Lexical Effects on the Phonetic Realization of English Segments

A thesis submitted in partial satisfaction of the requirements for the degree Master of Arts in Linguistics

by

Roger Billerey-Mosier

2000
The thesis of Roger Billerey-Mosier is approved.

___________________________________
Sun-Ah Jun

___________________________________
Carson Schütze

___________________________________
Patricia A. Keating, Committee Chair

University of California, Los Angeles

2000
# Table of Contents

**Table of Contents** ......................................................................................III

**List of Figures** .......................................................................................... V

**List of Tables** ............................................................................................. VI

**Acknowledgments** ................................................................................... VII

**Abstract of the Thesis** ............................................................................... VIII

**Introduction** ............................................................................................... 1

**Experiment I: Uniqueness Point** ...............................................................2

I. Background .................................................................................................. 2
   1. Uniqueness Point .......................................................................................2
   2. Givenness and Phonetic Realization ..................................................... 4
   3. Competition and Phonetic Realization ............................................... 8

II. Hypothesis and Experiment ....................................................................... 10

III. Procedure .................................................................................................. 11
   1. Word lists ...............................................................................................11
   2. Recording ...............................................................................................14
   3. Measurements .......................................................................................15

IV. Results ....................................................................................................... 16
   1. Results by Speaker ...............................................................................17
      1.1 Speaker 1 ........................................................................................17
      1.2 Speaker 2 ........................................................................................17
      1.3 Speaker 3 ........................................................................................18
      1.4 Across speakers .............................................................................18
   2. Results by Target Type .........................................................................19
      2.1 –board tokens ...............................................................................19
      2.2 –berry tokens ...............................................................................20
      2.3 –hand tokens ...............................................................................21
      2.4 –time tokens ...............................................................................22

V. Discussion ................................................................................................... 22
   1. Duration Differences .........................................................................23
   2. Formant Structure Differences .........................................................23
   3. General Discussion ..........................................................................24
LIST OF FIGURES

Figure 1. Target duration by type (speaker 1). No-noise condition (left), noise condition (right). ........................................................................................................................ .................. 17
Figure 2. Target duration by type (speaker 2). No-noise condition (left), noise condition (right). ........................................................................................................................ .................. 17
Figure 3. Target duration by type (speaker 3). No-noise condition (left), noise condition (right). ........................................................................................................................ .................. 18
Figure 4. Target duration by type (all speakers). No-noise condition (left), noise condition (right). ........................................................................................................................ .................. 18
Figure 5. Target duration in late-unique and early-unique words. Target="-board." No-noise condition (left), noise condition (right). ................................................................................ 19
Figure 6. Target F1 in late-unique and early-unique words. Target="-board." No-noise condition (left), noise condition (right). ................................................................................ 19
Figure 7. Target F2 in late-unique and early-unique words. Target="-board." No-noise condition (left), noise condition (right). ................................................................................ 20
Figure 8. Target duration in late-unique and early-unique words. Target="-berry." No-noise condition (left), noise condition (right). ................................................................................ 20
Figure 9. Target F1 in late-unique and early-unique words. Target="-berry." No-noise condition (left), noise condition (right). ................................................................................ 21
Figure 10. Target F2 in late-unique and early-unique words. Target="-berry." No-noise condition (left), noise condition (right). ................................................................................ 21
Figure 11. Target duration in late-unique and early-unique words. Target="-hand." No-noise condition (left), noise condition (right). ................................................................................ 22
Figure 12. Target duration in late-unique and early-unique words. Target="-time." No-noise condition (left), noise condition (right). ................................................................................ 22
Figure 13. Initial stop closure duration in nonwords and words................................. 35
Figure 14. Initial stop closure time in nonwords and words......................................... 36
Figure 15. Target vowel duration in nonwords and words................................................ 36
Figure 16. Target vowel F1 in nonwords and words......................................................... 37
Figure 17. Target vowel F2 in nonwords and words......................................................... 38
## List of Tables

Table 1. Complete list of target words and companions ............................................................ 13
Table 2. Summary of duration differences and their bearing on the shortening hypothesis .. 23
Table 3. Formant structure differences and their bearing on the assimilation to context hypothesis ................................................................. 24
Table 4. Formant structure differences and their bearing on the reduction to schwa hypothesis ................................................................. 24
Table 5. Complete list of target strings, words, nonwords, and cumulative word frequencies. 33
ACKNOWLEDGMENTS

I wish to thank my committee members (Sun-Ah Jun, Pat Keating, and Carson Schütze), as well as Anne Cutler, Richard Wright, Arty Samuel, Michael Vitevich, and Ellen Bard, for their input at various stages of the design and writing process.

I wish to thank Ha-ha the cat for her help in the preparation of the manuscript.

Special thanks to Jennifer.

All errors, omissions and misrepresentations are mine. Most remaining typos are Ha-ha’s.
This thesis reports two experiments investigating two aspects of the relationship between the phonetic realization of segments and the structure of the lexicon in English.

The first experiment attempts to determine whether English segments are realized differently depending on their position in a word relative to that word's uniqueness point, defined as the location of the segment which identifies the word uniquely against all the other words sharing the same initial string. No systematic differences according to position are found. This result is interpreted as showing that the domain of the H&H theory is larger than the word.

Experiment 2 investigates whether segments are realized differently depending on their occurrence within a word or a nonword. This experiment shows that the duration and formant structure of strings in words and nonwords do not differ systematically, providing
partial support for the syllabary assumption underlying Levelt et al.'s (1999) model of speech perception, and casting further doubt on the generality of frequency-driven reduction effects.

INTRODUCTION

It is well known that the acoustic realization of segments is highly variable. Factors contributing to systematic variation include the position of a word within an utterance, with edge effects such as pre-boundary lengthening or phrase-initial strengthening. In addition, it has been shown that speakers know a great deal about the perceptual relationships between words within their internal lexicon, in particular similar-sounding words (or, loosely, perceptual neighbors), and that these relationships affect their production in systematic ways.

Specifically, segments are reduced in predictable words (Lieberman [1963]), repetitions of earlier words (Fowler & Housum [1987]), or words with few perceptual neighbors (Wright [1997]). This may be explained by Lindblom's (1990) H&H (hypo- and hyper-articulation) theory. In Lindblom's model, a speaker will adjust his/her speech production between the poles of hyper-articulation, providing the listener with maximally distinctive linguistic units, and hypo-articulation, which minimizes articulatory effort. The introduction of a word, its unpredictability, or the need to discriminate against similar words may trigger hyper-articulation to facilitate identification; later repetitions, predictability and low confusability may trigger reduction.

This thesis reports the results of two experiments further investigating the relationship between the lexicon and the phonetic realization of English segments.

If a word's predictability within an utterance triggers reduction, the question comes to mind of whether predictability within a word will have the same effect. As words are confidently identified at their uniqueness point (Marslen-Wilson [1987]), defined as the
location of the segment which identifies the word uniquely against all the other words sharing the same initial string, segments occurring past that point within a word are arguably more predictable than earlier segments. The first experiment reported in this thesis determines whether reduction is indeed observed as a result of this predictability.

It has also been argued that word frequency triggers reduction. Segments in words might thus be reduced relative to nonwords, which have a frequency of zero. On the other hand, Levelt et al.'s (1999) argument that speakers access articulatory instructions for whole syllables predicts no differences based on word/nonword status. The second experiment in this thesis determines whether the duration and formant structure of strings in words and nonwords differ systematically.

**EXPERIMENT I: UNIQUENESS POINT**

**I. BACKGROUND**

**1. UNIQUENESS POINT**

The notion of uniqueness point has been refined and studied extensively in the work of Marslen-Wilson (1987, 1990, *inter alia*). A word's uniqueness point is the earliest point at which that word differs from all the words in the lexicon that share the same initial string. "All the words in the listener's mental lexicon that begin with [a given] initial sequence" are said to form a cohort (Marslen-Wilson [1987:78]). For example, the uniqueness point of the word *blueberry* occurs past the second [b], because of the words *bluebird, bluebonnet, blueblood* and *bluebook*, which share the initial string [blub]. A word like *strawberry*, on the other hand, may be identified uniquely at the [b], because there is no other single word beginning with
Marslen-Wilson (1987) found that "[t]he point at which a word becomes uniquely identifiable, as established through an analysis of that word's initial cohort, corresponds very well to the point at which listeners will confidently identify a word in [a] gating task." In other words, listeners are able to recognize a word after hearing only as much initial material as needed to differentiate it from other words in the lexicon.

The special status of a word's initial string has been confirmed in a number of studies. Within-word perceptual asymmetries have been found in written word recognition. Bruner & O'Dowd (1958) altered the spelling of written words at the beginning, in the middle and at the end of those words and found that word recognition was more disrupted the earlier the change occurred. Nooteboom (1981) presented listeners with halves of Dutch words that uniquely identified the words they were excised from. Words were recognized faster and more accurately when listeners heard the first half than when they heard the second half. In a phoneme restoration experiment, Samuel (1987) found stronger restoration effects by the third syllable of words with an early uniqueness point relative to words with a later uniqueness point; excised phonetic information was erroneously perceived as present more frequently past the carrier words' uniqueness point. Marslen-Wilson & Welsh (1978) studied fluent restorations of mispronounced words by speech shadowers, defined as "responses in which the subjects repeated the word in its normal form" after hearing it mispronounced, without any disruption of their own production. More fluent restorations were observed when the mispronunciation occurred in the third syllable than in the first syllable of the word (53% v. 45%). Assuming that the intended words were identified prior
to location of the mispronunciation in the third syllable, the authors interpret this result as evidence that

once a single word-choice has emerged, the recognition system will have achieved its primary goal, and a less detailed assessment of the remaining input for that word will be required. This will have the effect of making the system less sensitive to deviations that occur after the point of identification. (Marslen-Wilson & Welsh [1978:57])

This interpretation is consistent with Cole et al.'s (1978) finding that mispronunciations at the beginning of a monosyllabic word were detected more than twice as often as mispronunciations in the final segment.

These results indicate that a word's initial segments, at least up to that word's uniqueness point\(^1\), carry a different informational load (in the sense discussed below) from segments occurring later in the word, and that less attention is paid to the phonetic realization of segments occurring past a word's recognition point. This difference may be characterized as a contrast between identification and confirmation: early segments primarily identify a word uniquely against other words, and (in ideal circumstances) segments occurring past the identification point will essentially serve to confirm that identification.

2. GIVENNESS AND PHONETIC REALIZATION

The relationship between the phonetic realization of entire lexical items and givenness has been extensively studied. The term "givenness" has been used by various authors with slightly different definitions, and the one used here is that of predictability/recoverability. In this sense, a lexical item is given if "the speaker assumes that

\(^{1}\) Or recognition point, i.e. the point at which a word is accurately recognized; but the two are highly correlated (Marslen-Wilson [1987]).
the hearer can predict or could have predicted that a particular linguistic item will or would occur in a particular position within a sentence" (Prince [1981]).

I define, for the present purposes, the notion of informational load as follows. The informational load of a word is a value which varies inversely with the givenness of that word and the number of previous instances of that same lexical item in a given stretch of discourse. The informational load of a segment is a value which varies inversely with the predictability of that segment and with the number and frequency of words differing on that segment from the lexical item containing that segment. Numerous factors contribute to a word's predictability, such as the syntactic and semantic subcategorization frames of the words with which that word co-occurs; its occurrence within a cliché/proverb/maxim; multiple repetitions of that word in the utterance or discourse; theme-hood in the sense of "We are talking about X": words semantically related to X are more predictable than completely unrelated words. For example, if we are talking about a backyard, then "lawn", "flower" and "mulch" are more predictable than "muffler", "dashboard" and "seatbelt"; the reverse is true if we are talking about a car. Similarly, "book" is more predictable than "bill" as the object of "read" (Morton & Long [1976], Bolinger [1972]). Perceptually, semantically more predictable words (in "read a book"/"read a bill"-type pairs) are responded to faster in phoneme-monitoring tasks (Morton & Long [1976]; see Foss, Harwood & Blank [1980] for a review).

At the phrasal level, tone/accent assignment is highly correlated with the informational load of particular lexical items. Conversely, and more to the point, deaccenting and phonetic reduction (generally speaking, including shortening, reduced pitch range, etc.) are generally associated with parts of an utterance with a lesser informational load (Bolinger
[1972], Beckman & Pierrehumbert [1986]). Studying the realization of repeated words across a long stretch of natural speech, Fowler & Housum (1987) found that later repetitions of words were shortened relative to earlier productions. Studying Dutch tokens, Koopmans-van Beinum & van Bergem (1989) showed that vowels were less reduced in new words than in "old" words (i.e., words which had been introduced previously in the discourse). A possible reason for these observations is that a speaker's articulatory effort minimization strategies will be overridden locally by the need to maximize the listener's chance to identify a word as accurately as possible when it is first introduced in discourse, i.e., when its informational load is maximal; later occurrences of the same word will function in a similar way to anaphors, and merely adequate productions will still guarantee accurate identification. It has been argued that the acoustic reduction observed in later repetitions of a lexical item has the additional purpose of signaling to the listener that that lexical item is given (Fowler & Housum [1987], Sotillo [1997]).

In a famous experiment, Lieberman (1963) compared the intelligibility of words excised from productions in which they were more or less predictable (more, as in "A stitch in time saves nine," or less, as in "The number that you will hear is nine"). He found that more predictable words were less intelligible when excised and presented in isolation, and that they showed signs of acoustic reduction compared to the same words uttered in contexts in which they were unpredictable. I calculated from his data that words uttered in contexts in which they were unpredictable were on average 16.15% longer than the same

---

2 As the tokens measured were extracted from radio broadcasts, their prosodic position could not be controlled, which casts some doubt on the generality of the observation. However, the criteria for the selection of the tokens measured were very systematic, and shortening was observed for all speakers in all six independent samples, which suggests that the effect is robust.

3 The possible confounding effect of focus was only found in the example cited.
words uttered in predictable contexts. These results are consistent with the idea that a speaker has a sense of, and adapts his/her production to, a listener's immediate communicative needs in the way described above.

This idea is generalized in Linblom's (1990) sketch of the H&H theory. Lindblom suggests that a speaker will tailor his/her production to the listener's needs between the two poles of hypo- and hyper-speech ("H and H"). Listener-oriented hyper-speech maximizes distinctiveness; speaker-oriented hypo-speech minimizes articulatory effort:

In the **ideal case**, the speaker estimates the running contribution that signal-complementary processes will make during the course of an utterance, and dynamically tunes the production of its elements to the short-term demands for either output-oriented control (hyper-speech) or system-oriented control (hypo-speech). What he/she needs to control is -- not that linguistic units are actualized in terms of **physical invariants** (higher-order or whatever) -- but that their signal attributes possess **sufficient contrast**, that is discriminative power that is sufficient for lexical access (Lindblom [1990:405], original emphasis).

Speech production is viewed as a balance between the (articulatorily costly) need to pronounce linguistic units with enough distinctiveness for the listener to identify them, and the desire to reduce articulatory effort. The speaker must constantly assess his/her listener's needs in order to economize articulatory effort while still maintaining sufficient contrast between linguistic units.

While Lindblom proposes that this tuning occurs "between and within utterances," DeJong (1995) proposes a characterization of stress in English as local, within-word hyperarticulation. Van Heuven (1994) showed that, in Dutch, "at least some speakers have the means to express narrow focus on linguistic units below the level of the syllable" (e.g., when contrasting vowels as in *I said pool, not pull*).

---

4 However, Bradlow et al. (1999) have found that the relationship between intelligibility and articulation is not
Even though these researchers have slightly different concerns (informativeness for Lindblom, prominence for deJong and van Heuven), the notion of distinctiveness and its relationship with articulation is central in all three studies. This raises issue of whether the results summarized in the previous paragraphs can be extended to the within-word level, and specifically whether acoustic reduction can be observed in segments with a lesser informational load within lexical items. As words are confidently identified at their uniqueness point (Marslen-Wilson [1987]), segments occurring past that point are more predictable than earlier segments, assuming that those earlier segments have been perceived accurately: "[…] a second syllable is more predictable than a first syllable. A second syllable is constrained both by prior context and by a correct first syllable, while a first syllable is constrained only by prior context" (Cole & Jakimik [1978:151]). Since given and predictable words show signs of reduction, likewise segments past a word's uniqueness point may be reduced relative to the same phonological segments occurring before that point in other words.

3. COMPETITION AND PHONETIC REALIZATION

Another factor affecting the phonetic realization of segments is a word's perceptual similarity to other words in the lexicon. Differential patterns of phonetic realization depending on perceptual competition have been found by researchers investigating the Neighborhood Activation Model of word recognition (Luce [1986]). In this model, the relevant competitors of a given lexical item in recognition are elements of that word's similarity neighborhood, which overlaps but is not coextensive with its initial cohort. A lexical item's neighbor is a word differing from that lexical item by one segment. Items in a

as direct as previously assumed. This issue will be addressed in later paragraphs.
similarity neighborhood compete with each other in word recognition. The strength of a competing neighbor depends on its similarity with the target lexical item and on the relative frequencies of the target and that competitor. For the sake of comparison, the initial CV cohort of *king* includes *kit, kin, kiss, kilter, kiln, kink, kipper*, etc. and the similarity neighborhood of *king* includes *kit, kin*, and *kiss*, but also *ring, sing*, etc. A neighborhood is dense if it includes a large number of strong competitors, and sparse when it does not. Wright (1997, in press) showed that vowels in monosyllabic words with sparse similarity neighborhoods are centralized ("a well-known feature of reduction or hypo-articulation", Wright [1997]) compared to vowels in words with denser neighborhoods. Similarly, Goldinger & Van Summers (1989) found that the voice onset time difference within word pairs (bin/pin) from dense neighborhoods was larger than the difference within word pairs from sparse neighborhoods. These findings together indicate that phonetic material with a lesser informational load (in the sense given above) shows signs of articulatory reduction. Specifically, they show that speakers will reduce their articulatory effort when an intended word runs a lesser risk of being confused with other similar-sounding words. This result is not self-evident; another logical possibility (in addition to a complete absence of difference) would be greater variability without reduction in the phonetic realization of words in sparse neighborhoods relative to words in dense neighborhoods. That actual acoustic reduction is observed provides support for the functional interpretation of these findings described above, and suggests that similar results may be found for the same reasons in other contexts. The experiment described below examines one particular case.
II. HYPOTHESIS AND EXPERIMENT

The hypothesis tested below is that segments occurring after a word's uniqueness point, which have been argued to carry a lesser informational load, will show signs of reduction relative to the same phonological segments in the same segmental environment occurring before a word's uniqueness point. To test this hypothesis, sets of words containing particular segments located at (or before) and past a carrier word's uniqueness point (e.g., the underlined vowel in blueberry and strawberry, respectively) were read by speakers in normal speech and in a simulated conversational interaction in which they were asked to articulate as if they were talking to a foreigner or a hearing-impaired listener. The acoustic duration and formant structure of the target segments were then compared according to the segments' position relative to the carrier word's uniqueness point.

An important characteristic of the Neighborhood Activation Model experiments cited above is that they focus on monosyllabic words. This makes the computation of similarity neighborhoods both rigorous and tractable, but raises the issue of whether those results hold for the entire lexicon. This experiment attempts to integrate findings from both the Cohort model and the Neighborhood Activation Model of word recognition, by investigating whether acoustic differences similar to those found in monosyllables with similarity neighborhoods of different sizes can be found in longer words depending on the target segments' position relative to those words' uniqueness points. It should be noted that this experiment is not meant to test either model's predictions, since neither model explicitly predicts anything about production; rather, it is an attempt to find in longer words results comparable, but not identical, to the findings on monosyllables cited above. The
experimental protocol followed in this experiment is necessarily different from those in Wright (1997, in press) and Goldinger & Van Summers (1989), if only because of the nearly intractable problem of computing and comparing meaningful similarity neighborhoods for trisyllabic and longer words.

This experiment helps determine whether articulatory reduction always occurs when the informational load of segments decreases (assuming that this assessment of informational load is correct). In addition, it informs Lindblom's (1990) H&H theory by determining whether its domain of application can be as small as the word in a situation directly related to listeners' need for information.

III. PROCEDURE

1. WORD LISTS

A list of 43 words was compiled, 15 of which contained target segment sequences. In 7 of these (the late-unique words), the beginning of the target sequence coincided with the word's uniqueness point, and in the other 8 (the early-unique words), the beginning of the target sequence was located one segment past that point. The 28 remaining words are companions and were used as fillers. A companion is defined as a member of a target-carrying word's initial cohort, i.e., a word sharing the same first segments and stress pattern as the target-carrying word up to (and possibly including) the target vowel. The following examples illustrate each of the types described above:

**Target:** sequence [ɛɪʃ].
Early-unique target-carriers: cranberry and strawberry. These words are early-unique because they can be identified uniquely before the target string, as there is no other single word in English beginning with [kʌn] or [stɪb].

Late-unique target-carriers: blueberry and blackberry. These words share the same initial string as bluebottle, bluebell, bluebonnet, bluebird, bluebeard, blueblood and blackbird, blackbeard, blackbeetle, blackbook, blackball, respectively. They are late-unique because the point at which they can be identified uniquely is the beginning of the target string.

Companions: For target-carrier blueberry, companions were bluebird, bluebonnet, bluebook, bluebell, and bluebeard (blueblood was not included in the word list).

The following table is a comprehensive list of the words used in the study.
<table>
<thead>
<tr>
<th>Target</th>
<th>Early-unique</th>
<th>Late-unique</th>
<th>Companions</th>
</tr>
</thead>
<tbody>
<tr>
<td>[eɪ̞]</td>
<td>cranberry</td>
<td>blueberry</td>
<td>bluebeard</td>
</tr>
<tr>
<td></td>
<td>strawberry</td>
<td></td>
<td>bluebell</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>bluebird</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>bluebonnet</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>bluebottle</td>
</tr>
<tr>
<td>[ænd]</td>
<td>stagehand</td>
<td>longhand</td>
<td>longhaired</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>longhead</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>longhorn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>longhouse</td>
</tr>
<tr>
<td>[æjm]</td>
<td>bedtime</td>
<td>downtime</td>
<td>downtrend</td>
</tr>
<tr>
<td></td>
<td>lifetime</td>
<td></td>
<td>downturn</td>
</tr>
<tr>
<td></td>
<td>lunchtime</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nighttime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[oʊd]</td>
<td>chalkboard</td>
<td>blackboard</td>
<td>blackball</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>blackbeard</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>blackberry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>blackbird</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>blackbeetle</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>blackbook</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>blackboard</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Complete list of target words and companions.

The target words were embedded in carrier sentences of the type `<name> says <target> is a word`, e.g., *Sam says blueberry is a word*. The names used in the carrier sentences were all monosyllabic. No two adjacent sentences began with the same name. Four versions of the word list were prepared, each with a different random order for the carrier sentences. Each page began and ended with a sentence containing a word irrelevant to the study. Adjacent sentences did not contain words belonging to the same cohort (e.g., *bluebeard* did not precede or follow *bluebonnet*) except in one case, due to a design error. In that particular
case, the adjacent words were not target words, and thus any differences in production caused by their adjacency did not affect the results reported here. In addition, adjacent sentences did not contain words carrying the same target string, in order to avoid contrastive focus on the initial syllable, which would likely have reduced any potential production differences in the target string.

Given the structural description of the words required by the study, very few sets of words were available, and most of them exhibited characteristics that made it necessary to adopt the particular measurement criteria described below. Additionally, the relative frequencies of the target words and the members of their initial cohorts could not be matched. Lastly, cohort size could not be exactly matched either, but Marslen-Wilson (1987) showed that the size of the cohort does not matter in a lexical decision task, which yielded constant decision latency in spite of cohorts of different sizes:

For a distributed system like the cohort model, it need make no difference to the timing of the word-recognition decision whether two candidates have to be considered or two hundred. In either case, the timing of the selection process reflects the point at which a unique solution emerges. This is purely a matter of cohort structure, and has nothing to do with the number of alternatives per se. (Marslen-Wilson [1987:86])

2. RECORDING

Three paid participants recorded each of the four word lists four times. Recording sessions were separated by two to seven days. For each session, the subject sat in a soundproof room in the phonetics laboratory at UCLA. The order of the sequence of word lists was reversed for each session. The subjects were encouraged to take breaks every two pages of text or whenever necessary.
In two of the four sessions (the no-noise condition), the subjects were instructed to speak as naturally as possible. In the other two sessions, the subjects were instructed to speak as clearly as possible, as if to a foreigner or a hearing-impaired listener. In these two sessions (the noise condition), the subjects heard random noise played over headphones during recording, which all but suppressed auditory feedback on their own performance and forced them to speak more clearly and loudly than they normally would. Each word was recorded sixteen times, yielding eight tokens for each condition.

The subjects' productions were recorded to analog tape, then digitized at 12.5 kHz on a PC computer using CSL. Acoustic measurements were performed on a PC computer using PCQuirer.

3. Measurements

Acoustic reduction can be measured in various ways: Centralization of vowels in the vowel space or assimilation of vowels to their consonantal context, shown in their formant structures, or shortening. The results reported here are absolute duration measurements and the frequencies of the first two formants of the vowels in the target strings.

The target sequences measured included the entire sonorant sequence vowel+sonorant(+final vowel), ending at the beginning of the oral closure for [d] or the beginning of the intensity drop if /d/ was flapped (–board and –hand words), and at the beginning of the frication noise in the following [z] of is in the case of vowel- or sonorant-final words with no laryngealization (–time and –berry words). One speaker's vowels were systematically laryngealized word-finally and word-initially. Vowel-final words followed by vowel-initial words (blueberry is sequences) were thus joined by several periods of creaky
phonation. In these cases, the end of the word was measured at the center of the laryngealized stretch. This measurement procedure included more segmental material than a single vowel, but was more consistent than trying to determine where vowels ended and sonorants began.

First and second formants were measured at the beginning of the steady-state portion of the vowel in the target sequence. Formant frequencies were jointly determined with a 16-coefficient LPC magnitude spectrum and a spectrogram. Spurious formants (bandwidths greater than 350 Hz) were eliminated. However, the vowels in words ending in –band and –time were partially or completely nasalized, which obscured their formant structure. For this reason, formant data will only be reported for the –board and –berry words.

IV. RESULTS

All the results reported here are two-sample t-tests assuming unequal variances (two-tailed, \( \alpha = 0.05 \)).
1. RESULTS BY SPEAKER

1.1 Speaker 1

The target string was significantly longer in early unique -berry tokens in the no-noise condition (217ms v. 242ms, p<0.05), and in early unique -hand tokens in the noise condition (166ms v. 209ms, p<0.01).

![Figure 1. Target duration by type (speaker 1). No-noise condition (left), noise condition (right).](image)

1.2 Speaker 2

The target string duration was significantly shorter in early unique -berry tokens in the noise condition (349ms v. 309ms, p<0.005) and in early unique -time tokens in the noise condition (274ms v. 252ms, p=0.01). F2 was higher in early unique -board tokens in the noise condition (marginally significant; 846 Hz v. 902 Hz, p=0.10).

![Figure 2. Target duration by type (speaker 2). No-noise condition (left), noise condition (right).](image)
1.3 Speaker 3

The target string was significantly shorter in early unique -berry tokens in both conditions (no-noise: 325ms v. 306ms, p<0.05; noise: 401ms v. 372ms, p<0.05), and in early unique -hand tokens in the no-noise condition (209ms v. 186ms, p<0.05). F2 was lower in early unique -berry tokens in both conditions (marginally significant; no-noise: 1751 Hz v. 1708 Hz, p<0.10; noise: 1861 Hz v. 1821 Hz, p<0.10).

Figure 3. Target duration by type (speaker 3). No-noise condition (left), noise condition (right).

1.4 Across speakers

No significant or marginally significant differences in target string duration or formant frequencies were found across speakers.

Figure 4. Target duration by type (all speakers). No-noise condition (left), noise condition (right).
2. Results by Target Type

2.1 -board tokens

F2 was not reliably measurable for speaker 3, and so these data are not reported.

No significant duration differences were found for any speaker in either condition.

In one case, a formant frequency difference was marginally significant (speaker 2, noise condition: late unique F2=846 Hz, early unique F2=902 Hz, p=0.10).

Figure 5. Target duration in late-unique and early-unique words. Target="-board." No-noise condition (left), noise condition (right).

Figure 6. Target F1 in late-unique and early-unique words. Target="-board." No-noise condition (left), noise condition (right).
2.2 –berry tokens

The target string was significantly longer in early unique tokens for speaker 1 in the no-noise condition (217ms v. 242ms, p<0.05). It was significantly shorter in early unique tokens for speaker 2 in the noise condition (349ms v. 309ms, p<0.005), and for speaker 3 in both conditions (no-noise: 325ms v. 306ms, p<0.05; noise: 401ms v. 372ms, p<0.05).

F2 was significantly higher in early unique tokens for speaker 1 in the no-noise condition (1974 Hz v. 2069 Hz, p<0.05), and lower in early unique tokens for speaker 2 in the noise condition (significant; 2031 Hz v. 1985 Hz, p<0.05) and for speaker 3 in both
conditions (marginally significant; no-noise: 1751 Hz v. 1708 Hz, p<0.10; noise: 1861 Hz v. 1821 Hz, p<0.10).

![Figure 9](image9.png)

**Figure 9.** Target F1 in late-unique and early-unique words. Target="-berry." No-noise condition (left), noise condition (right).

![Figure 10](image10.png)

**Figure 10.** Target F2 in late-unique and early-unique words. Target="-berry." No-noise condition (left), noise condition (right).

### 2.3 –hand tokens

The target string was significantly longer in early unique tokens for speaker 1 in the noise condition (166ms v. 209ms, p<0.01), and significantly shorter in early unique tokens for speaker 3 in the no-noise condition (209ms v. 186ms, p<0.05).
Target duration in late-unique and early-unique words
Target = "-hand" Condition = No noise

Figure 11. Target duration in late-unique and early-unique words. Target="-hand." No-noise condition (left), noise condition (right).

2.4 –time tokens

The target string was longer in early unique tokens for speaker 1 in the no-noise condition (196ms v. 216ms, p<0.05), and shorter in early unique tokens for speaker 2 in the noise condition (274ms v. 252ms, p=0.01).

Figure 12. Target duration in late-unique and early-unique words. Target="-time." No-noise condition (left), noise condition (right).

V. DISCUSSION

The hypothesis tested here is that segments occurring past a carrier word's uniqueness point will be shortened and/or that their formant structure will be closer to that of schwa or to the characteristic frequencies of the neighboring consonants, compared to the same segments occurring before a carrier word's uniqueness point. Specifically, the
hypothesis is that shortening and formant structure reduction should be visible in target strings in early unique words.

1. DURATION DIFFERENCES

No significant differences in target string duration are found in 16 cases out of 24, and where a difference is found, it only supports the shortening hypothesis in 5 cases out of 8. In other words, the hypothesis is supported in only 5 out of 24 cases, or 21%. Table 2 below summarizes the eight significant duration differences observed.

<table>
<thead>
<tr>
<th>Shortening hypothesis supported/not supported</th>
<th>Tokens</th>
<th>Speaker</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>supported</td>
<td>-berry</td>
<td>2</td>
<td>noise</td>
</tr>
<tr>
<td>supported</td>
<td>-berry</td>
<td>3</td>
<td>both</td>
</tr>
<tr>
<td>supported</td>
<td>-hand</td>
<td>3</td>
<td>no-noise</td>
</tr>
<tr>
<td>supported</td>
<td>-time</td>
<td>2</td>
<td>noise</td>
</tr>
<tr>
<td>not supported</td>
<td>-berry</td>
<td>1</td>
<td>no-noise</td>
</tr>
<tr>
<td>not supported</td>
<td>-hand</td>
<td>1</td>
<td>noise</td>
</tr>
<tr>
<td>not supported</td>
<td>-time</td>
<td>1</td>
<td>no-noise</td>
</tr>
</tbody>
</table>

Table 2. Summary of duration differences and their bearing on the shortening hypothesis.

2. FORMANT STRUCTURE DIFFERENCES

As formant data are only available in post-labial context, the main difference we expect if early unique tokens assimilate more to their context is a lowering of F2 triggered by the labial.
As was the case above, few significant differences were observed, and none in the frequency of the first formant in any token for any speaker in any condition. Table 3 below summarizes the significant differences in F2 observed in the data. The hypothesis is only supported in 1 out of 11 cases, or 9% (3 out of 11, or 27%, if marginally significant differences are included).

<table>
<thead>
<tr>
<th>Assimilation to context hypothesis supported/not supported</th>
<th>Tokens</th>
<th>Speaker</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>supported</td>
<td>-berry</td>
<td>2</td>
<td>noise</td>
</tr>
<tr>
<td>supported (marginally)</td>
<td>-berry</td>
<td>3</td>
<td>both</td>
</tr>
<tr>
<td>not supported</td>
<td>-berry</td>
<td>1</td>
<td>no-noise</td>
</tr>
<tr>
<td>not supported (marginally)</td>
<td>-berry</td>
<td>2</td>
<td>noise</td>
</tr>
</tbody>
</table>

Table 3. Formant structure differences and their bearing on the assimilation to context hypothesis.

If reduction to schwa is the correct characterization of reduction, then we expect the F2 of [ɛ] in -*berry* tokens to be lowered and the F2 of [o] in -*board* tokens to be raised towards the F2 frequency of schwa (about 1500 Hz). The following table summarizes the significant differences in F2 observed in the data and their bearing on the reduction to schwa hypothesis. This hypothesis is thus only supported in 1 out of 11 cases, or 9% (4 out of 11, or 36%, if marginally significant differences are included).

<table>
<thead>
<tr>
<th>Reduction to schwa hypothesis supported/not supported</th>
<th>Tokens</th>
<th>Speaker</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>supported</td>
<td>-berry</td>
<td>2</td>
<td>noise</td>
</tr>
<tr>
<td>supported (marginally)</td>
<td>-berry</td>
<td>3</td>
<td>both</td>
</tr>
<tr>
<td>supported (marginally)</td>
<td>-berry</td>
<td>2</td>
<td>noise</td>
</tr>
<tr>
<td>not supported</td>
<td>-berry</td>
<td>1</td>
<td>no-noise</td>
</tr>
</tbody>
</table>

Table 4. Formant structure differences and their bearing on the reduction to schwa hypothesis.

3. General Discussion

The hypothesis that a sequence of segments occurring past a word's uniqueness point shows signs of reduction relative to an identical sequence located prior to that point in
another word is not supported by the data in the current experimental design. While a few
individual sets of tokens do exhibit significant and marginally significant differences in the
hypothesized direction, a number of tokens actually show significant differences in the
opposite direction, and most tokens do not exhibit any significant differences at all.

The perception literature summarized earlier unambiguously shows differences in the
processing of phonetic material depending on its location relative to a word's uniqueness
point. What the current results may suggest, then, is that speakers are not so attentive to
their listeners' perceptual processes as to modulate their production in the way hypothesized
here. This fact is not self-evident in the light of the results of Lieberman (1963), Fowler &
Housum (1987), and others discussed above. In addition, these results indicate that the
domain of the H&H theory of speech production as it relates to the notion of informational
load as defined here may be larger than the word, i.e., the phrase or utterance.

Bradlow & Pisoni (1999) showed that easy words (sparse similarity neighborhood,
high frequency relative to their neighbors) presented without any background noise and
transcribed by native speakers have higher overall intelligibility than hard words (dense
similarity neighborhood, low relative frequency). This result is somewhat surprising, given
that hard words are hyperarticulated (Wright [1997, in press]), and as such would be
expected to be more intelligible than easy words. The easy-hard effect observed by Bradlow
and Pisoni shows that "this hyperarticulation [is] not sufficient to overcome the effect of
lexical difficulty on the part of the listener," and fits in with the null result found above. In
simple terms, Bradlow & Pisoni (1999) showed that listeners do not (or cannot) necessarily
exploit perceptually the articulatory efforts of speakers; the first experiment here showed
that speakers do not necessarily exploit the location of a word's uniqueness point in production the way listeners do in processing.

More evidence can be found suggesting that speech production might be more egocentric than the H&H theory assumes it to be. Bard et al. (1989) found signs of degradation in second instances of memos recorded by speakers over their original production, in spite of the fact that their second production would actually be the only version heard by the memo's recipient. In an interactive communication situation in which speakers were giving each other directions based on geographical maps, Sotillo (1997) found that speakers reduced "the intelligibility of introductory tokens of landmark names when they gave instructions about a map they had seen previously to a new listener. Despite the fact that the entity was New for her listener, the speaker produced a less intelligible token than the one she produced in her first encounter with the map." She also found that assimilatory processes (e.g., /n/ realized as [m] before a labial, as in *The eggs have bee[m] broken*) were observed with the same frequency irrespective of whether the potentially assimilated tokens had many, few, or no competitors for recognition: "In other words, speakers hypoarticulate even when doing so results in a production that is lexically ambiguous; that is, speakers fail to maintain distinctiveness" (Sotillo [1997:270]). A possible interpretation of these findings is that, as noted by Anderson et al. (1997), developing and maintaining an accurate assessment of a listener's informational needs would require "extraordinary powers" on the part of the speaker, who is more likely to base such an assessment on his/her own model of the discourse situation. In addition, assuming that the task were possible, keeping an accurate record of the listener's needs may be cognitively
more costly (however that cost might be measured) than would be warranted by the articulatory effort saved via hypo-speech.

Another possible reason for the absence of effect observed here is that actual differences might not be apparent in the particular measurements made: it is well known that the acoustic record only provides indirect and incomplete articulatory information. In addition, no pitch measurements (peak frequency or location) were performed, and such pitch information could possibly have revealed systematic differences which were independent of duration and formant structure. In his study of the effect of focus on Dutch segments, van Heuven (1994) did not observe any differences of the sort that were studied here between focused and non-focused vowels. Rather, the differences he found were prosodic:

[...] at least some speakers have the means to express narrow focus on linguistic units below the level of the syllable. Crucially, such speakers do this by purely prosodic means, viz. by changing the properties of the accent-lending pitch contour (its shape and location) on the syllable that contains the contrasted segment, rather than by changing acoustic properties of the individual contrasted segment. (van Heuven [1994:95])

The experimental design itself may also have precluded the emergence of differences. First, the words were in a focused position in the carrier sentences, and segments in focused syllables are typically hyperarticulated relative to segments in non-focused syllables. However, the target strings themselves did not occur in stressed syllables, and so the effect of prosodic position is not self-evident; in addition, the tokens were ordered so as to minimize the likelihood of contrastive accent on the target strings; lastly, a similar objection could be leveled at Wright's (1997, in press) experiments, in which words were recorded in
isolation, a hyper-articulation-prone setting, yet he did find significant differences of the sort hypothesized here.

Second, the hypothesized functional explanation for the reduction effects investigated here relies on a speaker/listener interaction in a natural setting, but the tokens were recorded by speakers sitting alone in a soundproof room. That the simulated interaction in the noise condition did not yield different results from the no-noise condition cannot be taken as unmitigated confirmation of the null result reported here. Simulated interaction is not actual interaction; for example, it has been shown that the intonational characteristics of mother-to-child speech ("motherese") were dependent on the child's actual presence: "the full range of prosodic modifications in mothers' speech was evoked only in the presence of the infant," not when the mothers' speech was recorded in the child's absence for later presentation to the baby (Fernald & Simon [1984]). Again, however, Wright (1997, in press) did not attempt to simulate any kind of conversational interaction. Thus, the fact that significant differences were actually found occasionally, both supporting and refuting the hypothesis, suggests that a general difference would be observable if it existed, and that the hypothesis can be confidently rejected as a linguistic generalization.

We now turn to a different aspect of lexical informativeness, namely whether segments occur in an actual word or in a nonword.
EXPERIMENT II: SEGMENTS IN WORDS AND NONWORDS

I. BACKGROUND

Word frequency has long been proposed as a factor in articulatory reduction (Balota et al. [1989], Bybee [1994], Pierrehumbert [in press]). However, frequency effects on word production are not entirely reliable. Comparing the acoustic realization of segments in words and nonwords provides a good test. If frequency-driven reduction effects are pervasive, real words should show signs of reduction relative to nonwords (with a frequency of zero). In addition to testing the validity of frequency effects on word production, such a comparison should inform experimental methodology. Nonwords are often used in perception experiments, but acoustic measurements of the stimuli are rarely provided. If present, acoustic differences may be responsible for some of the effects observed and ascribed to other factors (such as lexical access).

The model of speech production described in Levelt et al. (1999) makes a different prediction about segment realization in words and nonwords. This model assumes that "a speaker has access to a repository of gestural scores for the frequently used syllables of the language." In this model, frequently used syllables are not generated on-line by segment concatenation; rather, "gestural scores" for whole syllables are accessed directly in speech production. It follows from this assumption that the phonetic realization of a given syllable should be independent of the lexical status of that syllable's carrier. If this assumption is correct, then no acoustic differences should be found between segments comprising a given
syllable, irrespective of whether that syllable occurs in words or nonwords (everything else being equal).

II. HYPOTHESES AND EXPERIMENT

Should differences be found between words and nonwords, at least two predictions can be made about the nature of those differences.

Johnson et al. (1993) found that listeners who were asked to change a speech synthesizer's settings to produce acceptable vowels systematically settled on "hyperarticulated" targets, i.e., "high vowels were higher, low vowels were lower, front vowels were farther front, and back vowels were farther back."\textsuperscript{5} They called this result the "hyperspace effect". Longer voice onset times in voiceless stops (and shorter VOTs in voiced stops) are commonly interpreted as signs of hyperarticulation (Keating et al. [in press], Jun [1993], among many others). Lastly, as vowel shortening is generally interpreted as a sign of hypoarticulation, longer duration may be seen as a sign of hyperarticulation. Thus, if words and nonwords are differentiated by the hyperarticulation of segments, we would expect to find one or more of the following differences co-occurring: lower F1 in [I], higher F1 in [Æ], higher F2 in [I] and/or [Æ], longer VOTs in voiceless stops, shorter VOTs in voiced stops, and longer vowels, in nonwords relative to words (on the assumption that words would show signs of frequency-driven reduction, as suggested by the research summarized above).

\textsuperscript{5} Their experiment was perceptual, and as such does not directly bear on the experiment reported here. Their study is cited because it provides information on the correlates of hyperarticulation in vowels.
Another potential difference between words and nonwords is the sensitivity of vowels to their context. Moon & Lindblom (1994), based on results from Lindblom (1963), analyze the differences observed between citation forms and clear speech forms (Chen [1980]) as differential assimilation to the neighboring consonants. This context dependence is "more limited for clear than for citation-form speech" (Moon & Lindblom [1994]). The hyperspace effect suggests that a hypoarticulated vowel should be less front/back/high/low than a non-hypoarticulated vowel, i.e., its formants should be closer to those of schwa. Moon & Lindblom (1994) and Lindblom (1963) predict that a non-hyperarticulated vowel's formants will be closer to the characteristic frequencies of neighboring consonants, which in some cases may pull the vowel's formants away from schwa: "for instance, in /b1b/, /d1d/, and /g1g/, a movement toward schwa would make the F1 of /i/ higher, whereas assimilation would lower it" (Moon & Lindblom [1994]). If frequency-driven reduction causes vowels to be more sensitive to context, we would expect vowels next to labial and/or velar consonants such as [p] and [k] to have a lower F2 in words relative to nonwords6.

In summary: the following experiment attempts to determine whether segments in words and nonwords are realized differently (as suggested by the research on word frequency summarized above) or identically (following an assumption in the work of Levelt and colleagues). If differences in vowel formant structures are found, the experiment also investigates whether they result from assimilation to consonantal context or from reduction to schwa; these differences are predicted to appear as acoustic reduction in real words relative to nonwords.

---

6 Lip-rounding (as in labial consonants) and tongue backing (as in velar consonants) both lower F2.
In order to test the above hypotheses, trochaic disyllabic words and nonwords containing target CV strings were recorded in carrier sentences of the type *The next word you will hear is* <target>. The acoustic characteristics of each segment in the target strings were compared between carrier words and nonwords.

### III. Procedure

#### 1. Word Lists

Four sets of one word and two nonwords (all disyllabic trochees) were constructed, each beginning with a target string of the form stop + vowel⁷. The words were matched for cumulative frequency (Kučera & Francis [1967]), defined as the sum of the frequencies of all inflected forms of a given word (singular, plural, and genitive). The following is a list of the words and nonwords used in the experiment, together with the target syllables and cumulative frequencies.

---

⁷ The term "string" is used here instead of "syllable" because the target strings are not necessarily whole phonological syllables. The syllabic affiliation of the consonant following the initial CV string is not a factor here, because it is the same within each set.
Whenever possible, the shape of a given nonword was designed to match the phonological shape of the word within its set. For example, *distump* and *distant* share the shape [target string] + [stə] + nasal stop + homorganic voiceless oral stop. The orthographic shape of the nonwords followed regular spelling-to-sound correspondence rules. The words and nonwords were embedded in carrier sentences of the type *The next word you will hear is <target>,* e.g., *The next word you will hear is pistip.*

### 2. RECORDING

Seven participants recorded each of the four sets four times in one recording session. The subjects were sitting in a soundproof room in the UCLA phonetics laboratory. The subjects were asked to speak naturally, and were shown by example that each nonword was trochaic (e.g., "the invented words in the list are pronounced like *Kansas*, not like *delete*"). All six subjects pronounced the nonwords identically. Four carrier sentence lists were constructed, each with a different pseudo-random order, which did not allow two adjacent words/nonwords belonging to the same set. Each page began and ended with a carrier sentence containing a disyllabic trochaic word or nonword (assigned randomly) that was not
relevant to the experiment. The subjects' productions were recorded to analog tape, then digitized at 11,025 Hz and analyzed on a PC computer using PCQuirer.

3. Measurements

The target syllables and carrier sentences were constructed so as to provide five unambiguous measures: closure duration and voice onset time (VOT) of the initial stop, vowel duration, and the vowel's first and second formants. Stop closure duration was measured from the end of the frication noise in *is* until the beginning of the stop's release burst. Closure duration was not recorded when the token included an audible pause. Voice onset time was measured from the beginning of a stop's release burst to the first peak of periodic activity in the waveform. Formants were measured with a 16-coefficient LPC magnitude spectrum and a spectrogram. Spurious formants (bandwidths greater than 350 Hz) were eliminated.

IV. Results

No statistically significant differences were found between nonwords within any set. As the point of linguistic interest being investigated is whether nonwords are acoustically different from real words in a systematic way, the two nonwords in each set were pooled and compared to the real word in that set. Where significant differences are found between words and nonwords, the nonwords in the relevant set will be examined individually.
1. INITIAL STOP CLOSURE DURATION

A near-significant difference in initial stop closure duration can be observed only in the cab-set. The closure duration for [k] is longer in nonwords than in real words (61.5ms v. 55.8ms, respectively, p=0.06). No other significant difference was found.

![Figure 13. Initial stop closure duration in nonwords and words.](image)

2. VOICE ONSET TIME

A near-significant difference was found in the pis-set. The VOT of initial [p] in nonwords was 45ms against 40ms in real words (p<0.06). The individual differences between pistol and each of the nonwords in its set approached significance only in the case of pistip (40ms v. 45ms, p<0.09), and as noted above, no significant difference was observed between the two nonwords pistump and pistip (44ms v. 45ms, p=0.69). No other significant difference was found.
3. TARGET VOWEL DURATION

No significant difference was found.

4. TARGET VOWEL F1, F2

No difference was found in the F1 of the target vowel within any set.
Three significant differences in F2 between vowels in words and nonwords were found. In the *cab* set and in the *pack* set, F2 is lower in nonwords than it is in words (*cab*: nonwords F2=1535 Hz, real words F2=1595 Hz, \( p < 0.05 \); *pack*: nonwords F2=1555 Hz, real words F2=1652 Hz, \( p < 0.001 \)). This is due to a significant difference between *cabin* and *cabble* (1598 Hz v. 1517 Hz, \( p < 0.02 \)). The difference between *cabin* and *cabbem* is not significant (\( p > 0.2 \)). Again, there is no difference between the two nonwords in the *cab* set (\( p > 0.32 \)).

In the *pis* set, F2 is higher in nonwords than it is in words (nonwords F2=1707 Hz, real words F2=1625 Hz, \( p < 0.04 \)). There is a significant difference between *pistol* and *pistip* (1625 Hz v. 1717 Hz, \( p < 0.04 \), and a near-significant difference between *pistol* and *pistump* (1625 Hz v. 1696 Hz, \( p < 0.1 \)), but no difference between the two nonwords (\( p > 0.56 \)).
V. DISCUSSION

No general pattern of hyperarticulation is observed. No significant difference in VOT was found in consonants. No difference was found in F1. The vowel [i] is farther front (its F2 is higher) in nonwords in the pis-set, which suggests that nonwords are hyperarticulated. However, the opposite result is found in both [i] and [æ] in nonwords in the cab- and pack- sets (their F2 is lower), suggesting that words are hyperarticulated relative to nonwords. The hyperarticulation hypothesis is thus not supported.

No differential pattern of assimilation to context was found either. The lowering of F2 observed in the cab- and pack- sets may indicate that the vowel is more strongly affected by its consonantal environment in nonwords than in words: backing triggered by the velar consonant and/or lip rounding due to the labial consonant would both lower F2 (this F2 lowering effect is seen when comparing F2 in the dis-set and in the pis-set, non-significantly in nonwords [1743 Hz v. 1707 Hz, p>0.25] and significantly in words [1734 Hz v. 1625 Hz,
p<0.03]). However, this would lead us to expect that F2 would also be lower in nonwords in the *pis*- set because of the labial consonant, and what we observe is the opposite.

In conclusion, whether a segment in a stressed syllable occurs in a word or a nonword does not systematically influence its acoustic realization.

The results reported here may be seen as partial support for the assumption of Levelt et al. (1999)'s model, in which a mental syllabary containing gestural scores for whole syllables used frequently is accessed directly during speech production. Another, non-theoretical implication of this finding is that results obtained in experiments using words and nonwords presented auditorily are not likely to be artifacts of the experimental stimuli.

However, the absence of difference may be interpreted differently. First, as in the first experiment, the acoustic measurements in this experiment may have missed actual articulatory differences; in addition, most of the reservations about the naturalness of the recording environment mentioned earlier still apply. Second, this result is also compatible with a model of speech production in which words and nonwords are generated segment by segment. Arguments and evidence supporting the syllabary model can be found in Levelt & Wheeldon (1994) and Levelt et al. (1999); they mostly have to do with a syllabary's efficiency in speech production. Lastly, a possible confound lies in the fact that nonwords do not differ from words only in frequency, but also qualitatively, as they are not actual words. However,

---

8 The near-significant difference observed in the *pis*- set does go in the direction of a hyperarticulation explanation. However, this explanation does not generalize to all sets.
9 The segments compared were embedded in the same phonological syllable in different tokens; the frequencies of the words used in the experiment were not particularly low; thus, neither were the frequencies of the syllables in which the target strings were embedded; consequently, any results found here apply to a discussion of Levelt et al.'s model.
the frequency difference between a word and a nonword remains true, and so the null result reported here reduces the likelihood of frequency-driven reduction in general and supports the interpretation by Wright (1997, in press) and Goldinger & Van Summers (1989) that the neighborhood effects they observed were due to the existence and distribution of competitors, as opposed to word frequency.

10 If the stimuli are recorded in sentential contexts, as in this experiment. Whether this would remain true in the case of stimuli recorded in isolation is an open question.
SUMMARY AND CONCLUSION

The first experiment reported here compared the realization of English segments depending on their location relative to a word's uniqueness point. The notion of informational load of a segment was defined as a value varying inversely with the predictability of that segment and with the number and frequency of words differing on that segment from the lexical item containing it. Based on the finding that words are confidently identified at their uniqueness point (Marslen-Wilson [1987]), it was argued that segments occurring past a word's uniqueness point were more predictable, and therefore that they had a lesser informational load, than segments occurring earlier in the word. As segments in predictable words (Lieberman [1963]), in later repetitions of previously introduced words (Fowler & Housum [1987]), and in words with few perceptual neighbors (Wright [1997, in press]) are reduced relative to words in other contexts, it was hypothesized that segments occurring past a word's uniqueness point would be reduced relative to earlier segments. However, the analysis of the acoustic characteristics of segments according to their position relative to a word's uniqueness point did not reveal any systematic pattern of differences.

This absence of difference can be accounted for in several ways. First, it may indicate that the functional hypothesis investigated here is indeed not supported because the domain of application of the H&H theory is larger than the word, i.e. the phrase or utterance. Second, as the hypothesized effect depends on an interaction between a speaker and a listener, the absence of difference observed may have been due to the lack of such an interaction. Third, the choice of measurements may have been inappropriate for the effect investigated, and the experimental materials may have been designed in a way that reduced
the likelihood of the hypothesized effect. In the light of experimental evidence (Sotillo [1997], Bard et al. [1989], Wright [1997, in press]), however, the position taken here is that the results are sufficiently robust for the hypothesis to be legitimately rejected.

In the second experiment, the acoustic characteristics of strings occurring in words and nonwords were compared. Word-frequency-triggered reduction data suggested that segments in words might be reduced relative to segments in nonwords, whereas an assumption underlying Levelt et al.'s (1999) model had as a consequence that those segment types should not differ systematically. The results did not reveal any systematic pattern of differences.

This result was interpreted as providing partial support for the syllabary assumption underlying Levelt et al.'s model, and as adding new data to the investigation of word frequency effects on word production: it casts further doubt on the generality of frequency-driven reduction effects, and suggests that Wright (1997, in press) and Goldinger & Van Summers (1989) correctly interpreted their results as being driven by similarity neighborhood structure and not raw word frequency.

The absence of uniqueness point effect on the production of segments suggests the following question: why should articulatory adjustments stop at the word level? Van Heuven (1994) noted that only a few of his speakers were able to reliably produce words with contrastive focus on individual segments. His result and mine taken together may be due to
the integrity of the word as a semantic, phonological and phonetic unit. Once a word has been selected and its phonological shape has been determined, only an exceptional situation, such as an imperious need to differentiate that word from a similar sounding competitor, would cause a speaker to break down the set of phonetic instructions generated from that word's phonological shape into a string of discrete units (e.g., phonemes) together with specific instructions about the realization of one or more of these units. If this picture is correct, then what the current experiment as well as Van Heuven's tell us is that the processing cost involved in such a procedure offsets what benefits may result from it in terms of articulatory effort preservation.

These observations suggest a possible follow-up experiment. The words used in the first experiment are not productively polymorphemic, as indicated by their stress patterns. In order to test whether the discussion above correctly identified the integrity of the word as an active factor in H&H-type articulatory adjustments, it may be fruitful to compare the realization of segments belonging to a root morpheme in a monomorphemic word and segments belonging to a non-root morpheme in a polymorphemic word relative to a word's uniqueness point. Whether articulatory differences describable in terms of hyper- or hypo-articulation are observed in root segments relative to non-root segments would help

---

11 Pierrehumbert's (in press) model is based on raw word frequency; it remains to be see whether her model would benefit from neighborhood structure data (relative frequencies of target words and their perceptual neighbors).

12 That morpheme should be as long as possible, in order to minimize the likelihood of a confound caused by the particular phonology of short morphemes. Segments in short inflectional morphemes (one or two phonemes) are typically more resistant to elision than segments in roots, which are characteristically long and thus allow for an easier recovery of segmental information (Casali [1997]), and this resistance may appear as surface hyperarticulation. While a comparative study of the articulation of segments in short morphemes and in roots is interesting, the introduction of a phonological confound into the proposed follow-up experiment would not be desirable.
determine which of the word or the morpheme is actually the smallest domain of application of the H&H theory.

The second experiment confirmed that raw word frequency is not a good predictor of the occurrence of reduction. On the other hand, the results in Fowler & Housum (1987) seem to have a connection with word frequency: the repeated words they analyzed may not have had a particularly high or low frequency, and nothing is known of their similarity neighborhoods; however, as they were repeated several times over a stretch of discourse, one could argue that their local frequency (as defined more extensively below), i.e. their frequency in the speech situation relative to the other words in the discourse and to their own raw frequency was abnormally high. The reduction observed might thus be perceptively accounted for in terms of this abnormal local frequency. This suggests that analyses of frequency effects on speech production or perception might benefit from the manipulation of a third component, in addition to raw frequency and frequency relative to a similarity neighborhood. This third component, namely the local frequency of a lexical item, empirically defined as a function of that lexical item's frequency relative to its similarity neighborhood, its frequency relative to other words within a given stretch of discourse, and its frequency within that stretch of discourse relative to its own raw frequency, can easily be varied by increasing or decreasing the number of occurrences of a particular lexical item in the relevant experimental situation, and its effects can then be investigated. Local frequency may provide a tool to unify analyses of frequency effects and priming effects on speech production and/or perception, priming being viewed as an increase in local frequency.
REFERENCES


