ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 4pSCb: Production and Perception I: Beyond the Speech Segment (Poster Session)

4pSCb3. Effects of stress on intervocalic stop lenition in American English
Dominique Bouavichith and Lisa Davidson*

*Corresponding author's address: Linguistics, New York University, New York, NY 10003, lisa.davidson@nyu.edu

Descriptions of English and other languages have claimed that intervocalic stops are often lenited to fricatives or approximants in connected speech, but relatively few acoustic analyses of factors that affect lenition have been reported for American English (cf. Lavoie 2001, Warner and Tucker 2011). In this analysis, intervocalic voiced stops produced in bi- and trisyllabic words during story reading are examined (participants N=13). The first result shows that American English speakers never lenite to fricatives, but rather produce approximants whenever lenition occurs. Second, stress plays an essential role: 51% of stops are lenited when stress is on the first syllable (e.g. "yoga"), but only 7% of stops lenite when stress is on the second syllable (e.g. "lagoon"). Overall, approximant productions are significantly higher for /d/ and /g/ as compared to /b/. For both stress placements, stop productions are longer and lower in intensity than approximant productions. These acoustic findings are partially consistent with Kingston's (2007) claim that lenition occurs to minimize interruptions to the prosodic unit and indicate that the current constituent is ongoing, but speakers may also use full and reduced stop variants to enhance cues to lexical stress.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION: INTERVOCALIC STOP LENITION IN AMERICAN ENGLISH

Previous work on lenition, or reduction, of intervocalic stops in American English has shown that speakers frequently produce not canonical stops with a closure phase and a burst, but rather an approximant with formant structure (Crystal and House, 1988a, Crystal and House, 1988b, Lavoie, 2001, Warner and Tucker, 2011). One claim that has been made for why such ‘weakening’ processes might be occurring is that there is systematic pressure to reduce articulatory effort when possible (Kirchner, 1998/2001, Kirchner, 2004). However, Kingston (2007) argues that a drive to reduce articulatory effort cannot be an appropriate account of intervocalic lenition processes, and provides evidence that vowel openness, which should be a predictor of where lenition appears under this hypothesis, does not adequately account for where lenition occurs. Instead, Kingston proposes that consonant reduction is compelled by speakers’ desire to minimize the interruption of the speech stream to indicate that the current prosodic constituent is ongoing (see also Harris, 2003). He reports some results from two speakers of South American Spanish showing that the degree of stop reduction is sensitive to prosodic boundaries, such that complete stops are produced more often at higher prosodic boundaries (i.e. subordinate clauses) than when they are within a phrase (see also Lavoie, 2001).

For American English, a very thorough phonetic study on the reduction or weakening of intervocalic voiced and voiceless stops is Warner and Tucker (2011). Their work examines stops in both post-stress position (e.g. “lobby”) and between unstressed syllables (e.g. “halibut”), since their main interest was on flapping environments in American English. Warner and Tucker examined a number of phonetic variables, both continuous and categorical, and they also varied the speech style, looking at word lists, reading passages, and spontaneous speech. Warner and Tucker found that there was generally more reduction as the speech style became more casual: consonantal durations became significantly shorter, the difference in intensity between the vowel and the reduced consonant decreased, the proportion of consonants with a burst decreased, and the proportion with a cessation of F2 and F3 also decreased.

Although Warner and Tucker (2011) provide a comprehensive picture of the phonetic implementation of all intervocalic stops in flapping environments, they do not examine whether there is any weakening to fricatives or approximants in consonants that precede stressed syllables within a word (e.g. “about”). This environment is important because a potential conflict between phonetic cues may arise. On one hand, there may be a general drive within connected speech to reduce stops to higher-intensity approximants within a prosodic unit (Kingston, 2007), but on the other hand, maintaining an acoustic distinction between a more stop-like production preceding a stressed syllable and a reduced, approximant variant following a stressed syllable may help the listener in ascertaining the stress pattern of the word. An extra cue to stress assignment besides the cues on the vowels themselves may be useful for lexical access or segmentation (e.g., Mattys, 2000). Results from Lavoie (2001) indicate that stress plays an important role in determining stop or approximant variants, but since her stimuli were embedded in carrier phrases, it is unclear whether the pattern scales up to long passages of connected speech.

In this paper, we examine whether lenition of intervocalic stops is a global process, or whether it is constrained by pressure to maintain a distinction between consonant articulations in stressed versus unstressed syllables. While a general drive to lenite consonants to approximants would follow from Kingston’s argument that speakers reduce consonants to indicate an ongoing prosodic constituent, it also makes a prediction that overall rates of reduction to approximants should be similar if they are not sensitive to stress. In this study, we focus on voiced stops.

EXPERIMENT

Methods

Participants

Participants were 13 native speakers of American English, six females and seven males between the ages of 18 and 25. All participants are college students from the American Midwest who volunteered their time.

Stimuli and Procedure

The target consonants in this study were the voiced stops /b d g/. These consonants were presented within multiple stimulus words for a total of 87 stimuli. There were two different stress environments: stress on the vowel
following the target consonant (e.g. obese, fidelity, together) (N = 40) and stress on the vowel preceding the target consonant (e.g. gazebo, judo, yoga) (N = 47). The unstressed vowels could either be a full vowel (as in obese), or a schwa (as in aback). The target words, which were real words of English or proper names, were embedded in sentences within semantically coherent short stories. The target words were positioned phrase-medially to avoid any prosodic, intonational, or glottalizing effects associated with phrase edges. Each word had only one repetition in the reading passages.

Participants read five passages at a comfortable reading rate. If mistakes were made, subjects were asked to repeat the entire phrase containing the stimulus after the passage had been read. Participants were recorded in a quiet room using a Shure WH30XLR cardioid condenser head-mounted microphone and a TASCAM DR-40 digital recorder. Recordings were segmented and analyzed using Praat and R.

Data coding

The target words were coded both to examine categorical variables and to take continuous measurements. For the categorical variables, a token was coded as containing a stop if there was a cessation of F2 and F3 during the consonant, giving rise to a clear period of silence (with voicing) between the flanking vowels. A token was coded for a fricative if there was aperiodic noise in addition to voicing during the intervocalic period; in the whole corpus, there was only one fricative production. If formant structure was retained throughout the intervocalic period, the token was coded as being produced as an approximant. Tokens that are phonologically produced as flaps in English—i.e. /d/ in post-stress position—were similarly coded as to whether they were stops or approximants. If there was a clear period of silence, albeit very short, they were coded as stops. If there was a short period of formant structure that decreased in intensity between the flanking vowels, it was coded as an approximant.

In addition to the categorical codes, the consonants were segmented in Praat in order to carry out duration and intensity analyses. For the tokens coded as stops, the closure portion was selected for both duration and intensity measures. For tokens coded as approximants, we placed segment boundaries where there was an abrupt decrease in intensity corresponding to the consonant, which again abruptly increased at the start of the next vowel. The average RMS intensity within the interval that was segmented out for the duration measure was calculated automatically using a Praat script.

Results

A total of 1005 tokens were analyzed, out of a possible 1131 (87 stimulus targets x 13 speakers). The remaining tokens were discarded either because of mispronunciation, incorrect stress placement, or because of reading disfluencies affecting the target segment. The first analysis is an examination of the categorical results for stops versus approximants. This analysis is carried out using a mixed effects binomial logistic regression implemented in R (Bates and Maechler, 2012, R Development Core Team, 2011) with random slopes were included for subjects and items. The binomial regression model included stress (syllable before or after the target consonant) and target consonant (/b d g/) as predictors. Results of the model show that there is a significant effect of both target consonant and stress, as shown in Table 1 and Figure 1.

| TABLE 1. Results of mixed effects binomial regression comparing proportions of stops and approximants. The baseline values for this analysis are Target = /b/ and Stress = preceding syllable. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Estimate        | Std. Err.       | z-value         | Pr(|z|)          |
| (Intercept)     | 0.7084          | 0.3544          | 1.999           | 0.0456          |
| Target(d)       | -1.3188         | 0.6577          | -2.005          | 0.0450 *        |
| Target(g)       | -1.2875         | 0.5096          | -2.527          | 0.0115 *        |
| Stress(following syllable) | 2.4779 | 0.9682 | 2.559 | 0.0105 * |
| Target(d):Stress(2) | 3.2443 | 1.4995 | 2.164 | 0.0305 * |
| Target(g):Stress(2) | 0.156 | 1.1179 | 0.14 | 0.889 |
The main effect for stress shows that there are significantly more approximant productions (i.e. utterances containing formants) when stress is on the syllable preceding the consonant than when stress is on the following syllable. Results for the target consonants show that there are more stop productions for /b/ than for the other two consonants. When stress is produced on the following syllable (e.g. 'lagoon'), speakers produce more than 90% of the /b/ and /d/ utterances as stops, and 79% of the /g/ utterances as stops. An analysis of the subset of the data consisting of only the stimuli with stress on the preceding vowel shows that there are significantly fewer stop productions for /d~ɾ/ (β = -1.33, z = -2.09, p = 0.03) and /g/ (β = -1.31, z = -2.61, p < 0.01) as compared to /b/. However, there are no significant differences between /d~ɾ/ and /g/ (β = -0.02, z = 0.04, p = 0.99). For the subset of the data containing only items with stress on the following vowel, results show that there are significantly fewer stop productions for /g/ as compared to /d/ (β = -4.96, z = -3.15, p = 0.004). No other results were significant.

For the duration and intensity analyses, instead of including the stress position as a factor in the model with both the target consonant and the realization of the consonant (as a stop or an approximant), we instead carried out one analysis for words with stress preceding the consonant and another for following stress. This is because stress position and the realization as a stop or an approximant are highly confounded, and these predictors should not be included in the same analyses. The analyses for both duration and intensity included target consonant (/b d g/) and realization (stop, approximant), and random slopes were included for subjects and items.

Duration results for the stimuli with a preceding stressed syllable (shown in Figure 2a) indicate that /d/ is significantly shorter than both /b/ (β = -0.018, z = -7.301, p < 0.001) and /g/ (β = -0.015, z = -6.184, p < 0.001), but /g/ and /b/ are not significantly different from one another. There is no significant effect of realization. An interaction between /g/ and realization (β = -0.0062, z = -2.79, p < 0.05) indicates that when /g/ is produced as a stop, it is significantly shorter than when it is produced as an approximant. Although /d/ is in the same direction, it is not significant; in general, the duration of /d/ is very short whether it is produced as a stop or as an approximant, which is consistent with flapping in this position (De Jong, 1998, Fukaya and Byrd, 2005, Herd et al., 2010, Warner et al., 2009, Zue and Laferreire, 1979).

The results for the analysis of stimuli with stress on the syllable following the intervocalic consonant indicates that none of the main effects or interactions in this analysis are significant. Recall that only 5% of /b/ tokens and 6% of /d/ tokens are realized as approximants in this position (along with 21% of /g/ tokens), so the N is too small to reveal any potential significant effects.

For intensity (shown in Figure 2b), the results indicate that when the preceding syllable is stressed, /d~ɾ/ is significantly shorter than /b/ (β = -2.934, z = -3.07, p = 0.006) and /g/ (β = -3.17, z = -2.95, p < 0.01), but there is no difference between /b/ and /g/. There is also a significant effect of realization, as stop productions are lower in intensity than approximants (β = -5.124, z = -6.231, p < 0.001). No interactions are significant.

When the following syllable is stressed, realization was significant (β = -3.79, z = -2.73, p < 0.01), as stops had lower intensity than approximants. Post-hoc Tukey tests indicate that there is also a significant difference between /g/ and /d/, since /g/ is significantly lower in intensity than /d/ (β = -5.23, z = -2.75, p < 0.02). There were no significant interactions.
**DISCUSSION**

The study reported here investigates the effect of stress on lenition of intervocalic voiced stops in American English. The comparison between preceding and following stress is of interest because if there is a conflict between the influence of stress and connected speech reduction, it would appear in this environment. The results in Figure 1 show that consonants in pre-stress environments are realized as stops in the vast majority of cases. In contrast, in post-stress position, /b/ and /g/ are realized as stops 53% and 40% of the time, respectively, while /d/ is produced as a stop with closure only 25% of the time. Fewer stop realizations for /d/ in the post-stress environment is expected since this is where flapping occurs in American English. That even the ‘stop’ categorizations of the flap are consistent with the phonetic variability of flapping reported in the literature is confirmed by the duration findings for /d/, which are shorter (M = 22ms) and higher in intensity (58dB) than either /b/ (M = 43ms, 55dB) or /g/ (M = 34ms, 54dB) (cf. Zue and Laferriere 1979, De Jong 1998, Fukaya and Byrd 2005, Warner et al. 2009).

These results indicate that lenition to approximants is effectively blocked when the intervocalic consonant precedes a stressed syllable. Because this is a word-internal position, it is not necessarily the case that it should be considered a strong prosodic position, but the very high proportion of stops in this environment may serve as a (redundant) cue to the listener regarding the position of main stress in the word (see Mattys, 2000 for other cues listeners may use). If a listener perceives a stop, especially a relatively long one, it indicates that a stressed syllable is very likely to follow. As also shown in other studies, when stops are realized with a closure and burst, they are significantly longer and have lower intensity when stress is on the syllable following the intervocalic stop (Crystal and House, 1988a, Crystal and House, 1988b, Lavoie, 2001, Turk, 1992). This evidence for the stronger phonetic implementation in pre-stress environments is analogous to greater aspiration in pre-stressed position for voiceless stops in American English (Cooper, 1991, Kahn, 1980, Lisker and Abramson, 1967). From a functional perspective, both the longer and less intense stop closures preceding stressed syllables should help listeners ascertain the correct stress placement within a lexical item.

The increased proportion of stops in pre-stress word-internal position, concomitant with a longer duration and lower intensity, is not fully consistent with Kingston’s (2007) claim that speakers strive to minimally disrupt an ongoing prosodic unit. If that were the case, it might be expected that even in pre-stress positions, there would be a substantial proportion of approximant productions of the underlying intervocalic stop, or at the very least, that they should be as short as or contain the higher intensity of the post-stressed consonants that are realized with a stop closure. Instead, there seems to be a trade-off between implementing adequate phonetic cues for marking lexical stress and reducing the interruptions to the high intensity string as best as possible. One possibility is that a dispersion-type effect is being implemented by the speakers (Flemming, 2004); when it is necessary to mark stress, speakers produce stops, but otherwise they do attempt to reduce interruptions to the extent possible, though lenition to an approximant is far from categorical. Such a strategy would make the stress pattern of the utterance maximally salient to the listener. This account could be further explored with a perception study that varies vocalic stress cues with the correct or incorrect consonantal implementations; this is an area for future research.
ACKNOWLEDGMENTS

We would like to thank Timothy Woitas for his role in designing and carrying out the pilot study for this work.

REFERENCES

Bates, D.M., and Maechler, M. 2012. lme4: Linear mixed-effects models using S4 classes.