce new stimuli which in addition to being controlled for reduction will also be controlled for frequency.

The speech stimuli in this dissertation research, though they replicated reduced speech, were controlled and produced in an experimental manner. If other speech styles were recorded and then used as targets, how would listeners respond? An experiment is currently in progress that uses speech from a spontaneous telephone conversation. Phrases from this conversation were excised. The same speaker was then invited back to read the excised phrases both in a careful and in a more casual manner (Mehta and Cutler, 1988). Using a cross-modal identity priming task it is possible that an effect of the spontaneous speech may arise. Additionally, other tasks such as auditory lexical decision and phoneme monitoring could be done using this same speech recording.

4.8 Final conclusions

Phonetic variation as found in various speech styles is a rich area for research on spoken word recognition. Most research on spoken word recognition has focused on what has been defined as laboratory speech. This work has shown that when phonetic variation, as it occurs in various speech styles, is investigated, differences in processing occur. This finding emphasizes the importance of research in processing and perception—and more generally, in psycholinguistics and phonetics—to investigate phonetic variation across and within speech styles.

The findings in this dissertation indicate more specifically that processing of reduced speech is inhibited by weakened acoustic information or mismatch to the underlying phonemic representation in the American English flap. Listeners use information about speech style to process the widely varying acoustic reflections of a segment in connected speech.

APPENDIX A STIMULI AND FILLERS FROM EXPERIMENT 1.

Lexical Decision Target Items:

Flap /d/ items

1. audit
2. bloody
3. body
4. border
5. buddy
6. cedar
7. cheddar
8. credit
9. cuddle
10. daddy
11. edit
12. fiddle
13. freedom
14. harder
15. huddle
16. hurdle
17. idle
18. judo
19. ladle
20. lady
21. middle
22. model
23. modern
24. moody
25. muddy
26. murder
27. needle
28. needy
29. odor
30. order
31. poodle
32. puddle
33. ready
34. saddle
35. spider
36. steady
37. study
38. sturdy
39. tidy
40. waddle

/g/ items

1. baggy
2. beggar
3. bogey
4. buggy
5. dagger
6. doggie
7. eager
8. eagle
9. fogey
10. foggy
11. giggle
12. goggles
13. google
14. groggy
15. gurgle
16. haggard
17. haggle
18. jogger
19. juggle
20. legal
21. logger
22. maggot
23. muggy
24. pagan
25. ragged
26. regal
27. rigor
28. rugged
29. saga
30. shaggy
31. smuggle
32. soggy
33. stagger
34. straggle
35. sugar
36. tiger
37. toga
38. trigger
39. vigor
40. yoga

Real word fillers (in English orthography)

1. badger
2. button
3. cable
4. copy
5. derby
6. dozen
7. driver
8. failure
9. flannel
10. kitty
11. office
12. passage
13. pencil
14. sorrow
15. stingy
16. story
17. ticket
18. volume
19. vowel
20. zipper

Non–word fillers (in IPA transcription)

1. aɪpi
2. baɪbɚ
3. banɪf
4. bɛzi
5. bibɪk
6. bidu
7. blavl̩
8. blɪki
9. bogə
10. boʊpl̩
11. brɛpi
12. bugo
13. bukɚ
14. buri
15. bʌbi
16. dæpə
17. dari
18. dibɚ
19. dibo
20. dræmpɪk
21. dufɚ
22. dʌpɚ
23. dʌsi
24. ʃigl̩
25. ʃodɚ
26. storpɚ
27. ɛpɚd
28. ɛsl̩
29. fædl̩
30. fɛʃɪt
31. fɛlpɚ
32. fetʃɪk
33. fɪpɚ
34. figɪd
35. fləmɪt
36. flɛntl̩
37. floɪmək
38. folbi
39. fotɪs
40. frænl̩
41. frarbl̩
42. fʌtʃɪs
43. fwapl̩
44. gado
45. gædʒɚ
46. gæmɪk
47. gapi
48. gaʊdɪk
49. gaʊmɚ
50. gɛvi
51. gifɚ
52. gɪpl̩
53. giθu
54. glʌzɚ
55. goɪdɪz
56. graθɪn
57. grʌml̩
58. gufə
59. guzəd
60. gʌbl̩
61. hagl̩
62. hɛgi
63. himɛk
64. hitʃl̩
65. hoʊdɪk
66. hundʒɪk
67. ɪdl̩
68. inəl
69. isət
70. jæfɚ
71. jaʃrən
72. japi
73. jatl̩
74. jɛlki
75. jimi
76. jʌbɚ
77. kabi
78. kæbl̩
79. keɪməs
80. kɛti
81. kiði
82. kigɚ
83. kinl̩
84. klæzɚ

85. kogə
86. koʊpɛm
87. koʊpl̩
88. kugə
89. kusə
90. kʌʃl̩
91. kʌtʃi
92. kwapl̩
93. lebi
94. leɪgɚ
95. lɚmɚ
96. lipl̩
97. lirdɚ
98. liti
99. loɪpɚ
100. luʃi
101. lukəm
102. luwɪʃ
103. lʌdi
104. lʌgl̩
105. mævl̩
106. mɪvi
107. muksi
108. muθɪdʒ
109. mʌnəm
110. nænɪd
111. naɪzɚ
112. nalɪk
113. neɪzɚ
114. nɛskɚ
115. nokl̩
116. nopə
117. nudʒl̩
118. nʌndl̩
119. orfəd
120. oʊksl̩
121. oʊndʒl̩
122. pabə
123. pagi
124. pɚgl̩
125. pɪdʒl̩
126. pɪmɚ
127. pɪndl̩
128. pirə
129. pɪtʃl̩
130. pɪvɚ
131. plʊpi
132. podɚ
133. poɪθli
134. priki
135. prʌnʃŋ
136. pugɚθ
137. pulə
138. rætʃi
139. rapi
140. rapoʊ
141. ribl̩
142. rotəm
143. rudʒik
144. sɛdɪk
145. sidl̩
146. sirgl̩
147. skaʊpi
148. slido
149. smudɚ
150. spifəʃ
151. stæpi
152. stani
153. stɪdʒɚ
154. stiʃəm
155. stɪkok
156. strati
157. struʃi
158. stʌldʒi
159. subi
160. subl̩
161. subŋ
162. susl̩
163. sʌnl̩
164. swukl̩
165. tægl̩
166. tagə
167. tʃælɚ
168. tʃɛmpli
169. tɛpəf
170. tʃupɚθ
171. tʃʌni
172. tidl̩
173. tilə
174. tipɚ
175. toʊvɪp
176. traɪmɚ
177. tramzi
178. tranə
179. trɛðɪk
180. trɪgl̩
181. trobɚ
182. trʌli
183. tukə
184. tʌndi
185. vemɚ
186. vɛti
187. vidɚ
188. ʌmpl̩
189. voɪkɚ
190. vʊdl̩
191. vupɚ
192. wæskɚ
193. wapɪt
194. warli
195. wekl̩
196. wɪdʒɚ
197. wɪlmɪk
198. zirɪg
199. θɛml̩
200. θrubl̩

APPENDIX B

Table B.1 Example lexical decision items with the total number of items occurring for /d/ data.

Frequencies for /d/ data					**Reduced**			**Unreduced**	
Number		872			38			40	
		Intensity Diff	Duration (s)		Intensity Diff.	Duration (s)		Intensity Diff.	Duration (s)
Mean		10.5836	0.0328		3.6331	0.0200		16.4332	0.0342
Percentile	10	4.0944	0.0506		1.5878	0.0278		12.2760	0.0447
	20	6.2920	0.0430		2.4271	0.0247		14.4292	0.0410
	30	7.9718	0.0379		2.7326	0.0225		14.7414	0.0369
	40	9.2258	0.0343		3.0192	0.0221		15.6442	0.0354
	50	10.4548	0.0307		3.2550	0.0194		16.4737	0.0325
	60	11.5259	0.0276		3.6313	0.0177		17.3885	0.0319
	70	12.6861	0.0240		4.2949	0.0164		17.7891	0.0302
	80	14.3296	0.0213		4.9201	0.0154		19.2077	0.0282
	90	16.9999	0.0175		6.4498	0.0134		20.8193	0.0248

Table B.2 Example lexical decision items with the total number of items occurring for /g/ data.

Frequencies for /g/ data					**Reduced**			**Unreduced**	
Number		457			39			40	
		Intensity Diff.	Duration (s)		Intensity Diff.	Duration (s)		Intensity Diff.	Duration (s)
Mean		17.3544	0.0602		5.5696	0.0278		21.0464	0.0550
Percentile	10	9.1386	0.0808		3.3000	0.0367		16.7977	0.0745
	20	11.6507	0.0737		4.0176	0.0330		18.5326	0.0670
	30	13.2894	0.0668		4.4953	0.0292		19.4484	0.0607
	40	14.7653	0.0631		5.2317	0.0277		19.8561	0.0586
	50	16.4476	0.0595		5.5992	0.0258		20.3864	0.0554
	60	17.9347	0.0555		6.1412	0.0244		21.4301	0.0502
	70	20.3603	0.0516		6.5177	0.0225		22.9428	0.0446
	80	23.0329	0.0450		6.7141	0.0210		23.6645	0.0412
	90	26.4989	0.0395		6.9024	0.0191		25.9642	0.0383

APPENDIX C STIMULI AND FILLERS FROM EXPERIMENT 1. NON–WORDS ARE IN ENGLISH ORTHOGRAPHY FOR VISUAL PRESENTATION.

Target Items

	Visual	Auditory	Auditory Control
1.	cuddle	cuddle	hayseed
2.	muddy	muddy	confess
3.	fiddle	fiddle	lasso
4.	credit	credit	belief
5.	edit	edit	incense
6.	middle	middle	beyond
7.	idle	idle	romance
8.	daddy	daddy	airy
9.	cedar	cedar	washer
10.	modern	modern	living
11.	bloody	bloody	rumor
12.	border	border	gather
13.	needy	needy	ashore
14.	cheddar	cheddar	bursar
15.	steady	steady	agent
16.	odor	odor	arrow
17.	ladle	ladle	pebble
18.	order	order	given
19.	audit	audit	armor
20.	hurdle	hurdle	granite
21.	murder	murder	region
22.	spider	spider	molten
23.	waddle	waddle	brownie
24.	needle	needle	orbit
25.	model	model	permit
26.	puddle	puddle	bracket
27.	body	body	across
28.	harder	harder	genius
29.	buddy	buddy	attach
30.	lady	lady	carry
31.	doggy	doggy	hermit
32.	beggar	beggar	torture
33.	tiger	tiger	revolt
34.	sugar	sugar	divine
35.	dagger	dagger	grumpy
36.	legal	legal	evil
37.	gurgle	gurgle	tiller
38.	rugged	rugged	essay
39.	muggy	muggy	wearer
40.	eagle	eagle	kidney
41.	vigor	vigor	python
42.	juggle	juggle	vomit
43.	foggy	foggy	prophet
44.	groggy	groggy	hockey
45.	straggle	straggle	gusto
46.	smuggle	smuggle	parcel
47.	jogger	jogger	casket
48.	ragged	ragged	devise
49.	eager	eager	excuse
50.	haggard	haggard	brochure
51.	google	google	mature
52.	maggot	maggot	daisy
53.	soggy	soggy	rover
54.	yoga	yoga	robot
55.	logger	logger	neutron
56.	goggles	goggles	coupon
57.	shaggy	shaggy	courtly
58.	toga	toga	fossil
59.	stagger	stagger	fable
60.	trigger	trigger	fabric

Filler Items

Phonological Overlap

	Visual	Auditory
1.	buker	cooker
2.	challer	caller
3.	chunny	choppy
4.	cutchy	patchy

5.	darry	starry
6.	faddle	fodder
7.	futchis	fascist
8.	gado	giddy
9.	gluzzer	laser
10.	goppy	gaudy
11.	hemered	custard
12.	koople	topple
13.	leerder	martyr
14.	lirmer	firmer
15.	luggle	toggle
16.	mavel	level
17.	mooksy	music
18.	nizer	miser
19.	nokel	snorkel
20.	nundle	handle
21.	nuppy	sappy
22.	pergle	purple
23.	pimmer	steamer
24.	piver	liver
25.	poithly	poison
26.	poola	cooler
27.	rasker	whisker
28.	roppy	reaper
29.	seedle	beetle
30.	seergle	gargle
31.	shoder	shorter
32.	sooben	pseudo
33.	stonny	stopper
34.	teeper	sneaker
35.	trauna	fauna
36.	umple	sample
37.	voiker	porker
38.	warly	early
39.	widger	pitcher
40.	yeemy	creamy

No overlap

	Visual	**Auditory**
41.	beedoo	bramble
42.	biber	burrow
43.	blicky	crabby
44.	bogga	diaper
45.	bonif	never
46.	clazzer	public
47.	couga	appear
48.	drampick	local
49.	dupper	money
50.	felper	value
51.	flentel	common
52.	gadger	pony
53.	gipple	color
54.	gowmer	nation
55.	loiper	running
56.	luddy	proper
57.	nallick	council
58.	pluppy	ballot
59.	poggy	demand
60.	preeky	borrow
61.	prunshin	offer
62.	quapel	crisis
63.	skoppy	valley
64.	speefish	pleasure
65.	stappy	explain
66.	steeshem	career
67.	stickock	jury
68.	sussle	humor
69.	taga	royal
70.	thoobel	honest
71.	touvip	talent
72.	traimer	rebel
73.	veeder	fever
74.	vooper	forum
75.	vudel	parlor
76.	wapit	rubber
77.	yashren	flavor
78.	yattle	soften
79.	yelcky	shallow
80.	yubber	sector
81.	frarble	concept

82.	noodgel	open
83.	orphad	simple
84.	pidgel	network
85.	flummit	budget
86.	gubble	shower
87.	ratchy	domain
88.	sleedoe	mango
89.	stidger	elder
90.	swookle	earring
91.	teela	cotton
92.	tepeph	hornet
93.	trully	cowboy
94.	yaffer	happy
95.	beebick	timber
96.	bezzy	pistol
97.	blavel	urban
98.	boogoe	sudden
99.	boorry	goalie
100.	bople	project
101.	breppy	theory
102.	chemply	normal
103.	chooperth	trophy
104.	copem	lizard
105.	dappa	mighty
106.	deeber	mortar
107.	doofer	rifle
108.	essel	future
109.	folby	lyric
110.	fuapple	mirror
111.	geefer	critic
112.	grathin	ballet
113.	heemek	kayak
114.	heetchel	moment
115.	hoedick	missile
116.	easot	nosy
117.	keddy	justice
118.	keeger	apple
119.	keennel	provide
120.	koosa	college
121.	kushel	puzzle
122.	layger	assort
123.	leedy	gamble
124.	leweesh	envoy
125.	lipple	acid
126.	looshy	shelter
127.	lukem	window
128.	mivy	regret
129.	moothidge	social
130.	munnem	cactus
131.	nanid	forest
132.	oaksel	curfew
133.	oengel	earthy
134.	pobba	buffer
135.	peera	pellet
136.	photis	figure
137.	pindel	founder
138.	pitchel	fellow
139.	poogerth	flipper
140.	roppoe	mascot
141.	rodem	tunnel
142.	sedick	modest
143.	sheegel	axis
144.	shorper	label
145.	smooder	bigot
146.	sooby	willow
147.	straddy	salad
148.	strooshy	central
149.	stuldgey	lemon
150.	sunnel	scandal
151.	teedel	rabbit
152.	themmel	blooper
153.	tramzy	machine
154.	triggle	arrest
155.	trober	basic
156.	tundy	factor
157.	vetty	paper
158.	weckel	treaty
159.	wilmick	vision
160.	zeerig	docent

Real–Word Fillers

	Visual	Auditory
161.	caper	badger
162.	noodle	button
163.	pigeon	cable
164.	fibber	copy
165.	swagger	derby
166.	orphan	dozen
167.	teaser	driver
168.	sleeper	failure
169.	hobby	faucet
170.	tender	flannel
171.	duster	kitty
172.	stacker	office
173.	marble	passage
174.	anal	pencil
175.	rabble	sorrow
176.	plumber	stingy
177.	bubble	story
178.	ratchet	ticket
179.	gravy	volume
180.	humble	vowel

Practice Items

	Visual	Auditory
1.	triggle	arrest
2.	belly	belly
3.	chubber	biscuit
4.	fuppy	funny
5.	harket	powder
6.	muddle	muddle
7.	solcor	popcorn
8.	wiggle	wiggle
9.	zipper	zipper

REFERENCES

Alphen, P. van and J. M. McQueen. (2006). The effect of Voice Onset Time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performancem*, 32, 178-196.

Amery, Rob. (2000). *Warrabarna Kaurna! Reclaiming an Australian Language*. Lisse, The Netherlands. Swets & Zeitlinger Pulishers.

Arai, T. (1999). 'A case study of spontaneous speech in Japanese', *Proc. of the International Congress of Phonetic Sciences (ICPhS)*, Vol. 1, pp. 615-618, San Francisco.

Bard, E. G., R. C. Shillcock and G. T. M. Altmann. (1988). The recognition of words after their acoustic offsets in spontaneous speech: effects of subsequent context. *Perception & Psychophysics*, 44(5), 395-408.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America,* 113, 1001-1024.

Bölte, J., & Coenen, E. (2002). Is phonological information mapped onto semantic information in a one-to-one manner? *Brain and Language, 81*, 384–397.

Byrd, Dani. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39-54.

Charles-Luce, J. (1997). Cognitive factors involved in preserving a phonemic contrast. *Language & Speech*, 40, 3, 229-248.

Cheng, Chin-chuan. (1968). *Mandarin Phonology*. University of Illinois Ph.D. dissertation.

Coenen, E., P. Zwitserlood, and J. Bölte. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* 16, pp. 535–564.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, 11(6), 1084-1089.

Cutler, Anne. (1985). "Cross-language psycholinguistics". *Linguistics*, 23, 659-667.

Cutler, Anne. (1996). The comparative study of spoken-language processing. *Fourth International Conference on Spoken Language Processing*, plenary lecture. http://www.asel.udel.edu/icslp/cdrom/vol1/1015/a1015.pdf

Cutler, Anne. (1997). The comparative perspective on spoken-language processing. *Speech Communication*, 21, 3-15.

Cutler, Anne. (1998). The recognition of spoken words with variable representation. In *Proceedings of the ESCA Workshop on Sound Patterns of Spontaneous Speech* (pp. 83–92). Aix–en–Provence: ESCA.

Cutler, Anne., & T. Otake. (1998). Assimilation of place in Japanese and Dutch. In *Proceedings of the Fifth International Conference on Spoken Language Processing* (pp. 1751-1754). Sydney: ICLSP.

Dahan, D., J. S. Magnuson, M. K. Tanenhaus, & E. M. Hogan. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507-534.

Dalby, J. (1986). *Phonetic Structure of Fast Speech in American English*. Indiana University Linguistics Club.

Davis, M. H., W. D. Marslen-Wilson, & M. G. Gaskell. (2002). Leading up the lexical garden-path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218-244.

de Jong, K. (1998). Stress-related variation in the articulation of coda alveolar flaps: Flapping revisited. *Journal of Phonetics,* 26, 283-310.

Dejean de la Bâtie, B., & D. C. Bradley. (1995). Resolving word boundaries in spoken French: Native and nonnative strategies. *Applied Psycholinguistics*, 16, 59-81.

Diehl, Randy L. and Margaret Walsh. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *The Journal of the Acoustic Society of America*, 85, 2154-2164.

Dressler, Wolfgang. (1972). *Allegoregeln rechtfertigen Lenteregeln. Sekundäre Phoneme des Bretonischen*. Innsbruck, Innsbrucker beiträge zur Sprachwissenschaft 9.

Dungan, Mary, Karen Morian, Benjamin V. Tucker and Natasha Warner. (2007). Fourth formant dip as a correlate of American English flaps, Presented at the *153rd Meeting of the Acoustical Society of America*, (Salt Lake City, UT).

Dupoux, E., C. Pallier, K. Kakehi, & J. Mehler. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, 5(16), 491-505.

Dupoux, E., K. Kakehi, Y. Hirose, C. Pallier, & J. Mehler. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1568-1578.

Eimas, P. D., and J. L. Miller. (1980). Contextual effects in infant speech perception. *Science*, 209, 1140-1141.

Ernestus, M. (2000). Voice Assimilation and Segment Reduction in Casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface. LOT, Utrecht.

Ernestus, M., R. H. Baayen, & R. Schreuder. (2002). The recognition of reduced word forms. *Brain and Language*, 81, 162–173.

Ernestus, Mirjam, Mybeth Lahey, Femke Verhees, and R. Harald Baayen. (2006). Lexical frequency and voice assimilation. *The Journal of the Acoustic Society of America*, 120, 1040–1051.

Fosler-Lussier, E., S. Greenberg, and N. Morgan. (1999). Incorporating Contextual Phonetics into Automatic Speech Recognition. ICPhS-99, San Francisco, CA, August 1999.

Fukaya, T., & D. Byrd. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association,* 35, 45-58.

Gaskell, M. G., and D. Marslen–Wilson. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144–158.

Gaskell, M. G., and D. Marslen–Wilson. (1998). Mechanisms of phonological inference in speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 380–396.

Godfrey, J., J. McDaniel, J. Holliman. (1992). SWITCHBOARD: a telephone corpus for research and development. *International Conference on Acoustics, Speech and Signal Processing (ICASSP) Proceedings*, San Francisco. pp. 517-520.

Goldinger, S. D. (1996a). Auditory lexical decision. *Language and Cognitive Processes*, 11, 559–567.

Goldinger, S. D. (1996b). 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, 251-279.

Gow D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28, 163-179.

Gow, D. W., and P. C. Gordon. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 344-359.

Greenberg, S. (1998). Recognition in a new key - Towards a science of spoken language. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP) Proceedings*. Seattle, pp. 1041-1045.

Greenberg, S. (1999). Speaking in shorthand - A syllable-centric perspective for understanding pronunciation variation. *Speech Communication.* 29 pp. 159-176.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab and H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines, Vol. 1: Speech perception*. New York: Academic Press, pp.187-294.

Grossberg, S. and C. W. Myers. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review.* 107, 735-767.

Hammond, Michael. (1999). *English Phonology*, Oxford University Press, Oxford.

Harris, James W. (1969). *Spanish Phonlogy*. Cambridge, MA. MIT Press.

Harris, James. (1983), Syllable Structure in Spanish: A Nonlinear Analysis. Cambridge, MA. MIT Press.

Hintzman, Douglas L. (1986). 'Schema Abstraction' in a multiple-trace memory model. *Psychological Review*. 93:328-338.

Holt, L. L. (2005). Temporally non-adjacent non-linguistic sounds affect speech categorization. *Psychol. Sci*. 16: 305–312

Hooper, J. B. (1978). Constraints on Schwa-Deletion in English. in *Recent Developments in Historical Phonology*, J. Fisiak (ed.) pp. 183-207. De Gruyter. Berlin.

Horna, Josephine. (1998). An Investigation into the Acoustics of American English Flaps, with a Secondary Emphasis on Spanish Flaps, in Fluent Speech. New York. Ph.D. Dissertation New York University.

Idsardi, W. J. and E-K. Sung. (2003). Flaps in American English and Korean: An Acoustic and Perceptual Study. *Generative Grammar in a Broader Perspective. Proceedings of the 4th GLOW in Asia.* Seoul: Hankook. 187-208.

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

Johnson, Keith. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.) *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*. Tokyo, Japan: The National International Institute for Japanese Language. pp. 29-54.

Jurafsky, D., A. Bell, M. Gregory, & W. Raymond. (2000). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In Bybee, Joan and Paul Hopper (eds.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Jusczyk, P. W., & P. A. Luce. (2002). Speech Perception and spoken word recognition: Past and present. *Ear & Hearing*, 23, 2-40.

Kahn, Daniel. (1976). *Syllable-based Generalizations in English Phonology*. Bloomington. Ph.D. dissertation Indiana University.

Kim, Chin-Wu. (1965). “On the Autonomy of the Tensity Feature in Stop Classification (with Special Reference to Korean Stops),” *Word* 21, 339–359.

Kim, Mi-Ryoung, and San Duanmu. (2004). “Tense” and “Lax” stops in Korean. *Journal of East Asian Linguistics*. 13, 59-104.

Kingston, John. (2003). Learning Foreign Vowels. *Language and Speech*. 46(2-3), 295-349.

Kiparsky, Paul. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry*, 10, 421–441.

Kirchner, Robert. (1998). *An Effort-Based Approach to Consonant Lenition*, Los Angeles, CA. Ph.D. Dissertation University of California, Los Angeles.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279-312.

Koster, C. J. (1987). *Word Recognition in Foreign and Native Language*. Foris, Dordrecht.

Kučera, H. and W. Francis. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Kuijpers, C. T. L., W. A. van Donselaar, and Anne Cutler. (1996). Phonological variation: Epenthesis and deletion of schwa in Dutch. In H.T. Bunnell, & W. Idsardi (Eds.), *Proceedings of the 4th International conference on spoken language processing (ICSLP) 1996*, pp. 94-97. New Castle DE: Citation Delaware.

Kuijpers, C., W. van Donselaar, and Anne Cutler. (1998). Phonological variation: Epenthesis and deletion of schwa in Dutch. *Proceedings of the Fourth International Conference on Spoken Language Processing*. Pp. 94-97. Philadelphia.

Labov, William. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Ladefoged, P. and D. E. Broadbent. (1957). Information conveyed by vowels. *Journal of the Acoustic Society of America.* 29, 98–104.

Ladefoged, Peter. (2003). *Phonetic data analysis: an introduction to instrumental phonetic fieldwork*. Oxford: Blackwells.

Ladefoged, Peter. (2006). *A course in phonetics (6th edition)*. Thomson/Wadsworth.

Lansford, Kaitlin, Julie Liss, Stephanie Spitzer, Sven Mattys, John Caviness. (2007). Word recognition in dysarthric speech: Evidence for the time-course hypothesis. Presentation at the Semiannual Conference of the Acoustic Society of America, Salt Lake City.

Liberman, A. M., P. C. Delattre, L. J. Gerstman, and F. S. Cooper. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of & Experimental Psychology*, 52, 127-137.

Lindblom, B. (1990). 'Explaining phonetic variation: a sketch of the H and H theory.' In Hardcastle, W. and Marchal, A. (editors), *Speech Production and Speech Modeling*, 403-439. Dordrecht: Kluwer Academic Publishers.

Liss, J. M., S. M. Spitzer, J. N. Caviness, and C. A. Adler. (2000). Lexical boundary decisions in the perception of hypokinetic and ataxic dysarthric speech. *Journal of the Acoustical Society of America*. 107, 3415-3424.

Liss, J. M., S. M. Spitzer, J. N. Caviness, and C. A. Adler. (2003). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America*. 112, 3022-3030.

Liss, J. M., S. M. Spitzer, J. N. Caviness, C. A. Adler and B. Edwards. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America.* 104, 2457-2466.

Lively, S. E., D. B. Pisoni, R. A. Yamada, Y. Tohkura, & T. Yamada. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96,* 2076-2087.

Lively, S. E., J. S. Logan, & D. B. Pisoni. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242-1255.

LoCasto, P. C. & C. M. Connine. (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics*, 64(2), 208-219

Logan, J. S., S. E. Lively, & D. B. Pisoni. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

Luce, P. A., & D. B. Pisoni. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1-36.

Luce, P. A., S. D. Goldinger, E. T. Auer, and M. S. Vitevitch. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62, 615-625.

Luce, P.A. (1986). Neighborhoods of words in the mental lexicon. Bloomington, IN. Ph.D. dissertation, Indiana University.

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

Marslen-Wilson, W. D. (1980). Speech understanding as a psychological process. In J.C. Simon (Ed.), *Spoken Language understanding and generation*. pp. 33–67. Dordrecht: Reidel.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altman (Ed.) *Cognitive models of speech processing*. pp. 148-147. Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & P. Zwitserlood. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585.

Marslen-Wilson, W., A. Nix, & G. Gaskell. (1995). Phonological variation in lexical access: abstractness, inference and English place assimilation. *Language and Cognitive Processes*, 10, 285–308.

Marslen-Wilson, W.D., & P. Warren. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101, 653–675.

Matter, J. F. (1986). A la recherché des frontiers oerdues. Utrecht. Ph.D. Dissertation University of Utrecht.

McAllister, J. (1991). The processing of lexically stressed syllables in read and spontaneous speech. *Language and Speech*, 34, 1-26.

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

McLennan, C. T., P. A. Luce, and J. Charles-Luce. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 539-553.

McQueen, J. M., D. Norris, & A. Cutler. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1363-1389.

McQueen, James & Taehong Cho (2003). The use of domain-initial strengthening in segmentation of continuous English speech. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*. Barcelona, Spain.

McQueen, James. (2007). Eight questions about spoken-word recognition. To appear in G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.

Mehta, G. and Anne Cutler. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech*, 31, 135-156.

Miller, J. L., and Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition*, 13, 135-165.

Miller, J. L., and Liberman, A. M. (1979). Some effects of later occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457–465.

Miller, J. L., and Voltaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics,* 46, 505–512.

Mitterer, H. (2006). Is Vowel Normalization Independent of Lexical Processing? *Phonetica*, 63, 209-229.

Mitterer, H. and M. Ernestus. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/–lenition in Dutch. *Journal of Phonetics,* 34, 73–103

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189-234.

Norris, D., A. Cutler, J. McQueen, and S. Butterfield. (2006). Phonological and conceptual activation is speech comprehension. *Cognitive Psychology*, 53, 146–193.

Nosofsky, R. (1985), Overall Similarity and the Identification of Separable-Dimension Stimuli: A Choice Model Analysis. *Perception & Psychophysics*, 38, 415-432.

Oostdijk, N. H. J. (2000). Building a corpus of spoken Dutch, In: P. Monchesi (ed.), *Computational Linguistics in the Netherlands 1999. Selected papers from the tenth CLIN meeting*. 147-158 Utrecht: Utrecht Institute of Linguistics, OTS.

Otake, T., K. Yoneyama, A. Cutler, and A. van der Lugt. (1996). The represebntation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, 100, 3831-3842.

Patterson, D. & Connine, C. M. (2001). A corpus analysis of variant frequency in American English flap production. *Phonetica*, 58(4), 254-275.

Patterson, D., LoCasto, P.C. & Connine, C. M. (2003). A corpus analysis of schwa vowel deletion frequency in American English. *Phonetica*, 60, 45-68.

Pérez, Patricia E. (1992). Gradient Sonority and Harmonic Foot Repair in English Syncope. *Coyote Papers: University of Arizona Working Papers in Linguistics.* 8, 118-142.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (1992). Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements, *Journal of the Acoustical Society of America,* 92, 3078-3096.

Picheny, M.A., Durlach, N.I., and Braida, L.D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28, 96-103.

Pickett, J. M. and Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*, 6, 151-164.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics*, 34, 314-322.

Pitt, M.A., Johnson, K., Hume, E., Kiesling, S. and Raymond, W. (2003). The ViC corpus of conversational speech. Manuscript submitted to *IEEE Transactions on Speech and Audio Processing: Special Issue on Spontaneous Speech Processing*.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America* 118 (4), p. 2561–2569.

Pollack, I., & Pickett, J. M. (1963). Intelligibility of excerpts from conversational speech. *Language and Speech*, 6, 165-171.

Racine, I. and Grosjean, F. (1997). La reconnaissance des mots en parole continue: Effacement du schwa et frontière lexicale. In: *La voyelle dans tous ses états. Actes des journées d'études linguistiques*. Nantes, 1997. Université de Nantes: Groupe AAI, Nantes.

Riehl, Anastasia K. (2003). American English Flapping: Perceptual and Acoustic Evidence Against Paradigm Uniformity with Phonetic Features. *Working Papers of the Cornell Phonetics Laboratory*, 15, 271-337.

Selkirk, E. (1982). The syllable. In Van der Hulst, H. & Smith, N. (eds.), *The Structure of Phonological Representations* (Part 2), 337–383. Dordrecht: Foris Publications.

Seneff, S. and V. Zue. (1988). "Transcription and alignment of the TIMIT database," *Getting Started With the DARPA TIMIT CD-ROM: An Acoustic Phonetic Continuous Speech Database*, J. S. Garofolo, Ed. Gaithersburg, MD: NIST.

Shames, George H. and Noma B. Anderson, Editors. (2002). *Human Communication Disorders: An Introduction* (6th edition). Boston: Allyn and Bacon.

Smiljanić, R. and Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America*, 118 (3), 1677-1688.

Smiljanić, R. and Bradlow, A. R. (2006). Word and Syllable Boundary Strengthening in English Clear Speech. *10th Conference on Laboratory Phonology*, Paris, France, June, 2006.

Smiljanić, R. and Bradlow, A. R. (in press). Stability of temporal contrasts across clear and conversational speech. *Journal of Phonetics.*

Spinelli, E., McQueen, J.M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, 48 (2), 233-254.

Steriade, Donca. (2000). Paradigm uniformity and the phonetics-phonology boundary. In Broe, M. B. & Pierrehumbert, J. B. (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 313–334. Cambridge: Cambridge University Press.

Streeter, L. A., & Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America*, 65, 1533–1541.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1074-1095.

Swinney, D., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition*, 7, 159–165.

Turk, A. E. (1992). The American English flapping rule and the effect of stress on stop consonant durations. *Cornell Working Papers in Phonetics* 7, 103–133.

van der Lugt, A. (1999). *From speech to words* (MPI Series in Psycholinguistics No. 13). Doctoral dissertation, University of Nijmegen, Nijmegen, the Netherlands.

Vitevitch, M. S. and Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science, 9*, 325-329.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*, 47-62.

Vroomen, J., & De Gelder, B. (1999). Lexical access of resyllabified words: Evidence from phoneme monitoring. *Memory and Cognition*, 27, 413-421.

Warner, Natasha and Benjamin V. Tucker. (2007). Categorical and Gradient Variability in Intervocalic Stops. Talk given at the annual meeting of the Linguistic Society of America.

Wescott, Roger W. (1965). Speech-Tempo and the Phonemics of Bini. *Journal of African Linguistics*. 4(3), 323-336.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics, 35,* 49-64.

Whalen, D. H. (1991a) Infrequent words are longer in duration than frequent words. *The Journal of the Acoustical Society of America* 90, 2311.

Whalen, D. H. (1991b). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics, 50,* 351-360.

Whalen, D. H., Catherine T. Best, and Julia Irwin. (1997. Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics 25,* 501-528.

Zue, V. and Lafferriere, M. (1979). Acoustical study of medial /t/, /d/ in American English. *The Journal of the Acoustical Society of America* 66, 1039–1050.

Zue, V., J. Glass, D. Goodine, H. Leung, M. Phillips, J. Polifroni, and S. Seneff. (1990). Recent progress on the summit system. In *Proceedings of the Third DARPA Speech and Natural Language Workshop*, pages 380-384.

Zwicky, Arnold M. (1972a). On Casual Speech. In Paul Peranteau, Judith N. Levi and Gloria C Phares (eds) *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*. 607-615.

Zwicky, Arnold M. (1972b). Note on a phonological hierarchy in English. in R. Stockwell and R. Macaulay, eds. Linguistic Change and Generative Theory. Indiana University Press.

Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes*, 11(6), 589-596.

Zwitserlood, P., and Coenen, E. (2000). Consequences of assimilation for word recognition and lexical representation. Presented at the SWAP workshop, Nijmegen, 1999.