Proceedings of Meetings on Acoustics

Volume 22 http://acousticalsociety.org/

168th Meeting of the Acoustical Society of America

Indianapolis, Indiana
27-31 October 2014

Speech Communication: Paper 2pSC1

Interactions among lexical and discourse characteristics in vowel production

Rachel Steindel Burdin, Rory Turnbull, and Cynthia G. Clopper
Department of Linguistics, Ohio State University, Columbus, OH; burdin@ling.osu.edu; turnbull@ling.osu.edu; clopper.1@osu.edu

Various factors are known to affect vowel production, including lexical frequency, neighborhood density, contextual predictability, mention in the discourse, and speaking style. This study explores interactions among all five of these factors on vowel duration and dispersion. Participants read paragraphs containing target words varying in frequency, density, and predictability. Each target word appeared twice in the paragraph. Participants read each paragraph twice: as if they were talking to a friend ("plain speech") and as if they were talking to a hearing-impaired or non-native interlocutor ("clear speech"). Measures of vowel duration and dispersion were obtained. Results revealed that high frequency words and words in plain speech were shorter and less dispersed in the vowel space than low frequency words and words in clear speech, and that second mentions were less dispersed in the vowel space than first mentions, as expected. A series of interactions among the lexical and discourse factors was also observed for both vowel duration and vowel dispersion, suggesting maximization of phonetic reduction in some contexts that impose limited processing costs on the listener and maximization of phonetic enhancement in other contexts that impose heavy processing costs on the listener.

Published by the Acoustical Society of America

© 2015 Acoustical Society of America [DOI: 10.1121/2.0000084]
Received 25 August 2015; Published 18 September 2015

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

1. Introduction

Numerous linguistic factors contribute to vowel variation in production, including properties of the words that vowels appear in and properties of the discourse in which those words appear. Lexical properties affecting vowel production include preceding and following segments, lexical stress, and word position (Klatt, 1976), as well as characteristics of the word itself, including its frequency and phonological similarity to other words (Munson & Solomon, 2004). Discourse properties affecting vowel production include prosodic structure, speaking rate, semantic predictability, and discourse mention (Klatt, 1976), as well as speaking style (Picheny, Durlach, & Braida, 1986).

One dimension of this vowel variation that has received considerable attention in the literature is phonetic reduction in both the temporal (i.e., duration) and spectral (i.e., vowel space dispersion) domains. The empirical evidence shows that high frequency words are reduced relative to low frequency words (Aylett & Turk, 2004; Gahl, Yao, & Johnson, 2012; Munson & Solomon, 2004), words with few phonologically similar neighbors are reduced relative to words with many neighbors (Munson & Solomon, 2004; Scarborough, 2010; Wright, 2004; cf. Gahl, 2015; Gahl et al., 2012), words that are semantically predictable in context are reduced relative to less predictable words (Clopper & Pierrehumbert, 2008; Hunnicutt, 1987; Lieberman, 1963), repeated words in a discourse are reduced relative to first mentions of words (Baker & Bradlow, 2009; Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Fowler & Housum, 1987), and speech directed towards an imagined friend or family member (i.e., plain lab speech) is reduced relative to speech directed towards an imagined hearing-impaired or non-native listener (i.e., clear lab speech; Ferguson & Kewley-Port, 2007; Picheny et al., 1986; Smiljanic & Bradlow, 2005, 2009).

Lindblom (1990) argued that this phenomenon of phonetic reduction arising from lexical and discourse factors reflects real-time adaptation by the talker to the needs of the listener. Specifically, according to his Hyper- and Hypo-speech (H&H) Theory, speech production reflects the outcome of competing production constraints to minimize effort and perception constraints to maximize communication. Thus, high frequency words are reduced in production relative to low frequency words because high frequency words are easier to identify in perception than low frequency words (Howes, 1957). Talkers can therefore conserve effort when producing high frequency words without sacrificing successful communication. By contrast, clear speech is unreduced in production relative to plain speech because clear speech is directed towards a non-native or hearing-impaired listener, who is assumed to have difficulty in speech perception. Thus, talkers must expend extra effort to make themselves understood to these listeners. Several more recent theories of phonetic reduction, including Aylett and Turk's (2004, 2006) smooth signal redundancy hypothesis and van Son and Pols' (2003) model of communicative efficiency, similarly attempt to provide a comprehensive account of phonetic reduction in terms of this competition between production-oriented and perception-oriented constraints.

Although this kind of comprehensive theory of phonetic reduction that can account for the effects of both lexical and discourse factors is desirable, recent research suggests that a more complex model may be necessary. For example, as Lindblom (1990) observed, temporal and spectral reduction are not perfectly correlated. In his analysis of clear vs. plain lab speech, he did not observe a linear relationship between vowel duration and dispersion, suggesting that phonetic reduction in the two domains may be independent (cf. Moon & Lindblom, 1994). This

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

independence of phonetic reduction across acoustic domains is further supported by mixed findings for temporal vs. spectral reduction for other linguistic factors. In particular, whereas temporal reduction is robustly predicted by high lexical frequency (Aylett & Turk, 2004; Gahl et al., 2012; Munson & Solomon, 2004), high semantic predictability (Clopper & Pierrehumbert, 2008; Hunnicutt, 1987; Lieberman, 1963), and second mention in a discourse (Baker & Bradlow, 2009; Bard et al., 2000; Fowler & Housum, 1987), the evidence for temporal reduction due to low phonological neighborhood density is fairly modest. Scarborough (2010) reported both temporal and spectral reduction for words with few neighbors relative to words with many neighbors, but Munson and Solomon (2004) and Munson (2007) obtained significant effects of phonological neighborhood density only in the spectral domain. These independent effects of temporal and spectral reduction minimally require a model in which the production and/or perception constraints impacting vowel variation differentially affect vowel duration and dispersion. However, these independent effects across acoustic domains could also imply that phonetic reduction is not a single process reflecting a single set of competing constraints on production and perception.

The notion of a single phonetic reduction process arising from competing constraints on production and perception is also potentially challenged by recent research demonstrating interactions among these factors in their effects on vowel duration. Interactions have been observed among lexical frequency, discourse mention, and speaking style (Baker & Bradlow, 2009), among lexical frequency and semantic predictability (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009), and among lexical frequency and neighborhood density (Turnbull, 2015; cf. Munson & Solomon, 2004). These interactions are typically interpreted as reflecting maximal reduction in some contexts and constraints on the lower bound of reduction in other contexts (e.g., Baker & Bradlow, 2009; Wright, 2004), suggesting that an additional set of constraints may be necessary to account for the combined effects of these lexical and discourse factors on phonetic reduction.

The goal of the current study was to obtain further empirical evidence for the independent and combined effects of lexical and discourse factors on temporal and spectral vowel reduction. Specifically, we examined the combined impact of lexical frequency, phonological neighborhood density, semantic predictability, discourse mention, and speaking style on vowel duration and dispersion in continuous read speech passages. The simultaneous manipulation of all five variables allows for a comprehensive exploration of the effects and interactions among these factors on both temporal and spectral vowel reduction. Given the previous findings, we expected to observe complex patterns of interactions among the five factors, as well as different patterns of interactions across the temporal and spectral measures.

2. Methods

2.1. Talkers

The read passages were produced by 30 monolingual native speakers of Midwestern American English (20 female, 10 male). The talkers ranged in age from 18-29 years old (*M* = 21, *SD* = 2). Half of the talkers of each gender (10 female, 5 male) were lifetime residents of the Midland dialect region, which encompasses the lower Midwestern United States. The other half of the talkers had lived exclusively in the Northern dialect region, which encompasses the upper Midwestern United States, until at least the age of 18 years old. The talkers were recruited from

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

the Ohio State University community and were recorded in Columbus, OH, in the Midland dialect region. Together, these 30 talkers comprise the Columbus component of the Ohio State Stories Corpus.

2.2. Materials

Each talker was recorded reading 30 short stories two times each, for a total of 60 stories per talker and 1,800 stories total. The short stories were written in the style of those developed by Baker and Bradlow (2009) and contained a total of 236 target words. The 236 target words included 234 unique targets (*hot* and *bad* were presented in two stories each) of which 232 were monosyllabic and two (*tunic* and *water*) were disyllabic. The stressed vowel in each target word was one of the set /i, ɛ, æ, ɑ, ɔ, u/. The target words were selected to achieve an approximate balance of high vs. low log frequency and high vs. low neighborhood density across the six vowel categories, to the extent possible given the constraints of the English lexicon. The log frequency and neighborhood density measures for each target word were obtained from the Hoosier Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984). For the purposes of target word selection, high frequency words were defined as those with a log frequency above 2.5 and high density words were defined as those with 12 or more neighbors. The distribution of target words by stressed vowel category, log frequency (high vs. low), and neighborhood density (high vs. low) is shown in Table 1. The complete list of target words is provided in the Appendix.

Vowel	Low Frequency High Density	Low Frequency Low Density	High Frequency High Density	High Frequency Low Density	Total
i	9	15	11	10	45
ɛ	7	12	10	10	39
æ	10	12	11	10	43
ɑ	9	12	10	3	34
ɔ	5	14	9	10	38
u	10	9	9	9	37
Total	50	74	60	52	236

Table 1. Number of target words with each stressed vowel for each combination of high and low lexical frequency and neighborhood density.

Each target word was presented twice within a single short story to allow for an examination of discourse mention (first vs. second). To increase the similarity of the prosodic context of the two mentions, both mentions of each target word were presented in similar positions in the sentence relative to punctuation commonly associated with prosodic phrase breaks, such as commas and periods.

The predictability of each mention of each target word was determined in a separate cloze task in which participants were visually presented with sentences missing a target word and were asked to provide the missing word in an open-set task. Given that some of the sentences contained more than one target word, 20 separate between-subject stimulus lists were created so that each participant saw a given sentence maximally one time. A total of 81 participants completed the cloze task and we obtained at least 8 responses to each target word (maximum: 81;

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

median: 32), with the exception of the second mentions of *top* and *dodge*, which were inadvertently left out of the cloze task materials. Cloze predictability was then defined as the proportion of correct responses (range: 0-1; median: 0.125). The cloze task confirmed that the predictability of the target words was relatively balanced across first and second mentions, as shown separately for each vowel in Table 2. For the purposes of describing the stimulus materials in Table 2, high predictability words were defined as those with a cloze predictability greater than or equal to 0.125. Each target word is represented twice in Table 2 (once per mention), except for *top* and *dodge*, for which predictability is not available for the second mention.

Vowel	Low Predictability First Mention	Low Predictability Second Mention	High Predictability First Mention	High Predictability Second Mention
i	22	27	23	18
ɛ	14	13	25	26
æ	20	20	23	23
ɑ	18	17	16	15
ɔ	15	23	23	15
u	18	22	19	15
Total	107	122	129	112

Table 2. Number of target words with each stressed vowel for each combination of high and low cloze predictability and first and second discourse mention. Each target word is represented twice in Table 2 (once per mention), except for *top* and *dodge*, for which predictability is not available for the second mention.

2.3. Recording procedure

Each talker was recorded reading the complete set of 30 short stories twice. The talker was seated at a laptop computer in a sound-attenuated booth. In the first block of the experiment, the short stories were presented one at a time on the laptop screen and the talker was asked to read the stories aloud as if talking to a friend, to elicit plain lab speech. The second block of the experiment was identical to the first, except that the talker was asked to read the stories aloud as if talking to someone who is hearing-impaired or a non-native speaker of English, to elicit clear lab speech. The plain speech block was always presented first so that any reduction in the second block due to repetition of the stories would not enhance the stylistic difference across blocks. In both blocks of the experiment, the presentation of the stories was self-paced and the talker could proceed to the next story by pressing the space bar. A reminder of the intended interlocutor was presented prior to each story to encourage maintenance of the target style throughout the block. The order of the stories was randomized separately for each talker for each block. A break was provided after 15 stories within each block and between the two blocks, during which the talker was invited to leave the booth if they wished.

The stories were digitally recorded using a Marantz PMD661 solid-state recorder and a Shure SM10A headset microphone. The recordings were digitized with a sampling rate of 44.1 kHz and 16-bit resolution.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

2.4. Acoustic analysis

To facilitate acoustic analysis of the target words, the recorded stories were segmented into word and phone units using the Penn Phonetics Lab Forced Aligner (Yuan & Liberman, 2008). The alignment of the target vowels was then hand-corrected by a team of research assistants overseen by the first author. Reliability of the hand-corrections was assessed by examining differences in extracted vowel duration and first and second formant frequencies for a set of five passages (one from each of five different talkers) that were hand-corrected by all of the research assistants. The distributions of the differences were all normally distributed around 0, suggesting that variation across research assistants was not biased towards larger or smaller values. The standard deviations of the difference scores were 69 ms for duration, 40 Hz for F1, and 88 Hz for F2, suggesting that variation across research assistants was also reasonably modest.[1]

The recordings produced a total of 28,320 target tokens (236 target words x 2 mentions x 2 styles x 30 talkers). From this set, 37 tokens (0.1%) were missing due to reading or recording errors, 343 tokens (1.2%) were excluded due to disfluencies in reading, 3,652 tokens (12.9%) were excluded for non-modal voicing that prevented accurate estimation of formant frequencies, and 167 second mentions (0.6%) were excluded because the associated first mention was either not produced, produced with a disfluency, or repeated due to a disfluency. Vowel duration and the first and second formant frequencies at five timepoints during the vowel (i.e., 20%, 35%, 50%, 65%, and 80% of the vowel duration) were extracted automatically from the remaining 24,121 tokens. The formant frequencies were converted to Bark (Traunmüller, 1990) for analysis. Trajectory length was calculated from the formant frequencies in both Hz and Bark for each token (Fox & Jacewicz, 2009). Duration outliers were identified separately for each talker as values that were more than 3 standard deviations from the talker's mean vowel duration. Formant frequency and trajectory outliers were identified separately for each vowel for each talker as values that were more than 3 standard deviations from the talker's mean of the target measure for that vowel. The 863 tokens which exhibited an outlier value for one or more of the seven measures (duration in ms; F1, F2, trajectory length in Hz; F1, F2, trajectory length in Bark) were excluded from further analysis. The 109 remaining second mentions of *top* and *dodge* were also excluded because cloze predictability was not available. The final analysis is therefore based on 23,149 tokens.

Temporal reduction was assessed using the vowel duration measure. To assess spectral reduction, dispersion was defined as the Euclidean distance in the F1 x F2 Bark space of the midpoint F1 and F2 of each vowel token from the center of the vowel space, defined separately for each talker as the grand mean of F1 and F2 over all six vowel categories.

2.5. Statistical analysis

The effects of lexical frequency, neighborhood density, cloze predictability, discourse mention, and speaking style on temporal and spectral vowel reduction were examined in a series of linear mixed-effects regression models. Vowel duration was the dependent variable in the models

---

[1] The relatively large standard deviation for duration reflects one measurement that resulted from a clear mislabeling of one section in one of the stories. The standard deviations of the difference scores without that measurement were 20 ms for duration, 32 Hz for F1, and 88 Hz for F2.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

exploring temporal reduction and vowel dispersion was the dependent variable in the models exploring spectral reduction. In both models, the fixed effects included all five linguistic factors, as well as all 2-, 3-, 4-, and 5-way interactions. Lexical frequency, neighborhood density, and cloze predictability were continuous variables and were centered prior to being entered in the models. Discourse mention and speaking style were categorical variables with two levels each. The reference level for discourse mention was first mention and the reference level for speaking style was clear speech. Vowel duration was also included as a continuous covariate fixed effect in the dispersion model to ensure that any observed dispersion effects could not be attributed simply to duration (Moon & Lindblom, 1994). In both models, the random effects were the maximal simple random effects structure justified by the design (Barr, Levy, Scheepers, & Tily, 2013) and consisted of random intercepts for word and talker as well as random by-talker slopes for lexical frequency, neighborhood density, cloze predictability, discourse mention, and speaking style and random by-word slopes for cloze predictability, discourse mention, and speaking style. Random slope interactions were not included to ensure model convergence. Given the large number of tokens included in the analysis, coefficient estimates with $|t| > 2$ were interpreted as significant ($p < .05$) because the t-distribution approaches a normal distribution with large degrees of freedom.

3. Results

3.1. Vowel duration

A summary of the model predicting vowel duration is shown in Table 3. The results of the analysis reveal significant effects of lexical frequency and speaking style as well as three significant interactions: lexical frequency x speaking style, neighborhood density x speaking style, and neighborhood density x discourse mention. None of the other main effects or interactions are significant.

The main effects of lexical frequency and speaking style are in the expected direction. High frequency words have shorter vowel durations than low frequency words and words in plain speech have shorter vowel durations than words in clear speech. These two main effects, as well as their interaction, are shown in Figure 1. The interaction reflects a greater effect of lexical frequency in clear speech than in plain speech such that the effect of speaking style is maximized for low frequency words.

The interaction between neighborhood density and speaking style is shown in Figure 2, with the x-axis scale reversed so that greater reduction is to the right as in Figure 1. The overall patterns are in the expected direction: words with few neighbors have shorter vowel durations than words with many neighbors and words in plain speech have shorter vowel durations than words in clear speech. The interaction reflects a greater effect of neighborhood density in plain speech than in clear speech, which is the opposite of the pattern observed for lexical frequency. That is, whereas the lexical frequency effect is enhanced in clear speech, the neighborhood density effect is enhanced in plain speech. This enhancement of the neighborhood density effect in plain speech relative to clear speech may provide an explanation for the mixed results reported in previous studies examining neighborhood density effects on temporal reduction (Munson, 2007; Munson & Solomon, 2004; Scarborough, 2010). In particular, neighborhood density effects on vowel duration may be more difficult to elicit in word list reading tasks (e.g., Munson, 2007; Munson & Solomon, 2004; Scarborough, 2010; Wright, 2004), which encourage

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

hyperarticulated citation form speech, than in more natural speech tasks, including the story reading task in the current study. Thus, the absence of a significant effect of neighborhood density on vowel duration in some of the previous studies (Munson, 2007; Munson & Solomon, 2004) may reflect a further attenuation of the magnitude of the effect in word list reading relative to clearly produced short stories.

Predictor	Estimate	t-value
Intercept	143.84	33.94
Frequency	-10.43	-3.04*
Density	0.60	1.90
Predictability	0.28	0.04
Mention (Second)	-2.07	-1.16
Style (Plain)	-15.81	-5.02*
Frequency x Density	0.11	0.28
Frequency x Predictability	-3.93	-0.40
Frequency x Mention	3.64	1.60
Frequency x Style	2.99	3.08*
Density x Predictability	0.23	0.30
Density x Mention	0.64	3.08*
Density x Style	0.28	3.08*
Predictability x Mention	-6.65	-0.93
Predictability x Style	-2.73	-1.05
Mention x Style	0.07	0.09
Frequency x Density x Predictability	0.58	0.57
Frequency x Density x Mention	-0.32	-1.20
Frequency x Density x Style	-0.06	-0.50
Frequency x Predictability x Mention	1.48	0.15
Frequency x Predictability x Style	2.34	0.66
Frequency x Mention x Style	0.23	0.22
Density x Predictability x Mention	-0.82	-1.12
Density x Predictability x Style	0.20	0.72
Density x Mention x Style	-0.15	-1.52
Predictability x Mention x Style	0.89	0.27
Frequency x Density x Predictability x Mention	-0.98	-1.03
Frequency x Density x Predictability x Style	-0.32	-0.81
Frequency x Density x Mention x Style	0.17	1.39
Frequency x Predictability x Mention x Style	-0.66	-0.15
Density x Predictability x Mention x Style	0.31	0.89
Frequency x Density x Predictability x Mention x Style	0.73	1.61

Table 3. Summary of the mixed-effects regression model predicting vowel duration. *p < .05.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

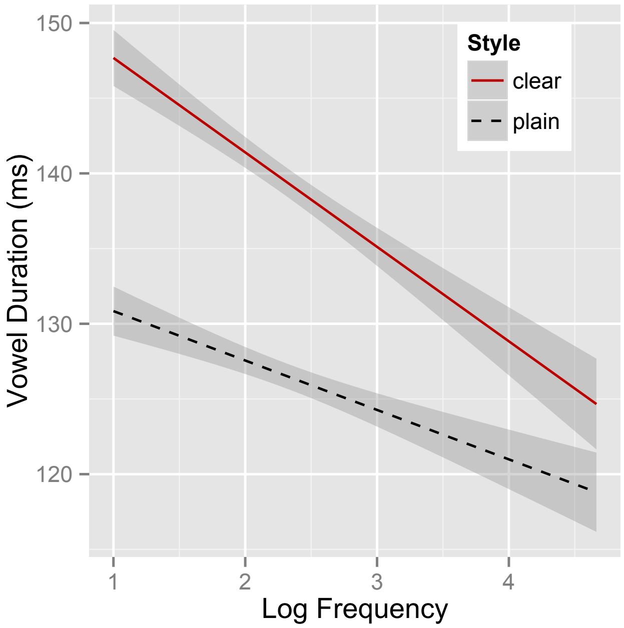

Figure 1. Effects of lexical frequency and speaking style on vowel duration in the Ohio State Stories Corpus.

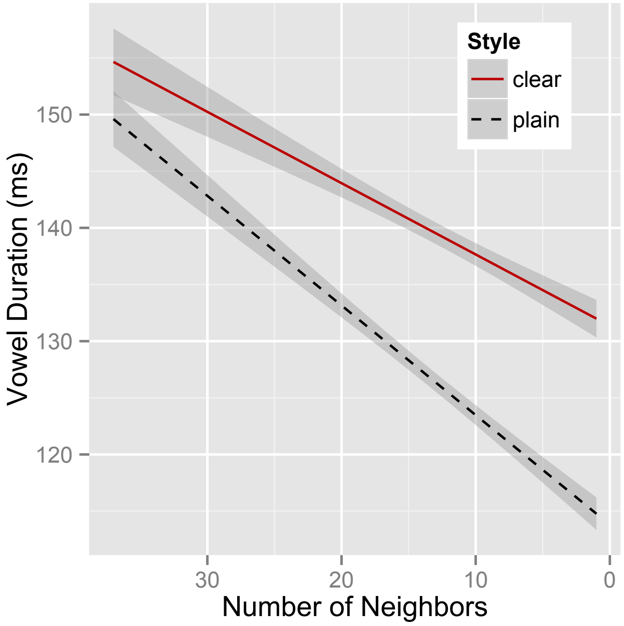

Figure 2. Effects of neighborhood density and speaking style on vowel duration in the Ohio State Stories Corpus. Note that the x-axis is reversed so that greater reduction is to the right.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

The interaction between neighborhood density and discourse mention is shown in Figure 3, with the x-axis scale reversed as in Figure 2. Consistent with the results of the statistical model, no clear effect of discourse mention can be observed. However, the interaction reflects a greater effect of neighborhood density for second mentions than first mentions. This result is comparable to the interaction between neighborhood density and speaking style: in both cases the effect of neighborhood density is greater in the context in which reduction is expected (plain speech and second mentions). The interaction also reveals a cross-over effect in which second mentions have longer vowels than first mentions for words with many phonological neighbors. This result is unexpected and may reflect other factors such as prosodic structure (Burdin & Clopper, 2015) or the nature of the relationship between the first and second mentions of the target words, including whether or not they involve the same referent and the amount of time between the two mentions (Fowler & Housum, 1987), neither of which were controlled in the short story materials or the statistical models.

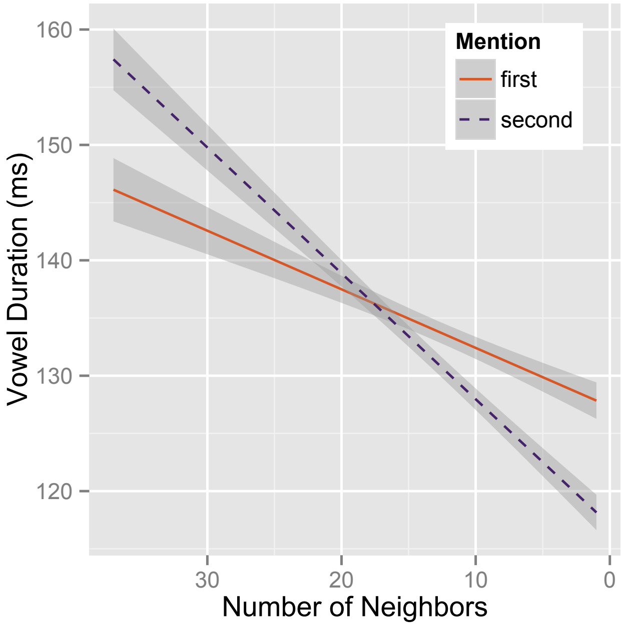

Figure 3. Effects of neighborhood density and discourse mention on vowel duration in the Ohio State Stories Corpus. Note that the x-axis is reversed so that greater reduction is to the right.

3.2. Vowel dispersion

A summary of the model predicting vowel dispersion is shown in Table 4. The results of the analysis reveal significant main effects of lexical frequency, discourse mention, and speaking style, as well as a significant four-way interaction between lexical frequency, neighborhood density, cloze predictability, and discourse mention. These effects are independent of a significant effect of the duration covariate which confirms that longer vowels are more dispersed than shorter vowels. No other main effects or interactions are significant.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

Predictor	Estimate	t-value
Intercept	1.9510	26.44
Frequency	-0.1535	-2.01*
Density	0.0035	0.49
Predictability	-0.0755	-1.59
Mention	-0.0497	-4.33*
Style	-0.1310	-6.88*
Duration	0.0019	23.55*
Frequency x Density	0.0014	0.16
Frequency x Predictability	0.0527	0.81
Frequency x Mention	0.0185	1.28
Frequency x Style	-0.0162	-1.40
Density x Predictability	0.0050	0.99
Density x Mention	-0.0003	-0.20
Density x Style	-0.0008	-0.74
Predictability x Mention	-0.0189	-0.41
Predictability x Style	0.0096	0.30
Mention x Style	-0.0108	-1.06
Frequency x Density x Predictability	0.0028	0.41
Frequency x Density x Mention	-0.0002	-0.10
Frequency x Density x Style	-0.0004	-0.29
Frequency x Predictability x Mention	0.0303	0.49
Frequency x Predictability x Style	0.0052	0.12
Frequency x Mention x Style	-0.0027	-0.20
Density x Predictability x Mention	0.0010	0.21
Density x Predictability x Style	-0.0021	-0.62
Density x Mention x Style	-0.0007	-0.60
Predictability x Mention x Style	0.0289	0.70
Frequency x Density x Predictability x Mention	-0.0128	-2.02*
Frequency x Density x Predictability x Style	-0.0003	-0.06
Frequency x Density x Mention x Style	0.0020	1.29
Frequency x Predictability x Mention x Style	-0.0483	-0.87
Density x Predictability x Mention x Style	-0.0045	-1.02
Frequency x Density x Predictability x Mention x Style	0.0043	0.75

Table 4. Summary of the mixed-effects regression model predicting vowel dispersion. *p < .05.

As in the vowel duration analysis, the main effects of lexical frequency and speaking style are in the expected direction, as shown by the coefficient estimates in Table 4. High frequency words have less dispersed vowels than low frequency words and words in plain speech have less dispersed vowels than words in clear speech. The main effect of discourse mention is also in the expected direction, as shown by the coefficient estimate in Table 4. Second mention words have less dispersed vowels than first mention words.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

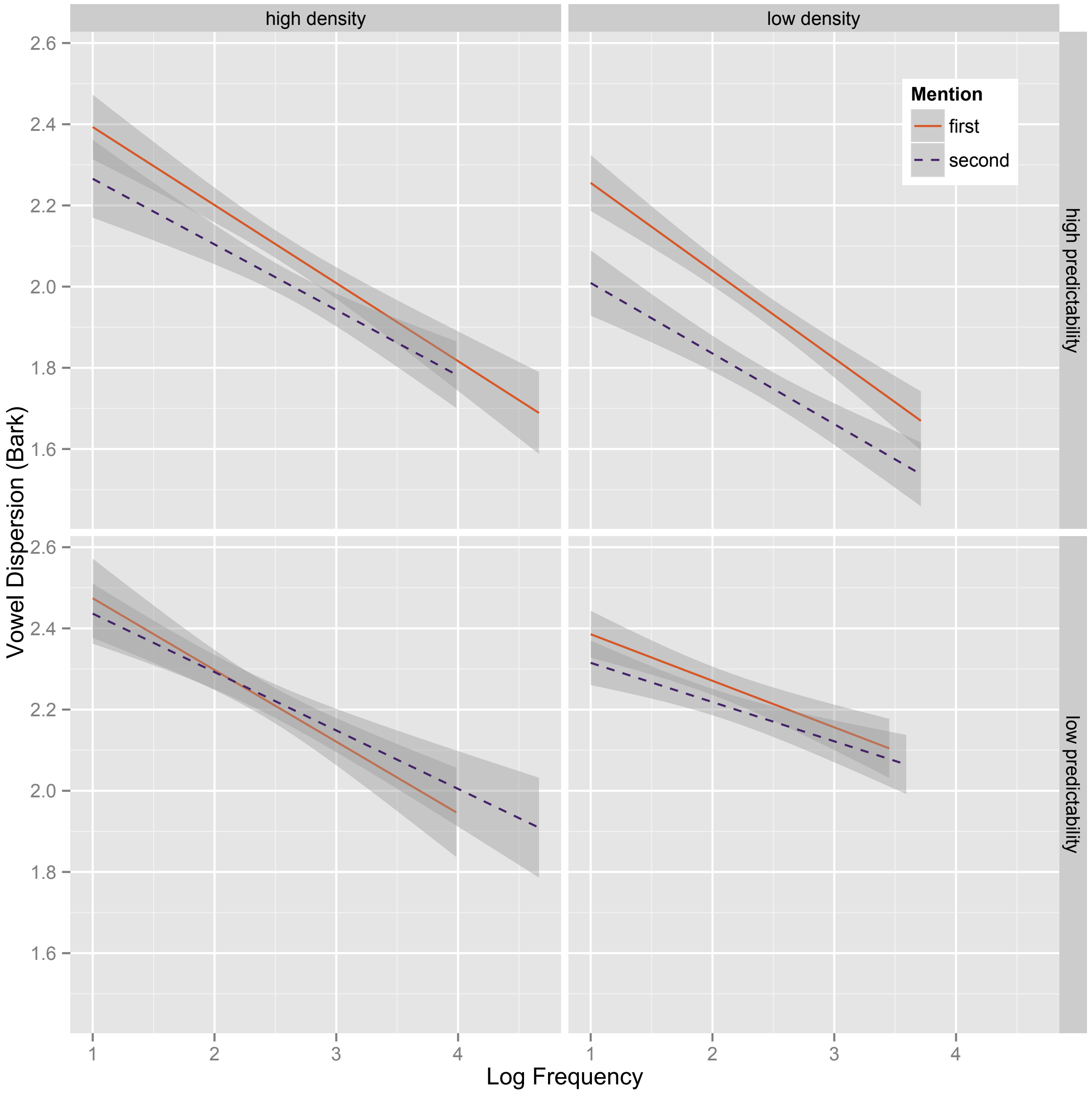

Figure 4. Effects of lexical frequency, neighborhood density, cloze predictability, and discourse mention on vowel dispersion in the Ohio State Stories Corpus. Note that neighborhood density and cloze predictability are shown here as binary factors, but were treated as continuous variables in the statistical analysis.

The four-way interaction between lexical frequency, neighborhood density, cloze predictability, and discourse mention is shown in Figure 4. For ease of illustration, the continuous neighborhood density and cloze predictability variables are treated as binary variables, as in Tables 1 and 2, respectively. The overall patterns for these two variables are in the expected direction: words with few neighbors have less dispersed vowels than words with many neighbors and high predictability words have less dispersed vowels than low predictability words.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

Consistent with the main effects observed in the model, high frequency words have less dispersed vowels than low frequency words for all combinations of neighborhood density, cloze predictability, and discourse mention. However, this effect of lexical frequency is greatest for the high density, high predictability words (top left panel of Figure 4) and smallest for the low density, low predictability words (bottom right panel of Figure 4). Similarly, second mention words have less dispersed vowels than first mention words overall, but this effect of discourse mention is greatest for the low density, high predictability words (top right panel of Figure 4) and smallest for the high density, low predictability words (bottom left panel of Figure 4). Taken together, these results suggest similar patterns of interaction for cloze predictability with lexical frequency and discourse mention: the effects of both lexical frequency and discourse mention are enhanced in high predictability words relative to low predictability words. In contrast, the results suggest opposite patterns of interaction for neighborhood density with lexical frequency and discourse mention: whereas the lexical frequency effect is enhanced for high density words, the discourse mention effect is enhanced for low density words.

4. Discussion

The results revealed significant effects of lexical frequency, neighborhood density, cloze predictability, discourse mention, and speaking style on vowel duration and/or dispersion. Lexical frequency and speaking style exhibited main effects for both acoustic measures and discourse mention exhibited a main effect on vowel dispersion. In all cases, the effects were in the expected direction: high frequency words had shorter and less dispersed vowels than low frequency words, second mention words had less dispersed vowels than first mention words, and words in plain speech had shorter and less dispersed vowels than words in clear speech. Neighborhood density and cloze predictability did not exhibit significant main effects in either analysis, but neighborhood density interacted significantly with discourse mention and speaking style in the duration analysis and both neighborhood density and cloze predictability contributed to a significant four-way interaction with lexical frequency and discourse mention in the dispersion analysis. Thus, all five of the variables that we examined contributed to variation in vowel duration and/or dispersion, consistent with the previous literature demonstrating the independent effects of these factors on vowel production.

However, the results of our analysis also revealed different patterns of main effects and interactions across the two acoustic domains. In particular, although main effects of lexical frequency and speaking style were observed for both vowel duration and dispersion, the main effect of discourse mention was only significant in the dispersion analysis. Similarly, although cloze predictability did not contribute to vowel duration, it was involved in a significant four-way interaction predicting vowel dispersion. Other patterns of interaction also differed across the two analyses. Whereas neighborhood density interacted separately with discourse mention and speaking style in the duration analysis, neighborhood density interacted with discourse mention, cloze predictability, and lexical frequency in a single four-way interaction in the dispersion analysis. Similarly, speaking style interacted separately with lexical frequency and neighborhood density in the duration analysis, but no significant interactions with speaking style were observed in the dispersion analysis. Given that the vowel dispersion results were independent of a significant effect of the vowel duration covariate (see Table 4), these different patterns of predictors across the two acoustic domains suggest that these five factors, independently and in combination, make different contributions to phonetic reduction in vowel duration and vowel

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

dispersion. Thus, the constraints on phonetic reduction may similarly vary across acoustic domains.

The patterns of interaction that we observed also point to the need for a more complex understanding of the factors at play in phonetic reduction. In the duration analysis, lexical frequency and speaking style exhibited a different pattern of interaction than neighborhood density and speaking style or neighborhood density and discourse mention. In particular, whereas the lexical frequency effect was enhanced in clear speech relative to plain speech, the neighborhood density effect was enhanced in plain speech and second mentions relative to clear speech and first mentions, respectively. Similarly, in the dispersion analysis, whereas the lexical frequency and discourse mention effects were enhanced in high predictability words relative to low predictability words and the discourse mention effect was enhanced for the low density words, the lexical frequency effect was enhanced for high density words.

From a listener-oriented perspective (e.g., Aylett & Turk, 2004; Lindblom, 1990), these various interactions reflect different patterns of enhancement that do not neatly map onto the presumed ease of processing for the listener.[2] In particular, with respect to ease of processing, low frequency words are harder to identify in perception than high frequency words (Howes, 1957), high density words are harder to identify than low density words (Luce & Pisoni, 1998), unpredictable words are harder to identify than predictable words (Miller & Isard, 1963), and new words are harder to identify than repeated words (Slowiaczek & Pisoni, 1986). Similarly, clear speech is directed towards listeners with presumed perceptual difficulty (Picheny et al., 1986). Thus, low frequency words, high density words, low predictability words, first mention words, and clear speech are all “hard” contexts for listeners, whereas high frequency words, low density words, high predictability words, second mention words, and plain speech are all “easy” contexts for listeners.

Table 5 presents a summary of the observed interactions in our study from the perspective of this listener-oriented “easy/hard” distinction. This summary reveals that whereas neighborhood density and discourse mention effects are enhanced in “easy” contexts, lexical frequency effects are more mixed. The lexical frequency effect is enhanced for dispersion in one “easy” context (high predictability), but it is also enhanced for dispersion in one “hard” context (high density) and for duration in one “hard” context (clear speech). The neighborhood density and discourse mention effects are consistent with maximum reduction in “easy” contexts (Baker & Bradlow, 2009), including maximum temporal reduction for low density, second mention, plain speech and maximum spectral reduction for low density, high predictability, second mentions. In contrast, the lexical frequency effects are consistent with a mixed pattern of maximum reduction in “easy” contexts and maximum enhancement in “hard” contexts (Wright, 2004), including maximum spectral reduction of high frequency, high predictability words, maximum spectral enhancement of low frequency, high density words, and maximum temporal enhancement of low frequency, clear speech.

---

[2] An alternative perspective is that processing difficulty for the talker, rather than the listener, underlies phonetic reduction effects (e.g., Bard et al., 2000; Bell et al., 2009). Our data cannot distinguish between these two perspectives and the following interpretation of our data would not be qualitatively different from the talker-oriented perspective.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

Acoustic domain	Effect of…	Is enhanced in the context of…	Easy/Hard
vowel duration	lexical frequency	clear speech	hard
vowel duration	neighborhood density	plain speech	easy
vowel duration	neighborhood density	second mention	easy
vowel dispersion	lexical frequency	high density	hard
vowel dispersion	lexical frequency	high predictability	easy
vowel dispersion	discourse mention	low density	easy
vowel dispersion	discourse mention	high predictability	easy

Table 5. Summary of the significant interactions from a listener-oriented perspective. Interactions resulting in enhancement of an effect in a "hard" listener context are highlighted.

Taken together, our results are consistent with a model of phonetic reduction in which vowels are temporally and spectrally reduced in contexts that impose relatively low processing costs on the listener (e.g., Aylett & Turk, 2004; Lindblom, 1990). The diverse patterns of interactions further suggest that phonetic reduction is maximized under certain combinations of "easy" contexts and phonetic enhancement is maximized under certain combinations of "hard" contexts. Moreover, these combinations of "easy" and "hard" contexts differ across the temporal and spectral domains for vowels. Thus, although these results are consistent with a listener-oriented model of phonetic reduction, they suggest that some refinement to such an approach is necessary to account for the observed differences in the magnitude of the effects across acoustic domains and across combinations of lexical and discourse factors.

Acknowledgements

The authors would like to thank Abby Walker for assistance with stimulus design, Anna Crabb, McKenna Reeher, and Shannon Melvin for assistance with data collection, and Anna Crabb, Jessica Hanson, Sarah Mabie, Rachel Monnin, Christine Prechtel, and Shannon Melvin for assistance with data annotation. This work was supported by the National Science Foundation (BCS-1056409).

Appendix

The 234 unique target words in the Ohio State Stories Corpus are: act, ask, axe, back, bad, bag, bath, beach, beak, beat, bed, beef, best, bet, black, bleed, booth, boots, boss, bought, box, brag, bread, breast, breath, brief, broth, bruise, cat, catch, caught, cause, chalk, cheap, cheese, chef, chess, chest, chief, chop, class, clause, claws, clogs, cloth, cops, cot, cough, crab, crash, crew, crop, cute, deaf, death, deck, depth, desk, dodge, dog, draft, dream, dress, drop, fact, faun, feast, feud, flaw, flax, fleas, fleece, flesh, floss, flute, food, fought, fox, fresh, frock, frog, fruit, gauze, glass, gnats, god, gone, goose, greed, grief, group, hat, haunt, head, heat, hot, huge, jaw, jazz, jeans, job, juice, June, knee, knots, law, lawn, leg, lodge, long, loose, lost, mask, match, math, mead, meat, mesh, mess, mob, mood, moon, moose, moss, moth, nap, neck, nest, net, news, niece, nude, off, ooze, ox, pact, pass, path, pet, please, pot, priest, proof, quest, raft, rat, rest, rock, root, rot, sad, sauce, scab, scheme, scrap, scratched, sect, seed, shed, sheep's, sheets, shock, shoe, shop, shot, shrew, sketch, ski, slapped, sled, sleeve, sleuth, sloth, smog, smooth, snag, snooze, soft, soothe, speech, speed, splash, squad, squeak, staff, step, stew, stock, stop, straw,

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

stream, street, stress, strong, suit, suite, swab, swan, swat, sweat, tag, talk, task, tax, teeth, test, theft, theme, thief, thought, thread, threat, top, toss, track, treat, tree, true, truth, tube, tunic, tweed, tweet, view, walk, wash, watch, water, weed, wheat, yacht, yeast, youth, zest.

References

Aylett, M., & Turk, A. E. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*, 31-56.

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

Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech, 52*, 391-413.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1-22.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational speech. *Journal of Memory and Language, 60*, 92-111.

Burdin, R. S., & Clopper, C. G. (2015). Phonetic reduction, vowel duration, and prosodic structure. *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, Scotland.

Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America, 124,* 1682-1688.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*, 1241-1255.

Fowler, C. A., & Housum, J. (1987). Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489-504.

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America, 126*, 2603-2618.

Gahl, S. (2015). Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics, 49*, 96-116.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

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

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America, 29*, 296-305.

Hunnicutt, S. (1987). Acoustic correlates of redundancy and intelligibility. *Quarterly Progress Status Report, Speech Transmission Lab, 28*(2-3), 7-14.

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

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

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

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

Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2*, 217-228.

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*, 40-55.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. In Cole, J., & Hualde, J. I. (Eds.), *Laboratory Phonology 9* (pp. 201-227). Berlin: Mouton de Gruyter.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of 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. *Research on Speech Perception Progress Report No. 10* (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana University.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*, 434-445.

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In Fougeron, C., Kühnert, B., D'Imperio, M., & Vallée, N. (Eds.), *Laboratory Phonology 10* (pp. 557-586). Berlin: De Gruyter Mouton.

Slowiaczek, L. M., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory and Cognition, 14*, 230-237.

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

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass, 3*, 236-264.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America, 88*, 97-100.

Turnbull, R. (2015). *Assessing the listener-oriented account of predictability-based phonetic reduction*. Doctoral dissertation, Ohio State University, Columbus, OH.

van Son, R. & Pols, L. C. W. (2003). Information structure and efficiency in speech production. *Proceedings of Eurospeech 2003* (pp. 769-772). Geneva, Switzerland.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In Local, J., Ogden, R., & Temple, R. (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp. 75-86). Cambridge: Cambridge University Press.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08,* 5687-5690.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 129.199.81.174 On: Thu, 01 Oct 2015 06:56:29