THE EFFECT OF SEMANTIC PREDICTABILITY ON VOWEL PRODUCTION WITH PURE WORD DEAFNESS

Charles B. Chang & Simon Fischer-Baum

Boston University, USA; Rice University, USA
cbchang@post.harvard.edu; simon.j.fischer-baum@rice.edu

ABSTRACT

Vowels tend to be reduced in words that are semantically predictable from context, an effect amenable to talker- or listener-oriented accounts of speech production. This study explored the role of perception in these accounts by testing for effects of semantic predictability on vowel production in the face of impaired speech perception (but otherwise normal hearing)—namely, in a patient with *pure word deafness*. Analysis of the patient's English vowels in read speech showed no effect of semantic predictability on vowel duration, but the expected effect on vowel dispersion: vowels tended to be less dispersed in predictable than in unpredictable words. Overall, these findings contradict listener-oriented accounts of reduction relying on stored exemplars or online perceptual modeling, suggesting instead that reduction arises due to talker-centric factors related to activation of long-term, abstract representations.

Keywords: vowel space reduction, semantic predictability, pure word deafness, Houston English.

1. INTRODUCTION

Production of speech sounds is influenced by many variables, such as a word's neighborhood density, its frequency, and its semantic predictability from the preceding context. In general, reduction in temporal and/or spectral dimensions occurs when a word exists in a dense neighborhood (cf. [20]), is highly frequent, and is highly predictable from the context [6, 7, 16, 24]. These reduction patterns are of interest to linguistic theory because they have informed our understanding of speech production—in particular, of the roles played by the talker and the listener. Many production patterns are amenable to talker- or listener-oriented accounts, which are not mutually exclusive (cf. the "H&H Theory" of [17]).

The semantic predictability effect is no exception to the two-sidedness of phonetic reduction patterns. On the one hand, this effect may be a talker-centric phenomenon: high predictability facilitates lexical access, leading to faster articulatory planning and, in turn, faster articulation of predictable words. On the other hand, the effect may be due to the listener: unpredictable words, whose interpretation does not benefit from contextual clues, are more likely to be misperceived by the listener, which leads to blocking of reduction and/or hyperarticulation in these words.

Although high predictability is generally associated with increased reduction, some studies have shown that factors influencing reduction interact with specific indexical factors [18, 19, 24], which can result in the opposite pattern (cf. [26]). For instance, when a dialectal vowel shift has the effect of making vowels more peripheral (e.g., /æ/-fronting in the Northern Cities Vowel Shift), high predictability leads to more temporal reduction, but *less* spectral reduction—an effect attributable to greater processing ease allowing for greater production of indexical features [5].

Cross-dialect variation in the semantic predictability effect was attributed by [5] to talker-oriented factors, but can also be explained in terms of listener-oriented considerations. Dialect-specific (i.e., non-standard) features such as /æ/-fronting may lower intelligibility for listeners of different dialectal backgrounds, so they may be selectively suppressed in low-predictability contexts for the benefit of the listener. In this way, listener-oriented factors could also lead to dialect-specific realizations of the semantic predictability effect.

In the current study, we capitalize on the etiology of a selective deficit in speech perception—a rare disorder called clinically *pure word deafness* (PWD)—to further explore the talker- and listener-oriented accounts of reduction, taking the semantic predictability effect as our test case. PWD is a condition that, in its canonical form, impairs speech perception while leaving non-speech perception and language more generally intact. PWD thus provides a unique opportunity to examine the role of perception in reduction, especially in relation to a listener-oriented account.

The perceptual impairment in PWD holds various possible consequences for reduction. If reduction arises because of talker-centric factors that are not based on perception (e.g., speed of lexical access), then a patient with PWD should show these same patterns. That is to say, insofar as perception is not necessary for production, PWD should not interfere with normative phonetic reduction. On the other hand, failure of a PWD patient to show reduction where expected would suggest either that the talker-oriented factors underlying reduction are dependent on perception, or that these factors are not in fact talker-oriented, but rather listener-oriented.

Listener-oriented factors could lead to context effects on production in one of two ways. First, in an exemplar framework [9, 11], the talker may draw upon stored exemplars of the target word whose associated features match those of the given utterance conditions; thus, for low-predictability contexts, exemplars from low-predictability contexts (which are relatively unreduced) would be the ones activated, thereby leading to a lack of reduction in these contexts. Second, the talker may engage in online perceptual modeling of the listener, monitoring her outgoing speech along with the context to estimate the likelihood of veridical perception and modulating articulation in order to maximize this likelihood.

These two routes of listener-oriented production lead to the same prediction for long-term cases of PWD: no contextual reduction. In the first instance, since exemplars are posited to decay over time, prolonged decay without perceptual replenishment by new exemplars should eventually leave no reduced exemplars to draw upon in high-predictability contexts (whereas unreduced exemplars may survive for longer due to their greater perceptual saliency). In the second instance, perceptual impairment should prevent active speech monitoring.

To test this prediction, we examined the speech of a long-term PWD patient, "NL" (a subject code; we use the same code as in [25] to make it clear that we are examining the same individual), to see whether or not he would show evidence of preserved reduction. Given the profile of his PWD (see Section 2.1), NL is an ideal candidate to test whether the semantic predictability effect in speech production occurs in the absence of speech perception. For nearly a decade, NL has been unable to perceive spoken words, without a corresponding impairment in speech production.

2. METHODS

2.1. Subject

NL is a native English-speaking male, who was in his late 60s at the time of testing. NL suffered a cerebrovascular accident (stroke) in July of 2005 (8.5 years prior to the onset of testing). The stroke resulted in damage to the left temporal and parietal lobes of the brain, but no appreciable damage to the right hemisphere or to cross-hemispheric connectivity (for further details, see [25]).

Background assessment revealed severe difficulties in single word comprehension, including impaired performance on minimal pair discrimination (82% accurate, PALPA 2; [13]), poor performance on spoken word/picture matching [4], and the ability to repeat only 10 of 120 spoken words on the Philadelphia word repetition task [23]. In contrast, non-speech auditory perception, as indexed by the ability to recognize environmental sounds [4], was no different from that of age-matched controls.[1] Furthermore, NL showed no obvious impairment in spoken language production, at least in terms of accuracy, scoring within normal range on the Philadelphia Picture Naming Task [22]. NL was also unimpaired in tests of semantic knowledge and written language comprehension, scoring within normal range on both the picture and written versions of the Pyramid and Palm Trees test [10].

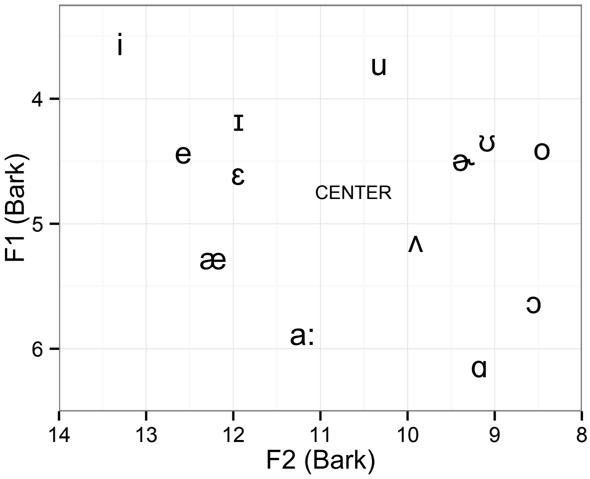
Raised and resident in the dialect region of Houston, Texas for most of his life, NL speaks a relatively conservative variety of Houston English, which manifests many (but not all) of the features described in the sociophonetic literature on Texas English and Southern English [8, 14, 15]. The characteristic Southern monophthongization of /ɑɪ/ is realized in an allophonic split: monophthongal [aː] before voiced consonants and word-finally (e.g., *nine*, *sky*, *smiled*), but diphthongal [aɪ] before voiceless consonants (e.g., *fight*, *Mike's*, *wipe*, *tighten*). On the other hand, the /ɑ/-/ɔ/ merger found throughout most of Texas [2] does not occur, as NL maintains the distinction between these vowels. This is shown in Fig. 1, which plots NL's (monophthongal) vowel space based on all the vowels carrying primary stress in content words in his read speech (see Section 2.2).

2.2. Materials

The materials for the reading corpus comprised 36 pairs of sentences created by [12]. The two sentences in each pair ended in the same critical word, but differed in terms of the predictability of that word given the preceding context: LOW (LP) vs. HIGH (HP). Sentence lengths in the LP condition ($M = 9.9$ words, $SD = 4.1$) and HP condition ($M = 9.8$ words, $SD = 4.3$) were not significantly different [$t(35) = 0.200, p = .843$]. An example sentence pair is given in (1) for the word *seeds*.

(1) a. You have considered the seeds. (LP)
b. Watermelons have lots of seeds. (HP)

Figure 1: NL's monophthongal vowel space (CENTER = mean F_1 and F_2 over vowels shown).



The 36 critical words were all monosyllabic and mostly consonant-final. Together they exemplified the vowels in 13 lexical sets of General American English [27]: /i ɪ e ɛ æ ɑ ɔ u ʌ ɪɹ oɹ ɑɪ aʊ/.

2.3. Procedure

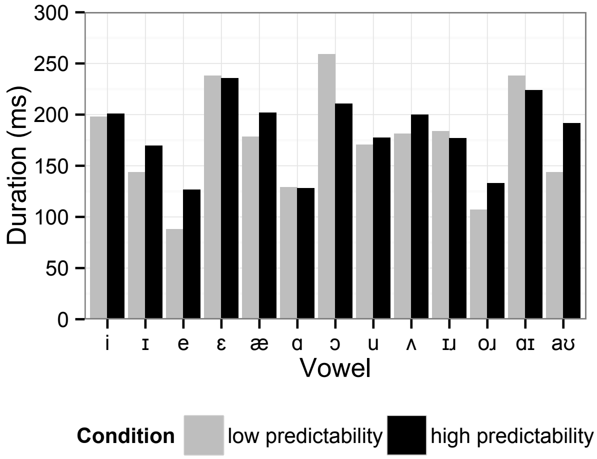
Testing was carried out in a soundproof room in two sessions approximately two months apart in 2013–14. The sentence materials were randomized and divided between the two sessions, such that only one sentence from a pair occurred in each session and each session consisted of an equal number of sentences from the LP and HP conditions. Sentences were presented one per page in size 24 Courier font. NL was given written instructions to read the sentences as quickly and accurately as possible, and his speech was recorded with a Samson C01U USB microphone directly into Praat [3] on a MacBook Pro.

2.4. Analysis

Recordings were annotated and analyzed acoustically in Praat. Vowel landmarks in each critical word were marked manually on a wide-band Fourier spectrogram (window length: 5 ms, dynamic range: 50 dB). Vowel onset was marked at the first time point where periodicity, high amplitude, and the first two formants (F_1 and F_2) were visible. Vowel offset was marked either at the last point with both periodicity and formants (obstruent- and vowel-final items) or at the drop in amplitude and onset of antiresonances for a final nasal or lateral.

Formants were measured over the middle 50 ms of each vowel demarcated in this way using linear predictive coding (LPC) analysis. Formant tracking was checked visually for every token, and for the few with tracking errors measurements were corrected by hand. On the basis of these measurements, a vowel dispersion measure was calculated for each vowel token as in [5], representing the Euclidean distance in F_1 x F_2 Bark space from that token to the center of the vowel space (see Fig. 1).

Figure 2: Mean duration, by vowel/condition.



3. RESULTS

3.1. Vowel duration

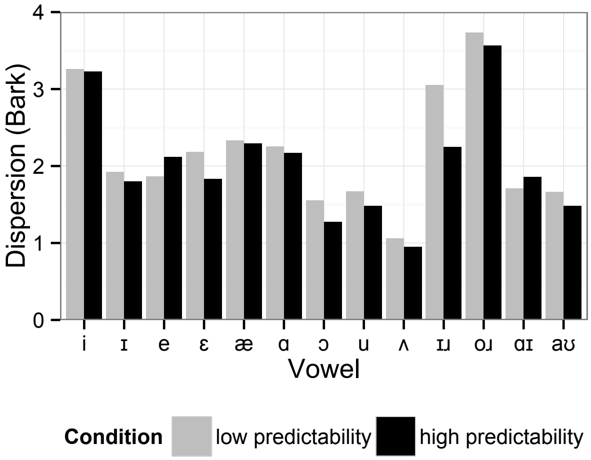
As shown in Fig. 2, semantic predictability did not have a consistent effect on vowel duration, although there was a tendency for vowels to be longer in the HP condition ($M_{HP} = 190$ ms; cf. $M_{LP} = 182$ ms). Compared to duration in the LP condition, duration in the HP condition was shorter for 3 vowels (/ɔ ɪɹ ɑɪ/), longer for 7 (/ɪ e æ oɹ u ʌ aʊ/), and virtually the same for 3 (/i ɛ ɑ/). A Wilcoxon signed-rank test showed that the difference between conditions was not significant [$W = 221, Z = -1.296, p = .199$].

3.2. Vowel dispersion

In contrast to the lack of temporal reduction, there was a strong trend toward spectral reduction in the expected direction: overall, vowels were less dispersed in the HP condition ($M_{HP} = 2.02$ Bark; cf. $M_{LP} = 2.10$ Bark). This was the case for 11 vowels (Fig. 3); the two exceptions were /e/ and /ɑɪ/, which showed the opposite effect. The reversal for /e/ may be attributed to a large duration difference that went in the unexpected direction (Fig. 2). However, this cannot account for the reversal for /ɑɪ/, since /ɑɪ/ showed the expected duration difference.

To better understand the patterning of /ɑɪ/, the

Figure 3: Mean dispersion, by vowel/condition.



data were further inspected by item. This analysis revealed that the overall reversal for /ɑɪ/ was due to three items with a final nasal (*line*, *nine*, *time*); the other two items (*fight*, *sky*) went in the expected direction. Since the former items are words in [aː], a front monophthong that is a salient marker of NL's dialect [1, 21], their failure to show spectral reduction may reflect an effect seen for /æ/ in [5]: enhancement of indexical features in contexts of high predictability. However, it is unclear why this enhancement is specific to words with a voiced coda and does not extend to the vowel-final word *sky*.[2]

Given the split in patterning of /ɑɪ/, the two conditions were compared twice (using the Wilcoxon signed-rank test): once including all items, and once including all items except those in [aː]. The difference between conditions did not reach significance in the first test [$W = 406, Z = 1.147, p = .129$], but did in the second [$W = 356, Z = 1.720, p = .044$; $\eta = .22$]. Together these results suggest that NL shows a semantic predictability effect on reduction, but in a dialect-specific manner for /ɑɪ/.

To facilitate comparison with previous findings, the mean decrease in dispersion from the LP to HP conditions was also calculated over the subset of vowels examined in [5]: /i æ ɑ ʌ/. For NL, the decrease in dispersion over these vowels was the same as for [5]'s Southern talkers (who, it should be noted, were female and mostly not from Texas): 0.07 Bark.

4. GENERAL DISCUSSION

To summarize, NL showed no evidence of temporal vowel reduction, but did show evidence of spectral vowel reduction in contexts of high semantic predictability. NL spectrally reduced to a similar degree as perceptually unimpaired talkers, and this spectral reduction occurred in spite of the lack of temporal reduction. In fact, 7 out of the 11 vowels that were spectrally reduced in the HP condition were actually longer in that condition, thus strengthening the finding of spectral reduction.

Although it should be borne in mind that this was a case study with relatively low power, the results are nevertheless challenging for listener-oriented accounts of phonetic reduction which understand reduction ultimately as a product of perception, since NL shows evidence of preserved reduction after years of living with PWD. Therefore, if reduction (or the lack thereof) is an audience-design phenomenon, it cannot be based on proximal speech perception (e.g., recently perceived exemplars of reduction or online perceptual modeling of the listener).

For NL, modeling of the listener could only be accomplished via linguistic representations that are both long-term and abstract. The observed production disparities across contexts could, for example, be derived by applying a rule of hyperarticulation to abstract representations in low-predictability environments. Such a rule, however, would need to be stochastic in nature in order to produce the appropriate variability across words and contexts in the size of the semantic predictability effect. In other words, if the relevant lexical representations are not themselves probabilistic (e.g., exemplar clouds), then the mechanisms that act on these representations to produce the semantic predictability effect must contain a probabilistic component.

On the other hand, a talker-oriented account of NL's preserved reduction is relatively straightforward. According to this account, production disparities arise across contexts because a biasing (high-predictability) context facilitates lexical access, speeding activation of the target word and, ultimately, its articulation. Crucially, such an effect could lead to the observed reduction patterns regardless of whether lexical representations are probabilistic or abstract. Moreover, under the assumption that articulatory plans associated with a word are independent of perception (at least once the word has been fully acquired), context effects on production are expected regardless of perceptual impairment.

In conclusion, the contributions of this study are twofold. First, it presents production data from a rare perceptual disorder that, while not incompatible with a listener-oriented account of reduction, more clearly support the talker-oriented account. Second, it provides evidence that lexical representations must include a long-term, abstract component that can serve speech production in perceptually impoverished circumstances.

5. REFERENCES

[1] Allbritten, R. 2011. *Sounding Southern: Phonetic Features and Dialect Perceptions*. PhD thesis, Georgetown University, Washington, DC.

[2] Bailey, G. 1991. Directions of change in Texas English. *Journal of American Culture* 14, 125–134.

[3] Boersma, P., Weenink, D. 2011. Praat: Doing phonetics by computer. Version 5.3.

[4] Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., Hodges, J. R. 2000. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* 38, 1207–1215.

[5] 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.

[6] Everett, C., Miller, Z., Nelson, K., Soare, V., Vinson, J. 2011. Reduction of Brazilian Portuguese vowels in semantically predictable contexts. In: Lee, W.-S., Zee, E., (eds), *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong, 651–654.

[7] 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.

[8] Gentry, E. 2006. Hovering between the South and the West: Houston's merged dialect. Paper presented at NWAV (New Ways of Analyzing Variation) 35, Columbus, OH.

[9] Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 195, 251–279.

[10] Howard, D., Patterson, K. 1992. *The Pyramids and Palm Trees Test*. Edmunds: Thames Valley Test Co.

[11] Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In: Johnson, K., Mullenix, J., (eds), *Talker Variability in Speech Processing*. San Diego: Academic Press, 145–165.

[12] Kalikow, D. N., Stevens, K. N., Elliott, L. L. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America* 61, 1337–1351.

[13] Kay, J., Lesser, R., Coltheart, M. 1992. *Psycholinguistic Assessments of Language Processing in Aphasia*. Hove: Lawrence Erlbaum.

[14] Koops, C. 2014. Iconization and the timing of Southern vowels: A case study of /æ/. *University of Pennsylvania Working Papers in Linguistics* 20, 81–90.

[15] Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton de Gruyter.

[16] Lieberman, P. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6, 172–187.

[17] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W. J., Marchal, A., (eds), *Speech Production and Speech Modeling*. Dordrecht: Kluwer, 403–439.

[18] Munson, B. 2007. Lexical access, lexical representation, and vowel production. In: Cole, J., Hualde, J. I., (eds), *Laboratory Phonology 9*. Berlin: Walter de Gruyter, 201–228.

[19] Munson, B. 2007. Lexical characteristics mediate the influence of sex and sex typicality on vowel-space size. In: Trouvain, J., Barry, W. J., (eds), *Proceedings of the 16th International Congress of Phonetic Sciences*. Dudweiler: Pirrot, 885–888.

[20] 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.

[21] Plichta, B., Preston, D. R. 2005. The /ay/s have it: The perception of /ay/ as a North-South stereotype in United States English. *Acta Linguistica Hafniensia* 37, 107–130.

[22] Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., Brecher, A. 1996. The Philadelphia naming test: Scoring and rationale. *Clinical Aphasiology* 24, 121–133.

[23] Saffran, E. M., Schwartz, M. F., Linebarger, M., Martin, N., Bochetto, P. 1988. *The Philadelphia Comprehension Battery*. Unpublished test battery.

[24] Scarborough, R. A. 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*. Berlin: Walter de Gruyter, 557–586.

[25] Slevc, L. R., Martin, R. C., Hamilton, A. C., Joanisse, M. F. 2011. Speech perception, rapid temporal processing, and the left hemisphere: A case study of unilateral pure word deafness. *Neuropsychologia* 49, 216–230.

[26] Turnbull, R., Clopper, C. G. 2013. Effects of semantic predictability and dialect variation on vowel production in clear and plain lab speech. *Proceedings of Meetings on Acoustics* 19, 060116.

[27] Wells, J. C. 1982. *Accents of English, Vol. 3: Beyond the British Isles*. Cambridge: Cambridge University Press.

---

[1] The main exception is with music: although NL performs at the same level as controls on consonance judgments, he has trouble discriminating melodic phrases. As [25] observe, however, "[w]hether this reflects problems with the processing of melody or harmonic structure, difficulty with musical memory, or simply is a function of [NL's] premorbid lack of musical interest and aptitude is, unfortunately, impossible to tell" (p. 226).

[2] One possible reason is a difference in the confusability of [aː] finally vs. non-finally. Compared to final position, in non-final position the monophthongal quality of [aː] may be more salient (and, thus, more indexical) because in that position [aː] is more confusable with nearby monophthongs (e.g., /ɑ æ/, which are either dispreferred or disallowed word-finally in American English).