Variability in Word Duration as a Function of Probability, Speech Style, and Prosody

Rachel E. Baker, Ann R. Bradlow

Northwestern University, Evanston, IL, USA

Key words

accenting, duration

clear speech

lexical frequency

second mention reduction

Abstract

This article examines how probability (lexical frequency and previous mention), speech style, and prosody affect word duration, and how these factors interact. Participants read controlled materials in clear and plain speech styles. As expected, more probable words (higher frequencies and second mentions) were significantly shorter than less probable words, and words in plain speech were significantly shorter than those in clear speech. Interestingly, we found second mention reduction effects in both clear and plain speech, indicating that while clear speech is hyper-articulated, this hyper-articulation does not override probabilistic effects on duration. We also found an interaction between mention and frequency, but only in plain speech. High frequency words allowed more second mention reduction than low frequency words in plain speech, revealing a tendency to hypo-articulate as much as possible when all factors support it. Finally, we found that first mentions were more likely to be accented than second mentions. However, when these differences in accent likelihood were controlled, a significant second mention reduction effect remained. This supports the concept of a direct link between probability and duration, rather than a relationship solely mediated by prosodic prominence.

1 Introduction and previous work

1.1
Introduction

Lindblom (1990) noted that words can be pronounced along a continuum from hyper-articulation to hypo-articulation. Hyper-articulation involves pronouncing words

Language and Speech

more clearly than they are normally pronounced, and is associated with various acoustic-phonetic features of enhanced speaker effort, such as longer durations and larger vowel spaces. Hypo-articulation involves pronouncing words less clearly than normal, and can involve features such as shorter durations, reduced vowel spaces and dropped phonemes. When and how speakers hyper- and hypo-articulate has been the topic of much recent research. For example, researchers have studied the effects of lexical probability[1] (e.g., Anderson & Howarth, 2002; Aylett & Turk, 2004, 2006; Fowler & Housum, 1987; Jurafsky, Bell, Gregory, & Raymond, 2001) and listener-oriented speech style modifications (e.g., Bradlow, 2002; Picheny, Durlach, & Braida, 1986; Smiljanic & Bradlow, 2005, 2008; Uchanski, 2005) on articulation level (the degree of hyper- or hypo-articulation). However, when viewed in combination, these studies raise some intriguing questions. How do potentially opposing probabilistic factors, such as lexical frequency and earlier mention in the discourse, interact? Do probabilistic effects on articulation level behave differently in different speech styles? Is there a direct link between lexical probability and articulation level, or is their relationship entirely mediated by prosodic prominence (as proposed by Aylett & Turk, 2004, 2006)? In this study we attempt to answer these questions.

1.2
Previous work on probabilistic effects and speech style

The main goal of oral communication is to pass information from the speaker to the listener. Lindblom (1990) points out that for this task to be successful, the listener must distinguish the speaker's actual words from all the other words he could have said. Lindblom's hyper- and hypo-articulation (H&H) theory states that the listener uses both the speech signal itself and knowledge of their language and the world (Lindblom's "signal complementary processes") to solve this problem. Therefore the speaker only needs to articulate clearly enough to ensure that the listener will be able to distinguish his/her intended words from other words, given the signal-independent information already at the listener's disposal. For example, there is more signal-independent information about the final word in (1) than in (2) (from Lieberman, 1963).

(1)   A stitch in time saves nine.

(2)   The number that you will hear is nine.

According to the H&H theory, the listener's knowledge of the saying (1) means that the speaker can hypo-articulate when pronouncing *nine* in this context because very little acoustic information is needed to distinguish this word from other possibilities. The most efficient way of speaking is to track the predicted signal-independent contribution and increase articulatory effort only in those cases when the signal-independent contribution is low. In addition, Lindblom divides constraints on the speech system into reception/output constraints and production/system constraints.

---

1   Lexical probability is determined by a number of factors, including how frequently a word is used in the language, and whether it has already been used in the discourse. Words that have been used recently are more likely than other words with similar meanings to be used again later in the discourse.

When reception constraints dominate, speakers produce hyper-speech, and when production constraints dominate, speakers produce hypo-speech. This idea captures the effects of speech style on articulation level. The interplay between perception and production constraints can be seen in a number of current models of speech production, including Stem-ML (Kochanski & Shih, 2001), the CHAM model (Oviatt, MacEachern, & Levow, 1998), van Son and van Santen's (2005) model of redundancy and articulation, Matthies, Perrier, Perkell, and Zandipour's (2001) study of the effects of speech style and rate on coarticulation, the Probabilistic Reduction Hypothesis (Jurafsky et al., 2001), and the Smooth Signal Redundancy Hypothesis (Aylett & Turk, 2004, 2006).

Although Lindblom (1990) never mentioned probability, the idea was implicit in his theory. A word's probability depends on signal-independent factors, such as lexical frequency and earlier use in the discourse. This probability influences where along the hyper-/hypo-articulation continuum it is pronounced (Aylett & Turk, 2004, 2006; Jurafsky et al., 2001). Jurafsky et al. offer the Probabilistic Reduction Hypothesis to describe the relationship between probability and articulation level. This hypothesis claims that word forms are reduced when they have a higher probability of occurrence. This concept is a component of the H&H theory because a higher probability means more signal-independent information, and therefore fewer constraints on the signal itself. The fact that there are fewer constraints on the signal allows the speaker to use less effort during the articulation of the word, leading to hypo-articulation. According to the Probabilistic Reduction Hypothesis, probability can be determined by neighboring words, syntactic and lexical structure, semantic factors, discourse factors (such as previous mention in the discourse), and frequency factors. Jurafsky et al. have examined a number of ways in which reduction is realized, including vowel centralization, final /t/ and /d/ deletion, and duration.

Aylett and Turk (2004, 2006) propose the Smooth Signal Redundancy Hypothesis to explain the relationship between a word's probability and its articulation level. They claim that two opposing constraints affect the care with which speakers articulate: producing robust communication, and efficiently expending articulatory effort. These constraints are analogous to Lindblom's (1990) reception and production constraints, respectively. In the Smooth Signal Redundancy Hypothesis, competition between the goals of communicating effectively and expending effort efficiently leads to an inverse relationship between an element's redundancy and the care with which speakers articulate it. In other words, less probable elements are articulated more carefully to increase the chance that they will be understood. This idea is equivalent to the Probabilistic Reduction Hypothesis, yet the Smooth Signal Redundancy Hypothesis goes one step further in proposing that speakers try to maintain smooth signal redundancy, or a roughly equal chance that each element will be understood. If a word is highly predictable from the preceding context and a speaker's pronunciation of it is relatively short, there is less information about the word in the speech stream itself, but more information in the preceding context. While having smooth signal redundancy as a goal is unique to Aylett and Turk's theory, smooth signal redundancy is a by-product of competition between the constraints in the H&H theory. Aylett and Turk claim that speakers maintain smooth signal redundancy because it is efficient and it ensures that the necessary amount of information is transmitted

in a noisy environment. They provide evidence based on syllable durations (Aylett & Turk, 2004) and vowel formants (Aylett & Turk, 2006) to support the Smooth Signal Redundancy Hypothesis.

A key distinguishing feature of the Smooth Signal Redundancy Hypothesis is that it claims that speakers use prosodic prominence to regulate smooth signal redundancy. For example, if a word is highly predictable from its context, a speaker would be more likely to de-accent this word, making it shorter than it would be if it were accented. In contrast, the Probabilistic Reduction Hypothesis does not mention prosodic prominence, and therefore allows a direct connection between a word's probability and its articulation level. According to the Smooth Signal Redundancy Hypothesis, the observed imperfect relationship between probability and prosodic prominence is a result of "both the indirect way in which redundancy influences the acoustic signal and learned, language-specific conventions about stress place-ment" (Aylett & Turk, 2004, p.34). It is important to note that in this theory prosodic prominence covers vowel reduction as well as phrasal and lexical stress. In addition, the relationship between probability and prosodic prominence can either be an online process or arise from a historical development in the language. One example of such a development is the tendency in English to put lexical stress on the first syllable of a word, which is the least predictable syllable. Aylett and Turk distinguish between reduced and full vowels, lexically stressed and unstressed syllables, and nuclear and non-nuclear phrasal stress. They claim that probability should not provide a unique contribution to a model explaining variance in articulation level, but rather that its contribution should be covered by the effects of prosody on articulation level. They found that the majority of the variance in syllable duration in their dataset that was accounted for by probability was also accounted for by prosody. However, they still found a significant independent contribution from probability. They also found a unique contribution of probability in models explaining vowel formant variance (Aylett & Turk, 2006). Van Son and van Santen (2005) argue against this aspect of the Smooth Signal Redundancy Hypothesis based on their observation of a correlation between consonant classes' normalized durations and the frequency of each class in a particular position within the word. This correlation was found in both stressed and unstressed positions. So consonants that were more predictable in some positions were shorter in those positions even after controlling for stress.

Speech style also plays a role in a speaker's choice of an articulation level between hyper- and hypo-articulation. Speakers use different speech styles in response to different listening conditions. When speakers believe listeners will not have trouble perceiving their speech, they tend to use a plain speech style in which they "globally" hypo-articulate for ease of articulation. However, when speakers believe their listeners might have difficulty perceiving their speech, they usually try to speak more clearly by "globally" hyper-articulating (for reviews of the clear speech research enterprise see Uchanski, 2005, and Smiljanik & Bradlow, 2009). Although global, utterance-level speech style at first seems unconnected to local, word-level probability, the two factors can be viewed as the same effect acting at different levels. A speaker's estimation of word probability is not independent of the communicative context; it is conditioned on the signal-independent information available to the listener. If a word's probability is low, the speaker must put more information in that word's signal

in order to communicate it effectively. Similarly, speech style is chosen based on a speaker's knowledge of his/her listener and the listening conditions. If the listener is a non-native speaker of the language, he/she brings less signal-independent knowledge to the conversation, so more information must be put in the signal itself. If a listener is hard of hearing, the speaker knows the signal being interpreted will be degraded, so he/she must compensate for this by speaking more clearly. In the non-native speaker situation there is less signal-independent information available throughout the entire dialogue, and in the hard of hearing situation the overall level of signal-information needs to be higher than it would normally be.

Although there is a sizable body of research on the effects of probability and speech style on articulation level, few studies have examined how such factors interact with each other. It is possible that each factor plays an equal role in the final articulation of a word. But it is also possible that some factors are more influential than others, so a stronger factor might nullify the impact of a weaker factor. As a word can be highly probable according to one factor (e.g., lexical frequency) and highly improbable according to another factor (e.g., conditional probability based on the preceding word), these factors can work in opposite directions, potentially canceling each other out. If they are both working in the same direction, their effects may be additive, multiplicative, or one effect may be much larger than the other, hiding the effect of the weaker factor. Moreover, the general requirements for more information in the signal at a global level (i.e., the requirements that promote the use of a clear speaking style) could override local probabilistic effects, such as the effects of lexical frequency, conditional probability, and previous mention. Jurafsky et al. (2001) simultaneously investigated the effects of lexical frequency, conditional probability of the word given the following word, conditional probability of the word given the preceding word, and the joint probability of the word and its preceding word. However, they did not directly examine whether one probability factor increased or decreased the effects of any of the other probability factors.

In this study we use lexical frequency and previous mention in the discourse as measures of a word's local probability and we vary speaking style (plain versus clear) as a means of manipulating global hypo-/hyper-articulation. We then examine the combined effects of these factors, namely lexical frequency, previous mention, and speaking style, on word duration as an index of articulation level (i.e., hypo-/hyper-articulation).

A number of studies have shown that higher frequency words tend to have shorter durations (Aylett & Turk, 2004; Bell et al., 2002; Jurafsky et al., 2001). Jurafsky et al. found that high frequency words were 18% shorter than low frequency words, a difference that was highly significant. Bell et al. studied a number of factors affecting a word's probability, including conditional and joint probabilities with previous and following words, semantic relatedness, and repetition, and found that lexical frequency had the strongest individual effect on word duration after all the other factors had been accounted for. In their study, high frequency words were 20% shorter than low frequency words. Aylett and Turk found that syllables in high frequency words had significantly shorter durations than those in low frequency words, even after controlling for the number of phonemes in the syllable.

Second mention reduction is another example of speakers reducing more predictable words. When English speakers repeat a word in a discourse, the second mention tends to be reduced (shorter and less intelligible) relative to the first mention (Fowler & Housum, 1987). Fowler (1988) showed that this effect is not simply articulatory priming, as it does not appear for words primed by a homophone in paragraphs, or for repeated words in word lists. This effect appears relatively robust, and second mention reduction has been found even when the second mention is produced by a different speaker than the first mention (Anderson & Howarth, 2002). Speakers also produce less intelligible second mentions of words even when they know that the listener has changed since the speaker produced the first mention of the word (Bard et al., 2000). However, there are still some situations in which second mention reduction is not produced. Bard, Lowe, and Altmann (1989) provide evidence that second mention reduction occurs when the two mentions refer to the same entity, but not when the second mention refers to a new entity of the same sort as the first mention. In addition, when Fowler, Levy, and Brown (1997) asked participants to describe a television show, they found second mention reduction within a description of a single scene, but not when the two mentions appeared in descriptions of two different scenes separated by metanarrative statements such as "in the next scene."

Both frequency and mention are word-level effects that license hypo-articulation for more predictable words. In contrast, clear speech is a discourse-level effect that requires hyper-articulation. Clear speech has been shown to be more intelligible than plain speech for multiple listener populations including normal hearing, hearing impaired, elderly, non-native speaker, and children with and without learning impairments (Chen, 1980; Helfer, 1998; Picheny, Durlach, & Braida, 1985). The acoustic-phonetic features of clear speech when compared to plain speech are numerous and affect almost all the dimensions known to be important for speech production and perception. These include both temporal and spectral dimensions at segmental and suprasegmental levels. Recent cross-language work has shown cross-language similarities and differences indicating that clear speech production is guided by both general, auditory-perceptual factors and language-specific, phonological-structural factors (Smiljanic & Bradlow, 2005, 2008). Clear speech can involve significantly longer speech sound durations than plain speech (Picheny et al., 1986; Smiljanic & Bradlow, 2008). Specifically, vowels in stressed syllables in clear speech tend to be longer than their counterparts in plain speech (Bradlow, 2002; Picheny et al., 1986; Smiljanic & Bradlow, 2008). Unvoiced stops also have longer voice onset times (VOTs) in clear speech than in plain speech (Chen, 1980; Picheny et al., 1986; Smiljanic & Bradlow, 2008). In addition, clear speech tends to have less alveolar flapping, (Bradlow, Krause, & Hayes, 2003; Picheny et al., 1986), fewer instances of stop burst elimination, (Bradlow et al., 2003; Picheny et al., 1986), and less reduction of unstressed vowels to schwas (Picheny et al., 1986; Smiljanic & Bradlow, 2008).

The current study examines how two word-level probabilistic factors, frequency and mention, interact with each other to determine a word's articulation level as realized through word duration. We also look at word durations for identical materials produced under clear and plain speech conditions. We specifically examine whether word-level probabilistic effects are the same or different in the two speech styles.

Finally, we look at whether the connection between probability and articulation level is direct or indirect (mediated through variations in prosodic prominence). This question is broken down into two parts: does probability affect prosodic prominence, and does probability affect articulation level when prosodic prominence is controlled?

1.3
Predictions

We expect to replicate earlier findings of shorter durations for higher frequency words, second mentions, and words produced in plain speech than for lower frequency words, first mentions, and words produced in clear speech, respectively. In addition, this study examines how probabilistic factors interact with each other and with clear speech. Three hypotheses regarding clear speech are examined. All three predict that durations in clear speech should be longer than in plain speech, but make different predictions regarding how frequency and second mention reduction affect a word's articulation level in clear speech.

1.3.1
Maximum Hyper-articulation Clear Speech Hypothesis: Clear speech is maximally hyper-articulated

In this case, clear speech should nullify other factors that affect articulation level in plain speech, including frequency and second mention reduction. Under this scenario, clear speech would appear to operate at a higher level than probabilistic effects, so general, auditory-perceptual considerations would override the linguistic-structural factors that operate at the discourse and lexical levels.

1.3.2
Many Factors Clear Speech Hypothesis: Clear speech is just one of many factors affecting articulation level

In this case, a number of factors should affect articulation level in clear speech, including frequency and second mention reduction. Under this scenario, clear speech would appear to operate at a level where general, auditory-perceptual considerations are integrated with linguistic-structural factors from the discourse and lexical levels.

1.3.3
Maximum Discourse Information Clear Speech Hypothesis: The goal of clear speech is communicating maximum information about the discourse history, not hyper-articulation

In this case, second mention reduction should appear (and possibly even be enhanced) in clear speech, because the distinction between first and second mentions of words communicates discourse information to the listener. However, there is no useful information for the current discourse history in the distinction between words with high and low frequencies of usage in the language, so lexical frequency effects on articulation level should be lost. Under this scenario, as with the "many factors clear speech hypothesis" above, clear speech would appear to operate at a level where general, auditory-perceptual considerations and linguistic-structural factors are integrated. However, in this case, clear speech interacts with linguistic-structural factors from the discourse level but not with those from the lexical level.

In addition to studying the interactions between probabilistic factors and speech style, this experiment examines how probabilistic factors affecting articulation level interact with each other.

1.3.4
Interaction Hypothesis 1: Probabilistic factors have additive effects on articulation level

This hypothesis predicts no interactions between second mention reduction and frequency effects. High frequency words should undergo no more or less second mention reduction than low frequency words. In this scenario, all probabilistic factors that affect articulation level are separate. Their interaction is simply the result of the fact that they affect the same acoustic dimensions (e.g., duration and vowel space).

1.3.5
Interaction Hypothesis 2: Probabilistic factors have interactive effects on articulation level

This hypothesis predicts that the articulation level of a word cannot be determined by adding up the effects of each probabilistic factor, but instead, the effects of one factor could be increased or decreased by another factor. These interactions could appear in clear speech, plain speech, or both. In this scenario, a word's probability is treated holistically. In other words, if multiple factors make a word probable, it is easier to predict than if it is probable according to one factor (e.g., lexical frequency) and improbable according to another (e.g., preceding context). Therefore those words that are probable by multiple factors can be hypo-articulated more than words that are probable by one factor but improbable by another (interacting) factor.

2 Methods

2.1
Participants

Six students at Northwestern University, USA (three male and three female) ranging in age from 21 to 49 participated in this experiment. Each was paid $5 for his or her participation. All were native speakers of American English, and none had any reported speech or hearing impairment. Only one participant reported being bilingual in English and another language (French) although his language background indicated a strong English dominance.

2.2
Stimuli

Five paragraphs containing 59 repeated mentions of words were written for the experiment. These paragraphs appear in Appendix A. The paragraphs range from 6 to 12 sentences long, with an average length of 8.6 sentences. They were designed to ensure that the repeated mentions of words appeared in equivalent phonetic and prosodic contexts. A number of entire phrases (e.g., *beets and string beans*) were repeated, so the words contained in these phrases could appear in identical or near-identical contexts. As most punctuation marks are accompanied by prosodic phrase breaks

(Taylor & Black, 1998), both mentions of each word appeared in identical positions relative to periods. Both mentions were either sentence-medial or sentence-final. Both mentions also almost always appeared in identical positions in relation to commas, so both members of a pair were either non-adjacent to any punctuation, immediately preceding a comma, immediately following a comma, or sentence-final.

Many of the target words contain point vowels (/i/, /a/, /u/), allowing for future analyses of the vowel space area. The repeated words include nouns, verbs, adjectives, pronouns, determiners, prepositions, and conjunctions. The frequencies of the target words were taken from the British National Corpus (BNC). The BNC is a 100 million word corpus consisting of samples of written and spoken British English from a variety of sources (*British National Corpus*, 2007). The target words range in frequency from four (*meet*-N) to 2,886,105 (*of*), with a mean of 130,268.9 and a median of 1694. All target words and their frequencies are listed in Appendix B. The distance between the two mentions of target words ranged from four to 156 words.

2.3
Procedure

Participants were told that they would be reading five paragraphs twice, in two different speech styles. Half the participants read all the paragraphs in clear speech first, and half read them in plain speech first. The plain speech instructions stated: "Please read the paragraphs as if you are talking to someone familiar with your voice and speech patterns, like a friend." The clear speech instructions stated: "Please read the paragraphs very clearly, as if you are talking to a listener with a hearing loss, or to a non-native speaker learning your language." Before each paragraph, participants were reminded of the speech style they were trying to achieve. Every participant read the paragraphs in a different order, but the order of paragraphs was the same in the two speech styles for each participant.

Recordings were made in a soundproof booth on an AKG C420 Headset Cardioid Condenser Mic. They were stored as.wav files and analyzed using Praat (Boersma & Weenink, 2004).

2.4
Duration measurements

All duration measurements were made by the first author, RB. Particular acoustic features, such as the start of frication or a stop burst, were chosen to mark the start and end of each word. These start and end points were marked on a Praat text grid, and a Praat script calculated the target word durations from this text grid. A second labeler (MB) measured a subset of the target words to check the reliability of the duration measurements. The subset included 182 target words, nearly a quarter of all 742 target words in the analysis. The reliability checking subset consisted of eight paragraphs, with examples from each of the six speakers and each of the five paragraph types. No speaker or paragraph was included more than twice. Half of the paragraphs in the subset were spoken in a clear speech style, and half in a plain style. Pairs of words in the two sets differed by an average of 17.3 ms. The correlation between the sets was 0.96, which was highly significant, $t(180) = 46.47$, $p < .0001$.

2.5

Disfluencies

All paragraph recordings containing major disfluencies (repetitions of phrases or halting speech throughout the paragraph), or disfluencies on or around a target word were removed from the analysis. To maintain equivalence between the clear and plain conditions, both versions of any unusable paragraph were removed. For example, Speaker 3 repeated the phrase "when Bobbie skied near enough" in her plain reading of Paragraph 2. As this phrase contains the target words *Bobbie* and *skied*, both her clear and plain readings of Paragraph 2 were removed from the analysis. This measure was taken because it has been shown that words in disfluent contexts tend to have longer durations than words in fluent contexts (Bell et al., 2003). It was important to minimize the participants' familiarity with the paragraphs to encourage them to treat the first mention of each word as a true first mention. The drawback of this is that participants produced a large number of disfluencies, which resulted in the loss of data. In total, 14 of the 30 paragraphs were removed from the analysis. Because the same paragraphs were removed for the same speakers in clear and plain speech, there are matched datasets for the two speech styles. One participant had all of his paragraphs retained, and one participant had all but one of her paragraphs removed. All other participants fell between these two extremes. Each paragraph had usable recordings from at least two speakers, but no paragraph had usable recordings from every speaker.

2.6

Reduction ratios

Degree of second mention reduction is difficult to compare across speech styles because of the generally longer word durations associated with clear speech. Greater duration differences are expected in clear speech because the actual word durations are greater. To deal with this problem, ratios of each word's first mention duration divided by its second mention duration were used to analyze the amount of reduction in clear and plain speech.

2.7

Prosodic analysis

Prosodic breaks and the presence of pitch accents on target words were determined by the first author, RB, after listening to the recordings and examining their waveforms, spectrograms, and F0 contours using Praat. Breaks with a ToBI break index of 3 or 4 (intermediate or intonational phrase breaks) were counted as prosodic breaks. A second labeler (JG), naïve to the purposes of the study, carried out the same prosodic analysis on the subset of the data used for duration measurements reliability checking. The two researchers agreed on the accents for 162 out of 182 target words, resulting in 89% agreement on the presence or absence of pitch accents on target words. The two researchers agreed about prosodic break context for only 63% of the target words. JG was more likely to posit prosodic breaks than RB. However, they agreed on whether the first and second mention break context matched for 80% of the target words. Agreement on whether the contexts match is more important for this study because the break data were only used to eliminate words for which the different mentions were produced in different break contexts (e.g., one was followed by a break and another was not).

Table 1

Word duration statistics (in milliseconds) averaging over all speakers, by speech style and mention (Clear 1 = clear speech, first mention, Clear 2 = clear speech, second mention, Plain 1 = plain speech, first mention, Plain 2 = plain speech, second mention, Clear Ratio = first mention duration divided by second mention duration in clear speech, Plain Ratio = first mention duration divided by second mention duration in plain speech)

	Clear 1	Clear 2	Plain 1	Plain 2	Clear Ratio	Plain Ratio
$n = 59$ (word tokens in each condition)						
Mean	404	380	322	306	1.074	1.073
Median	406	372	318	298	1.063	1.071
Std. dev.	136	130	122	123	0.126	0.144
Min.	142	112	95	76	0.838	0.739
Max.	740	695	614	614	1.463	1.628

3 Results

3.1
Replications

Two-tailed paired Wilcoxon signed rank tests, pooling across speakers, were run on the duration data. As predicted, comparisons of clear and plain speech showed that durations in the clear speech condition were significantly longer than those in the plain speech condition for first mentions, $W = 0$, $p < .0001$, and second mentions, $W = 51$, $p < .0001$. Also as predicted, comparisons of first and second mention durations showed significantly longer durations for first mentions in both speech styles (plain speech: $W = 1361.5$, $p < .0005$, clear speech: $W = 1392$, $p < .0001$). The significant second mention reduction and the significantly longer durations found in clear speech can be seen in Table 1.[2] Individual analyses found significant second mention reduction for four out of six speakers in plain speech and four out of six speakers in clear speech, $p < .05$. They also found significantly longer durations in clear speech for both first and second mentions for all speakers, $p < .05$.

3.2
Reanalysis accounting for unequal phrasing

3.2.1
Background

It is possible that some of the second mention reduction effect in clear speech is due to the fact that clear speech generally has more prosodic breaks than plain speech. In this

---

2 These effects of second mention reduction and clear speech reported for the main dataset also appeared in the subset of measurements performed by MB: clear speech for first mentions: $U = 576$, $p < .0005$; clear speech for second mentions: $U = 554$, $p < .0005$; Second mention reduction in clear speech: $W = 739$, $p < .05$; Second mention reduction in plain speech: $W = 758.5$, $p < .01$.

experiment, speakers produced an average of 18.83 prosodic breaks per paragraph in clear speech, while they only produced an average of 13.25 in plain speech. A one-tailed paired Wilcoxon signed rank test, averaging over speakers, showed this difference to be significant, $W = 15$, $p < .05$. Phrase-final lengthening before prosodic breaks is a well-studied phenomenon (Klatt, 1975), and Bell et al. (2002) found longer durations for utterance-initial and -final words than for utterance-medial words. It is possible that participants were more careful about distinguishing between the speech styles at the beginning of each paragraph than at the end, when they might have slipped into their natural style of read speech. The combination of more prosodic breaks in clear speech and shifting speech styles could lead to more phrase-final lengthening at the beginning of clear speech paragraphs than at the end. Some of the target words would be affected by this phrase-final lengthening, resulting in an inflated second mention reduction effect in clear speech.

In order to eliminate this possibility, the duration measurements were reanalyzed after removing the data for words with mentions appearing in different prosodic contexts. For each fluent paragraph, each speaker produced four mentions of every target word (Clear1, Clear2, Plain1, and Plain2). Each mention was coded for prosodic context as (1) preceded and followed by a break, (2) only preceded by a break, (3) only followed by a break, or (4) not adjacent to a break. The duration data for a speaker was only included in a word's average durations if that speaker produced all four mentions of the word in the same prosodic context. For example, Speaker 5 put prosodic breaks after both mentions of *beets* in clear speech but after neither mention in plain speech. This means that he did not produce all four mentions of this word in the same prosodic context, and therefore these measurements were not included in the mean duration calculation for the word *beets* in the revised dataset.

3.2.2
Results of reanalysis
Thirty out of 185 sets of words (16.2%) were removed from the old dataset to create the new dataset. The results of the reanalysis were similar to the results of the original analysis. Comparisons of clear and plain speech durations showed that durations in the clear speech condition were still significantly longer than those in the plain speech condition for both first mentions (two-tailed paired Wilcoxon signed rank test, $W = 9$, $p < .0001$) and second mentions (two-tailed paired Wilcoxon signed rank test, $W = 4$, $p < .0001$). Comparisons of first and second mention durations also still showed significantly longer durations for first mentions in both speech styles (plain speech: two-tailed paired Wilcoxon signed rank test, $W = 1228$, $p < .0005$, clear speech: two-tailed paired Wilcoxon signed rank test, $W = 1181$, $p < .001$).[3] These effects can be seen in Table 2.

---

3   To check whether applying a more inclusive criterion for breaks would affect our results, we tested for second mention reduction after removing all words for which JG reported a break context mismatch. Two-tailed paired Wilcoxon signed rank tests revealed that significant second mention reduction still appeared in both clear speech, $W = 586$, $p < .05$, and plain speech, $W = 452$, $p < .005$, even after removing all words for which JG reported a break mismatch.

Table 2

Word duration statistics (in milliseconds) without boundary mismatch averaging over all speakers, by speech style and mention (Clear 1 = clear speech, first mention, Clear 2 = clear speech, second mention, Plain 1 = plain speech, first mention, Plain 2 = plain speech, second mention, Clear Ratio = first mention duration divided by second mention duration in clear speech, Plain Ratio = first mention duration divided by second mention duration in plain speech)

	Clear 1	*Clear 2*	*Plain 1*	*Plain 2*	*Clear Ratio*	*Plain Ratio*
n = 55 (word tokens in each condition)						
Mean	396	374	324	304	1.075	1.083
Median	392	358	320	297	1.065	1.072
Std. dev.	134	131	130	122	0.134	0.157
Min.	142	112	95	73	0.780	0.788
Max.	740	695	673	614	1.463	1.758

3.3
Reanalysis accounting for unequal accenting

3.3.1
Background

These results raise the question of how speakers are controlling the articulation levels of individual words. They may be adjusting the likelihood that a word will be accented based on its probability, or they may be adjusting the word's duration independently of prosodic prominence. These two possibilities can be examined in a controlled environment using the second mention reduction phenomenon. Second mention reduction may be a by-product of the fact that speakers tend to accent first mentions of words because they communicate new information, and de-accent second mentions because they tend to be old information (Brown, 1983). Accented words tend to have longer durations than unaccented words (Klatt, 1976). The other possibility is that mention, along with many others factors, including information status, lexical frequency, and conditional probability, influences a word's articulation level along a continuum ranging from hyper- to hypo-articulation. Under this account, there is variation within the sets of accented and unaccented words. Therefore, even if both mentions of a word are accented, or both mentions are unaccented, they can still exhibit second mention reduction. It is even possible that different mechanisms are used in clear and plain speech. To examine this question, we compared the number of accented first mentions to the number of accented second mentions. We then reanalyzed the data after controlling for accent status. Every word in the paragraphs used in the original analysis was coded as accented or unaccented (as described above in Section 2.7). First and second mention durations were compared after removing any sets of words for which the accent status was not consistent across all four mentions (Clear1, Clear2, Plain1, and Plain2). For example, Speaker 4 accented his first mention of the word *piece* in clear speech, but de-accented all other mentions of the word. Because

Table 3

Percent of word tokens accented, averaging over words

Mention	Clear	Plain
1st mention	79.15%	64.18%
2nd mention	63.73%	49.1%

of this, Speaker 4's durations for *piece* were not included when calculating the mean durations for this word.

Because words with longer durations are more likely to be judged as accented, by removing sets of words with mismatched accent statuses we are biasing our results toward more equal first and second mention durations. This reduces the likelihood that we will find second mention reduction.

3.3.2
Accent analysis

Sign tests were used to examine whether first mentions and words produced in clear speech were more likely to be accented. Because each word did not have the same number of tokens in the analysis (due to disfluencies) we calculated the percentage of tokens of each word that were accented. For example, four speakers' paragraphs containing the word *alley* were included in the analysis. For each of the individual mentions of *alley* (Clear1, Clear2, Plain1, and Plain2) we counted the number of speakers who accented it, then calculated the percentage of times it was accented. Three of the four speakers accented *alley* when they first mentioned it in the clear speech condition, so it had a 75% accenting rate in the Clear1 category. One-tailed sign tests were used to compare the accenting percentages of first and second mentions and clear and plain speech styles. Words were significantly more likely to be accented in clear speech than in plain speech for both first mentions, $p < .05$, and second mentions, $p < .01$. First mentions were also significantly more likely to be accented than second mentions in both clear speech, $p < .05$, and plain speech, $p < .05$. The mean accenting percent in each of the four categories can be seen in Table 3.

3.3.3
Results of reanalysis

Eighty-five out of 185 sets of words (46%) were removed for the reanalysis. Significant second mention reduction remained after this reanalysis. Comparisons of first and second mention durations showed that first mentions were still significantly longer than second mentions in both speech styles (plain speech: two-tailed paired Wilcoxon signed rank test, $W = 770.5$, $p < .005$, clear speech: two-tailed paired Wilcoxon signed rank test, $W = 760$, $p < .005$), despite the bias toward first and second mention equality inherent in this reanalysis. In addition, the new dataset still had a significant clear speech effect, with longer durations in clear speech than in plain speech (first mentions: two-tailed paired Wilcoxon signed rank test, $W = 990$, $p < .0001$, second mentions: two-tailed paired Wilcoxon signed rank test, $W = 956$, $p < .0001$). These differences can be seen in Table 4.

Table 4

Word duration statistics (in milliseconds) without accent mismatch averaging over all speakers, by speech style and mention (Clear 1 = clear speech, first mention, Clear 2 = clear speech, second mention, Plain 1 = plain speech, first mention, Plain 2 = plain speech, second mention, Clear Ratio = first mention duration divided by second mention duration in clear speech, Plain Ratio = first mention duration divided by second mention duration in plain speech)

	Clear 1	*Clear 2*	*Plain 1*	*Plain 2*	*Clear Ratio*	*Plain Ratio*
n = 44 (word tokens in each condition)						
Mean	409	383	324	304	1.089	1.106
Median	401	368	324	280	1.081	1.093
Std. dev.	135	132	118	124	0.134	0.178
Min.	142	112	95	76	0.834	0.763
Max.	704	669	590	562	1.443	1.685

3.4

Frequency

A partial correlation was used to analyze the relationship between frequency and duration. The partial correlation controlled for word length, measured as number of phonemes.[4] A partial correlation can be used when two independent variables (e.g., frequency and length in phonemes) are correlated with one another. The contribution of one independent variable (here, word length) is removed from the target-independent variable (frequency) and the dependent variable (duration) to determine the effect of the target independent variable alone on the dependent variable (Tabachnick & Fidell, 2007). Log frequency was used in this analysis instead of actual frequency because the distribution of target word frequencies was highly skewed, with only a few high frequency words and many low frequency words. As a result, frequency effects were investigated using a partial Pearson correlation run on log frequency and first mention duration. Significant negative correlations were found in the plain, $r = -0.37$, $t(56) = -2.98$, $p < .005$, $r^2 = 0.137$, and clear, $r = -0.451$, $t(56) = -3.78$, $p < .0005$, $r^2; = 0.204$, conditions. These correlations indicate that higher frequency words tended to have shorter durations even when the effect of word length is controlled for. These results are in line with previous research on the relationship between frequency and duration (Aylett & Turk, 2004; Bell et al., 2002; Jurafsky et al., 2001). The replication of earlier findings shows that the materials and measurements in this study are behaving as expected. The frequency effects in clear speech extend these previous findings by showing that not all words in clear speech are maximally hyper-articulated.

---

4 Although number of phonemes is an imperfect measure of word length, larger units such as syllables fail to capture the variation in length between words with the same syllable count. Smaller units, such as feature changes are more likely to vary between speakers.

3.5
Frequency and second mention reduction

In order to examine the relationship between frequency and amount of second mention reduction, Pearson correlations were run on log frequency and second mention reduction ratios (first mention duration divided by second mention duration) in both conditions. Word length was not controlled in these correlations because any effect of word length would lead longer words (a trait associated with low frequency) to exhibit more second mention reduction than shorter words because they have more segments that can each be reduced or deleted. In contrast, the tendency to reduce highly predictable words leads us to predict more second mention reduction for high frequency words (which are generally shorter than low frequency words) because they are the more predictable. A significant positive correlation between log frequency and second mention reduction ratio was found in plain speech, $r = 0.292$, $t(57) = 2.31$, $p < .05$,[5] indicating that high frequency words exhibited more second mention reduction than low frequency words. No significant correlation was found between log frequency and second mention reduction ratio in clear speech, $r = -0.058$, $t(57) = -0.44$, $p = .66$. The difference between these two correlations is significant in a one-sided z-test, $z = 1.899$, $p < .05$. The difference between the clear and plain speech correlations cannot be attributed to insufficient power to find this effect in clear speech, as the plain speech correlation is positive, while the clear speech correlation is negative.

4 Discussion

These results replicate and extend previous findings about the effects of speech style, repeated mention, and lexical frequency on word duration. They replicate earlier findings that clear speech involves longer durations than plain speech (Picheny et al., 1986). They provide further confirmation of the second mention reduction phenomenon (Fowler & Housum, 1987), and show that it appears in both plain and clear styles of read speech. The first reanalysis demonstrates that second mention reduction in clear speech is not a result of the larger number of phrase breaks associated with clear speech. The accent analysis shows that words in clear speech and first mentions are more likely to be accented than words in plain speech and second mentions. This difference between first and second mentions could explain the second mention reduction effect. However, the second reanalysis, which included only sets of words that were either all accented or all unaccented, shows that second mention reduction is not simply a by-product of de-accenting old information. In addition, the results replicate the finding that, all else being equal, high frequency words tend to have shorter durations than low frequency words (Aylett & Turk, 2004; Bell et al., 2002; Jurafsky et al., 2001), and furthermore show that this effect also appears in both clear and plain read speech styles. Finally, we found that high frequency words exhibit more second mention reduction than low frequency words in plain speech, but not in clear speech.

---

5 This correlation was strongly driven by the word *and*, which had the highest second mention reduction ratio (1.685) and one of the highest frequencies (2,621,900) in the experiment. After removing this outlier, the correlation between log frequency and second mention reduction ratio in plain speech was no longer significant, $r = 0.127$, $t(56) = 0.96$, $p = .34$.

The results of this experiment support the Many Factors Clear Speech Hypothesis, which states that clear speech is just one of many factors affecting articulation level. As expected, word durations in clear speech were significantly longer than durations in plain speech, showing that clear speech involves hyper-articulation. However, effects of both second mention reduction and frequency on word duration were found in clear speech, indicating that clear speech does not involve maximal hyper-articulation and thereby disproving the Maximum Hyperarticulation Clear Speech Hypothesis. Instead, clear speech is one of many factors affecting word duration, and it is a mode of speech production that retains discourse-level structure (such as old vs. new information) conveyed by articulation level. The fact that second mention reduction and frequency effects both appear in clear speech shows that speakers do not distinguish between contrasts that carry information about the discourse history, such as old versus new information, and contrasts that carry information about the lexicon, such as high versus low frequency words, in clear speech, disproving the Maximum Discourse Information Clear Speech Hypothesis. As the frequency effect is maintained in clear speech, it seems listeners are scaling up the articulation levels they would have applied in plain speech rather than only maintaining or enhancing contrasts that are relevant to the current discourse.

Although second mention reduction and frequency effects were both found in clear and in plain speech, high frequency words showed more second mention reduction than low frequency words in plain speech, but this interaction was not found in clear speech. In other words, speakers reduce the most when all factors support reduction in plain speech, yet there appears to be a limit on word duration reduction in clear speech. The Probabilistic Reduction Hypothesis, which claims that more probable words are reduced, offers no explanation of this difference between clear and plain speech, because it simply describes factors that cause reduction. It offers no opposing factor to counter reduction. However, the Smooth Signal Redundancy Hypothesis can explain the clear speech floor effect because it has a constant redundancy level (based on both lexical predictability and articulation level) as a goal. This model allows the stipulation that clear speech requires a higher level of redundancy than plain speech. In plain speech, speakers can accept the risks associated with extreme hypo-articulation of highly predictable words. However, in clear speech the listener's ability to comprehend the speech is already believed to be compromised, so some basic level of articulation must be maintained for all words regardless of how predictable they are.

Our examination of whether probabilistic effects are prosodically mediated suggests that there is a connection between probability and articulation level independent of prosodic prominence. Aylett and Turk (2004, 2006) also found a small but significant effect of probabilistic factors on duration, independent of prosodic factors. They claim that these results could be due to imperfect probabilistic or prosodic models, but acknowledge that they could signify a direct connection between probability and articulation level. One important distinction between our study and theirs is that they looked at levels of prosodic prominence other than phrasal stress, including vowel type (reduced vs. unreduced vowels), lexical stress, and nuclear versus nonnuclear phrasal accents. In our analysis of second mention reduction, vowel type and lexical stress were controlled because we were comparing two productions of the same word. We did not examine or control for whether accented words contained nuclear

accents, but this factor did not provide a significant unique contribution to Aylett and Turk's (2004) regression models of duration. Therefore, it seems unlikely that the results of our reanalysis would be different if we used a more detailed prosodic model. Our finding that first mentions are more likely to be accented than second mentions supports their idea that a word's probability influences the likelihood that it will be accented. However, after these accent differences were controlled, the highly significant second mention reduction effect remained. This fact provides evidence for a direct link between a word's probability and its articulation level that exists independent of prosodic prominence.

This work has opened up a new avenue of research examining the interactions between different types of probability factors and speech styles. One possibility is to look at how lexical probability and speech style affect articulatory features other than word duration, such as vowel space and amplitude. Aylett and Turk (2006) found that vowels in more probable contexts are generally centralized relative to vowels in less probable contexts. We would similarly expect to find more centralized point vowels and lower peak amplitudes for second mentions and high frequency words because these words are more predictable. Future work should also study these effects in a more realistic environment, using natural rather than read speech. Such experiments can examine whether there are different types of clear speech. Is clear speech directed at a non-native speaker different from clear speech directed at someone with a hearing problem? However, such naturalistic experiments also introduce confounds because the target words do not appear in controlled environments. Finally, these effects should be studied in other languages and dialects and in the speech of non-native English speakers to discover the universality and nature of these phenomena. We predict that a word's probability should affect its articulation level in other languages because probabilistic reduction is the most robust and efficient way of communicating orally, regardless of the language. However, speakers of some languages and dialects do not de-accent second mentions like speakers of American English tend to. For example, repeated mentions of words can be accented in Indian English and many Romance languages, such as Romanian and Italian (Ladd, 1996). In addition, some languages, such as Korean, mark information structure through phrasing rather than through pitch accents. Although we found that second mention reduction appeared in American English even when accent status was controlled, low probability words were still more likely to be accented, which contributed to probabilistic reduction. It would be interesting to see if such effects appear in languages and dialects in which second mentions are not de-accented (e.g., Indian English, Italian), and in languages in which focus is not marked through pitch-accents (e.g., Korean). If so, it would support our conclusion that these effects appear at a gradient level separate from prosodic prominence. If we do find differences in probabilistic reduction between languages, it would be especially interesting to examine the performance of non-native speakers of American English. They may behave as they do in their native language, they may behave like native American English speakers, or they may do something in between. Their behavior might reflect the higher processing load of speaking a second language, which could weaken the effects of discourse history on articulation level. It might also reflect their lack of familiarity with American English, so the pattern of lexical
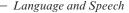
frequency effects might be different than that of native English speakers, reflecting their unique experience with American English.

5 Conclusion

One goal of this study was to examine how factors governing articulation level interact with one another. No single factor examined in this study overwhelmed any of the other factors, but rather, all the factors contributed to each word's articulation level. Interestingly, an interaction between mention, frequency, and speech style was found. Words were most hypo-articulated when all three factors supported hypo-articulation. However, clear speech seemed to impose a limit on the amount of hypo-articulation allowed, so high frequency words did not exhibit more second mention reduction than low frequency words in this speech style. In contrast, in plain speech, high frequency words were subject to more second mention reduction than low frequency words because the combination of the two probabilistic factors made second mentions of high frequency words highly predictable. This three-way interaction was taken as support for the Smooth Signal Redundancy Hypothesis, which stresses the importance of producing robust communication in addition to efficiently expending articulatory effort. The other goal of this study was to test whether probabilistic effects are mediated through prosodic prominence. We found that although lower probability words (first mentions) are more likely to be accented, the inverse relationship between probability and duration remains after differences in prosodic prominence are controlled. This finding is consistent with the pattern of results reported by Aylett and Turk (2004, 2006) in providing evidence against a strict interpretation of the Smooth Signal Redundancy Hypothesis' claim that the link between probability and articulation level is prosodically mediated.

References

ANDERSON, A. H., & HOWARTH, B. (2002). Referential form and word duration in video-mediated and face-to-face dialogues. In J. Bos, M. E. Foster, & C. Matheson (Eds.), *Proceedings of the Sixth Workshop on the Semantics and Pragmatics of Dialogue (EDILOG 2002), Edinburgh, U.K., 4–6 September 2002*, (pp.13–20). Edinburgh: Cognitive Science Centre, University of Edinburgh.

AYLETT, M., & TURK, A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech. *Language and Speech*, **47**, 31–56.

AYLETT, M., & TURK, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, **119**, 3048–3058.

BARD, E. G., ANDERSON, A. H., SOTILLO, C., AYLETT, M., DOHERTY-SNEDDON, G., & NEWLANDS, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, **42**, 1–22.

BARD, E. G., LOWE, A., & ALTMANN, G. (1989). The effects of repetition on words in recorded dictations. In J. Tubach & J. Mariani (Eds.), *Proceedings of EUROSPEECH-1989 First European Conference on Speech Communication and Technology (EUROSPEECH '89), Paris, France, September 27–29, 1989* (pp.573–576). Paris: European Speech Communication Association.

BELL, A., GREGORY, M. L., JURAFSKY, D., GIRARD, C., BRENIER, J., & IKENO, A. (2002). Which predictability measures affect content word durations? In W. Byrne, E. Fosler-Lussier, & D. Jurafsky (Eds.), *Proceedings of the Workshop on Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology (PMLA)* (pp.65–70). Estes Park, CO, USA.

BELL, A., JURAFSKY, D., FOSLER-LUSSIER, E., GIRAND, C., GREGORY, M., & GILDEA, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, **113**, 1001–1024.

BOERSMA, P., & WEENINK, D. (2004). Praat: Doing phonetics by computer (Version 4.2.25) [Computer program]. Retrieved December 15, 2004, from http://www.praat.org

BRADLOW, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp.241–273). Berlin and New York: Mouton de Gruyter.

BRADLOW, A. R., KRAUSE, N., & HAYES, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, **46**, 80–97.

British National Corpus, version 3 (BNC XML Edition) (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved December 2, 2007, from http://www.natcorp.ox.ac.uk

BROWN, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp.67–77). Heidelberg: Springer-Verlag.

CHEN, F. R. (1980). *Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level*. Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

FOWLER, C. A. (1988). Differential shortening of repeated context words produced in various communicative contexts. *Language and Speech*, **31**, 307–319.

FOWLER, C. A., & HOUSUM, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, **26**, 489–504.

FOWLER, C. A., LEVY, E. T., & BROWN, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language*, **37**, 24–40.

HELFER, K. S. (1998). Auditory and auditory-visual recognition of clear and conversational speech by older adults. *Journal of the American Academy of Audiology*, **9**, 234–242.

JURAFSKY, D., BELL, A., GREGORY, M., & RAYMOND, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp.229–254). Amsterdam: Benjamin.

KLATT, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129–140.

KLATT, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208–1221.

KOCHANSKI, G., & SHIH, C. (2001). Automated modelling of Chinese intonation in continuous speech. In P. Dalsgaard, B. Lindberg, H. Benner, & Z. Tan (Eds.), *Proceedings of EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event, Aalborg, Denmark, September 3–7, 2001* (pp.911–914). Aalborg, Denmark.

LADD, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

LIEBERMAN, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, **6**(3), 172–187.

LINDBLOM, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp.403–439). Dordrecht: Kluwer.

MATTHIES, M., PERRIER, P., PERKELL, J. S., & ZANDIPOUR, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research*, **44**, 340–353.

OVIATT, S., MacEACHERN, M., & LEVOW, G.-A. (1998). Predicting hyperarticulate speech during human–computer error resolution. *Speech Communication*, **24**, 87–110.

PICHENY, M. A., DURLACH, N. I., & 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.

PICHENY, M. A., DURLACH, N. I., & BRAIDA, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, **29**, 434–446.

SMILJANIC, R., & 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.

SMILJANIC, R., & BRADLOW, A. R. (2008). Stability of temporal contrasts across clear and conversational speech. *Journal of Phonetics*, **36,** 91–113.

SMILJANIC, R., & BRADLOW, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistic and Language Compass*, **3**(1), 236–264.

TABACHNICK, B. G., & FIDELL, L. S. (2007). *Using multivariate statistics: Fifth edition*. Boston, MA: Pearson.

TAYLOR P., & BLACK A. W. (1998). Assigning phrase breaks from part-of-speech sequences. *Computer Speech and Language*, **12**, 99–117.

UCHANSKI, R. M. (2005). Clear speech. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp.207–235). Malden, MA: Blackwell Publishing.

van SON, R. J. J. H., & van SANTEN, J. P. H. (2005). Duration and spectral balance of inter-vocalic consonants: A case for efficient communication. *Speech Communication*, **47**, 100–123.

Appendix A: Experiment paragraphs with target words underlined and second mentions italicized

Paragraph 1

In today's show we will learn how to make a soup with beets and string beans. The first step is to wash *and* slice the *beets* and *string beans*. Next you fry the onions and toss a piece of ginger into the pot. Make sure the *piece of ginger* isn't too big, or its strong flavor could overpower the subtle *flavor* from the string beans. Finally, add the beets and string beans, and pour the meat stock through a sieve into the *pot*. Nearly any type of *meat stock* will do, but an old chicken carcass boiled down to meat stock works best. Let it simmer for an hour, and at the end you will have a fantastic *soup* that your whole family will enjoy!

Paragraph 2

My cousin Sue competes in snowsports, and was eager to win first place at the ski meet this year. She had been training hard after her broken leg healed. Her nemesis Bobbie had caused her *broken leg* just before the *ski meet* last year in a suspicious collision, although *Sue* couldn't prove it wasn't an accident. This year, Sue was taking no chances. When *Bobbie* skied near enough, Sue threw a pole out in front of her. Bobbie *skied* straight into a tree in confusion and broke her arm.

Best of all, when they investigated the regrettable *collision*, everyone thought it was an *accident*!

Paragraph 3

During his sophomore year of college, <u>Bob</u> Andrews was having a really hard time. He was flunking all his <u>classes</u>, and he only had enough <u>money</u> to <u>buy</u> Top Ramen noodles. He also broke up with his <u>girlfriend</u>. Then one day, he decided to become a <u>Linguistics</u> major. Suddenly everything changed. He loved the *Linguistics classes*, and started getting all As. He <u>worked in</u> a <u>lab</u> for a semester, where he earned enough *money* to *buy* <u>steak</u> <u>every week</u>. A girl named <u>Judy</u> *worked in* the *lab* too, and he was instantly attracted to her. He got close to her by inviting her over for *steak* at his house *every week*. After only a couple of these dates she became his *girlfriend*. Last year, *Bob* and *Judy* got married. Bob's life is great, and he owes it all to Linguistics.

Paragraph 4

On <u>her</u> <u>birthday</u>, <u>Dottie</u> showed up early at the <u>zoo</u> for work. She was in charge of the <u>birds</u> and the <u>baby animals</u>. The <u>cuckoos</u> and the <u>geese</u> were *her* <u>favorite</u> *birds*, and the <u>zebras</u> and the <u>baboons</u> were her *favorite baby animals*. She went to the bird house first, and was surprised to see that the *cuckoos* and the *geese* were missing and her <u>assistant</u> <u>Lucy</u> wasn't there. She hurried over to the baby animal area and was upset to find that the *zebras* and the *baboons* were gone too, and her other *assistant* Lisa was nowhere to be found. She muttered, "I can't run the *zoo* by myself", and ran back to her office to find out what was going on. But when she opened the door, *Dottie* heard a honk and a bunch of voices yelling "Surprise!" She realized with joy that not only had *Lucy* conspired with Lisa to throw a surprise party for her *birthday*, but they had even brought along all of her favorite animals!

Paragraph 5

If you want to go to <u>Gina's Pizza Shop</u>, I can tell you the best way to get there. Go straight down this street and follow the signs for the <u>Johnson Expressway</u>. However, don't actually go onto the *Johnson Expressway*. When you get to the on-ramp, take a left onto <u>Cleveland</u> Street, the main street in town. You'll go past a big school called *Cleveland* High School, right between a <u>church</u> with a yellow door and a *church* with a <u>blue steeple</u>. There is a small <u>alley</u> just past the church with the *blue steeple*. Take this *alley* for several blocks, and turn left on the third <u>road</u> you come to. Eventually, the *road* will split in two. Take Fillmore Boulevard, which is the one on the right. A block and a half later you'll see the sign for *Gina's Pizza Shop*, also known as the best pizza place in town.

Appendix B: Target words and their BNC frequencies

Word	BNC frequency	Word	BNC frequency	Word	BNC frequency
accident	6298[a]	cuckoos	85	money	36,560
alley	604	Dottie	55	of	2,886,105[j]
and	2,621,900	every	39,146	piece	8883[a]
animals	8500	expressway	32	pizza	480
assistant	1375[a]	favorite	963[c,f]	pot	1695[a]
baboons	82	flavor	1352[a,f]	road	18,997[a]
baby	8604[a]	geese	394	shop	9126[a]
beans	1296	Gina's	341[g]	ski	464[a]
beets	5[a]	ginger	120[a]	skied	61
birds	5603[b]	girlfriend	1130	soup	1282
birthday	3174	her	203,369[h]	steak	466
blue	7713[c]	in	1,691,158[i]	steeple	84[a]
Bob	3709[d]	Johnson	3031	stock	6869[a]
Bobbie	179[d]	Judy	423[d]	string	2433[a]
broken	2524[c]	lab	663[a]	Sue	1527[d]
buy	1,2041[e]	leg	5026[a]	week	31,677[a]
church	19961[a]	Linguistics	784	worked	12,372
classes	5979[b]	Lucy	2507[d]	zebras	51[b]
Cleveland	1694	meat	3518[a]	zoo	760
collision	629	meet	4[a]		

Note: For words used as multiple parts of speech, we used the only frequencies for the part of speech used in the experiment paragraphs. For words spelled differently in the U.S.A. and the U.K., we added together the frequencies of both spellings.

[a] Frequency of the singular common noun
[b] Frequency of the plural common noun
[c] Frequency of the adjective
[d] Frequency of the proper noun
[e] Frequencies of finite + infinitive verbs
[f] Combined UK and US spellings
[g] Used count for "Gina"
[h] Frequency of the possessive determiner pronoun
[i] Frequency of the preposition
[j] Frequency of the preposition "of"