PATTERNS OF VOICING IN AMERICAN ENGLISH VOICED OBSTRUENTS IN CONNECTED SPEECH

Lisa Davidson

New York University
lisa.davidson@nyu.edu

ABSTRACT

Previous research shows that the implementation of obstruent voicing in American English is strongly influenced by adjacent sounds and prosodic factors, but studies mainly focus on words in isolation and carrier phrases. This study examines the voicing of stops and fricatives in phrase-medial position in the connected speech of 37 speakers. Results indicate that stops devoice most often word-initially or next to other obstruents (regardless of voicing). Fricatives devoice word-finally, but are less affected by the adjacent sound. An analysis of where voicing is realized in the constriction interval shows that bleed from a preceding sonorant is common, but voicing beginning partway through the constriction interval is almost non-existent. Aerodynamic and articulatory implications of the results are discussed.

Keywords: voicing, stops, fricatives, English

1. INTRODUCTION

Phonetic research on English voiced obstruents has demonstrated that the implementation of voicing is not a straightforward phonetic property of these consonants. Despite a designation of [+voice] in most phonological frameworks, the actual surface realization of voicing in English obstruents depends on factors such as word position, stress patterns, and adjacent phonemes.

For example, studies have reported that post-pausal voiced stops can sometimes be produced with prevoicing before the release burst, but more frequently contain no voicing at all and short-lag voice onset time (VOT) [4, 6, 7, 16]. In contrast, stops in intervocalic position show dramatically higher rates of voicing; for example, Westbury [16] found that 89% of intervocalic stops are fully voiced in American English, and Suomi [14] reported 93% for British English. Stops in word-final position tend to be partially devoiced [4], and are also affected by adjacent stops, which tend to cause greater full devoicing [16]. Fricatives are more likely to be voiced in word-initial position than stops are [4, 5], but are most often devoiced in word-final position [5, 12]. Adjacent obstruents lead to partial instead of full voicing [12].

While previous research has laid the groundwork for understanding the implementation of voicing in American English voiced obstruents, the current study aims to address several shortcomings. First, most previous studies used words in isolation or in short carrier phrases, which may not reflect how voicing is implemented in connected speech. Second, studies tend to focus on word-initial or final position (with special emphasis on pause-adjacent cases), and data on word-medial cases is limited.

Third, when partial voicing is discussed, the total proportion of voicing is reported, but not where in the obstruent constriction it is actually realized. For example, it could be present at the beginning of the constriction interval and die out before the end of the closure, it could appear partway through the constriction and continue past the end of the obstruent, or it could be present, then die out, and then pick up again. Reporting only on what percentage of the interval is voiced obscures a better understanding of the articulatory and aerodynamic influences on the implementation of partial voicing.

The goal of the current study is to use a relatively large corpus of connected (read) speech to address both the segmental and prosodic factors that condition voicing during the constriction of voiced obstruents, and what shape the typical partial voicing patterns take for stops and fricatives produced in connected speech. We focus on the subset of obstruents that are in phrase-medial position (i.e., always adjacent to another speech sound), though they can appear in initial, medial, or final position in the word.

2. METHOD

2.1. Participants

The data for this study comes from 37 speakers in two previously published studies [2, 3]. The 13 speakers in [2] were college students in the American Midwest, between 18-25 years old. The 24 speakers in [3] were all college students in New York City, ranging in age from 18-25.

2.2. Materials

The current corpus of voiced obstruents comes from the three short stories read by the participants in [3] and the five short stories in [2]. The recordings were made in quiet or soundproof rooms with solid state digital recorders and a high quality Shure cardioid condenser head-mounted microphone. The targets consisted of all instances of the stops /b d ɡ/, fricatives /v ð z ʒ/, and the affricate /dʒ/, except in cases where obstruents were excluded due to difficulty in segmentation: (1) /ð/ in function words (e.g. *there*, *that*) and /d/ and /v/ in *and* and *of*, respectively, (2) /d/ before /t, d, ð/, (3) /z, ʒ/ before a sibilant fricative, and (4) a voiced stop before another stop if the target lacked a release burst. For analysis, affricates were grouped with stops with the matching voicing specification.

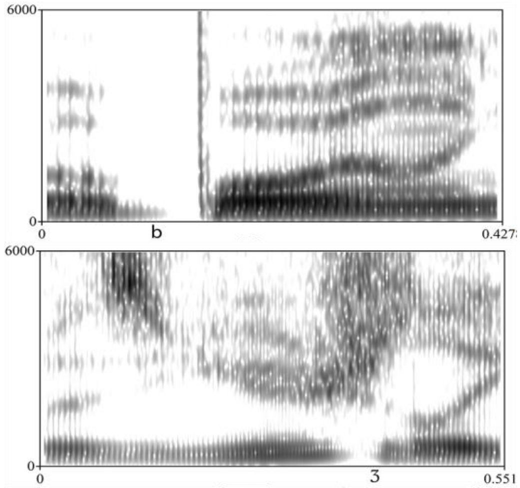
Praat textgrids were created for each story using the Penn Forced Aligner [19], which were used to segment the target obstruents, followed by manual adjustment. The obstruents were coded as to whether they were produced canonically (i.e. with a stop closure or period of frication), or whether they were reduced to an approximant or spirant, or glottalized. Only obstruents that were produced in their canonical realizations were included in this study.

Surrounding segmental contexts and word position was determined by converting the stories into the Carnegie Mellon (CMU) pronunciation dictionary transcription system. First, the sounds preceding and following the target obstruent were obtained, and classified as sonorants (vowels or approximants), nasals, voiced fricatives, voiceless fricatives, voiced stops, or voiceless stops. Second, phrase position was determined, using the presence of a comma or period in the transcript as a proxy for a pause delimiting utterance-initial and final positions. Third, the stress of the preceding and following vowels was determined from CMU stress conventions. Finally, target obstruents in phrase-medial position were coded for word position (initial, medial or final).

Once the canonically produced obstruents were identified, the *fraction of locally unvoiced frames* measure in Praat's Voice Report was used to obtain the proportion of voicing in the constriction. The amount of voicing during the closure is analysed in two ways: as a ternary classification of *voicing degree* and as a measure of *partial voicing shape*. For voicing degree, each obstruent was classified as to whether it was voiced (greater than 90% of the constriction was identified as voiced by the Voice Report), unvoiced (less than 10% of the constriction was identified as voiced) or partially voiced (between 10-90% of the constriction was voiced).

For the partial voicing shape measure, the period of obstruent constriction was divided into thirds and labelled for the following patterns, again using Praat's Voice Report measure: (1) *bleed*, where the proportion of voicing decreased from the 1st to the 3rd interval (usually being completely absent in the 3rd interval), (2) *trough*, where there was a greater proportion of voicing in the 1st and 3rd intervals than in the 2nd, (3) *negative VOT*, where the proportion of voicing increased from the 1st to the 3rd interval (typically being completely absent in the 1st interval) and (4) *hump*, where the proportion of voicing increased from the 1st to the 2nd interval and then decreased again. 72% of targets were produced with the bleed pattern and 24% with the trough pattern. Only 3% of tokens showed the hump pattern and 1% were produced with negative VOT. The common *bleed* and *trough* types are illustrated in Figure 1.

Figure 1: Partial voicing shapes *bleed* (top, in 'a **b**oiling') and *trough* (bottom, in 'was u**s**ually')



3. RESULTS

3.1. Categorical ternary voicing measure

The following analyses focus on phrase-medial obstruents realized with full voicing (N = 3145), full devoicing (N = 1760), or partial voicing (N = 4678). The first analysis examines the effects of word position, preceding segment, and obstruent duration on voicing degree. Following segment was not included in the model because it caused the model to not converge; the decision to retain the preceding segment as the more influential environment was made based on discussion in [4, 8].

A mixed effects multinomial logistic regression was carried out using *polytomous* in R [1]. The fixed effects included manner (stop, fricative), word position (initial, medial, final), preceding and following stress (stressed, unstressed), preceding segment (sonorants, nasals, voiced and voiceless fricatives, voiced and voiceless stops), and a numeric predictor of duration. Interaction terms for manner and preceding segment (except for voiceless fricatives, which did not occur before voiced fricatives), and random intercepts for words and speakers were also included.

Results for this analysis are in Table 1. Word position and manner can be summarized as having a similar overall effect: there is less full voicing and more complete devoicing for (1) initial and final word position as compared to medial word position, and (2) stops as compared to fricatives. Figure 2 shows that word-final fricatives and word-initial stops have the most devoicing. For stress, there is significantly more full voicing when stress precedes the obstruents, and less full voicing accompanied by more partial voicing and devoicing when stress follows the obstruents. The obstruent duration predictor indicates that longer obstruents are significantly less likely to be fully voiced and more likely to be partially voiced.

The preceding segment effects are shown in Figure 3. The only type of sound to significantly increase full voicing is a nasal, accompanied by a decrease in partial voicing. All of the remaining preceding sounds—voiced and voiceless stops and fricatives—lead to a significant decrease in fully voiced and partially voiced tokens. However, the interaction terms indicate that the preceding segment has substantially different effects on stops and fricatives. Compared to fricatives, nasals and preceding voiced stops predict significantly more fully voiced stops, and nasals predict significantly fewer partially voiced and unvoiced stops. Voiced fricatives and voiceless stops predict significantly less full voicing and significantly more unvoiced stops as compared to their effects on fricatives.

Figure 2: Proportions of unvoiced, partially voiced, and fully voiced obstruents by word position.

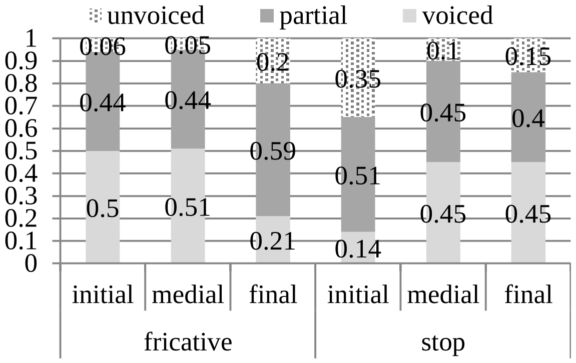
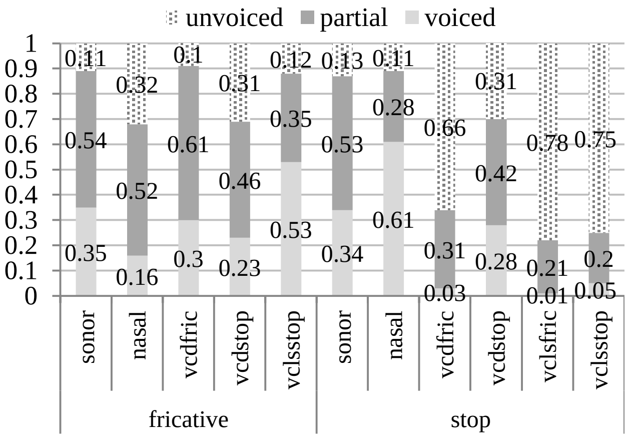


Figure 3: Proportions of unvoiced, partially voiced, and fully voiced obstruents by preceding segment.



3.2. Partial voicing shape

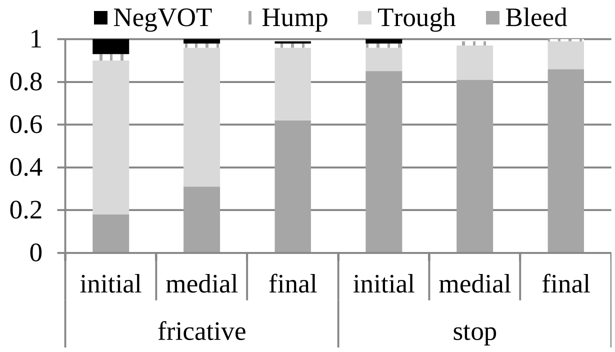
While Praat may report that an obstruent has a partially voiced constriction portion, this reveals nothing about where in the interval the voicing is present. Using the criteria for voicing shape from Sec. 2.1, a binomial logistic regression was carried out comparing the two main patterns of trough and bleed. This analysis had fixed effects of manner (stop, fricative) and word position (initial, medial, final). Words were included as random factors with random intercepts, and subjects with random slopes and intercepts. These results are shown in Figure 4.

There was a significant effect of manner, indicating that there were fewer troughs and more

Table 1: Multinomial mixed effects regression coefficients in log-odds for voicing categories. * indicates significance p< .05, ** indicates p< .001

	voiced	partial	unvoiced
(Intercept)	0.129	-1.655**	-2.746**
Manner (stop)	-0.171**	0.06	0.203*
WdPos (final)	-0.231**	0.057	0.607**
WdPos (initial)	-0.452**	0.066	0.473**
PrecStress (stress)	0.422**	-0.035	-0.03
FollStress (stress)	-0.212**	0.236**	0.414**
PrecSeg (nasal)	-0.602**	-0.208	0.655**
PrecSeg (vcd fric)	-0.017	-0.003	-0.202
PrecSeg (vcls fric)	-3.883**	-1.151**	1.445**
PrecSeg (vcd stop)	-0.347*	-0.278*	0.69**
PrecSeg (vcls stop)	0.371*	-0.379	-0.258
Obstruent Duration	-20.55**	14.174**	-1.777
Man (stop):PrecSeg (nas)	0.841**	-0.306*	-0.926**
Man (stop):PrecSeg (vcdfric)	-2.146**	-0.584*	1.685**
Man (stop):PrecSeg (vcdstop)	0.476*	-0.039	-0.161
Man (stop):PrecSeg (vclsstop)	-2.284**	-0.723	1.622*

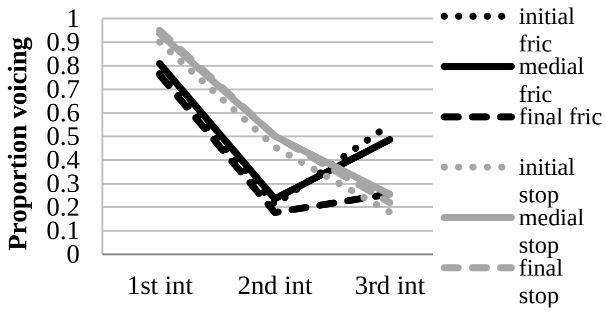
Figure 4: Proportion of voicing shapes for fricatives and stops by word position.



bleed patterns for stops than for fricatives (β = -1.49, $p < 0.001$). There was also a significant effect of position for both medial (β = 1.43, $p < 0.001$) and initial position (β = 2.44, $p < 0.001$), indicating that there were significantly more instances of bleed in final position, though this effect is mainly carried by the fricatives. A significant interaction between manner and word position corresponds to the smaller proportion of bleed in initial (β = -1.2, $p < 0.001$) and medial position (β = -2.64, $p < 0.001$) for fricatives than for stops.

The results for partial voicing shapes are further illuminated by Figure 5, showing the mean proportion of voicing for each of the three intervals. The numerical patterns for stops (in gray) reflect the bleed pattern. The U-shaped pattern for fricatives (in black), demonstrates why trough is the most common pattern for fricatives.

Figure 5: Proportion of voicing by interval (int). 'Initial, medial, final' refer to word position.



4. DISCUSSION

The results for the ternary classification as fully voiced, partially voiced, or unvoiced indicate that stops and fricatives are affected differently by word position. Whereas stops are more often devoiced word-initially, fricative devoicing occurs word-finally, despite both obstruents being flanked by another sound in this data (usually sonorants, which are the most common adjacent sound). These patterns are similar to those reported for utterance-initial and final positions for stops and fricatives, respectively [4, 5, 12, 18]. A possible interpretation of these results is that listeners may be extending the probabilistic cues to voicing that are phonetically natural at phrase edges, so that consistency between both phrase-edge and phrase-medial positions can assist with word segmentation.

Results for stress show that when the vowel preceding the obstruent is stressed, the target obstruent is almost twice as likely to be fully voiced compared to when the vowel is unstressed, and the opposite is true for the following vowel. This is compatible with work on consonant reduction in the American English flapping position, which shows that voiced consonants weaken (become voiced, or approximants) in post-stress environments [2, 15].

As for preceding context, one generalization is that voiced stops preceded by obstruents in general are realized with more complete devoicing than their fricative counterparts. This pattern is consistent with previous observations that as periods of obstruction lengthen, the closure in the second part of the sequence is less likely to remain voiced since the pressure requirements to retain phonation are difficult to sustain without active vocal tract expansion [10, 11, 18]. In contrast, the lack of this effect for fricatives may indicate that speakers are weakening the frication (but not eliminating it) to assist in the maintenance of voicing [5, 9, 13]. The strength of frication noise was not analysed here, but future study would help to clarify this point.

Finally, the majority patterns of bleed (for stops) and trough (for fricatives) both take advantage of the phonation from the preceding sonorants to ensure that voicing is present during the constriction, possibly by implementing manoeuvres to expand the supralaryngeal cavity to prolong voicing [11, 17] (as compared to voiceless stops, which have been shown to have shorter durations of carryover voicing [4, 14]). The greater frequency of the trough pattern for fricatives may again be attributable to a gradual weakening of frication noise during the transition to the following sound, which would allow for voicing to increase toward the end of the fricative. Notably, the 'negative VOT' pattern does not occur for stops, which is not surprising as it is least likely that speakers will initiate phonation precisely when oral pressure is highest.

Voiced obstruents in American English are an interesting test case for examining articulatory and aerodynamic effects on the implementation of voicing, since the constrictions in English obstruents are not obligatorily (fully) voiced. As a consequence, the environments that naturally promote the prolongation of phonation and those that curtail it are made evident.

5. ACKNOWLEDGMENTS

Thanks to James Whang, Steven Foley, and Benjamin Gittelson for their help in data analysis. I also thank Gillian Gallagher, Suzy Ahn and Sang-Im Lee-Kim for their helpful comments. This research was supported by funds from New York University.

6. REFERENCES

[1] Arppe, A., 2013. polytomous: Polytomous logistic regression for fixed and mixed effects, 0.1.6, R package.

[2] Bouavichith, D., Davidson, L., 2013. Segmental and prosodic effects on intervocalic voiced stop reduction in connected speech, *Phonetica* 70, 182-206.

[3] Davidson, L., Erker, D., 2014. Hiatus resolution in American English: the case against glide insertion, *Lg* 90, 482-514.

[4] Docherty, G., 1992 *The Timing of Voicing in British English Obstruents*. Berlin: Foris Publications.

[5] Haggard, M., 1978. The devoicing of voiced fricatives, *J. Phon.* 6, 95-102.

[6] Keating, P., 1984. Phonetic and phonological representation of stop consonant voicing, *Lg* 60, 286-319.

[7] Lisker, L., Abramson, A., 1964. A cross-language study of voicing in initial stops: acoustical measurements, *Word* 20, 384-422.

[8] Möbius, B., 2004. Corpus-based investigations on the phonetics of consonant voicing, *Folia Linguistica* 38, 5-26.

[9] Ohala, J., 1983. The origin of sound patterns in vocal tract constraints. In MacNeilage, P., (ed.), *The Production of Speech*. New York: Springer-Verlag, 189-216.

[10] Ohala, J., Riordan, C. J., 1979. Passive vocal tract enlargement during voiced stops. In Wolf, J. J. and Klatt, D., (eds), *Speech Communication Papers*. New York: Acoustical Society of America, 89-92.

[11] Proctor, M., Shadle, C., Iskarous, K., 2010. Pharyngeal articulation in the production of voiced and voiceless fricatives, *J. Acoust. Soc. Am.* 127, 1507-1518.

[12] Smith, C., 1997. The devoicing of /z/ in American English: effects of local and prosodic context, *J. Phon.* 25, 471-500.

[13] Stevens, K., Blumstein, S., Glicksman, L., Burton, M., Kurowski, K., 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters, *J. Acoust. Soc. Am.* 91, 2979-3000.

[14] Suomi, K., 1980 *Voicing in English and Finnish stops*, 10. Turku, Finland: University of Turku.

[15] Warner, N., Tucker, B., 2011. Phonetic variability of stops and flaps in spontaneous and careful speech, *J. Acoust. Soc. Am.* 130, 1606-1617.

[16] Westbury, J., 1979 *Aspects of the temporal control of voicing in consonant clusters in English*. Austin: University of Texas, Austin.

[17] Westbury, J., 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing, *J. Acoust. Soc. Am.* 91, 2903-2910.

[18] Westbury, J., Keating, P., 1986. On the naturalness of stop consonant voicing, *J. Ling.* 22, 145-166.

[19] Yuan, J., Liberman, M., 2008. Speaker identification on the SCOTUS corpus, *Proceedings of Acoustics '08*, 5687-5690.