

Lexical Access, Lexical Representation, and Vowel Production

Benjamin Munson

Abstract

Previous research (Munson and Solomon 2004; Wright 2004) has shown that both phonological neighborhood density and word frequency influence the size of the acoustic F1/F2 vowel space. This study investigated whether these influences reflect the role of real-time lexical access processes on articulation. Specifically, it examined the influence of word frequency and phonological neighborhood density on vowel acoustics in a condition in which lexical access was stressed (the *immediate-response* condition) and one in which it was facilitated (the *long-delay* condition). Frequency and neighborhood density had a large influence on vowel-space dispersion in the immediate-response condition. In contrast, the influence of word frequency was not significant in the long-delay condition. Multiple regressions showed that the reduced influence of frequency in the long-delay condition could be accounted for by response-time differences between the two conditions. The reduced effect of frequency in the long-delay condition suggests that its influence on vowel articulation is mediated by the process of lexical access. The influence of neighborhood density on articulation, however, could not be accounted for by differences in real-time speech production processes.

1. Introduction

The phonetic reflexes of phonological categories vary systematically. One of the challenges to the fields of laboratory phonology and psycholinguistic has been to determine the factors that condition the phonetic reflexes of phonological categories in different words and utterances, and to design models of speech production and phonetic implementation¹ that account for this systematic variation. Many of the factors that influence phonetic outcomes are well known in the field of laboratory phonology. It is axiomatic that phonetic context influences articulation. For example, the sound /s/ is articulated with different degrees of lip rounding in sV, spV, and swV contexts, with the articulatory characteristics of the adjacent

segment having predictable effects on the resulting acoustic signal (Munson 2004). Structural influences on phonetic outcomes extend to higher levels of phonological organization: the same sound may be articulated differently depending on the prosodic structure of the utterance in which it is embedded. For example, vowels in phrase-final words are typically lengthened relative to the same vowels in phrase-initial or phrase-medial position; voice-onset times are typically longer for voiceless stops at the left edge of a prosodic phrase, relative to the same stops in other prosodic contexts. The effects of prosodic structure on articulation and acoustics are present across typologically diverse languages (e.g., Keating et al. 2004).

One of the goals of models of speech production is to provide an explanation for phonetic variation due to segmental and prosodic structure. Current models address these phenomena straightforwardly. Modular feed-forward models of speech production (e.g., Levelt, Roelofs, and Meyer 1999; Shattuck-Hufnagel 1992) posit that categorical representations of the phonemes that comprise a word are copied to a phonological buffer after the semantic representation of the word has been activated. Items in this buffer can be modified based on their segmental context, and the prosodic structure of the utterances in which they are embedded. These structurally coded categorical units are then sent to an articulatory system, which translates them into continuous articulatory commands. Models of phonetic implementation (e.g., Keating 1990) allow the phonetic specification of phonological categories to be influenced by these same contextual factors. Thus, the effects of phonetic context and prosodic structure on articulation can be accounted for by models that allow plans for articulation, as well as the execution of these plans, to make reference to higher-level structural variables.

Some differences in articulation, however, are not related to structural variables like phonetic context and position in phrase. A growing body of literature has documented the existence of word-specific phonetic patterns. Instances of the 'same' phonological category may differ systematically across the words in which they are attested, as a function of a variety of factors. These phenomena have the potential to challenge current models of speech production. It is important to understand the factors that condition non-structural influences on phonetic outcomes, so that we may gauge the extent to which current models might need to be modified to account for them.

Two non-structural variables that affect phonetic outcomes are word frequency and phonological neighborhood density. Research by Bybee (2001), Jurafsky, Bell, and Girard (2002), and Munson and Solomon (2004) has shown that words' frequency of usage predicts patterns of phonological reduction. In general, words that are frequently used tend to show more reductive phonological changes (such as deletions, cluster reductions, vowel reduction, etc.) than words that are used infrequently. For example, speakers are much more likely to shorten or delete the schwa from the second syllable of the high-frequency word *memory* than from the phonologically similar low-frequency word *mammary* (Bybee 2001). The influence of frequency of usage extends beyond categorical phonological alternations. Munson and Solomon found that the acoustic vowel spaces associated with high-frequency words are less dispersed than those associated with low-frequency words. Jurafsky et al. showed that a lemma's frequency of occurrence affects its phonetic outcome, even when different lemmas are associated with the same lexeme (i.e., demonstrative pronoun *that* versus complementizer *that*). These findings complement an extensive literature demonstrating incontrovertibly that frequency affects lexical access and word recognition (e.g., Goldinger et al. 1996).

Another example of a word-specific pattern unrelated to phonetic or prosodic context is neighborhood density. Neighborhood density refers to the number of words that differ from a target word. In the most commonly used metric, it is measured as the number of real words that can be created by adding, changing, or deleting a phoneme from a target word (e.g., Luce and Pisoni 1998). Monosyllabic words of English vary greatly in their neighborhood density. A number of studies have shown that words with higher neighborhood densities are more difficult to perceive and recognize than those with lower neighborhood densities. This is true for a variety of populations (e.g., Luce and Pisoni 1998).

Wright (2004) examined the acoustic characteristics of vowels in words with high phonological neighborhood densities relative to the same vowels in words with low neighborhood densities. Wright examined the degree of dispersion in the F1/F2 acoustic vowel spaces of high- and low-neighborhood density words, by measuring the average Euclidian distance from the arithmetic center of the F1/F2 space. He found that the acoustic vowel spaces associated with high-neighborhood density words were more dispersed than those for low-neighborhood density words. In the stimuli that he used, neighborhood density co-varied with word frequency: high-density words were lower in frequency than low-frequency words.

Recently, Munson and Solomon (2004) replicated this finding using a subset of Wright's stimuli. They also showed that the vowel-space dispersion that Wright noted was not an artifact of the longer durations of the vowels comprising the high- and low-neighborhood density words. Previous research (e.g., Moon and Lindblom 1994) found positive correlations between vowel duration and peripherality of vowels in the F1/F2 space. Because he did not report vowel duration, it was unclear whether Wright's original result was a consequence of the durations of the vowels in high- and low-neighborhood density words; Munson and Solomon showed that it was not. In addition, Munson and Solomon conducted a second experiment with a new set of stimuli in which density did not co-vary with word frequency. In this experiment, participants read 80 words that varied in both word frequency and neighborhood density. Again, vowels in words with high neighborhood densities were more expanded than those in low-neighborhood density words. Munson and Solomon also found that high-frequency words were produced with less-expanded vowel spaces than low-frequency words. These factors did not interact significantly, although the absolute difference in vowel-space dispersion as a function of neighborhood density was greater for low-frequency items than for high-frequency ones.

There are a number of explanations for the influence of word frequency and phonological neighborhood density on the acoustic characteristics of vowels. Some of these appeal to aspects of modular, feed-forward theories of speech production and phonetic implementation. Other explanations appeal to aspects of speech production that generally have been considered outside of the realm of these theories. An example of a latter type of explanation stems from the *Hyperspeech and Hypospeech* (H & H) theory of Lindblom (1990). Lindblom proposed that articulation exists on a continuum from highly effortful, clearly articulated speech (*hyperspeech*) to less effortful, less clearly produced speech (*hypospeech*). Lindblom proposed that talkers use hyperspeech when they presume a listener to need an atypically clear signal. Hypospeech is used in other situations, and serves to minimize the talker's effort. This theory accounts for observations that articulation may be dependent on unique speaking demands, such as the hyperarticulation that occurs when speaking to someone who is presumed to need an atypically clear speech signal, or in the presence of increased background noise (Bradlow 2002). In this framework, the effect of neighborhood density on vowel-space dispersion would occur because of talkers' tacit knowledge of the perceptual difficulties associated with high-

neighborhood density and low-frequency words. More-dispersed vowel spaces in these types of words would presumably reflect a partial compensation to the perceptual difficulties that they challenge, as expanded vowel spaces have been shown to be associated with more intelligible speech (Bradlow, Toretta, and Pisoni 1996). Indeed, this is the explanation that Wright originally advanced for his findings. This explanation does not require a fundamental modification to modular, feed-forward theories of speech production and phonetic implementation. Rather, it would only require the assumption that speakers are able to monitor the impact of environmental variables on communication tasks. This assumption is not incompatible with current architectures (Levelt 1989).

A second class of explanations posits that the influence of word frequency and neighborhood density on vowel-space dispersion is due to their influence on long-term representations. Most models of spoken-word recognition posit some form of interactive activation, such that a word that is recognized and retrieved faster and more accurately is posited to have a higher resting activation level than one that is not. As pointed out by Pierrehumbert (2002), current modular feed-forward models of production do not make reference to resting activation levels and their influence on articulation. However, they would require only minor modifications to do so, by assuming that a gradient value representing the word's resting activation level could be passed into the phonological buffer. This gradient value would scale the articulation plan for that word appropriately, with higher resting activation levels associated with a value scaling down the articulation.

There is, however, a third explanation for the effects of frequency and neighborhood density on vowel-space dispersion that is subtly different from the previous one. This explanation is suggested by Bell et al.'s (2003) study of the influence of real-time planning difficulties on vowel articulation. Bell et al.'s study was motivated by the work of Shriberg (1995) and Fox Tree and Clark (1997), who noted that words surrounding spontaneous disfluencies tend to be hyperarticulated. For example, the word *the* tends to be pronounced with the full, citation-form vowel /i/ when it is adjacent to a disfluency, such as a hesitation, filler, repetition, or revision. Bell et al. examined variation in the pronunciation of ten commonly occurring function words in a large corpus of phonetically transcribed spontaneous conversations. One of the dependent measures that these authors analyzed was vowel reduction, as assessed by the phonetic transcriptions. Bell et al. found a strong effect of disfluencies on the

occurrence of vowel reduction in this corpus: words occurring adjacent to disfluencies were more likely to contain their citation-form vowel (e.g., [æ] for *and*, [ɪ] for *in*) rather than a reduced (e.g., [ə], [ɪ̃]) or deleted vowel.

Bell et al.'s finding suggests that there is a relationship between ease of lexical access and hyperarticulation. Bell et al. did not study the influence of disfluencies on the articulation of vowels in content words. Unlike function words, content words do not show the categorical alternations in vowel quality that are characteristic of some content words (i.e., the pronunciation of [ði] or [ðə] for *the*). Previous studies may not have detected an influence of disfluencies on content-word articulation because they utilized transcription data, rather than acoustic measures. Given Bell et al.'s findings, we may see relationships between other hyperarticulation phenomena and ease of lexical access. For example, the influence of word frequency and neighborhood density on the acoustic characteristics of vowels may represent the influence of real-time planning on variation in articulation. Just as content words are hyperarticulated when adjacent to disfluencies, the expanded vowel spaces of the low-frequency and high-neighborhood density words may reflect the fact that they occur in words that are difficult to access, rather than an active modification to maintain clarity.

This explanation differs fundamentally from hypotheses about resting activation levels. It proposes that the hyperarticulation arises not from characteristics of the long-term representation, but from the process of lexical access itself. That is, the gradient scaling factor referred to above would reflect the ease with which an item was accessed from long-term memory, rather than its resting activation level, although the two are presumed to be correlated in most speaking situations. This was the solution hypothesized by Pierrehumbert (2002) to account for the influence of frequency and neighborhood density on phonological and phonetic variation.

Pierrehumbert's proposal was made under the assumption that neighborhood density and frequency both influence ease of lexical access. Indeed, a great deal of evidence supports the hypothesis that frequency affects lexical access. In contrast, relatively less work has examined the influence of neighborhood density on speech production. The results of these studies have provided contradictory evidence. Vitevitch (2002) found that pictures representing objects with high-neighborhood density names were named with shorter latencies than those representing objects with low-neighborhood density names. He also found that high-neighborhood

density words elicited fewer errors in a tongue-twister task than low-neighborhood density words. In contrast, Amico, Charles-Luce, and McEldowney (2004) showed that adults read high-density words less quickly and with longer durations than low-density words in different conditions of semantic priming.

The purpose of this study is to examine experimentally whether planning difficulties mediate the influence of neighborhood density and word frequency on vowel articulation. This is accomplished by examining this relationship in two conditions, one in which lexical access is facilitated, and one in which it is not. If planning difficulties mediate the effects of word frequency and neighborhood density on vowel dispersion, then we should see a reduced or absent effect in the facilitated condition relative to the non-facilitated condition.

In this paper, ease of lexical access is varied using a modification of a paradigm developed by Balota and Chumbly (1985), developed to assess whether the influence of word frequency on reaction times resulted from its influence on the lexical access and recognition² or the phonological planning stages of production. Balota and Chumbly constructed a task in which delay intervals were enforced between the orthographic presentation of words varying in frequency, and a response signal indicating that participants should produce the word. Balota and Chumbly found that the influence of word frequency on reading times was present, albeit reduced, in the delayed-response conditions (those in which at least a 1000 ms delay was enforced). This result suggested to Balota and Chumbly that the facilitative effect of word frequency was due to its influence on both lexical access and phonological encoding.

This paper uses a modified version of this paradigm. In the *immediate-response* condition, participants were presented with a word simultaneous with a response signal. In the *long-delay* condition, an interval of 1000 ms was enforced between the presentation of the word and the onset of the participants' responses. Following Balota and Chumbly, productions in the immediate-response condition are presumed to reflect the combined influence of lexical access and phonological encoding on articulation. Those in the long-delay condition are presumed to have occurred after lexical access has taken place, and reflect the influence of phonological encoding only. If the effects of neighborhood density and word frequency on vowel articulation are due to the mediating effect of lexical access, then we should expect to see a reduced effect of these factors in the long-delay condition. In addition, we examine the relation between lexical access and

vowel-space dispersion by examining correlations between vowel-space dispersion measures and response latency, which is presumably related to the relative difficulty of lexical access.

2. Experiment

2.1. Participants

Fifteen individuals from the University of Minnesota community participated in this study. Nine of the participants were women. They ranged in age from 20;5 (years;months) to 25;4 ($M=22;3$, $SD=1;3$). All participants were native speakers of English. Participants reported no history of speech, language, or hearing disorders. All participants passed a pure-tone hearing screening at 0.5, 1, 2, and 4 kHz at 20 dB HL (ANSI 1989) in one ear, or reported a normal audiometric evaluation within the past 3 months. All had normal or corrected-to-normal vision. Subjects received \$5.00 for participating in this experiment, which lasted approximately 30 minutes.

2.2. Stimuli

The stimuli used in this experiment are presented in Table 1. The stimuli consisted of 20 quadruplets of high- and low-neighborhood density, high- and low-frequency CVC words. Word frequency was taken from Kucera and Francis (1967). Kucera and Francis measured the frequency of words in written texts; frequencies from this corpus are often used in psycholinguistics research (Pisoni et al. 1985). Neighborhood density was taken from the values in the Hoosier Mental Lexicon (HML, Pisoni et al. 1985). Neighborhood density was measured as the number of real words that could be created by adding, deleting, or substituting a phoneme in a target word, weighted to the frequency of the words in the neighborhood.³ Stimuli were selected by taking a median split of the entire set of CVC words in the HML based on word frequency, then taking a median split within each of the two lists based on frequency-weighted neighborhood density. A multivariate analysis of variance showed that the differences in neighborhood density were statistically equivalent for high- and low-

frequency words, and the differences in word frequency were statistically equivalent for high- and low-neighborhood density words. Each quadruplet contained the same vowel. Within each quadruplet, voicing of the final consonant was consistent across the four items. Moreover, close attention was paid to balancing the manner and voicing of the consonants comprising the four word types across the 20 quadruplets, as these have been shown in previous work to influence vowel duration. The four lists of words did not differ significantly in the manner of the initial or final consonants or the voicing of the initial consonant ($\chi^2[6] = 1.7, p > 0.05$ for initial consonant manner, $\chi^2[6] = 4.2, p > 0.05$ for final consonant manner; $\chi^2[3] < 1, p > 0.05$ for initial consonant voicing). The lists were fully balanced for final consonant voicing. Previous research has shown that back vowels are subject to fronting in the environment of alveolar obstruents. Multiple χ^2 tests showed that the four lists did not differ in the distribution of alveolar versus non-alveolar consonants in word-initial position, ($\chi^2[3] = 3.8, p > 0.05$) or final position, ($\chi^2[3] = 4.8, p > 0.05$). Importantly, the small differences in occurrence of alveolar consonants that did occur across conditions were not systematically related to neighborhood density. Instead, they appeared to be related to the higher occurrence of final alveolar consonants in high-frequency words than in low-frequency words.

Table 1. Stimuli

Vowel	HF/HD	HF/LD	LF/HD	LF/LD
a	GOT	DOCK	DOT	MOP
a	LOCK	ROCK	KNOCK	SOCK
a	POT	TOP	COT	COP
æ	BAD	BAG	DAD	DAB
æ	SAD	SANG	FAD	SAG
æ	HALF	LAUGH	MASH	RASH
ɛ	GET	DEATH	DEBT	DEAF
ɛ	BET	CHECK	PET	PEP
eɪ	SAVE	GAVE	CAGE	BATHE
eɪ	GAME	GAIN	DAME	BABE
eɪ	TAPE	SHAPE	CAKE	NAPE
i	BEAT	BEACH	BEAK	LEACH
i	TEAM	SCENE	KEEN	SIEGE
i	MEAN	BEAM	BEAN	GENE

10 Benjamin Munson

ou	NOTE	WROTE	MOAT	ROPE
ou	ROSE	KNOWN	MOAN	ROBE
ou	BONE	LOAN	ROAM	DOMe
u	YOUTH	SUIT	BOOT	HOOT
u	MOON	ROOM	WOMB	TUNE
u	FOOT	PUT	HOOK	HOOF

Note: HF=high frequency, LF=low frequency, HD=high density, LD=low density

2.3. Procedures

Testing took place in a double-walled sound-attenuated room. Stimuli were presented on a 17-inch video monitor. During each trial, the sequence *XXX*, written in 18-point courier font, appeared at the center of the video screen for 1.5 seconds. This display was then replaced with the target word, written in 36-point courier font. In half of the trials, a 100 ms. 1000 Hz pure-tone was played concurrent with the presentation of the word. In the other half of the trials, the tone was played 1000 ms. after the word was presented. Participants were instructed to read the word as soon as they heard the tone. Each word was presented at each delay interval three times in randomized order, for a maximum total of 480 productions per subject. The experiment was preceded by a practice block, which consisted of 8 items that were not used in the experiment. The experiment was self-paced; participants pressed a button on a button-box to advance items.

The data were recorded on a Marantz CDW300 CD recorder, at a sampling rate of 44.1 kHz, with 16-bit quantization, and an anti-aliasing filter with a cutoff frequency of 22.05 kHz. Participants wore an AKG-C420 head-mounted condenser microphone, attached to a Rolls phantom power source.

2.4. Analysis

Data were transferred to a personal computer for acoustic analysis. The Praat v. 4.0.7 signal-processing software (Boersma and Weenink 2002) was used for analysis. Prior to analysis, tokens were excluded if they were judged to be produced disfluently, if they occurred prior to the response signal, or if there was a reading error. Errors occurred in 1.7% of items.

The onset and offset of the vowel was marked in Praat, using standard segmentation criteria (see Munson and Solomon 2004, for a detailed description of segmentation criteria). A second individual re-measured 50 tokens (1% of the data) to assess measurement reliability. These tokens were equally distributed among the 15 talkers, and among the different experimental conditions. The range in duration between these measures and the original measures was -11 ms. to 12 ms.; the average absolute difference was 10 ms.

Vowel formants were measured automatically using an LPC algorithm. Values that were greater than 100 Hz away from the mean values reported by Hillenbrand et al. (1995) were hand-checked for accuracy and re-measured where necessary. As in Wright (2004) and Munson and Solomon (2004) formant frequencies were converted to the Bark scale prior to analysis. Vowel-space dispersion was calculated using a measure from Bradlow, Toretta, and Pisoni (1996), as the mean Euclidian distance from the center of the participants' F1/F2 space. Dispersion was calculated separately for the four lists of words in both the immediate-response and long-delay conditions.⁴

In addition to vowel-space dispersion, we measured vowel duration and response latency, so that we could assess the extent to which vowel-space dispersion was related to duration and relative difficulty of lexical access within and across the different experimental conditions. Vowel durations were measured automatically based on the labels used to extract vowel formants. Response latency was measured as the interval from the offset of the response tone to the onset of acoustic evidence of the response. These were hand-marked and extracted using Praat.

2.5. Results

Prior to completing statistical analyses, Kolmogorov-Smirnov tests of normality were computed to ensure that data met normality assumptions required to calculate parametric statistics. In cases where the data did not meet these assumptions, the appropriate nonparametric statistics were used.

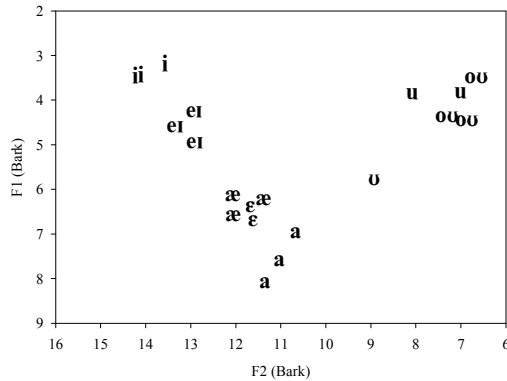


Figure 2. Representative Vowel Space for High-Neighborhood density Low-Frequency Words

Finally, there was a significant three-way interaction among delay condition, frequency, and neighborhood density, $F[1,14] = 4.5$, $p = 0.05$. This interaction is shown graphically in Figures 3 and 4, which plot average vowel-space dispersion in high- and low-frequency words in the immediate-response and long-delay conditions. High-density words are shown in Figure 3; low-density words are shown in Figure 4.

The three-way interaction was examined with post-hoc tests of simple main effects. Results of these tests found that the influence of neighborhood density on vowel-space dispersion was robustly present in both the immediate-response and long-delay conditions ($F[1,14] = 8.9$, $p < 0.01$ for the long-delay condition; $F[1,14] = 5.8$, $p < 0.01$ for the immediate-response condition). The effect of frequency, however, was significant for the immediate-response condition only ($F[1,14] = 18.7$, $p < 0.01$). The effect did not achieve significance in the long-delay condition ($F[1,14] = 3$, $p > 0.05$). Although the interaction between frequency and neighborhood density did not achieve statistical significance in the long-delay condition, inspection of Figure 3 shows that the influence of frequency on vowel-space dispersion in this condition was greater for low-density words (mean difference = 0.152 Bark) than for high-density words (mean difference = 0.023 Bark).

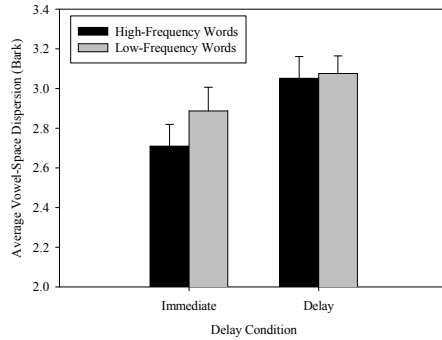


Figure 3. Average Vowel-Space Dispersion for Vowels in HD words

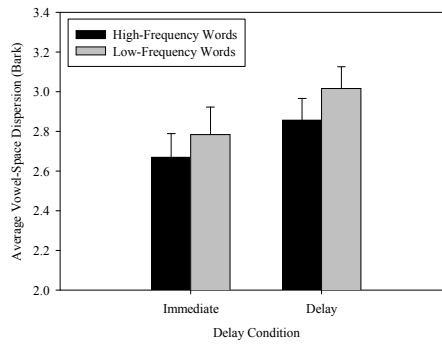


Figure 4. Average Vowel-Space Dispersion for Vowels in LD Words.

Figures 1 and 2 suggest that not all vowels were affected equally by the experimental manipulations. To examine this systematically, a series of multivariate analyses of variance (MANOVAs) was conducted examining the influence of word frequency, neighborhood density, and delay condition on each of the individual vowels (/i/, /e/, /ɛ/, /æ/, /a/, /o/, /ʊ/, and /u/). The dependent measure in these MANOVAs was the average bark-transformed F1 and F2 frequencies. No influence of frequency, density, or delay condition was found for the vowel /i/. A pattern of significant effects mirroring that of the summary measure of vowel-space dispersion was found for F2 only for the vowels /eɪ/, /oʊ/, and /ʊ/. That is, the F2 values for these vowels were affected by frequency and density, and the influence of frequency on formant values did not achieve statistical significance in the long-delay condition. A pattern of significant effects mirroring that of the summary measure of vowel-space dispersion was found for the F1 and

F2 of the vowels / ϵ /, / æ /, / a /, and / u /. This analysis suggests the influence of frequency, density, and delay interval on vowel-space dispersion is due neither to the influence of these variables on the acoustic characteristics of a single vowel, nor to it influencing all vowels equally.

These findings are broadly consistent with Balota and Chumbly's (1985) data, which showed the difference in response latencies between high- and low-frequency words to be decreased in a long-delay condition relative to an immediate-response condition. This result does not, however, suggest that the influence of neighborhood density on vowel-space dispersion is the result of difficulties in lexical access. This influence was reliably present in both the immediate-response and long-delay conditions. In fact, the raw difference between high- and low-neighborhood density words was slightly larger in the 'facilitated' long-delay condition relative to the 'non-facilitated' immediate-response condition. However, the data are compatible with the idea that the influence of word frequency on usage is, at least in part, due to the effect of real-time lexical access. The influence of this variable was attenuated in the long-delay condition, particularly for high-density words. This relationship is explored further in 2.5.4.

2.5.2. Duration

The next analysis examined whether the influence of delay, frequency, and neighborhood density on vowel-space dispersion was related to the duration of the vowels in the different experimental conditions. Previous work by Moon and Lindblom (1994) found positive correlations between vowel-space dispersion and vowel duration. Munson and Solomon (2004) found that vowels in high-frequency words were shorter than the same vowels in low-frequency words; however, correlations between vowel duration and vowel-space dispersion were not significant, suggesting that frequency influenced the two variables independently. However, Munson and Solomon examined vowels produced in an immediate-response condition only. The influence of delay on vowel-space dispersion in the current study may have been due to differences in vowel duration between the two conditions.

Vowel duration was measured automatically in Praat. Average vowel durations were submitted to a three-factor within-subjects ANOVA, using the same factors as in the ANOVA for vowel-space dispersion. A significant main effect of frequency was found, $F[1,15]=19$, $p < 0.01$. As in

Munson and Solomon (2004), vowels in low-frequency words were produced with slightly longer durations (215 ms) than those in high-frequency words (208 ms). There was also a small but significant effect of delay on vowel duration, $F[1,14] = 6.2$, $p < 0.05$. Vowels in words produced in the long-delay condition were slightly longer (215 ms) than those in the immediate-response condition (211 ms). Neighborhood density did not affect vowel duration, nor were there any interactions among factors.

Despite the small absolute size of these differences, there was remarkable consistency in the direction of these effects across the 15 participants. Although the influence of frequency and delay condition on vowel duration are exceedingly small, the results of this analysis cannot rule out that the effects of frequency and delay condition on vowel-space dispersion are potentially attributable to differences in vowel duration. These possible associations are explored in section 5.2.4 using multiple regression. Mean values for the eight experimental conditions are presented in Table 2.

Delay Condition	High Frequency		Low Frequency	
	High Density	Low Density	High Density	Low Density
Immediate Response (SD)	207 (43)	206 (45)	211 (45)	210 (48)
Long-Delay (SD)	212 (49)	211 (52)	215 (46)	214 (49)

Table 2. Vowel Duration in the Eight Experimental Conditions

2.5.3. Latency

The next analysis examined whether response latency differed systematically as a function of frequency, neighborhood density, and delay interval. Previous research on the influence of frequency and neighborhood density on vowel-space dispersion did not examine this variable. In this investigation, response latency measures were used two ways. First, they

were used to ensure that participants responded to the experimental manipulations made in this study similarly to the way that participants in previous research have. That is, we used these data to verify that word frequency and neighborhood density affected processing times as they had in previous studies (e.g., Luce and Pisoni 1998) and whether processing times in the long-delay condition indeed showed a facilitative effect of allowing 1000 ms to elapse between the presentation of the stimulus and the onset of the response. Second, they were used as a continuous measure of difficulty of lexical access. That is, they were used to examine whether variability in vowel-space dispersion within and across conditions was due to variability in the ease of lexical access, as indicated by measures of response latency.

As predicted, a large, significant main effect of delay interval on response latencies was found, $F[1,14] = 27.6$, $p < 0.01$. Mean response latency in the long-delay condition was 357 ms, as opposed to 448 ms in the immediate-response condition. As in Balota and Chumbly (1985), we interpret this 91 ms difference to reflect the additional time required to conduct lexical access in the immediate-response condition.

There was also a three-way interaction among frequency, neighborhood density, and delay condition, $F[1,14] = 4.5$, $p = 0.05$. This three-way interaction arose because there was a significant two-way interaction between frequency and neighborhood density in the immediate-response condition but not the long-delay condition. In the immediate-response condition, response latencies mirrored those seen in previous research (e.g., Vitevitch and Luce 1999): high-frequency, low-neighborhood density words were read with shorter response latencies ($M = 441$ ms) than low-frequency, high-neighborhood density words ($M = 453$ ms). This difference was significant in a Wilcoxon signed ranks test, $z = -1.99$, $p < 0.05$. In contrast, high-frequency, low-neighborhood density words were read with slightly longer latencies ($M = 365$ ms) than low-frequency, high-neighborhood density words ($M = 359$ ms) in the long-delay condition. This difference, however, was not statistically significant in a Wilcoxon signed ranks test, $z = 1.1$, $p > 0.05$.

Table 3. Response Latency in the Eight Experimental Conditions

Delay Condition	High Frequency		Low Frequency	
	High Density	Low Density	High Density	Low Density
Immediate Response (SD)	447 (181)	442 (183)	453 (180)	452 (181)
Long-Delay (SD)	349 (151)	365 (167)	359 (158)	353 (137)

2.5.4. Relations among measures

The last set of analyses examined whether the effects of frequency, neighborhood density, and delay on vowel-space dispersion were mediated by vowel duration and response latency. This was examined using a series of hierarchical multiple regressions. In each hierarchical regression, the dependent measure was vowel-space dispersion. The independent measures were (a) a continuous measure of either vowel duration or response latency, and (b) a dummy-coded variable representing one of the experimental conditions.

The first regression analysis examined whether the influence of delay interval on vowel-space dispersion was due to the differences in vowel durations between the two conditions. For this analysis, each subject's vowel-space dispersion and vowel duration measures for the four frequency/neighborhood density combinations were averaged separately for the immediate-response and long-delay conditions. On the first step, duration did not account for a significant proportion of variance in vowel-space dispersion ($B < 0.001$, $SE B = 0.001$, $\beta = 0.022$, $t < 1$, $p > 0.05$). However, on the second step, a binary dummy-coded variable representing delay interval did account for a significant proportion of variance, ($B = 0.237$, $SE B = 0.080$, $\beta = 0.264$, $t = 2.96$, $p < 0.01$). Together, these variables predicted a significant proportion of variance in vowel-space dispersion, $F[2,27] = 4.46$, $p < 0.05$.

The second analysis examined whether neighborhood density affected vowel-space dispersion beyond the effect that it had on lexical access, as measured by response latencies. For this analysis, each subject's vowel-space dispersion and response latency measures for the four

frequency/delay condition combinations were averaged separately for the high- and low-density words. On the first step, latency was marginally significantly related to vowel-space dispersion, ($B = 0.0005$, $SE B < 0.001$, $\beta = 0.214$, $t = 1.96$, $p = 0.09$). On the second step, a binary dummy-coded variable representing neighborhood density accounted for a significant proportion of variance, ($B = 0.217$, $SE B = 0.106$, $\beta = 0.256$, $t = 2.05$, $p < 0.05$). Together, these variables predicted a significant proportion of variance in vowel-space dispersion, $F[2,27] = 3.59$, $p < 0.05$.

The final regression examined whether frequency affected vowel-space dispersion beyond the effect that it had on lexical access, as measured by response latencies. For this analysis, each subject's vowel-space dispersion and response latency measures for the four density/delay condition combinations were averaged separately for the high- and low-frequency words. On the first step, response latency accounted for a significant proportion of variance in vowel-space dispersion ($B = 0.0005$, $SE B < 0.001$, $\beta = 0.201$, $t = 2.24$, $p < 0.05$). However, on the second step, a binary variable dummy-coding frequency did not account for a significant proportion of variance, ($B = -0.117$, $SE B = 0.080$, $\beta = -0.130$, $t = 1.45$, $p > 0.05$). Together, these variables predicted a significant proportion of variance in vowel-space dispersion, $F[2,27] = 3.30$, $p < 0.05$.

These regression analyses suggest three things: first, that the effect of delay interval on vowel-space dispersion is not an artifact of the longer vowel durations noted in that condition; second, that the differences in vowel-space dispersion as a function of neighborhood density are not an artifact of lexical access (as measured by response latency); and third, that the differences in vowel-space dispersion as a function of word frequency are likely related to lexical access. The latter two findings complement the results of the ANOVAs, which found a significant effect of neighborhood density in both the facilitated and non-facilitated lexical-access conditions; but a significant effect of frequency in the non-facilitated condition only.

3. Conclusion and Discussion

3.1. Hypothesis revisited

The results of this experiment do not support the hypothesis that the influence of phonological neighborhood density on vowel-space dispersion

is a consequence of difficulties in accessing words residing in dense phonological neighborhoods. This conclusion is based on two different findings. First, the influence of neighborhood density on vowel-space dispersion was not reduced in a condition in which lexical access was facilitated. When we enforced a 1000 ms delay between the presentation of a word to be read and the onset of the talker's response, response latencies decreased by almost 100 ms, suggesting that lexical access was significantly facilitated in that condition. However, this facilitation in lexical access did not remove or attenuate the influence of neighborhood density on vowel-space dispersion. Second, regression analysis found that a variable dummy-coding neighborhood density predicted a significant proportion of variance in vowel-space dispersion beyond what was accounted for by a measure of lexical access, response latency.

As in Munson and Solomon (2004), this study found an effect of word frequency on vowel-space dispersion. This effect may be related at least partially to the influence of real-time lexical access processes on production. Again, two findings support this hypothesis. First, this effect was greatly attenuated in the long-delay condition. As in Balota and Chumbly (1985), performance in long-delay conditions is presumed to reflect less of the contribution of lexical access than performance in immediate-response conditions. Second, this effect was predicted completely by a continuous measure of ease of lexical access, response latency.

3.2. Implications and Ongoing Research

The findings in this paper suggest that modular, feed-forward theories of speech production need to be modified two ways to account for the effects of frequency and neighborhood density on vowel production. First, the influence of neighborhood density on vowel production is best accounted for by claiming that neighborhood density is coded during speech production. This coding may reflect a property of the long-term representation, or it may reflect summed activation that is spread to words in the target word's phonological neighborhood. In contrast, the influence of word frequency on vowel articulation can be accounted for by positing that speech production takes into account the ease with which an item was accessed from memory.

The fact that word frequency and neighborhood density influence vowel-space dispersion differently may be due to the different cognitive processes that cause lexical access difficulties for high-density and low-frequency words. The effect of word frequency on lexical access is typically attributed to resting activation levels. High-frequency words are associated with higher resting activation levels than low-frequency words. Consequently, they require less input activation (i.e., weaker activation from high-level conceptual representations) than low-frequency words to be activated. The influence of density, in contrast, is ascribed to processes that happen after a lexical item has been activated. High-density words are presumed to spread activation to the many phonologically similar words in their neighborhood. These words in turn must be suppressed before the target word can be produced or perceived. The access difficulties associated with high-density words reflect the number of words in the neighborhood that must be suppressed before the target word is articulated. The interaction between frequency and delay may reflect the fact that the 1000 ms delay interval allowed enough time for the low-frequency words to receive activation equivalent to that of high-frequency words, but is not enough time to suppress the activation associated with the neighbors of high-density words. If so, we predict that the effect of neighborhood density would attenuate at even longer delay intervals than examined in this study.

Our results support the hypothesis that the influence of word frequency on phonetic variation (as shown by Bybee 2001 and Jurafsky et al. 2002, among others) is due to real-time planning processes. That is, the influence of word frequency on phonetic outcomes may be related to its influence on lexical access, rather than to its influence on aspects of words' long-term phonological representations. Future research utilizing laboratory manipulations of ease of lexical access should examine other frequency-related phonetic alternations (e.g., the influence of word frequency on weak-syllable deletion, as in Bybee 2001) to examine the extent to which frequency effects on phonetic outcomes can be attributed to real-time processing.

The data in this paper, however, do not rule out the possibility that the influence of neighborhood density is due to a process that occurs separate from the higher-level cognitive-linguistic processing that underlies speech production. Perhaps this effect represents an active attempt to maximize the clarity of perceptually difficult words, as proposed by Wright (2004). If this were the case, then we would expect that the difference in vowel-space

dispersion would be exaggerated in difficult listening situations and, potentially, attenuated in optimal listening situations.

There is another alternative account of this phenomenon that was not considered by Wright, Munson and Solomon, or in the current paper. This explanation is categorically different from those proposed above, in that it appeals to exemplar models of production, rather than modular feed-forward models. Namely, this effect may reflect speakers' selective encoding of clearly produced exemplars of the easily confusable high-neighborhood density words, as proposed by Pierrehumbert (2002). Exemplar models of production are outlined by Pierrehumbert (2002). Exemplar models of perception propose that people's long-term representations for linguistic items code a great deal of token-specific detail. Episodic detail is encoded provided that the token that is being perceived can be discriminated from tokens already encoded in memory. Exemplar models of production propose that similarly detailed episodic representations influence speech production. A full discussion of exemplar models of production is beyond the scope of this paper. However, these models are constructed such that they provide a unified account both for cases in which there is an apparent influence of episodic perceptual detail on production (e.g., Goldinger and Azuma 2004), as well as for the cases in which talkers apparently ignore such detail.

Pierrehumbert (2002) proposed that talkers select an exemplar of a word from memory, and then use it (averaged with similar exemplars) as a goal for speech production. In her framework, the influence of density on production might arise indirectly, as a consequence of the influence of density on perception. All words, regardless of whether they are high- or low-density, are produced with some degree of deviation from the target form due to random variation in the articulatory system. When a low-density word is produced with random deviation, listeners are likely to encode it, given that there are few phonetically similar words in memory with which it might be confused. When a high-density word is produced with deviation, however, it is less likely to be encoded in memory, as it might be confused with other words in the neighborhood. If these episodic traces were to be used as production goals, then the vowel-space differences between high- and low-density words would arise naturally from the impact that neighborhood density has on perception.

Pierrehumbert's selective-encoding hypothesis makes predictions about the emergence of the influence of neighborhood density on vowel-space dispersion in children. Corpus studies of children's lexicons have shown

that young children are less likely than older children and adults to know a cohort of words constituting a high-density neighborhood (e.g., Charles-Luce and Luce 1995). Recently, Munson, Swenson, and Manthei (2005) showed that these differences have predictable effects on processing times: young children show a much smaller influence of neighborhood density on repetition latencies for real words than older children. Following Pierrehumbert, we would predict that children's lexicons would support more encoding of 'sloppy' productions of all words, even those that would be in high-density neighborhoods for adults, given that children's lexicons do not demonstrate the variation in density that adults' do. Hence, young children should show a smaller influence of density on vowel-space dispersion than older children or adults.

There is some support for this hypothesis from post-hoc analyses of a subset of the data collected by Munson, Swenson, and Manthei. Vowel-space dispersion was measured for a subset of five high-density words (*bath, bone, cook, kick, ride*) and five low-density words matched for phonetic content (*catch, home, foot, fish, time*) produced by 12 younger children (approximately 4-year-old) and 12 older children (approximately 7-year-old). As in Munson, Swenson, and Manthei, response latency and vowel duration did not differ as a function of neighborhood density in the younger children in a Wilcoxon signed ranks test ($z < 1$, $p > 0.05$ for both tests). In the older children, these variables did differ significantly, ($z = -2.21$, $p < 0.03$ for duration, $z = -2.05$, $p < 0.05$ for latency). There was no influence of neighborhood density on vowel-space dispersion (measured using Bradlow, Toretta, and Pisoni's method), for the younger children ($z < 1$, $p > 0.05$), but the older children showed an adult-like pattern, ($z = -1.95$, $p = 0.05$). This result is predicted by Pierrehumbert's selective-encoding hypothesis. Currently, we are following up on this result with a prospective study utilizing more stimuli and a larger group of participants. A robust finding that age mediates the influence of neighborhood density on vowel-space dispersion would support Pierrehumbert's hypothesis that selective encoding of clearly produced high-density words partially underlies the influence of neighborhood density on vowel-space dispersion.

Acknowledgements

This research was funded by NIH grant R03 DC005702 to Benjamin Munson. I thank Mary Beckman, Stefan Frisch, Matt Goldrick, Robert Kirchner, Janet Pierrehumbert, Nancy Pearl Solomon and Richard Wright for useful comments on

this work. As always, I acknowledge the hard work of Pauline Welby in writing Praat scripts to automate the acoustic analyses, and for useful comments on my work in progress.

Notes

1. This discussion makes a distinction between models of the process in which the semantic representation of a word leads to the selection of the phonological units that comprise it (*speech production*) and the process of translation of those into continuous plans for articulation (*phonetic implementation*).
2. As in Balota and Chumbly (1985), we use the term *lexical access* to encompass both word recognition processes and the process of accessing a word's lemma.
3. Munson and Solomon (2004) reported that the criterion used to determine high- and low-density words was raw ND, when, in fact, the criterion was frequency-weighted ND. Multivariate ANOVA indicated significant differences between the high- and low-density lists for both frequency-weighted ND ($F[1,78] = 55.6, p < 0.001$) and raw ND ($F[1,78] = 5.1, p = 0.026$). The high- and low-density lists were completely separated in terms of frequency-weighted ND, whereas some overlap occurred between lists in terms of raw NDs of individual items.
4. As in Munson and Solomon (2004), the use of Bark-transformed formant values and a fully within-subjects statistical design obviated the need for further statistical normalization of formant-frequency values.

References

- Amico, Patricia, Jan Charles-Luce, and Elizabeth McEldowney
2004 The relation between semantics and lexical properties in spoken word production [abstract]. *Journal of the Acoustical Society of America* 114: 2429.
- American National Standards Institute
1989 *Specifications for Audiometers*. Washington, DC: American National Standards Institute.
- Balota, David A., and James I. Chumbly
1985 The locus of word-frequency effects in the pronunciation task. *Journal of Memory and Language* 24: 89-106.
- Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea

- 2003 Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113: 1001-1024.
- Boersma, Paul, and David Weenink
2002 *Praat version 4.0.7 [Computer Software]*. Institute of Phonetic Sciences, Amsterdam, The Netherlands.
- Bradlow, Ann R.
2002 Confluent talker- and listener-related forces in clear speech production. In *Papers in Laboratory Phonology VII*, Carlos Gussenhoven and Natasha Warner (eds.), 241-273. New York: Mouton de Gruyter.
- Bradlow, Ann R, Gina Torretta, and David Pisoni
1996 Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20: 255-272.
- Bybee, Joan
2001 *Phonology and Language Use*. Cambridge, UK: Cambridge University Press.
- Charles-Luce, Jan, and Paul A. Luce
1995 An examination of similarity neighborhoods in young children's receptive vocabularies. *Journal of Child Language* 22: 727-735.
- Fox Tree, Jean E, and Herbert H. Clark
1997 Pronouncing "the" as "thee" to signal problems in speaking. *Cognition* 62: 151-167.
- Goldinger, Steven A., and Tamiko Azuma
2004 Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review* 11: 716-722.
- Goldinger, Steven A., David B. Pisoni, and Paul A. Luce
1996 Speech perception and spoken word recognition: research and theory. In *Principles of Experimental Phonetics*, Norman J. Lass, (ed.), 277-327. St. Louis: Moseby.
- Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler
1995 Acoustic characteristics of American English Vowels. *Journal of the Acoustical Society of America* 97: 3099-3111.
- Jurafsky, Daniel, Alan Bell, and Cynthia Girand
2002 The role of the lemma in form variation. In *Papers in Laboratory Phonology VII*, Carlos Gussenhoven and Natasha Warner (eds.), 1-34. New York: Mouton de Gruyter.
- Keating, Patricia A.
1990 The window model of coarticulation: articulatory evidence. In *Papers in Laboratory Phonology I*, J. Kingston and M. Beckman (eds.), 451-470. Cambridge: Cambridge University Press.

- Keating, Patricia A., Taehong Cho, Cecile Fougeron, and Chai-Shun Hsu
2004 Domain-initial articulatory strengthening in four languages. In *Phonetic Interpretation: Papers in Laboratory Phonology 6*, John Local, Richard Ogden, Roslyn Temple (eds.), 143-161. Cambridge: Cambridge University Press.
- Kucera, Henry, and Nelson Francis
1967 *A Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Levelt, Willem J.M.
1989 *Speaking: From Intention to Articulation*. Cambridge, MA: MIT press.
- Levelt, Willem J.M., Ardi Roelofs, and Antje S. Meyer
1999 A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1-75.
- Lindblom, Bjorn
1990 Explaining phonetic variation: a sketch of the H and H theory. In *Speech Production and Speech Modeling*, W. Hardcastle and A. Marchal (eds.), 403-439. Dordrecht: Kluwer.
- Luce, Paul A., and Pisoni B. David
1998 Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19: 1-36.
- Moon, Seung-Jae, and Bjorn Lindblom
1994 Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America* 96: 40-55.
- Munson, Benjamin
2004 Variability in /s/ production in children and adults: evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research* 47: 69-80.
- Munson, Benjamin, and Nancy Pearl Solomon
2004 The influence of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47: 1048-1058.
- Munson, Benjamin, Cyndie L. Swenson, and Shayla C. Manthei
2005 Lexical and phonological organization in children: evidence from repetition tasks. *Journal of Speech, Language, and Hearing Research* 48: 108-124.
- Pierrehumbert, Janet B.
2002 Word-specific phonetics. In *Papers in Laboratory Phonology VII*, Carlos Gussenhoven, Tanya Rietvelt, and Natasha Warner (eds.), 101-139. New York: Mouton de Gruyter.
- Pisoni, David B., Howard Nusbaum, Paul A. Luce, and Louisa Slowiaczek

- 1985 Speech perception, word recognition, and the structure of the lexicon. *Speech Communication* 4: 75-95.
- Shattuck-Hufnagel, Stefanie
- 1992 The role of word structure in segmental serial ordering. *Cognition* 42: 213-259.
- Shriberg, Elizabeth
- 1995 Acoustic properties of disfluent repetitions. *Proceedings of the International Congress of Phonetic Sciences* 4: 384-387, Stockholm, Sweden.
- Vitevitch, Michael
- 2002 The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28: 735-747.
- Wright, Richard A.
- 2004 Factors of lexical competition in vowel articulation. In *Papers in Laboratory Phonology VI*, John J. Local, Richard Ogden, and Rosalyn Temple (eds.), Cambridge: Cambridge University Press.
- Zwicker, E., and E. Terhardt
- 1980 Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America* 68: 1523-1524.