Phonetic reduction in the "Easy/Hard" database without "difficulty"

Susanne Gahl
gahl@berkeley.edu

Abstract

A widely-cited study investigating effects of recognition difficulty on the phonetic realization of words (Wright, 1997, 2004) described vowel dispersion in a subset of the Easy/Hard database (Torretta, 1995). The core finding was that vowel dispersion, i.e. distance from the center of the talker's F1/F2 space, was greater in words that represent difficult recognition targets, due to competition from other words in the lexicon ('hard' words) than in easy recognition targets ('easy' words). The goal of the current study was to test whether the pattern observed in the subset extended to a larger portion of the Easy/Hard database, and whether the effect persisted when controlling for known other determinants of vowel dispersion. Extending the investigation to all monophthongs in the database, we find that recognition difficulty fails to have a significant effect on dispersion when effects of segmental context are brought under statistical control. We conclude that the pattern of dispersion in Wright (1997) is not due to lexical competition, but is most likely due to segmental context, particularly place of articulation of consonants preceding and following the target vowels. We discuss the implications of this reanalysis for studies of pronunciation variation.

1. Introduction

Patterns of pronunciation variation in spoken language are at the focus of converging research programs in Phonetics, Sociolinguistics, and Psycholinguistics. Each of these has identified a multitude of factors giving rise to phonetic variation (see e.g. Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fosler-Lussier & Morgan, 1999; Hay & Jannedy, 2006). One challenge for current research is to find out which of the potential sources of variation give rise to observed patterns, generally and in a given sample.

Some sources of variation are essentially uncontroversial: Clearly, segmental context affects the realization of segments, as does speaking rate (Farnetani, 1997). Deep theoretical questions remain open, for example concerning the nature of articulatory targets for production (Johnson, Flemming, & Wright, 1993), the relationship between speaking rate and articulatory movement (Moon & Lindblom, 1994), and the degree of abstractness of phonological representations (Pierrehumbert, 2001). Yet, it is clear beyond doubt that some articulatory variability is in some manner attributable to segmental context and speaking rate.

Other sources of variation remain controversial, in the sense that there has yet to emerge a consensus about the scope of the proposed explanations. One current area of active research that is rife with controversy concerns the role of clarity and intelligibity in pronunciation, and more generally, the relationship between production and perception, and the role of talker's models of listener's needs in pronunciation (Arnold, 2008; Aylett & Turk, 2006; Gahl, Yao, & Johnson, 2012; Galati & Brennan, 2010; Jaeger, 2010; Stent, Huffman, & Brennan, 2008).

One of the earliest studies examining the relationship of recognition difficulty and pronunciation focused on vowel dispersion (1997; 2004; henceforth W1997 and W2004).[1] Vowel dispersion (and its opposite, vowel centralization) refers to the distance between vowel tokens and the center of a talker's vowel space. Following Bradlow et al. (Bradlow, Torretta, & Pisoni, 1996), W1997 measured vowel dispersion as the Euclidean distance between vowel tokens and the center of each talker's F1/F2 space. High vowel dispersion is a feature of 'clear speech', i.e. the speaking style talkers adopt when asked to speak clearly. That fact makes vowel dispersion a suitable tool for examining the role of intelligibility in the realization of words, both because dispersion affects intelligibility, and because speakers are able to adjust their speech so as to increase or decrease dispersion. W1997 examined two groups of words, respectively "easy" and "hard" targets for recognition, based on prior research by (Luce & Pisoni, 1987; Luce, Pisoni, & Goldinger, 1990; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). W1997 found vowel dispersion to be greater in 'hard' words than in 'easy' words. That conclusion was based on a subset of a database of "easy" and "hard" words (Torretta, 1995).

The W1997 study has had a very significant impact. The idea that talkers aim to optimize overall intelligibility while preserving effort, by expending articulatory effort on those words that might be difficult to understand, is a cornerstone of theories relating speech production, speech perception, and informativity (Aylett & Turk, 2004; Lindblom, 1990). Its central finding has been widely accepted, replicated, and extended to other aspects of pronunciation besides vowel dispersion (Kilanski, 2009; Munson, 2007; Munson & Solomon, 2004; Scarborough, 2010). Yet, the role of recognition difficulty in pronunciation variation remains controversial (Arnold, 2008; Galati & Brennan, 2010). Furthermore, a recent study (Gahl et al., 2012) examining a much larger set of words than the one discussed in W1997 reported results that appear to run counter to the pattern observed in W1997. Clearly, many questions about the role of recognition difficulty in pronunciation remain unanswered and controversial.

A number of relatively uncontroversial factors are known to affect vowel dispersion. Among these are vowel duration, speaking rate, talker sex, and segmental context. Other things being equal, vowel spaces tend to be more compact, i.e. less dispersed, at faster speaking rates (Lindblom, 1983; Moon & Lindblom, 1994), in the speech of male vs. female talkers (Byrd, 1994), and in the vicinity of alveolar consonants (Farnetani, 1997). The effect of alveolar consonants arises since these consonants require tongue positions near the alveolar area, as do centralized vs. highly peripheral vowels. Manner of articulation of adjacent consonants may also be predictive of vowel dispersion, as a result of nasalization and coloring due to pre- or postvocalic nasals and liquids.

Previous further research suggests a way in which these articulatory and indexical factors may interact with recognition difficulty to affect vowel dispersion: Rather than increasing dispersion for ‘hard’ targets across the board, talkers may expend effort to decrease the amount of context-dependent undershoot in ‘hard’ words, thereby keeping these words from being even harder to understand (Flemming, 2010; Lindblom, 1990). If talkers indeed aim to optimize a trade-off between effort and intelligibility, then they might expend effort in order to minimize target undershoot in ‘hard’ words, while allowing target undershoot in ‘easy’ words, for example as a result of coarticulatory influences of neighboring alveolar consonants. As a result, vowels in ‘easy’ words would tend to be centralized in the vicinity of alveolar consonants, but vowels in ‘hard’ words need not.

W1997 sought to control for segmental context, by focusing on a set of ‘easy’ and ‘hard’ words matched for segmental factors affecting dispersion. Therefore, the analysis in W1997 targeted a subset of a larger database, known as the Easy/Hard database (Torretta, 1995). The Easy/Hard database contains 150 lexical items, spoken at three different speaking rates by 10 different talkers. W1997 included all ten talkers, but only 68 lexical items, produced at one level of speaking rate (the 'medium' rate), for a total of 680 tokens out of the 4500-token database. The fact that the analyses focused on a subset of the database invites the question whether the observed effect extends to other words in the database and whether it holds at all three speaking rates. A related question concerns whether the differences between the 'easy' and 'hard' words are primarily attributable to recognition difficulty; or whether they reflected other determinants of pronunciation variation.

The goal of the present study was to ask whether the pattern in W1997 might have resulted from other factors besides recognition difficulty – specifically, factors that are uncontroversially thought to affect vowel dispersion. A related goal was to examine whether the observed pattern extended to other parts of the Easy/Hard database. To preview our results: we successfully replicate W1997, both in the W1997 subset and in the Easy/Hard database as a whole, but find that, when segmental context is brought under statistical control, there is no evidence for recognition difficulty modulating vowel dispersion in the Easy/Hard database.

2. Data and Methods

The database is described in Torretta (1995). The full database consists of 4500 audio files, representing 150 word types, read by ten talkers at three speaking rates. The recordings were made at the Speech Research Laboratory at Indiana University. No information about the talkers' linguistic background given in Torretta (1995). W1997 states that the talkers represented a variety of dialects, all characterized as “General American English”, and that “all the dialects had the same vowel-quality categories in all of the stimuli”.

The word lists were constructed on the basis of previous research on word recognition, specifically of the effects of lexical familiarity, lexical frequency, and phonological neighborhood structure on recognition difficulty (Luce & Pisoni, 1987; Luce et al., 1990; Pisoni et al., 1985). Phonological neighborhood structure is captured by two related variables: (a) phonological neighborhood density and (b) neighborhood frequency. Phonological neighbors are words in the lexicon that differ from a target by addition, deletion, or substitution of one phoneme. For example, the neighbors of *pat* include the words *cat, pot, spat,* and *pan*. Phonological neighborhood density refers to the number of neighbors of a target. Neighborhood frequency was defined as the mean word frequency of a target's neighbors. The 150 word types consisted of two sets of 75 words, termed 'easy' and 'hard', on the basis of recognition difficulty. “Easy” words, i.e. easy targets for recognition, were high-frequency words facing little competition from their neighbors, i.e. with low neighborhood density and low neighborhood frequency, relative to the target frequency. “Hard” words, i.e. difficult recognition targets, were low-frequency words with many neighbors and high neighborhood frequency. Lexical familiarity was held constant across the two groups: both groups had very high familiarity ratings (greater than 6.7 on a seven-point scale). Estimates of familiarity, frequency, and phonological neighborhood structure were based on the Hoosier Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984).

According to Torretta (1995), the words were presented in random order on a computer monitor. Talkers were instructed to read the words “in a normal speaking voice” three times, at three different speaking rates. The utterances were low-pass filtered at 10kHz and digitized at a sampling rate of 22050 kHz. The speech files were edited so as to remove periods of silence on either side of the words and to check the waveform. The files were rerecorded “in appropriate cases (e.g. speaking level too loud/too soft)” (Torretta, 1995). No information is available about how many trials were repeated for this reason.

The subset of the Easy/Hard database analyzed in W1997 consisted of 68 words (34 'easy' ones and 34 'hard' ones) from the original set of 150. As in the full database, the 'easy' words had few neighbors and were of high lexical frequency relative to their neighbors, whereas the 'hard' words had many neighbors and were of low lexical frequency relative to their neighbors. Since the words in the two sets could not be homophones - which would entail having identical phonological neighborhood structures -, matching segmental context perfectly was impossible. Given that constraint, inclusion in the analysis in W1997 was based on the desire to avoid “consonantal contexts that could result in vowel colouring” and to observe target vowels “in similar contexts in 'easy' and 'hard' words” (Wright, 2004, p. 79). On the basis of those criteria, words with postvocalic /l/ and /r/ were excluded in W1997 and postvocalic nasals were matched across the two sets.

In our reanalysis of the audio files, one of the audio files was found to be corrupt and had to be excluded from the analysis. Another file had to be excluded because the talker produced the word *mail* instead of the target *mall*. To facilitate the analysis of a larger data set, we used automatic alignment and formant extraction. The audio files for the Torretta Easy/Hard database were aligned with the broad transcriptions of the words at the phone level using the Penn Phonetics Lab Forced Aligner Toolkit (Yuan & Liberman, 2008). The start and end times of each vowel phone was obtained from the alignment results, and a portion of each token's audio file was extracted, starting 40ms before the start time and ending 40ms after the end time of the vowel. This audio was downsampled to 12 KHz and analyzed by the Watanabe and Ueda formant tracker (Ueda, Hamakawa, Sakata, Hario, & Watanabe, 2007; Watanabe, 2001). Measurements for F1-F4 and F0 were recorded for the analysis frame occurring at the temporal midpoint of the vowel. In four cases, the automatic tracking resulted in F1 values of zero Hz or below 2 on the Bark scale. In seven cases, formant tracking errors resulted in missing F1 or F2 values. These tokens were excluded from further analysis.

There were a small number of discrepancies between the audio files and the description of the data base in Torretta (1995). The first discrepancy was that was that three tokens of the word “job” were coded as containing the vowel [o], despite the fact that none of the talkers pronounced *job* with that vowel (as might be the case if they were referring to the biblical figure Job). These three tokens were re-coded as containing the same vowel as the words *shop, watch, cod, knob,* and *wad*. Secondly, the item *bag* appears in Torretta (1995) and Wright (1997, 2004), but the corresponding recordings appeared to be the word *hag* for all talkers, with a period of audible frication before the vowel. It is no longer possible to recover whether the discrepancy is due to an error in stimulus description, stimulus presentation, participant error, or some other factor. The word *wrong* appeared on both the 'easy' list and the 'hard' list. For this reason, it was excluded from all current analyses (the item was also excluded in W1997 or W2004).

There was also a discrepancy between the word lists in W1997 vs. W2004. W2004 lists the orthographic form *caught*, but not *cot*, whereas W1997 lists *cot*, but not *caught*. Torretta (1995) lists the item in question as *cot*. In the current study, the spelling *cot* (and the corresponding lexical frequency) is assumed, since Torretta (1995) and W1997 were closer in time to the data collection phase, and given that the item in question appears on the list of “hard” words: Since *caught* has a fairly high lexical frequency, it is unlikely that it would have met the inclusion criteria for the “hard” set.

Diphthongs are transcribed as single segments by the aligner. Since a diphthong's temporal midpoint does not provide a reliable estimate of the formant characteristics of the diphthong as a whole, words containing diphthongs were excluded from the analysis. There were 8 items including diphthongs in the W1997/2004 subset. Excluding words with diphthongs left 60 types (600 tokens) in the set for the replication of W1997, and 125 types (3734 tokens) for the analysis of all monophthongs.[2]

Table 1 shows the mean values and the ranges of the frequency and neighborhood density of all word types in the Easy/Hard database, and of the subsets analyzed in W1997 and in the current study. For the most part, the mean of the lexical frequency and neighborhood density of easy vs. hard words are similar across samples. An exception is the lexical frequency of ‘hard’ words, which is lower in the sample for the current study than either W1997 or Torretta (1995). Since low lexical frequency is one of the characteristics of the ‘hard’ group, the lower frequency of the ‘hard’ words in the current study should, if anything, aid in replicating any effect of recognition difficulty.

Table 1. *Mean (range) of lexical frequency and neighborhood density of ‘easy’ and ‘hard’ words in Torretta (1995), Wright (1997) and the current study*

	Lexical frequency		Neighborhood density	
	Easy	Hard	Easy	Hard
Torretta (1995) n = 150	384.84 (0.59-5654.73)	10.73 (0.31 – 171.45)	14.47 (1-31)	27.75 (8-45)
Wright (1997) n = 68	218.25 (13.98-1167.82)	12.05 (0.31-171.45)	14.0 (4-28)	26.91 (8-43)
Current study n = 125	434.31 (0.59-5654.73)	8.88 (0.31-42.73)	14.95 (1-31)	28.25 (16-45)

Vowel dispersion was calculated as the Euclidean distance between the point defined by the F1 and F2 (Bark) values of each vowel token and the talker's average F1 and F2 (Bark) values, following the method adopted in Wright and proposed in Bradlow et al. (1996). To trace the method used in W1997 as closely as possible, talker-specific vowel centers were calculated using only the words that also entered into Wright's analysis for the purposes of the replication (Study 1). For the analysis of the larger set (Study 2), talker-specific vowel centers were calculated using all analyzable monophthongs.

For the statistical analysis of the larger dataset, we fitted linear mixed-effects regression models, using modeling tools now commonly used in analyses of pronunciation variation and many other psycholinguistic variables (H. Baayen, 2008; H. Baayen, Davidson, & Bates, 2008; H. Baayen, Tweedie, & Schreuder, 2002). All statistical analyses were performed using R (R Development Core Team, 2008) and the R packages languageR (H. Baayen, 2008), and lme4 (Bates & Maechler, 2010). Normality and homogeneity of the residuals were checked by visual inspection of plots of residuals against fitted values.

In multivariate regression modeling without random effects, all categorical predictors are treated as fixed effects: All levels of fixed effects are known ahead of the analysis, and estimated coefficients describing the relationship between each level of the predictor and the outcome variable can vary freely. An example of a fixed effect in our models is Speaking rate, with the levels Fast, Medium, and Slow. The model treats the estimates for these three levels as three separate parameters, without imposing any constraints on how the three levels might differ from one another. Random effects, by contrast, are categorical predictors whose values are treated as random samples from a larger population. An example of a random effect in our models is Talker: The assumption is that talker-specific "baseline" values of vowel dispersion (the random intercept), as well as talker-specific variation associated with other variables (random slopes) represent normally distributed random variables with means equal to the population mean for each (intercept and slope) and unknown variance, estimated by the model. Treating talker as a random effect has two implications: First, it means that the modeling results can be understood as statements about the whole population of talkers, not just the particular talkers who participated in the experiment. Secondly, it imposes a constraint on the by-talker (and by-word) estimates: Talker-specific adjustments cannot differ arbitrarily from one another, but are assumed to represent samples drawn from a normal distribution representing the population of all possible talkers.

Based on prior research on vowel dispersion, we included the following factors as fixed effects in our analysis: Vowel type (i.e. phoneme), Speaking rate, Place of preceding consonant (a binary factor distinguishing alveolar vs. all other places of articulation), Place of following consonant (alveolar vs. all others), Manner of preceding consonant (approximants vs. nasals vs. obstruents), Sex, and Difficulty ('easy' vs. 'hard', based on the classification in Torretta, 1995). If the relationship between recognition difficulty and other factors affecting vowel dispersion is as outlined in Lindblom (Lindblom, 1990; Moon & Lindblom, 1994) and Flemming (Flemming, 2010), i.e. if talkers expend effort to keep vowels in 'hard' words from being difficult to recognize, then effects of recognition difficulty on vowel dispersion might manifest themselves not only as statistical main effects of Difficulty, but also as a significant interaction between place of articulation and Difficulty. To check whether this is the case, we tested for the presence of significant interactions between Place of articulation (alveolar vs. other) and Difficulty.

Following the recommendations in Baayen & Milin (2010), we refrained from excluding extreme values of the dependent variable as outliers ahead of the analysis and instead excluded outliers identified based on model criticism. Specifically, we excluded cases that were associated with large residuals (more than 2.5 SDs from zero) in the initial modeling phase and then refitted the model without those cases, to prevent these observations from unduly influencing estimates of coefficients. The outcome variable (dispersion) was centered around its mean, by subtracting the mean from each observed value.

3. Results

3.1. *Replicating Wright (1997, 2004)*

To gauge the degree of consistency between the current study and the earlier analyses, we first sought to replicate W1997/2004, restricting our attention to words that were included in that study (excluding diphthongs), and following the analytical steps in W1997/2004 before extending the analysis to other parts of the database or taking additional predictors of dispersion into account.

Average dispersion in the W1997 part of the data set, pooled across all talkers and vowel types, was higher for the hard words than for easy words (1.999 vs. 1.816, n = 599), as in W1997. Since the overall means represent group averages of dispersion values, rather than averaged differences between the pairs of words, they do not necessarily indicate the effect of the easy/hard manipulation. Therefore, W1997 analyzed the results with a repeated-measures ANOVA, examining main effects of lexical category (i.e. 'easy' vs. 'hard', termed “Difficulty” here), talker, and vowel type, as well as the interaction between lexical category and vowel type. Applying the same repeated-measures ANOVA model to the automatically extracted formant values, we observed the same pattern of results as that reported in W1997: Assuming an alpha level of .05, there was a significant main effects of Difficulty ($F(1,399) = 67.67$, $p < .0001$). There was also a significant interaction of Difficulty by Vowel type ($F(9,399) = 6.32$, $p < .0001$).

To explore differences across vowels further, W1997 also reported the average dispersion for 'easy' and 'hard' words separately for each vowel. Figure 1, analogous to Figure 4.3 in W2004 shows the by-vowel averages for the the easy vs. hard words.

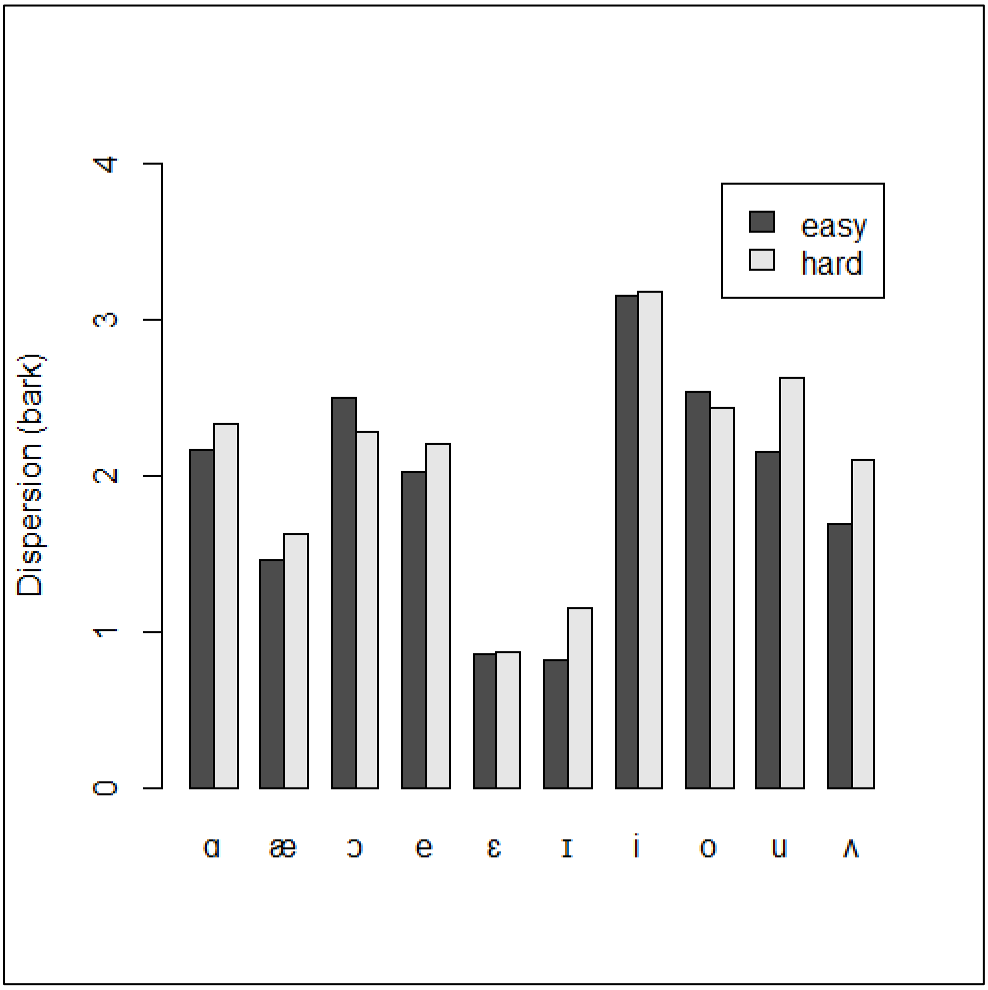

Figure 1: Vowel dispersion in ‘easy’ vs. ‘hard’ words (n = 600)

For eight of the ten vowel types, dispersion was greater for the “hard' words than the “easy” words. For some vowels, the observed difference not only followed the pattern observed in W1997, but was numerically larger. For example, average dispersion in easy vs. hard words containing the vowel [ʌ] differed by ca. 0.1 Bark in Wright's report, and by ca. 0.4 Bark (1.69 vs. 2.10) in the current study. The exceptions were the vowels /ɔ/ and /o/. The latter of these may be diphthongal, which would render the automatic extracted formant values doubtful. The vowel [ɔ] patterned in the opposite direction from that observed by W1997 (2.50 for the ‘easy’ condition vs. 2.28 for the ‘hard’ condition). It should be noted that this vowel was represented by a single pair of words (*wash* and *cot*) in the W1997 subset.[3]

W1997 reported that the effect was “clearest” for the vowels /i, æ, ɑ, ɔ, u/. The current study recovers this effect for four of those vowels – the four that are represented by more than one item per condition. Overall, then, the differences in methods did not prevent a successful replication.

The overall mean of dispersion in W1997/2004 was approximately 2.3 for the hard words and 1.8 for the easy words (according to Figure 4.1 in W2004), vs. 2.0 and 1.8 in the current study. It would appear, then, that the absolute degree of dispersion is lower in the current study, particularly for the “hard” words. One possible reason for this is that measurements for W1997 were made at the point of maximal displacement, i.e. the point “when F1 and F2 are the most characteristic for that particular vowel” (Wright 2004: 80). For vowels characterized by especially high or especially low F1 or F2 or F1/F2 ratios, that criterion often means that the measurement represents the extreme F1 and/or F2 for a given vowel token (though not always, since in cases “[w]here F1 and F2 were not in agreement, F1 was taken as the point of reference and F2 was measured at that point”). As a result, the 'maximal displacement' criterion favors extreme distances from the center of vowel space as the point where measurements are taken. The automatic formant extraction used in the current study does not favor extreme points and thus will tend to reduce the mean values for classes of vowels that are most likely to contribute extreme points - vowels in the 'hard' words, assuming the effect observed by Wright is present. Nevertheless, the overall pattern was similar: Dispersion was higher on average in 'hard' words compared to 'easy' words.

3.2. *The rest of the data*

We now turn to the analysis of the larger dataset, i.e. the set of all analyzable monophthongs in the Easy/Hard database. The larger and more segmentally varied dataset enables us to probe whether previously-observed generalizations about vowel dispersion, for example the effects of place of articulation or neighboring consonants, are borne out in the data. With that information in hand, we can examine the effects of recognition difficulty further.

To make sure that any differences between our overall conclusions and those of W1997 were not due to extraneous effects of speaking rate, or more generally a failure to replicate the earlier analysis, we first checked whether the pattern observed in W1997 held at all three speaking rates, for the items included in W1997. That appeared to be the case: ANOVAs of the data from the ‘slow’ and ‘fast’ conditions produced results that were parallel to the pattern at the medium rate, with significant effects of Vowel type, Talker, and, critically, Difficulty. In both the ‘fast’ and the ‘slow’ condition, there were significant main effects of Vowel type, and Difficulty (Vowel $F(9,399) = 46.73$, $p < .001$; Difficulty $F(1,399) = 35.69$, $p < .001$ in the ‘fast’ condition; Vowel $F(9,398) = 57.63$, $p < .001$; Difficulty $F(1,398) = 17.36$, p = .004 in the ‘slow’ condition). There was also a significant interaction of Difficulty by Vowel type (in the ‘fast’ condition F(9,399) = 6.63, p < .001; in the ‘slow’ condition F(9,398) = 3.20, p = .002.). Most important for the replication is the presence of a significant main effect of Difficulty: At each level of speaking rate, vowel dispersion was significantly higher in the hard words than the easy ones.

As shown in Table 2, the pattern of vowel dispersion in easy vs. hard words was not entirely uniform at all speaking rates. For example, only the five vowels /æ, ɔ, i, o, ʌ/ showed greater dispersion in ‘hard’ words than ‘easy’ ones at all three speaking rates. Nevertheless, it appears that the pattern reported in W1997 holds at all three speaking rates, when one considers Difficulty (easy/hard), Speaking rate, Vowel type, and Talker as the only predictors.

Table 2. *Vowel dispersion (in Bark) at three speaking rate.*

	Slow		Medium		Fast	
Vowel	easy	hard	easy	hard	easy	hard
ɑ	2.43	2.32	2.24	2.19	2.18	2.16
æ	1.87	1.94	1.61	1.79	1.63	1.68
ɔ	2.65	3.54	2.46	3.22	2.40	3.00
ɛ	1.31	1.24	1.19	1.13	1.22	1.07
e	2.49	2.57	2.22	2.30	2.04	2.01
ɪ	1.39	1.46	1.25	1.32	1.23	1.16
i	3.23	3.39	3.01	3.30	2.73	3.15
o	2.67	3.03	2.24	2.59	2.29	2.55
ʊ	2.55	NA	2.55	NA	2.54	NA
u	2.88	2.02	2.59	1.96	2.56	1.88
ʌ	1.50	1.66	1.38	1.65	1.31	1.56

To understand the pattern of vowel dispersion in the larger set more fully, we fit mixed-effects regression models controlling for factors affecting vowel dispersion, as indicated by prior research: Vowel type, Speaking rate (slow/medium/fast), Sex, Place (alveolar vs. other) and Manner or articulation of adjacent consonants, and Difficulty. In order to ascertain the individual contribution of the various predictors of dispersion, we first fitted a model (the “full” model) containing all predictors. Outliers, defined as observations with standardized residuals greater than 2.5 standard deviations, were removed. The number of observations removed by this criterion was 83, i.e. 2.2% of the observations in the sample. We refit the model without the outliers and ascertained the contribution of each predictor in turn in a series of leave-one-out model comparisons, each comparing the full model to a model without the predictor in question. The model is summarized in Tables 3 and 4. The p-values associated with the difference in log likelihood in the leave-one-out models vs. the full model are reported in the column labeled "p(chi square)" in Table 3.

The model confirmed previously reported effects of neighboring consonants and of speaking rate, but did not reveal any significant effect of Difficulty. Dispersion varied significantly across vowel types, as one would expect. In addition, dispersion was greater at the medium speaking rate than at the fast speaking rate (the baseline level for modeling purposes) and greatest at the slow speaking rate. Dispersion was also significantly greater in vowels immediately before or after non-alveolar consonants. Target vowels preceded or followed by nasals and (oral) obstruents were centralized compared to vowels preceded or followed by approximants. We did not observe significant effect of Sex. We suspect that this might be due to the presence of the random effect for Talker in the model: Since there were only five talkers of each sex, differences between male and female talkers are likely modeled as individual differences across talkers. Crucially, there was no significant effect of Difficulty (beta = -.05, pMCMC = .50, pChisq = .50). Figure 2 shows the partial effects plots, i.e. the model predictions for each predictor in the model when all other predictors are kept constant. The pattern of results that is reflected in the plots is that the model recovers effects of articulation of neighboring consonants and of speaking rate, but gives no indication that Difficulty affects dispersion when other predictors are taken into account.

Table 3. *Estimated coefficients, standard errors, t-values and p-values, based on MCMC sampling, and log likelihood tests for fixed effects of the model of vowel dispersion (n = 3734), and predicted ranges. Model R-square = .81*

		Beta (SE)	t	pMCMC(>|t|)	p(chisq)	Predicted range
(Intercept)		0.2797 (0.17536)	1.595	.1107		
Vowel	æ	-0.3729 (0.13151)	-2.835	.0046	< .0001	2.00
	ʌ	-0.6196 (0.10726)	-5.777	< .0001		
	ɔ	0.5116 (0.16955)	3.017	0.0026		
	ɛ	-0.8372 (0.16486)	-5.078	< 0.0001		
	e	0.0704 (0.13278)	0.530	0.5961		
	ɪ	-0.9912 (0.12559)	-7.892	< 0.0001		
	i	1.0116 (0.13525)	7.480	< 0.0001		
	o	0.2814 (0.15953)	1.764	0.0778		
	ʊ	-0.1336 (0.23867)	-0.560	0.5757		
	u	0.1378 (0.14590)	0.944	0.3450		
Speaking rate	Medium	0.1083 (0.01570)	6.896	< 0.0001	< .0001	0.29
	Slow	0.2948 (0.01571)	18.768	< 0.0001		
Place_before	Non-alveolar	0.2923 (0.07531)	3.882	0.0001	.0002	0.29
Place_after	Non-alveolar	0.1925 (0.07765)	2.479	0.0132	.0145	0.19
Manner_before	Nasal	0.2438 (0.13097)	1.861	.0628	.1750	0.24
	Obstruent	0.0996 (0.08274)	1.204	.2288		
Manner_after	Nasal	-0.4074 (0.13097)	-3.111	0.0019	< .0001	0.73
	Obstruent	-0.7251 (0.11599)	-6.251	< 0.0001		
Sex	Male	0.0411 (0.10982)	0.374	0.7085	.7095	0.04
Difficulty	Hard	-0.0503 (0.07388)	-0.681	0.4956	.4958	0.05

Table 4. *Summary of random effects in the model*

Random effect	Variance	SD	MCMC median	HPD95lower	HPD95upper
Talker (intercept)	0.029752	0.17249	0.1818	0.1136	0.2753
Word (intercept)	0.121207	0.34815	0.2620	0.2370	0.2890
Residual	0.150190	0.38754	0.3937	0.3845	0.4029

The model included random intercepts, to capture the fact that there is individual variation in vowel spaces, but did not include random slopes for the effect of Difficulty, which would capture individual differences in the degree to which talkers adjust their pronunciation as a function of recognition difficulty. We did explore this possibility, by including by-talker random slopes for Difficulty; the resulting model showed the identical pattern of significant effects of Place and Manner of articulation, Vowel type, and Speaking rate; and the same pattern of non-significance of Sex and Difficulty. That model did not substantially increase model goodness-of-fit, so we report the simpler model here.

Just as the random effect for Talker might prevent Sex from producing a significant effect, the random effect for Word might prevent Difficulty from producing a significant effect. To test this possibility, we refit the model again, this time without a random effect for Word. In a model without a random effect for Word, the contrast between ‘easy’ and ‘hard’ words yielded a marginally significant effect – in the opposite direction from what one would expect based on the analysis in W1997: Predicted dispersion for ‘hard’ words was lower than for ‘easy’ words: beta = -0.0354 (SE = -0.0353), pMCMC = 0.0770, chisq = 3.1346, p(chisq) = .07665. This result suggests that the failure of Difficulty to produce a significant effect in the full model is not due to the presence of the random effect for Word.

Since we suspected that Difficulty was in competition with other predictors in the model, we explored the possibility that order of entry into the model might have kept the effect of Difficulty from revealing itself. To give Difficulty a chance to account for variability in dispersion without facing competition from other fixed effects (save those variables also included or kept constant in W1997), we fitted a model with only the random effects (Talker and Word), Vowel type, and Speaking rate, to which we then added Difficulty. The coefficient of Difficulty in the resulting model was insignificantly different from zero, and model comparison did not reveal any significant model improvement when Difficulty was added to the baseline model (chi square = 0.49, p = .48; beta = 0.0586 (SE = 0.05855), pMCMC = .48.

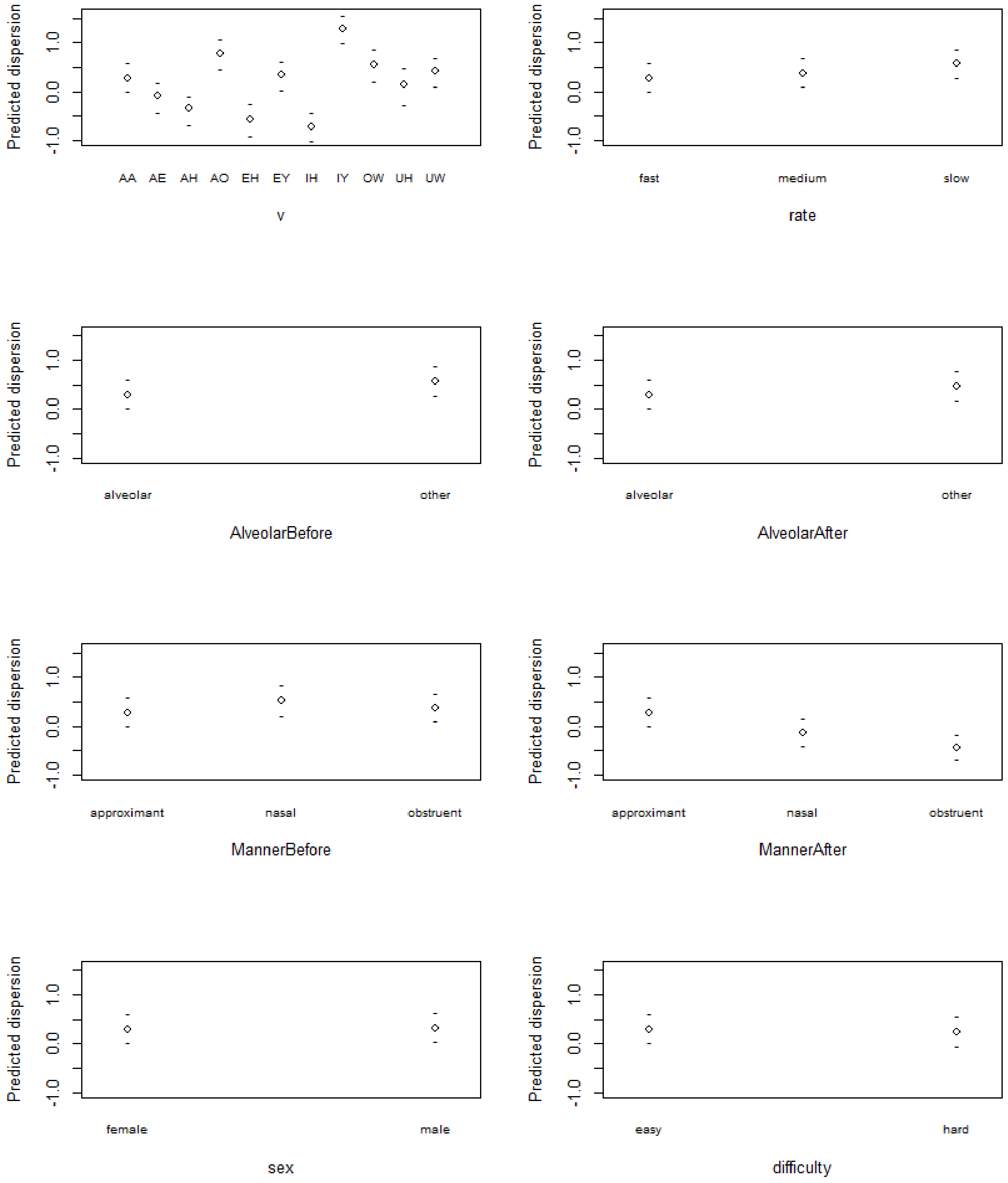

Figure 2: Partial effects in the model of vowel dispersion in the Easy/Hard database

We also explored the possibility that Vowel type might interact with Speaking rate, or with Difficulty. Testing these interactions was motivated by the observation that vowels change far more than others as a function of speaking and difficulty (Lindblom, 1964; Moon & Lindblom, 1994; Wright, 1997). The interaction of Vowel type by Rate was significant, reflecting the fact that some vowels, such as /i/, allow for greater variability in dispersion and consequently change to a greater degree as a function of speaking rate. As in the model without the Vowel x Rate interaction, Difficulty failed to yield a significant effect. The pattern of significance of the other predictors (Place and Manner, and Sex) was unchanged. Including the interaction in the model caused R^2, i.e. the proportion of variability explained, to increase slightly, from .81 to .82. Since this increase is so slight, and since the behavior of the variables of interest remained unchanged, we report the model without the Vowel x Rate contrasts, in the interest of readability.

The words *full, pull,* and *put* were the only words in the database containing the vowel /ʊ/. Since all three of these were in the "easy" set, testing the interaction of Vowel type with Difficulty was only possible after excluding *full, pull,* and *put.* We therefore refitted the model to the database without /ʊ/, this time including interactions of Vowel type with Speaking Rate and with Difficulty. Log likelihood tests indicated a significant interaction of Vowel type by Speaking rate (χ^2= 333.24, p < .001), but no significant interaction of Vowel type by Difficulty (χ^2= 8.9, p = .45). The Difficulty contrast did not yield a significant effect in any of the Models. The pattern of significances of all other predictors was likewise unchanged.

We suspected that vowels flanked by alveolars on both sides might tend to be more centralized than vowels adjacent to only one alveolar consonant. To test whether this was the case, we also entered an interaction of the two factors coding the presence of alveolars before and after the target vowel. That interaction did not lead to significant model improvement, based on model comparison, and we refrained from exploring it further.

Since the voicing of consonants affects vowel duration (Peterson & Lehiste, 1960), which in turn affects vowel dispersion, we also explored the effect of voicing of both the preceding and following consonants. Neither factor yielded significant effects when added to the model, perhaps because of the presence of Manner of articulation in the model: Since all approximants in the sample are voiced, effects of voicing may have been captured by the Manner variable.

In summary, the models discussed so far do not provide any evidence suggesting that vowel dispersion varied as a function of Difficulty, at least not in the manner outlined in W1997: There was no significant effect of Difficulty in the models of the larger data set. This was the case both when other predictors of dispersion, such as place and manner of articulation of surrounding consonants, were controlled and when they were kept out of the model to give Difficulty a chance to account for as much of the variability in dispersion as possible.

3.3 *Testing the interaction of Difficulty and segmental context*

So far, we have followed the analytical strategy in W1997, by testing whether dispersion is higher in ‘hard’ words, as suggested in W1997. As mentioned above, there is another possible way in which Difficulty might affect dispersion: Talkers might expend additional articulatory effort to reach articulatory targets in ‘hard’ words, so as to produce vowels that are unobscured by coarticulation. In ‘easy’ words, talkers might allow dispersion to vary primarily as a function of vowel target and contextual factors (e.g. reduced dispersion near alveolar consonants). To test whether this was the case, we fitted a new set of models, this time including interactions of Difficulty with each of the other fixed effects, i.e. Place and Manner of articulation of surrounding consonants, Sex, and Speaking rate.

The only interactions that gave rise to significant interactions were those of Difficulty with Place of articulation of the preceding consonant and Manner of articulation of the following consonant. The interactions of Difficulty with Speaking rate, Place of articulation of the following consonant, Manner of articulation of the preceding consonant, and Sex were non-significant. The model R^2 was .81, i.e. unchanged compared to the model without the interactions.

To give the effect of Difficulty another chance to reveal itself, we followed up with separate analyses of the ‘easy’ and ‘hard’ words. We fitted the same predictors to each subset that were in the full model. An effect of Difficulty on talkers’ efforts to counteract coarticulation and assimilation should produce a situation in which place and manner of articulation of surrounding consonants affects dispersion in ‘easy’ words to a greater degree than in ‘hard’ words.

The results did not bear this out. The presence of alveolar consonants before the target vowel affected vowels in ‘easy’ words, but not in ‘hard’ ones (with dispersion being greater after non-alveolars), but did not yield a significant effect in the ‘hard’ words. Manner of articulation of the consonant following the target vowel likewise had a significant effect on 'easy' words, but not 'hard' ones, with target vowels in ‘easy’ words being more centralized before nasals and obstruents than before approximants. On the other hand, place of articulation of the consonant following the target vowel had a significant effect for both sets of words (greater dispersion before peripheral consonants). Additionally, there was a significant effect of manner of articulation of the consonant preceding the vowel for the 'hard' set, but not for the 'easy' set, contrary to what one would expect if talkers attempt to correct for coarticulatory effects when producing words that might be difficulty to understand.

On balance, the analysis of the larger data set does not provide evidence for an effect of Difficulty on dispersion along the lines argued in W1997 and subsequent studies, but leaves open the possibility that recognition difficulty might modulate effects of consonant-vowel coarticulation. This result underscores the importance of stimulus selection: Since consonantal context affects dispersion, the larger data set may not be the ideal testing ground for probing effects of recognition difficulty. W1997 restricted the focus of analysis to a subset of the data base precisely for this reason, by matching segmental properties of the stimuli across the Easy/Hard condition. To understand whether Difficulty affected dispersion in the matched subset, we now turn to a re-analysis of the subset in W1997, using the information about the effects of other variables gleaned from the larger data set.

3.4 *Returning to the W1997 subset: Does Difficulty affect dispersion when other factors are controlled?*

To test whether the effect of Difficulty persisted when segmental properties were controlled, we refit the model of dispersion to the set of monophthongs in W1997. After removing outliers identified based on model residuals (n = 34, i.e. 1.9% of the data), we refit the model. The model is summarized in Tables 5 and 6.

Table 5. *Estimated coefficients, standard errors, t-values and p-values, based on MCMC sampling, and log likelihood tests for fixed effects of the model of vowel dispersion (n = 581), and predicted ranges. Model* $R^2 = .80$

		Beta (SE)	t	pMCMC(>|t||)	p(chisq)	Predicted range
(Intercept)		0.40715	2.897	0.0066		
Vowel	æ	-0.66821	-6.214	0.0001	< .0001	2.34
	ʌ	-0.66140	-6.472	0.0001		
	ɔ	0.08813	0.593	0.5582		
	ɛ	-1.33658	-12.823	0.0001		
	e	-0.23227	-2.435	0.0162		
	ɪ	-1.39292	-13.953	0.0001		
	i	0.94242	8.914	0.0001		
	o	0.37601	3.067	0.0040		
	u	0.18482	1.220	0.2296		
Speaking rate	Medium	0.09764	4.396	0.0001	< .0001	0.26
	Slow	0.26488	11.920	0.0001		
Place_before	Non-alveolar	0.18853	2.985	0.0048	.0040	0.19
Place_after	Non-alveolar	0.18691	2.975	0.0040	.0041	0.19
Manner_before	Nasal	0.10903	1.016	0.3052	.1306	0.17
	Obstruent	-0.06493	-0.948	0.3634		
Manner_after	Obstruent	-0.40744	-5.769	0.0001	< .0001	0.41
Sex	Male	0.05688	0.582	0.5960	.5639	0.06
Difficulty	Hard	0.06134	1.103	0.2824	.2725	0.06

Table 6. *Summary of random effects in the model*

Random effect	Variance	SD	MCMC median	HPD95lower	HPD95upper
Talker (intercept)	0.023071	0.1519	0.1635	0.0968	0.2799
Word (intercept)	0.025952	0.1611	0.1628	0.1322	0.1990
Residual	0.144726	0.3804	0.3824	0.3698	0.3958

As in the model of the larger dataset, Vowel Type, Speaking rate, Place of articulation of consonants before and after the target vowel were associated with significant effects, in the same direction as in the model of the larger set. Manner of articulation of consonants preceding or following the target vowel was not associated with a significant effect. There were no significant effects of talker sex or Difficulty. In summary, the model recovers the (fairly uncontroversial) effects of place of articulation, vowel type, and speaking rate in the subset of the data examined in W1997, but does not provide evidence for an effect of recognition difficulty on dispersion.

4. General Discussion

We analyzed vowel dispersion in the Easy/Hard database (Torretta, 1995), with the aim of understanding the role of recognition difficulty in pronunciation variation. A seminal study (Wright, 1997, 2004) reported that vowel dispersion was greater in words that are 'hard' targets for recognition than in words that are 'easy' targets for recognition. The main finding in the current study is that recognition difficulty fails to have a significant effect on vowel dispersion in the Easy/Hard database when segmental properties and other known determinants of dispersion are controlled.

The current investigation is subject to some of the same limitations as W1997, in that we had no control over the words that were included in the stimulus lists. As a result, a number of combinations of factors were underrepresented in the data or not represented at all. With a more varied and more balanced set of words, many more interactions could usefully be explored, such as the three-way interactions of Vowel type, Place of articulation, and

Difficulty. However, the Easy/Hard database, although sizable, was not constructed with that aim in mind and does not contain word types representing all possible combinations of all relevant factors. We refrained from exploring any such interactions further, given the limitations of the data base.

Another limitation of the current study is that we did not investigate individual differences in the degree to which talkers are affected by the various predictors of dispersion – including, but not limited to, recognition difficulty. Doing so is possible in principle, by fitting models with by-talker random slopes for each of the fixed effects. However, fitting such models to the existing database featuring 10 talkers would almost certainly result in overfitting.

Another limitation concerns the treatment of recognition difficulty in the database: Lexical frequency and neighborhood density were treated as binary variables (high vs. low frequency and density) and were manipulated simultaneously, but the two variables represent two distinct gradient properties. It is tempting to enter frequency and neighborhood density as two continuous variables into the model, in place of the binary variable capturing the easy vs. hard distinction. Doing so might shed light on the individual contribution of frequency and neighborhood density. However, the design of the Easy/Hard database meant that modeling frequency and neighborhood density as continuous predictors would be problematic: The words in the database were selected so as to represent extremes along the dimensions of frequency and neighborhood density, so the mid ranges of both of these variables is underrepresented in the database. This precludes using the database to investigate the effects of the frequency and neighborhood density continua.

Recent literature on effects of neighborhood density on pronunciation variation presents some puzzling discrepancies, quite apart from fundamental theoretical differences of interpretation: Beginning with W1997, several studies observed recognition difficulty (operationalized as a combination of high neighborhood density with low target word frequency, or as high neighborhood density independent of frequency) to be associated with increased vowel dispersion (Kilanski, 2009; Munson, 2007; Munson & Solomon, 2004) or with other factors affecting clarity (Baese-Berk & Goldrick, 2009; Scarborough, 2005, 2010). On the other hand, high neighborhood density has been found to be associated with durational shortening of words in connected speech (Gahl et al., 2012; Kilanski, 2009) and with vowel centralization (Gahl et al., 2012), despite the fact that shortening and centralization tend to be associated with decreased intelligibility, other things being equal. One contribution of the current study is to remove one part of the puzzle and show that it probably belongs to a different puzzle altogether.

5. Conclusion

The analysis presented here casts doubt on a widely accepted interpretation of a study of vowel dispersion (Wright, 1997, 2004). More broadly, we take the current study as telling a cautionary tale: Finding out which of the myriad known sources of pronunciation variation best explain systematic patterns – generally, and in a given sample – poses a significant empirical and theoretical challenge in understanding variation. Understanding theoretical implications of any given empirical finding requires considering alternative explanations, which may come from a range of disciplines, including articulatory or perceptual Phonetics, Sociolinguistics, and Psycholinguistics. Sometimes, the data may reflect patterns that are relatively uncontroversial, rather than favoring one side or another of a theoretical divide.

6. Notes

* Acknowledgments: I thank David Pisoni and Luis Hernandez at the Speech Research Laboratory at Indiana University for making the audio files available to me.

1. The report was first published as a technical report (Wright 1997) and subsequently as a contribution to an edited collection (Wright, 2004). Wright (2004) reflects some slight changes in wording and formatting. Comments and analyses mentioning W1997 and W2004 in the current manuscript apply to both versions, except where noted otherwise.

2. The vowels /o/ and /e/ are realized as monophthongs in some varieties of American English and as diphthongs in others (see e.g. Thomas, 2001). We conducted all analyses here with and without /o/ and /e/. The pattern of results remained unchanged regardless of whether these vowels were included. For the sake of comparison to W1997, we report the results of the analysis including /o/ and /e/.

3. It should be noted that these means are based on averaging the absolute dispersion values across talkers, rather than on any contrast an individual talker might make distinguishing *wash* vs. *cot*. More generally, the by-vowel means do not necessarily reflect the effect of the easy vs. hard distinction (or of any other lexical or segmental property). One way to investigate whether the overall pattern holds across talkers is to conduct a by-item (rather than just by-subject) ANOVA, along the lines proposed in Clark (1973). W1997 did not report such an analysis, and since the goal of this part of the current analysis is to ascertain whether we were able to replicate W1997, we refrained from any further ANOVA based on the factorial experimental design. To take by-talker and by-vowel variation into account, we instead fit mixed-effect regression models,.

References

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23*(4), 495-527.

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence and Duration in Spontaneous Speech. Language and Speech. *Language and Speech, 47*(1), 31-56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America, 119*(5), 3048-3058.

Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to Statistics using R*. Cambridge: Cambridge University Press.

Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Baayen, H., Tweedie, F. J., & Schreuder, R. (2002). The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon. *Brain and Language.Special Issue: Mental lexicon II, 81*(1-3), 55-65.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12-28.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*, 527-554.

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-33. http://CRAN.R-project.org/package=lme4.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123-147.

Bradlow, A. R., Torretta, G., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.

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

Clark, H. H. (1973). The Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

Farnetani, E. (1997). Coarticulation and connected speech processes. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of Phonetic sciences* (pp. 371-404). Oxford: Blackwell.

Flemming, E. (2010). Modeling listeners: Comments on Pluymaekers et al. and Scarborough. In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 587-606). Berlin: Mouton De Gruyter.

Fosler-Lussier, E., & Morgan, N. (1999). Effects of speaking rate and word predictability on conversational pronunciations. *Speech Communication, 29*, 137-158.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*(4), 789-806.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35-51.

Hay, J., & Jannedy, S. (Eds.). (2006). *Modelling Sociophonetic Variation. Special issue of Journal of Phonetics*.

Jaeger, F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*(1), 23-62.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated *Language, 69*, 505-528.

Kilanski, K. (2009). *The effects of token frequency and phonological neighborhood density on native and non-native speech production.* University of Washington, Seattle.

Lindblom, B. (1964). Articulatory activity in vowels. *Speech, Music and Hearing, Quarterly Progress Report, 1*, 1-15.

Lindblom, B. (1983). Economy of speech gestures. In P. MacNeilage (Ed.), *The production of speech* (pp. 217-245). New York: Springer.

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

Luce, P. A., & Pisoni, D. B. (1987). Neigborhoods of words in the mental lexicon. In *Research on Speech Perception Technical Report No. 6*. Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives.* (pp. 122-147). Cambridge, MA, US: The MIT Press.

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*(1), 40-55.

Munson, B. (2007). Lexical Access, Lexical Representation, and Vowel Production. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9: Phonology and Phonetics* (pp. 201-227). Berlin: Mouton de Gruyter.

Munson, B., & Solomon, N. P. (2004). The Effect of Phonological Neighborhood Density on Vowel Articulation. *Speech, Language, and Hearing Research, 47*, 1048-1058.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words*. Bloomington, IN: Psychology Department, Indiana University.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32*, 693-703.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. [References]. In *Bybee, Joan (Ed); Hopper, Paul. (2001). Frequency and the emergence of linguistic structure. Typological studies in language, vol. 45.* (pp. 137-157). Amsterdam, Netherlands: John Benjamins Publishing Company.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*, 75-95.

R Development Core Team. (2008). R: A language and environment fo statistical computing. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Scarborough, R. A. (2005). Coarticulation and the Structure of the Lexicon (Doctoral dissertation, UCLA Department of Linguistics, 2004). *Dissertation Abstracts International, A 66*(02), 574.

Scarborough, R. A. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In L. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Laboratory Phonology 10* (pp. 557-586). Berlin, New York: De Gruyter Mouton.

Stent, A. J., Huffman, M. K., & Brennan, S. E. (Writer) (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation, *Speech Communication*. Netherlands: Elsevier Science.

Thomas, E. (2001). *An acoustic analysis of vowel variation in New World English*. Durham, NC: Duke University Press.

Torretta, G. (1995). The "easy-hard" word multi-talker speech database: An initial report. In *Research on Spoken Language Processing Progress Report* (Vol. 20, pp. 321-333). Bloomington, IN: Speech Research Laboratory, Indiana University.

Ueda, Y., Hamakawa, T., Sakata, T., Hario, S., & Watanabe, A. (2007). A real-time formant tracker based on the inverse filter control method. *Acoustical Science and Technology of the Acoustical Science of Japan, 28*(4), 271-274.

Watanabe, A. (2001). Formant estimation method using inverse filter control. *IEEE Transactions on Speech and Audio Processing 9*(4), 317-326.

Wright, R. (1997). Lexical competition and reduction in speech: A preliminary report. *Indiana University Research on Spoken Language Processing Progress Report No. 21*, 471-485.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden & R. Temple (Eds.), *Papers in Laboratory Phonology VI* (pp. 26-50). Cambridge: Cambridge University Press.

Yuan, J., & Liberman, M. (2008). *Speaker identification on the SCOTUS corpus.* Paper presented at the Proceedings of Acoustics 2008.