Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 2pSCb: Speech Intelligibility (Poster Session)

2pSCb8. Effects of semantic predictability and dialect variation on vowel production in clear and plain lab speech

Rory Turnbull* and Cynthia G. Clopper

***Corresponding author's address: Department of Linguistics, Ohio State University, 222 Oxley Hall, Columbus, OH 43215, turnbull@ling.osu.edu**

Speech addressed to a non-native or hearing impaired listener features longer, more peripheral vowels. In addition, more extreme dialect-specific forms are produced in semantically predictable contexts, and less extreme forms (more standard forms) in unpredictable contexts. This study investigated the interactions between predictability and speaking style on Southern American English monophthongization of the vowel /aj/. The Midland dialect of American English served as the comparison. Participants read a set of sentences with monosyllabic target words in sentence-final position. Target words varied in semantic predictability based on the preceding sentential context. Each set of sentences was produced twice by each participant - first as if talking to a friend ("plain" speech) and again as if talking to a non-native or hearing impaired listener ("clear" speech). The duration, dispersion, and trajectory length of the vowel in each target word were measured. Preliminary results suggest that, as expected, Southern /aj/ has a shorter trajectory length than Midland /aj/, and in both dialects, /aj/ has a shorter trajectory length in clear speech than plain speech. However, these processes do not interact with each other or with semantic predictability, suggesting that style and predictability effects are independent of the realization of some dialect variants.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800652]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Vowel production is known to be influenced by a number of diverse factors. Among these are effects of speech style and semantic predictability. Speech in a "clear" style, which is addressed to a non-native or hearing-impaired listener, has been observed to feature longer, more peripheral vowels than speech in a "plain" style (or "lab speech"), which is addressed to a friend (e.g. Picheny *et al.*, 1985; Smiljanić and Bradlow, 2009). In semantically predictable contexts, speech has been shown to be less intelligible (Lieberman, 1963) and vowels are shorter and more centralized (Aylett and Turk, 2006).

Mitsch and Clopper (2012) examined the production of the vowel /æ/ by talkers of the Northern and Midland dialects of American English. This vowel is more fronted and diphthongal in the Northern dialect than in the Midland dialect (Labov *et al.*, 2006). Mitsch and Clopper's (2012) task involved participants speaking in first a plain and then a clear style, and was designed to investigate potential interactions between style and dialect. While their results confirmed previous findings of style—longer, more peripheral vowels in clear speech—and regional dialect effects—a more fronted and diphthongal /æ/ for Northerners—they did not observe any interactions between these two effects. On the other hand, the predictability effect has been found to interact with dialect variation. Clopper and Pierrehumbert (2008) found that more extreme dialect-specific forms are produced in semantically predictable contexts, and less extreme forms (more standard forms) are produced in unpredictable contexts. However, it is not known how semantic predictability and speaking style interact together with regards to their effects on dialect-specific vowel production. The goal of the current study was to explore the potential interactions among these factors.

The differences between /æ/ in the Midland and Northern dialects are not particularly socially salient (Labov, 2010, Ch. 10), and the Northern dialect itself is not fully established in folk discourse on 'accents' (Campbell-Kibler, 2012). The lack of an interaction in Mitsch and Clopper's (2012) results could be due to the Northern talkers' non-awareness that they even *have* a non-standard accent. For these reasons, the present study investigated potential interactions between predictability and speaking style on more socially salient variables: the Southern vowel shift and monophthongization of the vowel /aj/ in Southern American English. The Southern dialect is characterized by /aj/ monophthongization and the Southern vowel shift, which involves peripheralization of /ɛ/ and centralization of /ej/, among other effects not relevant to the present study (Wolfram and Schilling-Estes, 1998). Thus, the vowels under analysis in the present study were /ej, ɛ, aj, ɑ/. The monophthongization of /aj/ has been shown to be a very salient variable for indexing Southern speech and Southern identity, both for Southerners and non-Southerners alike (Allbritten, 2011; Plichta and Preston, 2005; Thomas, 2003). These facts make this sociolinguistic variable an excellent point of comparison in the investigation of dialect-specific vowel production in different speaking styles and predictability contexts.

METHODS

Participants

Sixteen Ohio State University students, nine from the Midland dialect region (six females) and seven from the Southern dialect region (four females), participated in the production study for partial course credit or $10. All participants were monolingual native speakers of American English with monolingual English-speaking parents. They had all lived in their respective dialect region from birth to adulthood . None reported a history of speech, language, or hearing disorders.

Materials

Thirty-two matched high predictability (HP) and low predictability (LP) sentences were selected from the SPIN set (Kalikow *et al.*, 1977). A list of all stimulus sentences is given in Table 1. Note that for each sentence pair the target (final) word is the same in both sentences but the context differs in how well it predicts the target word. Sentences were selected to have 8 target words for each of the 4 vowels /ej, ɛ, aj, ɑ/.

TABLE 1: List of all stimulus sentences used in the experiment.

Vowel	Low predictability	High predictability
/ej/	Tom wants to know about the cake	For your birthday I baked a cake
	Tom discussed the hay	The farmer baled the hay
	I hope Paul asked about the mate	The lonely bird searched for its mate
	She's glad Jane asked about the drain	Ruth poured the water down the drain
	She might have discussed the ape	A chimpanzee is an ape
	Mr. Smith knew about the bay	The boat sailed across the bay
	The girl should consider the flame	The candle burned with a bright flame
	Mr. Smith spoke about the aid	The nurse gave him first aid
/ɛ/	Paul should know about the net	He caught the fish in his net
	He hears she asked about the deck	The sailor swabbed the deck
	We hear they asked about the shed	To store his wood he built a shed
	Ruth has discussed the peg	A round hole won't take a square peg
	They heard I asked about the bet	The gambler lost the bet
	He could discuss the bread	Spread some butter on your bread
	They heard I called about the pet	My son has a dog for a pet
	We will consider the debt	The poor man was deeply in debt
/aj/	The old man discussed the dive	The airplane went into a dive
	They might have considered the hive	The honey bees swarmed around the hive
	I am thinking about the knife	I cut my finger with a knife
	She couldn't discuss the pine	The furniture was made out of pine
	I've spoken about the pile	The sand was heaped in a pile
	Ruth must have known about the pie	For dessert he had apple pie
	We are speaking about the prize	Her entry should win first prize
	Mary can't consider the tide	We swam at the beach at high tide
/ɑ/	The girl knows about the swamps	Crocodiles live in muddy swamps
	She's spoken about the bomb	The airplane dropped a bomb
	We hear you called about the lock	This key won't fit in the lock
	Tom will discuss the cot	Harry slept on the folding cot
	He doesn't discuss the mop	Wash the floor with a mop
	We spoke about the knob	Unlock the door and turn the knob
	The man should discuss the ox	The plow was pulled by an ox
	The woman considered the notch	Tighten the belt by a notch

Procedure

Participants were asked to read all 32 sentences aloud one at a time as they were presented on a computer screen. In the first part of the experiment, participants were asked to read the set of sentences "as if you are talking to a friend", i.e. in a plain style. In the second part, participants were asked to read the set of sentences again, this time "as if you are talking to someone who is hearing-impaired or a non-native speaker of English", i.e. in a clear style.

Within each part, the sentences were blocked by semantic predictability, the order of which was counterbalanced across participants. Therefore, half of the participants read the LP sentences first, and half read the HP sentences first. Within each block, the sentences were presented in a different random order for each participant. All participants produced utterances in the plain style first, so that any reduction effects due to repetition would be countered by the clear speaking style. With 8 words × 4 vowels × 2 predictability conditions (HP and LP) × 2 style conditions (plain and clear), each participant produced a total of 128 target sentences. The productions were recorded at a sampling frequency of 44.1kHz using a Shure SM10A head-mounted microphone, set approximately 2.5cm from the lips, connected directly to a Marantz PMD661 digital recorder. The experiment was carried out in a double-wall sound attenuated booth.

Acoustic analysis

Four different measures of vowel reduction were taken—duration, vowel space area, vowel repulsion, and formant trajectory length. The duration of the vowel in each target word was measured from waveform and spectrogram displays.

To assess dispersion in the vowel space, the frequencies of the first two formants were measured at the midpoint of each vowel, using 10th-order LPC analysis with a sample window of 25ms. The total size of the vowel space can be assessed by treating it as a polygon of N points with an area A as defined in Equation 1, where the first and last points are the same, i.e. $F1_N F2_N = F1_0 F2_0$.

$$A = \frac{1}{2}\sum_{n=0}^{N-1} F1_n F2_{n+1} - F1_{n+1} F2_n \quad (1)$$

Using this method it is possible to compare the size of the vowel space, and thus dispersion, across different conditions. A more disperse vowel space with more peripheral vowels will have a larger area than a less disperse vowel space with less peripheral vowels.

An additional measure of dispersion was taken: a variety of repulsive force, defined in Equation 2. This definition differs from that of Liljencrants and Lindblom (1972) and Wright (2004) in that it is the raw sum of the distance measures, rather than the inverse sum.

$$E' = \sum_{i=1}^{N-1}\sum_{j=0}^{i-1}\sqrt{(F1_i - F1_j)^2 + (F2_i - F2_j)^2} \quad (2)$$

While both of these methods estimate vowel dispersion, they do so in different ways. The vowel area measure is based on geometric relations, treating the vowels as vertices of a polygon. As such it is a useful metric for the overall size of a vowel space, and is not sensitive to individual relationships between vowels in the vowel space. The repulsive force measure, on the other hand, does consider the distances between each pair of vowels, and provides an estimate of the extent to which vowels are crowded in the space. Thus, the area measure is a useful assessment of the overall size of the space, and the repulsion measure is a useful assessment of how 'crowded' the space is.

For each vowel, a measure of formant trajectory length (TL) was also taken to examine effects of dialect, style, and predictability on diphthongization (Fox and Jacewicz, 2009). Trajectory length is a measure of how much the formants of a vowel move during the vowel's production, and is calculated from formant frequencies measured at five time points within the vowel (20%, 35%, 50%, 65%, and 80%), as shown in Equation 3.

$$TL = \sum_{i=1}^{4}\sqrt{(F1_i - F1_{i+1})^2 + (F2_i - F2_{i+1})^2} \quad (3)$$

Longer TLs indicate more diphthongal vowel production, whereas shorter TLs indicate more monophthongal vowel production.

We expected clear speech and the LP condition to feature longer and more dispersed vowels than plain speech and the HP condition, respectively. Vowels in the HP condition were also expected to show more dialect-specific productions than vowels in the LP condition—thus, for Southern participants, more monophthongization of /aj/ (i.e. shorter TL), a more front /ɛ/, and a more back /ej/. If speaking style similarly affects dialect variation, we would also expect more dialect-specific productions in plain speech than clear speech. The productions from Midland speakers served as a control for comparison.

Data Summary

Experimenter error led to one sentence pair being presented incorrectly, so these tokens were subsequently excluded. All disfluent utterances or restarts were also excluded from analysis. A total of 78 tokens were excluded, leaving 1,970 usable tokens which were analyzed below.

RESULTS

Duration

Figure 1 shows boxplots of vowel duration, by predictability and style condition. As expected, clear speech features longer vowels than plain speech; however there is little difference between the HP and LP conditions.

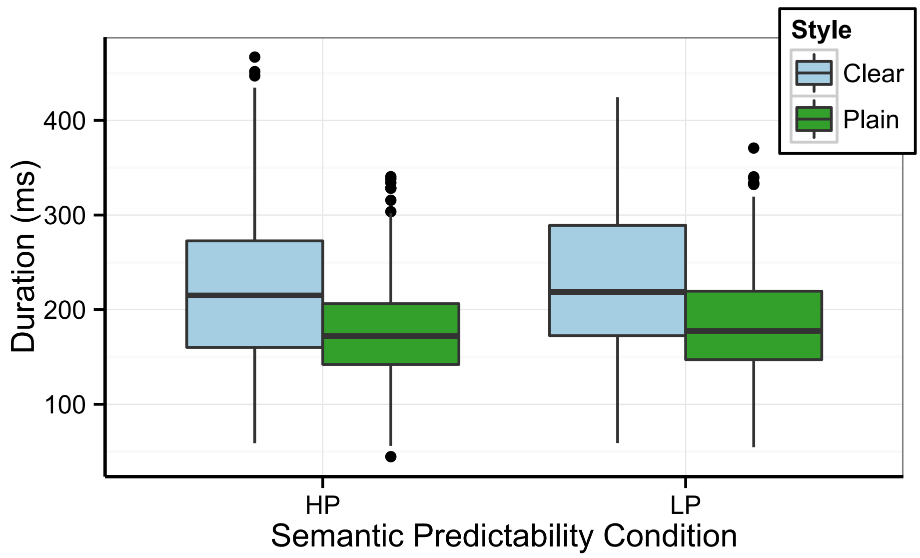

FIGURE 1: Boxplot of vowel duration by speech style by semantic predictability.

Vowel Space Area

The vowel space area was calculated for each participant in each of the four conditions. Figure 2 shows mean vowel spaces for plain and clear speech conditions, for all participants. The upper two panels show vowel spaces for Midland participants in the HP condition (left) and the LP condition (right); the lower two panels show vowel spaces for Southern participants in the HP (left) and LP (right) conditions.

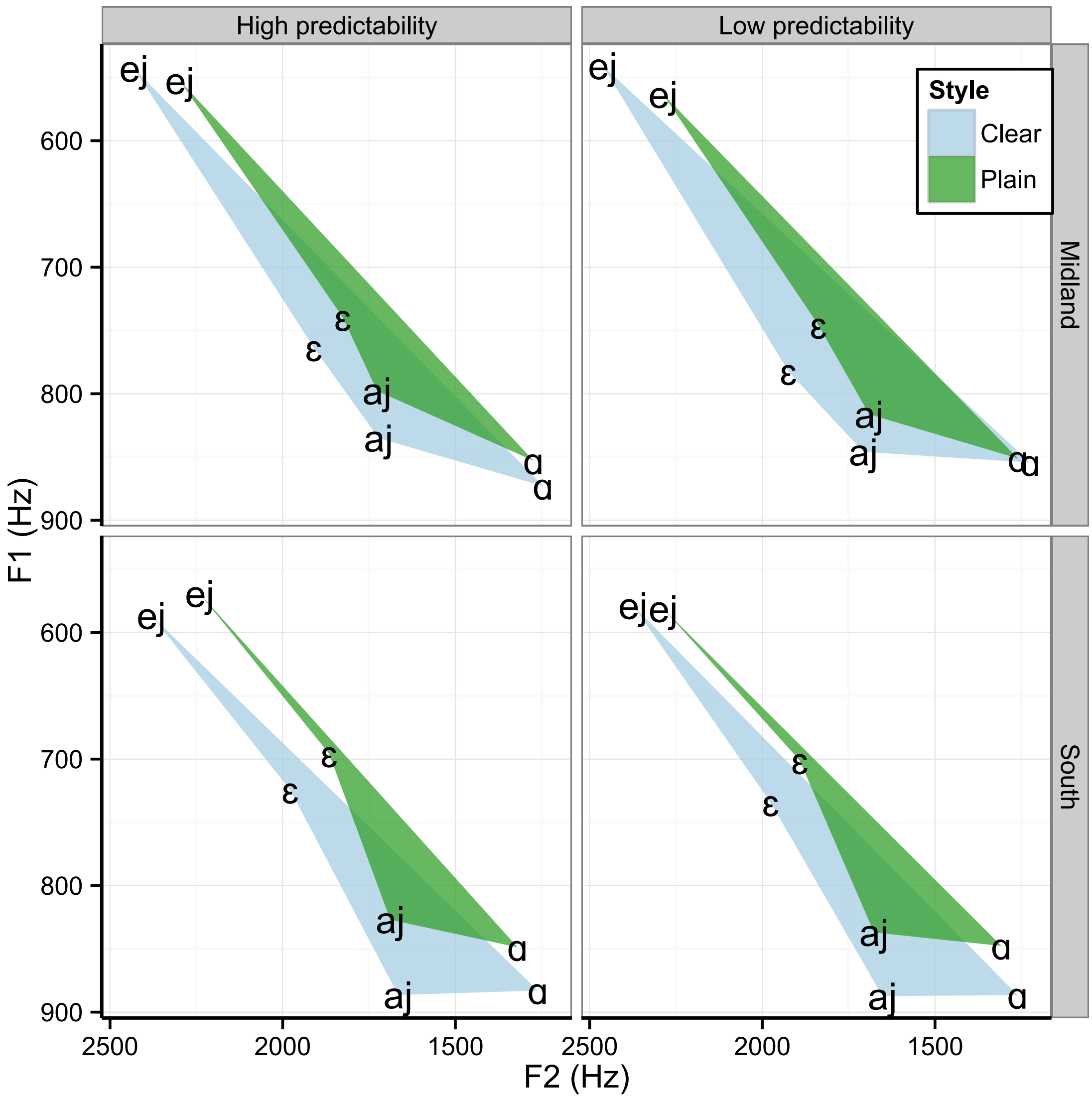

FIGURE 2: Mean vowel spaces for plain and clear speech conditions.

A linear mixed effects model was constructed to predict vowel space area with dialect, style, predictability, and all two- and three-way interactions as fixed effects and speaker identity as a random effect with random intercepts. A significant effect of style was observed ($\beta = -23658.3$, $t = -3.25$, $p_{MCMC} = 0.015$), such that the plain speech condition had an overall smaller vowel space area ($M = 37,251\text{Hz}^2$) compared to the clear speech condition ($M = 65,892\text{Hz}^2$). Additionally, a significant effect of predictability was observed ($\beta = 21899.1$, $t = 3.01$, $p_{MCMC} = 0.024$), such that the LP condition had an overall larger vowel space area ($M = 56,114\text{Hz}^2$) compared to the HP condition ($M = 47028\text{Hz}^2$). These results suggest vowel space reduction in plain speech relative to clear speech and for vowels in HP contexts relative to LP contexts, as expected. No other significant effects or interactions were observed.

Repulsive Force

A linear mixed effects model was constructed to predict repulsion, with style, predictability, dialect, and their 2- and 3-way interactions as fixed effects and participant as a random effect with random intercepts. A significant effect of style was observed ($\beta = -1283.21$, $t = -6.95$,

$p_{MCMC} < 0.001$), such that plain speech had a significantly lower repulsion (i.e. less dispersed vowels, $M = 6637$Hz) than clear speech ($M = 7923$Hz) as expected. No other significant effects or interactions were observed.

Trajectory Length

Figure 3 shows boxplots of formant TL for each vowel, split by speech style and dialect. A linear mixed-effects regression model was constructed, with TL as the dependent variable; vowel identity (treatment coded with monophthongal /ɑ/ as the reference vowel), speech style, predictability, dialect, and their 2- and 3-way interactions as fixed effects; and word and participant identity as random effects with random intercepts. The model revealed several significant effects, all relating to vowel identity and its interactions. These effects can be seen in Figure 3.

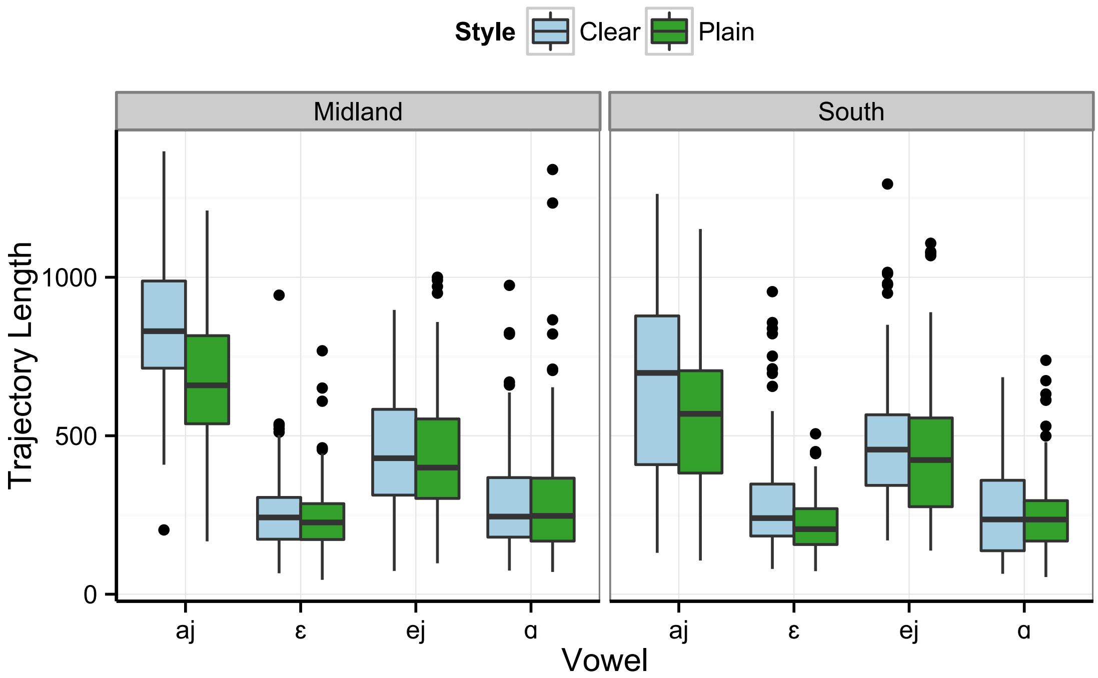

FIGURE 3: Boxplots of trajectory length per vowel, split by speech style and dialect.

There was a main effect of vowel, such that the TLs of /aj/ ($\beta = 542.12$, $t = 12.38$, $p_{MCMC} < 0.001$) and /ej/ ($\beta = 134.73$, $t = 3$, $p_{MCMC} = 0.002$) were significantly longer than that of /ɑ/. This result is expected, due to the fact that /ej/ and /aj/ are diphthongs and therefore are expected to have a longer TL. An interaction between vowel identity and speech style was observed, such that the TL of /aj/ was longer in clear speech than in plain speech ($\beta = -126.09$, $t = -3.7$, $p_{MCMC} < 0.001$). This interaction was not observed for any other vowels. Finally, an interaction between vowel identity and dialect was observed. The vowel /aj/ had a significantly shorter TL for Southern participants than for Midland participants ($\beta = -167.75$, $t = -4.64$, $p_{MCMC} < 0.001$). Additionally, the vowel /ej/ had a significantly longer TL for Southern participants than for Midland participants ($\beta = 90.865$, $t = 2.51$, $p_{MCMC} = 0.014$); the vowel /ɛ/ also had a longer TL for Southern participants, but this difference did not reach significance ($p_{MCMC} = 0.077$). No three-way interactions were observed, nor any main effects of dialect or predictability on TL.

Conclusion

As expected, Southern /aj/ had a shorter trajectory length than Midland /aj/, and in both dialects, /aj/ had a shorter trajectory length in clear speech than plain speech. Likewise, greater vowel duration, vowel space area, and vowel repulsion were observed in clear speech than in plain speech. A predictability effect was observed for vowel space area, but not for the other reduction measures. These results confirm previous findings about the effects of speaking style and predictability on vowel production. However, no interactions between dialect and predictability or speaking style were observed, suggesting that these effects are independent of the realization of these dialect variants.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation Grant BCS-1056409 to Ohio State University. We are grateful to Anna Crabb for assistance with data collection and coding.

References

Allbritten, R. (**2011**). "Sounding southern: Phonetic features and dialect perceptions", Ph.D. thesis, Georgetown University, Washington, DC.

Aylett, M. and Turk, A. E. (**2006**). "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei", J. Acoust. Soc. Am. **119**, 3048–3058.

Campbell-Kibler, K. (**2012**). "Contestation and enregisterment in Ohio's imagined dialects", J. Engl. Linguistics. **40**, 281–305.

Clopper, C. G. and Pierrehumbert, J. (**2008**). "Effects of semantic predictability and regional dialect on vowel space reduction", J. Acoust. Soc. Am. **124**, 1682–1688.

Fox, R. A. and Jacewicz, E. (**2009**). "Cross-dialectal variation in formant dynamics of American English vowels", J. Acoust. Soc. Am. **126**, 2603–2618.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability", J. Acoust. Soc. Am. **61**, 1337–1351.

Labov, W. (**2010**). *Principles of Linguistic Change*, volume III: Cognitive and Cultural Factors (Wiley-Blackwell, Oxford).

Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology and Sond Change* (Mouton de Gruyter, New York).

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

Liljencrants, J. and Lindblom, B. (**1972**). "Numerical simulation of vowel quality systems: The role of perceptual contrast", Lang. **48**, 839–862.

Mitsch, J. and Clopper, C. G. (**2012**). "Stylistic variation for perceived listener need: Social variation in clear and plain lab speech", Poster presented at NWAV41.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech", J. Speech Hear. Res. **28**, 93–103.

Plichta, B. and Preston, D. (**2005**). “The /ay/s have it: The perception of /ay/ as a north-south stereotype in US English”, Acta Linguistica Hafniensia. **37**, 107–130.

Smiljanić, R. and Bradlow, A. R. (**2009**). “Speaking and hearing clearly: Talker and listener factors in speaking style changes”, Lang. Linguistics Compass. **3**, 236–264.

Thomas, E. (**2003**). “Secrets revealed by southern vowel shifting”, Am. Speech. **78**, 150–170.

Wolfram, W. and Schilling-Estes, N. (**1998**). *American English: Dialects and Variation* (Blackwell, Malden, MA).

Wright, R. (**2004**). “Factors of lexical competition in vowel articulation”, in *Papers in Laboratory Phonology IV: Phonetic Interpretation*, edited by J. Local, R. Ogden, and R. Temple, 75–85 (Cambridge University Press, Cambridge).