Variability in the implementation of voicing in American English obstruents

Lisa Davidson
New York University

Running head: Voicing in American English obstruents

Address for correspondence:
Department of Linguistics
10 Washington Place
New York, NY 10003
212-992-8761
lisa.davidson@nyu.edu
Abstract

Previous research has shown that in languages like English, the implementation of voicing in voiced obstruents is affected by linguistic factors such as utterance position, stress, and the adjacent sound. The goal of the current study is to extend previous findings in two ways: (1) investigate the production of voicing in connected read speech instead of in isolation/carrier sentences, and (2) understand the implementation of partial voicing by examining where in the constriction voicing appears or dies out. The current study examines the voicing of stops and fricatives in the connected read speech of 37 speakers. Results confirm that phrase position, word position, lexical stress, and the manner and voicing of the adjacent sound condition the prevalence of voicing, but they have different effects on stops and fricatives. The 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 (i.e., negative voice onset time) is much rarer. The acoustic, articulatory, and aerodynamic sources of the patterns of phonation found in connected speech are discussed.
1. Introduction

Voicing in English obstruents is a much discussed topic in the phonetic literature. However, while many studies focus on voice onset timing (VOT) as a cue to voicing in stops (e.g., Lisker & Abramson, 1967; Zlatin, 1979), and others on cues such as the length of the preceding vowel (e.g., Raphael, 1972) or F0 at the obstruent offset (e.g., Hanson, 2009; Ohde, 1984), fewer studies report details about the realization of phonation during the constriction period of voiced obstruents. Yet, as discussed in Gonet (2010), popular textbooks of English phonetics nevertheless promulgate incomplete and anecdotal descriptions of how phonation during stop closures and frication periods is assumed to be implemented. According to these textbook authors, obstruents are only fully voiced in intervocalic position (and presumably word-medially, given examples such as ‘lagging’ and ‘saving’) (Cruttenden, 2008; Davenport & Hannahs, 2010) and are likely to have little phonation at all in initial or final position (e.g., ‘bill’, ‘done’, ‘cub’, ‘lid’) (e.g., Ball & Rahilly, 1999; Catford, 2001; Cruttenden, 2008; Davenport & Hannahs, 2010). For most of these descriptions, the authors do not distinguish word-initial and final from phrase-initial and final positions, though the examples suggest that these claims about voicing are pertinent primarily to words produced in isolation.

Taken together, these descriptions raise more questions than they answer, since there are many conditioning factors such as word position, phrase position, adjacent consonants, and lexical stress that could affect the proportion of voicing present during obstruent constrictions. Thus, one goal of this paper is to examine the effect of these variables on where and how much voicing is produced during obstruent constrictions (stop closures and the frication portion of fricatives) in American English connected speech. (Note that throughout this paper, ‘voicing’ will be used synonymously with ‘phonation’, to indicate the presence of vocal fold vibration
during obstruent constrictions.) Another goal is to gain a clearer understanding of how ‘partial voicing’ is implemented. As will be reviewed in more detail in the following section, a number of studies have observed that English voiced obstruents may be partially voiced in some environments. Most accounts of partial voicing seem to treat it as phonation lasting less than 100% of the closure duration, but a major drawback of this usage is that it does not characterize where voicing occurs during the closure. In this study a large number of obstruents are examined to address not only which segmental and prosodic factors condition voicing during the constriction of voiced obstruents, but also what shape the typical partial voicing patterns take for stops and fricatives produced in connected (read) speech.

1.1. The realization of voicing in English obstruents: findings and limitations

In spite of the relative vagueness in textbooks, a number of studies provide information aimed at quantifying the proportion of voicing in English obstruents. A particularly extensive study is Docherty (1992), who examined the production of phonation in obstruents in Southern British English. The target stimuli were produced in three environments—in isolation, in a “voiced carrier” phrase that provided a vowel-adjacent context on the left for initial obstruents and on the right for final obstruents, and in a “voiceless carrier” phrase that likewise provided a voiceless obstruent context. Results for voiced stops in word-initial position showed that they were completely devoiced in isolation and in the voiceless carrier phrase, and that the vast majority of tokens in the voiced carrier phrase were partially voiced (with some tokens being fully voiced). Word-initial stops were often voiced for more than half of the closure (but the location of this voicing within the constriction was not investigated). Final stops showed a
somewhat different pattern; even in isolation and the voiceless carrier phrase, the large majority of tokens showed at least partial and occasionally full voicing.

Fricatives in word-initial position (excluding /ð/, which patterned almost completely like a voiced stop) were more evenly split among completely voiced, completely voiceless, and partially voiced, except that the voiced carrier phrase tended to condition more completely voiced tokens. Voiced fricatives in word-final position were most often produced with partial voicing, though the overall amount of voicing was lower than for initial fricatives, with an increase in full voicing in the voiced carrier phrase. From this data, Docherty concludes that the definition of the phonological category ‘VOICED’ is more complex than had been thought, and that “many aspects of the realization of voicing timing are not predictable from the nature of the voicing category (129).”

While the results in Docherty (1992) are extremely informative, there are still some limitations to this study. First, carrier phrases were used, which may not reflect how voicing is implemented in more connected speech (i.e. reading of semantically coherent text, or spontaneous speech). Second, when Docherty calculated the proportion of voicing, he collapsed fully voiced and partially voiced tokens, which may obscure how partial voicing is actually realized. Third, as explained previously, Docherty only considered the total proportion of voicing, but not where in the obstruent constriction it is actually realized (though an impressionistic account of this is discussed in his Chapter 5). Fourth, these data focus on obstruents in word-initial and final position, but do not include results for word-medial obstruents. In fact, these limitations generally hold true for all of the following studies that will be reviewed, except where noted. Finally, Docherty’s data are for Southern British English.
Beyond Docherty, other studies provide some conflicting information about the realization of voicing in voiced stops or fricatives. Starting with voicing in post-pausal, word-initial position (often referred to as ‘prevoicing’), Lisker and Abramson (1964; 1967) observed stops with both prevoicing and short-lag voice onset time in American English, but the prevoiced tokens were effectively found only for one of their four speakers. Similarly, Keating (1984) found that words with post-pausal initial /b/ read by two speakers were occasionally produced with prevoicing, but by and large they were implemented with short lag VOT and no phonation during the closure. On the other hand, other authors have found that half or more of post-pausal voiced stops are produced with some prevoicing (Flege, 1982; Smith, 1978; Westbury, 1979).

As for intervocalic position, most studies have shown dramatically higher rates of full voicing for American English, often between 80-90% (Flege & Brown, 1982; Keating, 1984; Westbury, 1979). The patterns for British English are similar (Suomi, 1980), but Suomi also found that even when a stop is between two vowels, word position matters; when words were produced in a carrier phrase, intervocalic stops at the end of a word were partially voiced substantially less often than intervocalic word-initial stops. Similarly, Jacewicz et al. (2009) found that the degree of emphasis of a word in a phrase affected the proportion of voicing produced for /b/ in inter-sonorant position; the proportion of voicing was greater for words with less emphasis than words with more emphasis. Interestingly, this result was true for the speakers from Wisconsin in this study, but not the speakers from North Carolina, who had very high rates of voicing in this environment regardless of target word emphasis. This is a clue that prosodic factors are one consideration that can have a potential effect on how much voicing is realized in English obstruents.
Like the studies of stops, research dealing specifically with fricatives in English confirms that the implementation of voicing depends on whether a fricative is adjacent to a vowel or sonorant, or to an obstruent or a pause, regardless of its voicing specification (Gonet, 2010; Haggard, 1978; Smith, 1997; Stevens, Blumstein, Glicksman, Burton, & Kurowski, 1992). For example, Smith (1997) found that word-final American English /z/ was much more likely to be fully voiced if it was between two sonorants (including vowels), as compared to preceding a voiced stop. Regarding adjacent stops, there was still more partial voicing for adjacent voiced stops than voiceless ones. As for word position, sentence final /z/ was always devoiced, whereas phrase-medial /z/ (regardless of word position) was dependent on the surrounding context, as already described. Very similar patterns were also reported in Haggard (1978), who examined all of the British English fricatives. Haggard includes the additional information that the fricative /θ/ is generally more voiced in all environments than all other fricatives, and /z/ is mostly devoiced when between stressed vowels. Taken together, these studies demonstrate that fricatives lose some degree of voicing next to pauses or obstruents, and that lexical stress likely also plays a role in how voicing is realized.

In general, the number of studies that provide some information about the realization of voiced obstruents in connected speech contexts is small. For English, Gonet and Święciński (2012) examine the voiced obstruents of four speakers of British English taken from interviews in word-initial, word-medial, and word-final contexts. They measure the total percentage of completely devoiced tokens of fricative and obstruent constrictions. Gonet and Święciński find a number of variables that condition devoicing: word-initial position, a preceding or following obstruent, being in a stressed syllable, and coda position.
In addition, although these studies do not include any data on English, work by Shih and Möbius and colleagues is worth mentioning since these studies attempt to quantify how much voicing occurs at 10-11 points during the normalized duration of consonant constrictions (Möbius, 2004; Shih & Möbius, 1998; Shih, Möbius, & Narasimhan, 1999). The languages under investigation in the studies include one speaker of German, Mandarin, Hindi, Spanish and Italian, who is producing read speech. As Möbius (2004) focuses only on German, which is more similar to English with respect to phonological obstruent voicing categories than any of the other languages (despite some previously reported differences in phonetic implementation, Keating, 1984), the results from this study bear some relevance to the current investigation. The data from the 10 time points indicates that when this speaker’s voiced stops are preceded by a sonorant, their probability of being voiced at each time point is much higher than when preceded by a voiceless obstruent. Although the right-side context was not specifically examined, Möbius notes that there may be some occasional anticipatory prevoicing before sonorants. The report for voicing for fricatives demonstrates that when preceded by a sonorant, the voicing probability for the strident fricatives starts high and decreases, while the probability of voicing for /v/ stays consistently high, at nearly 100%. When preceded by a voiceless obstruent, however, the stridents are nearly voiceless throughout, whereas /v/ generally starts out devoiced and rise to a 90% probability of voicing by the last time point. These results show that not only does the surrounding context affect the proportion of voicing found in the constriction of a voiced obstruent, but that the understanding of ‘partial voicing’ goes beyond knowing simply what proportion of the whole obstruent constriction is produced with voicing.

In the present study, a detailed analysis of the realization of voicing in American English phonologically voiced obstruents in connected (read) speech is undertaken. The goal of this
paper is to assess the influence that several linguistic factors have on the implementation of phonation using a relatively large set of data, with a specific focus on phrase position, word position, stress, and preceding context. Since voiced obstruents are not obligatorily phonated in American English, it is an interesting language for investigating the articulatory, aerodynamic, and possible perceptual forces that influence how phonation is implemented in obstruents, which are typically considered inhospitable for voicing (e.g., Ohala, 1983). It is hypothesized that the results will generally conform to the previous reports of words read in isolation or carrier sentences; that is, environments such as preceding and following pauses or voiceless sounds should lead to substantially less phonation in the target obstruent, whereas surrounding sonorants should lead to higher rates of partial and full voicing. However, it is expected that there will also be interactions between surrounding context and phrase position, as stronger prosodic positions and word boundaries may also condition greater amounts of partial or total devoicing. In addition to quantifying the proportion of voiced, devoiced, and partially voiced obstruents that are produced, we also investigate the shape of the voicing—that is, where in the obstruent constriction the voicing is produced. There are a number of options for how partial voicing could be 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. These patterns too should be affected by adjacent segmental and prosodic contexts.

It should be noted that this research will not examine to any great extent differences in duration or overall proportion of voicing as a function of place of articulation, as these issues have been discussed in considerable detail in previous literature (Docherty, 1992; Haggard, 1978; Ohala & Riordan, 1979; Pape, Mooshammer, Hoole, & Fuchs, 2006; Westbury, 1979,
1983; Westbury & Keating, 1986). Likewise, though Jacewicz et al. (2009) found an effect of sentential prominence on the implementation of phonation for at least some of their speakers, a systematic investigation of the interaction between intonation and voicing cannot be carried out at this time since the corpus currently does not contain a ToBI prosodic transcription (Beckman & Ayers Elam, 1997). This is an area for future development.

2. Methods

2.1. Participants

The data for this study comes from two previously published studies: Bouavichith and Davidson (2013), which focused on intervocalic stop reduction, and Davidson and Erker (2014), which examined hiatus resolution. The 13 speakers in Bouavichith and Davidson (six female and seven male) were all between the ages of 18 and 25. The speakers were college students in the Midwest. The 24 speakers (17 female and 7 male) collected for Davidson and Erker were all college students in New York City, ranging in age from 18-25. Most speakers were originally from the Northeast (mid-Atlantic and New England), but there were also speakers originally from the Midwest (Chicago, Michigan, Minnesota), and one each from Georgia, Texas, and New Mexico. No participant reported a history of speech or hearing disorders, except one who reported speech therapy as a small child for the misarticulation of /s/. Since there was no evidence of a remaining issue and because childhood difficulty with place of articulation was unrelated to the question being studied, this participant’s data was retained. All speakers were compensated for their participation.

2.2. Materials
The corpus of voiced obstruents collected for the study comes from the three short stories read by the participants in Davidson and Erker (2014) and the five short stories in Bouavichith and Davidson (2013). The recordings were made in a sound proof room with a Marantz PMD-660 digital solid state recorder (Davidson and Erker) or a quiet room with no background noise with a TASCAM DR-40 digital solid state recorder (Bouavichith and Davidson), both using a Shure WH30XLR cardioid condenser head-mounted microphone. The targets consisted of all instances of the stops /b d ɡ/, fricatives /v ʒ z/, and the affricate /dʒ/. Because the stories were designed for different experiments, the number of different obstruents and their positions in the phrases and words were not controlled. There were a number of exceptions that led to the exclusion of certain voiced obstruents. First, /ð/ in function words (e.g., there, that) and /d/ and /v/ in and and of, respectively, were not included, since they are often so reduced as to be absent on the spectrogram. Second, /d/ was excluded when it was before /t, d, ŋ/, since /d/ was never released in this environment and therefore could not be distinguished from the following sound. Similarly, /z, ʒ/ before a sibilant fricative was also excluded for the same reason. Third, a voiced stop before another stop was only included if it was released before the closure of the following stop began; otherwise, it would not be possible to distinguish between the end of the first stop and the beginning of the second stop.

Praat textgrids were created for each story using the Penn Forced Aligner (Yuan & Liberman, 2008), which were used to segment the stops and fricatives identified as target sounds. Research assistants then adjusted each boundary by hand to segment the frication period of the fricatives and the closure period of the stops. To simplify the analysis of the affricates, which would have been too small a category on their own, only the closure portion was segmented and affricates were subsequently grouped with the stop. For both types of targets, when adjacent
segments were vowels or sonorant consonants, the offset of the second formant was used to
determine the left edge of the frication or closure boundaries. The right edge of the fricatives was
also taken to be the onset of F2. While this segmenting convention could potentially increase the
proportion of tokens that are coded as having some voicing during the closure, it was adopted
because it is often difficult to determine a distinction between F1 and F0 on a waveform and
spectrogram and we did not want to erroneously exclude tokens that do contain voicing. For
stops, the onset of the burst was used as the right edge of the closure, except where there was no
visible burst, in which case the onset of F2 was taken as the edge of the sound. When a stop was
adjacent to a stop or a fricative, a boundary was placed at the offset of the stop burst or frication
and the onset of the closure. In the case of adjacent fricatives (almost always a sibilant-non-
sibilant boundary, as sibilant-sibilant sequences were excluded), a boundary was placed where
there was an abrupt change between the high intensity of the sibilant and the low intensity of the
non-sibilant. A marked change in the concentration of the frication noise was also usually
evident at this juncture.

In addition to the segmentation of the frication or closure period, each target sound was
also categorically coded as to whether it was produced in canonical fashion (i.e. stops and
affricates with a period of closure and fricatives with aperiodic noise), or whether it was reduced
to an approximant, or to a fricative (in the case of the stops). Sounds were also coded as to
whether they were glottalized, which was operationalized by being surrounded on both sides by
the irregular glottal striations that are characteristic of glottalization (e.g., Dilley, Shattuck-
Hufnagel, & Ostendorf, 1996; Garellek, 2012). Only obstruents that were produced in their
canonical realizations were included in this study, and those that were weakened or glottalized
were not analyzed since the purpose of the study is to examine the distribution of voicing in
obstruents. Each file was double checked by the author to ensure that the segmentation criteria had been met, and that reduced and glottalized implementations were properly coded.

In order to investigate the effect of the surrounding context, word and phrase position, and lexical stress on voicing in the target obstruents, environmental information was also collected. A Python script was written to convert each of the stories into the Carnegie Mellon University (CMU) pronouncing dictionary transcription system and then to search these transcripts to keep track of a number of variables. First, the sounds preceding and following the target obstruent were obtained, and classified as to whether they were sonorants (vowels or approximants), nasals, voiced fricatives, voiceless fricatives, voiced stops, or voiceless stops. Affricates were grouped with stops with the matching voicing specification to remain consistent with the grouping of the target affricates with the stops. Second, using the CMU notations that indicate stress (a ‘1’ following the vowel when it is stressed, and ‘0’ when it is unstressed), the target obstruents were also coded for whether the preceding and following vowels were stressed or unstressed. Third, the obstruents were coded for word position (word-initial, medial or final). Fourth, the phrase position of the obstruents were determined, namely, whether the obstruent was phrase-initial (preceded by a pause), phrase-medial (in the middle of an utterance, regardless of the word position) or phrase final (at the end of the utterance). The determination of phrase/utterance-initial and phrase/utterance-final was based on whether there was a comma, period, exclamation mark or colon in the transcript.

Due to the large number of speakers, and obstruents that met the criteria for inclusion in the study (N = 10,610), it was not possible to investigate each surrounding context to ensure, for example, that the adjacent consonant was not deleted, or that stress was produced as would be expected from the CMU transcription, or whether there was a disfluency that introduced a pause
in the middle of a sentence. It was possible, however, to use the small pause designation (‘sp’) that is inserted by the Penn Forced Aligner when a short silence is present between two expected phonemes to determine how often a pause preceded or followed a target voiced obstruent. This is helpful for verifying the appropriate classification of the phrase position of the voiced obstruents. The output of a Python script showed that an ‘sp’ was present 87% of the time when relevant punctuation was next to a voiced obstruent in the story that the participants’ read. A rough examination of the output suggested that the errors were randomly distributed over speakers and items, and that no target item was systematically produced without an adjacent pause, making it unlikely that these misclassifications substantially affect the results. While the large number of tokens should mitigate these concerns to some extent, the limitations of this method should be considered in the interpretation of the results.

Once the canonically produced obstruents were identified, a Praat script was used to measure the duration of the obstruent interval and the proportion of voicing in the interval. The latter measurement was done only for target obstruents in phrase-medial position, since acoustic data alone do not provide information about where stops begin, or where they end if they are unreleased. (While the proportion of voicing in fricatives could have been measured regardless of phrase position, that measurement was not included here in order to be able to compare stops and fricatives.) The voicing measure was obtained using a Praat script that implemented the Voice Report (Boersma & Weenink, 2013), which contains a measure called the fraction of locally unvoiced frames. Praat’s defaults were used, and the pitch settings were optimized for voice analysis as described in the Praat manual (see also Eager, 2015, Bárkányi & Kiss, 2009 for use of this measure in obstruent voicing). This measure reports the proportion of voicelessness in an interval, which was converted to a proportion of voicing. In addition to assessing the
proportion of voicing for the whole duration of the frication or closure, the duration of each fricative and stop was also divided into thirds in order to determine the shape of voicing, that is, for example, whether the proportion of voicing steadily increased or decreased over each of the three intervals, or whether it decreased from the first to the second and then increased again.

For the voiced obstruents, the amount of voicing during the closure is analyzed in two ways, both as a categorical classification and as a measure of partial voicing. For the categorical measure, each obstruent was classified as to whether it was voiced (greater than 90% of the interval was identified as voiced by the Praat Voice Report), unvoiced (less than 10% of the interval was identified as voiced) or partially voiced (between 10-90% of the interval was voiced). This range was chosen (instead of 0% and 100%) so as to not exclude cases from being fully unvoiced or voiced by just one glottal pulse.

The partial voicing measure, in conjunction with the division of the obstruent interval into thirds, is used to further examine the amount and shape of voicing present for those obstruents that fall into the partially voiced category. By examining the smaller intervals, we can determine, for example, whether there is voicing ‘bleed’ from a preceding sound, or perhaps whether there is ‘classic negative VOT’, which would be evident if voicing is not present in the first interval, but it begins instead in the second or third intervals. This measure is described in more detail in Section 3.2.

3. Results

3.1. Categorical voicing measure

The analyses in this section examine the role of segmental and prosodic variables on the proportion of obstruents realized with full voicing (N = 3200), full devoicing (N = 2434), or
partial voicing (N = 4976) as defined in Section 2.2. As previous accounts of English phonetics have often stated that voiced obstruents—especially stops—in utterance-initial position are frequently completely devoiced, the first analysis focuses on the role of phrase position on the implementation of voicing. In phrase-initial stops, it is not possible to know whether voicing coincides with the onset of the closure or whether it begins after the closure, just as it is not possible to know whether the end of voicing coincides with the end of a stop closure in final position if the stop is not released. Therefore, for the analysis of phrase position, voicing is treated as a binomial variable with the levels ‘voiced’ and ‘unvoiced’, which represents the presence of voicing of any duration versus the complete devoicing. This is a simple model that includes only manner of the target obstruent and phrase position, since variables such as preceding or following stress, or preceding segment will not apply to all of the phrase positions. A more complex model including prosodic and segmental factors will follow using only phrase medial stimuli.

The analysis of phrase position is a binomial mixed effects regression using lme4 in R (Bates, Maechler, Bolker, & Walker, 2014), with fixed effects of obstruent manner (stop, fricative) and phrase position (initial, medial, or final). Note that no phrase-initial voiced fricatives appear in the corpus, so the levels of phrase position for fricatives are limited to medial and final. Words and speakers are treated as random effects, with random intercepts for words and random slopes (for manner and phrase position) and intercepts for speakers.

Results show that overall, significantly fewer stops are produced with voicing than fricatives (β = -0.54, p < 0.001), though this result can only be cautiously interpreted since there were no phrase-initial fricatives. Initial and final position have significantly fewer voiced tokens than medial position (initial: β = -1.87, p < 0.001, final: β = -3.09, p < 0.001), and there are also
significantly more voiced tokens in final position than in initial position ($\beta = 1.23, p < 0.006$). An interaction between manner and position indicates that stops are voiced significantly more often than fricatives in final position ($\beta = 1.79, p < 0.001$). These results are given in Table 1.

**TABLE 1 HERE**

The next analysis examines the effects of word position, preceding stress, following stress, preceding segment, and obstruent duration on whether the target obstruents are realized with full, partial, or no voicing.\(^1\) This analysis was carried out over a subset of the data containing only obstruents in *phrase-medial* position in order to ensure that none of the variables were undefined for some levels (e.g., preceding stress in phrase-initial position). Since the dependent variable in this analysis contains three levels, a mixed effects multinomial logistic regression was carried out using the package *polytomous* in R (Arppe, 2013). The fixed effects included target manner (stop, fricative), word position (initial, medial, final), preceding stress (stressed, unstressed), following stress (stressed, unstressed), preceding segment (sonorants [defined as vowels and approximants], nasals, voiced fricatives, voiceless fricatives, voiced stops, and voiceless stops), and a numeric predictor of obstruent duration. The duration predictor was included since previous research has shown that there is a greater proportion of voicing in an obstruent constriction when the duration is shorter (e.g., Ohala & Riordan, 1979; Westbury &

---

\(^1\) The initial intention was to include both preceding segment and following segment in the analysis. However, the model containing both of these factors did not converge. In order to determine which factor to ultimately include, one model was run containing preceding segment (and the remaining fixed factors, as listed) and a separate one in which following segment was substituted for preceding segment. Whereas the preceding segment led to significant effects for all segment types for all three voicing categories, the following segment was not significant for any segment type for the partial voicing category, or for one segment type for the voiced category or two types for the voiceless category. The significant effects that were found can be summarized as showing that all following obstruents (regardless of their own voicing specifications) led to less full voicing in the target obstruents (following voiced fricatives: $\beta = -0.254$, voiced stops: $\beta = -0.32$, both $p < 0.05$, voiceless fricatives: $\beta = -0.67$, voiceless stops: -1.98, both $p < 0.001$). Because of the overall greater presence of significant effects for preceding segment, we focus on that variable in this analysis. Möbius (2004) also reported that the preceding context was of greater significance in German as well.
Keating, 1986). Since visual inspection of the results for preceding segment indicated that there were substantial differences between stops and fricatives, interaction terms for manner and preceding segment were also included (for preceding segments except voiceless fricatives, which did not occur before voiced fricatives in the corpus). The baseline values for the predictors were fricative, medial word position, unstressed (for preceding and following stress), and sonorant for preceding segment. Random intercepts for words and speakers are included, but random slopes were not added as the model failed to converge when they were included.

Results for this analysis are given in Table 2, and the fixed effects are illustrated in Figures 1-3. This analysis indicates that word position and manner can be summarized as having a similar overall effect in phrase medial position: the significant effects show that 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. Partial voicing is unaffected by these manipulations. For the lexical stress factors, there is significantly more full voicing when stress precedes the obstruents, and significantly less full voicing accompanied by more partial voicing and full devoicing when stress follows the obstruents. The linear predictor of obstruent duration indicates that longer obstruents are significantly less likely to be fully voiced and more likely to be partially voiced, but there is no significant effect for full devoicing.

The preceding segment has a number of different effects, as shown in Figure 3. For the main effects, the only type of sound to significantly increase full voicing is a nasal (e.g., ‘Canadian border’), accompanied by a decrease in partial voicing. All of the remaining preceding sounds—voiced and voiceless stops and fricatives (e.g., voiced fricative before voiced stop: ‘jazz band’, voiceless fricative before voiced stop: ‘si[ks] days’)—lead to a significant decrease in fully voiced and partially voiced tokens (except the full voicing coefficient for preceding voiced
stops, which is negative but not significant). 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.

TABLE 2
FIGURES 1-3 HERE

The categorical division of voicing types into fully voiced, partially voiced, and fully voiceless is useful because similar divisions have been carried out in previous studies (e.g., Docherty, 1992). The results of this analysis are consistent with both previous impressionistic and quantitative analyses, though other studies have mainly focused on target obstruents in carrier words or phrases, and/or on varieties of British English. Stops exhibit significantly less full voicing than fricatives, and obstruents at the beginning and ends of phrases are completely devoiced significantly more often than they are in phrase-medial position. A similar pattern is mirrored for initial and final word position as compared to word-medial position when the obstruents are in phrase-medial position. Obstruents that are preceded by a stressed vowel are fully voiced significantly more often. The primary influence of the immediately preceding segment is that adjacent obstruents usually cause the target obstruent to be fully devoiced more often, even in cases when the neighboring segment is (underlyingly specified as) voiced. As noted in footnote 1, although the effect of the following segment on the whole was much less
informative, the significant effect that did emerge is that a following obstruent (regardless of whether it is underlyingly voiced or voiceless) also significantly decreased the probability that the preceding obstruent would be completely voiced.

By examining only the divisions of voicing into voiced, voiceless, and partially voiced, an important aspect of the implementation of voicing in voiced obstruents in American English is obscured. When obstruents are partially voiced—that is, Praat has determined there is voicing in 10%-90% of the closure or frication interval—it is not clear what shape the partial voicing takes on. That is, there are several logical possibilities for how voicing could be implemented during an obstruent. These are detailed in the next section, which examines the distribution of voicing in obstruents that were classified as partially voiced.

3.2. Partial voicing shape

While Praat may report that an obstruent in a particular utterance has a constriction portion that is, say, 45% voiced, this reveals nothing about where in the interval the voicing is present. For example, the proportion of voicing in the constriction for partially voiced phrase-medial obstruents is shown in Figure 4. In phrase-medial position, the amounts of voicing in word-initial and medial position are very similar for both stops and fricatives, though the difference between stops and fricatives is more obvious for final position. However, as will be shown below, when the location of the voicing is taken into account, there are striking differences between the patterns of voicing for stops as compared to fricatives, and with respect to word position as well.

FIGURE 4 HERE
There are several possibilities for how partial voicing can be implemented in an obstruent. The first possibility can be referred to as *bleed*, to indicate voicing that continues from the preceding sonorant and then dissipates before the following stop release or end of frication. This is illustrated in Figure 5a. The next type can be called the *trough* pattern, which refers to the case where voicing continues from the preceding sound, then dies out, and then reappears before the stop release or end of frication in a fricative (see Figure 5b). The third type is referred to as *negative VOT*, after classic descriptions of negative voice onset time as starting in the middle of the closure or frication period and continuing into the following sound (see Figure 5c). The final type can be called *hump*, to indicate that there is no voicing at the beginning or end of the closure or frication period, but it is present in the middle of the interval (see Figure 5d).

**FIGURE 5a-5d HERE**

Using the division of the stop closure or frication interval into thirds, each obstruent classified as partially voiced as described in Section 2.2 was further categorized for the shape of the voicing. This analysis is again limited to obstruents in phrase-medial position since the proportion of voicing in phrase-initial and final position often cannot be conclusively determined. An obstruent was characterized as *bleed* if the proportion of voicing decreased from the first to the third interval (being completely absent in the third interval in a large proportion of cases), and as *negative VOT* if the proportion of voicing increased from the first to the third interval (again, typically being completely absent in the first interval). An obstruent was labeled as *trough* if there was a greater proportion of voicing in the first and third intervals than in the second, and as *hump* if the proportion of voicing increased from the first to the second interval.
and then decreased again. As shown in more detail in Figure 6, 72% of phrase-medial obstruents were produced with the bleed pattern and 24% with the trough pattern. Overall, only 3% of tokens showed the hump pattern and only 1% were produced with the negative VOT pattern.

FIGURE 6 HERE

As is evident from Figure 6, there is a difference in the voicing shape patterns corresponding to stops and fricatives. To analyze this more thoroughly, a statistical analysis was carried out comparing only the trough and bleed patterns. Negative VOT and hump patterns were not included in the analysis since they were such a small proportion of the data. This analysis was a binomial logistic regression with fixed effects of manner (stop, fricative) and word position (initial, medial, final). The reference values were fricatives and final position. Words were included as random factors with random intercepts, and subjects with random slopes and intercepts for all of the fixed effects. Results showed that there was a significant effect of manner, indicating that there were fewer trough patterns and more 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 than in either medial or initial 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 position (β = -1.2, p < 0.001) and medial position (β = -2.64, p < 0.001) for fricatives than for stops. The proportions of bleed are similar for all word positions for stops (ranging from 81%-86%).

In addition to the categorical analysis of voicing shape, an examination of the continuous proportion of voicing over the three intervals of the constriction provides further information
about the amount of voicing that English speakers produce, and how the proportion changes over
the course of the obstruent. For this analysis, a mixed effects linear regression was carried out
using \texttt{lme4} in R on the subset of words in phrase-medial position. The dependent variable is
proportion of voicing, and the fixed effects are manner (stop, fricative), word position (initial,
medial, final), interval (first, second, third), and obstruent duration. Manner, word position and
interval were fully crossed. Obstruent duration was entered as a main effect and as two-way
interactions with manner, word position and interval, in order to avoid a proliferation of three
and four-way interactions. Words were included as random factors with random intercepts, and
subjects with random slopes and intercepts for manner and word position. The fixed effects are
sum coded. The baseline for the model was stops, first interval, and word-initial position. The
linear regression, unlike the logistic and multinomial models above, does not return a p-value as
it is currently implemented in \texttt{lme4}. Therefore, following the guidelines established in Gelman
and Hill (2006), and widely used in psychological and linguistic research, a factor was
considered significant if the absolute value of the t-value was greater than 2. The proportion of
voicing in the three intervals for each manner and word position are shown in Figure 7.

FIGURE 7 HERE

Results of the mixed effects analysis are shown in Table 3. First there is a significant
effect of interval: both the second interval and third interval have significantly less voicing than
the first interval. Although there are no significant main effects of word position, a significant
two-way interaction between the third interval and medial position and three-way interactions
between manner, the third interval and both medial and final word positions clarify the results. In
the third interval, there is significantly more voicing for fricatives in initial and medial word
position than there is for all stops and for fricatives in final word position. While there is no main effect of manner, there is a significant two-way interaction between manner and the second interval, which reflects the lower proportion of voicing for fricatives in all word positions than all stops in the second timeframe. There is also a two-way interaction between manner and the third interval, as well as the significant three-way interaction already noted. These interactions arise because the proportion of voicing goes up for the fricatives between the second and third interval for fricatives, whereas it continues decreasing for the stops. The significant main effect of duration indicates that overall, shorter obstruents are produced with more voicing, which is consistent with the analysis of the categorical labels in Section 3.1. The only significant interaction for duration is with manner, which indicates that shorter fricative durations have a lower proportion of voicing than shorter stop durations do.

TABLE 3 HERE

The results for the proportion of voicing across the constriction duration divided into thirds indicate that broadly, stops reflect the bleed pattern and fricatives reflect the trough pattern, except for fricatives in word-final position, which increase in voicing only a negligible amount from the second interval to the third. Possible reasons for these patterns will be discussed in the General Discussion. It is worth noting that these patterns are largely upheld by each stop and fricative consonant (treating affricates as part of the stop category), with the exception of /ð/, which patterns more like a stop in word-medial position (it was not present in any tokens in initial or final position). For the sake of illustration, the proportions of voicing in each interval for each consonant are provided in Table 4. The behavior of /ð/ is consistent with Docherty’s
findings for /ð/ compared to other voiced fricatives, and in American English may be further attributable to the fact that /ð/ is often produced as a stop in many dialects (Zhao, 2010).

TABLE 4 HERE

4. Discussion

The implementation of phonation in American English obstruents is a complex issue that can be best understood when myriad factors are considered, including word position, phrase position, relative location of stress and the difference between stops and fricatives. Moreover, further fine details of phonation are revealed when more typical categorizations such as ‘fully voiced’, ‘partially voiced’, and ‘completely devoiced’ are broken down into the extent of voicing present in smaller intervals. In this study, the proportion of phonation in obstruents in phrase-medial position was examined by dividing the obstruent closure into three intervals, which revealed that there are substantial differences in how voicing is realized in fricatives and stops. The influence of the variables examined in this study are further discussed in sequence.

4.1. Effects of linguistic factors on the implementation of voicing

*Effects of phrase position.* To start with phrase position, phrase-initial stops in (American) English have been much studied in previous research, both since they often defy their phonological specification for voicing and because they are prevalent in the laboratory speech that phoneticians often work with (Keating, 1984; Lisker & Abramson, 1967; Smith, 1978; Westbury, 1979). Variation in the amount of voicing preceding the burst (i.e.,
‘prevoicing’) has been found, but all researchers have previously shown that phrase-initial voiced stops in American English can be produced both with some phonation preceding the release, or without any phonation at all. The proportion of voicing implemented in phrase-initial position in the connected speech examined in this study (25% of tokens had some amount of prevoicing) is on the low end of previous findings, some of which have reported over 50%. In phrase-final position, the proportion of stops containing some voicing during the closure rose considerably (57%). For stops, when they are preceded by a pause in initial position or unreleased in utterance final position, it is not possible to tell where voicing begins relative to the moment of closure of the stop on the basis of acoustic data alone. Westbury (1983) makes the brief observation for utterance initial stops based on a cinefluorographic recording of one speaker that voice onset followed the oral occlusion, but consistently preceded release by 60-90ms. This suggests that initial stops can in fact manifest the negative VOT pattern described in Section 3.2, but the rest of the data in the current study indicate that this is the primary location where this implementation occurs in American English. As Solé (2011) succinctly explains, voicing in utterance initial position is hampered because oral and subglottal pressure increase simultaneously instead of generating higher subglottal than oral pressure, which is necessary for voicing to arise if no additional vocal cavity adjustments occur (see also Slifka, 2000; Westbury & Keating, 1986). In final position, maintaining phonation into the closure is easier because the stop is most likely preceded by a sonorant, which means that voicing is already occurring when the closure phase begins.

Unfortunately, no phrase-initial voiced fricatives were observed in the corpus. In phrase-final position, the proportion of fricatives containing some voicing was low (35%). This is similar to Smith’s (1997) findings that phrase-final /z/ was always devoiced by her speakers, but
differences are likely due to experimental factors such as the inclusion of /v/ in the category (i.e., Haggard (1987) found that utterance-final /v/ was voiced slightly more often than /z/), the more connected nature of the speech of whole read paragraphs (as opposed to single unconnected sentences), and a greater number of participants. Smith’s airflow data shows that the rate of airflow for /z/ before a pause is consistently low for her speakers, which can be argued to be occurring because the phrase-final transglottal pressure is too low to maintain voicing. Note that the pressure requirements for phrase-final fricatives may be more precise than those for final stops (Ohala, 1983; Rodgers & Fuchs, 2010 for pressure data on voiced final stops), which have some voicing about twice as often as the fricatives do.

Effects of word position. Unlike phrase-initial and final obstruents, the investigation of obstruents in phrase-medial position allows for a more balanced examination of phonotactic and positional effects on the implementation of phonation. To begin with word position, Figure 1 shows that there are differences based on consonant manner. For fricatives, initial and medial position pattern together, with about half of the tokens being produced with full voicing, and 45% with partial voicing. Word-final fricatives, on the other hand, are only fully voiced 21% of the time. The pattern for stops is the opposite, with medial and final stops patterning together with full voicing 45% of the time and initial stops having much less full voicing (14%) and more complete devoicing (35%). Except for the absent data for phrase-initial fricatives, the general pattern for the phrase-medial obstruents in word-edge positions matches those of the phrase-edge findings. This is potentially unexpected, since the surrounding environment in phrase-medial position frequently contains adjacent sonorants which should facilitate voicing across the board. One possible interpretation of these results is that listeners may be able to use the probabilistic cues to voicing that are consistent in phrase-edge and phrase-medial positions to assist with word
segmentation (see also Padgett & Myers, 2014 for related comments on domain generalization). This idea could be tested with a perceptual follow-up study.

A comparison of phonation in phrase-medial versus phrase-initial stops can also inform questions about domain-initial strengthening on the implementation of acoustic cues. While this study cannot directly address the distinction between voiceless (fortis) versus voiced (lenis) stops in various prosodic positions in English, it is possible to discuss the ramifications of domain position within the class of voiced stops (recall that utterance-initial fricatives were not in the corpus). One theory of domain-initial strengthening, which has been referred to as the Uniform Strengthening Account (Cole, Kim, Choi, & Hasegawa-Johnson, 2007; Kuzla & Ernestus, 2011), would predict that voiced stops at higher domains in English should have more fortis-like properties, given that the spatio-temporal expansion that has often been found for English domain-initial stops (e.g., Cho & Keating, 2009; Fougeron & Keating, 1997) is less hospitable for certain properties of voiced/lenis stops, including phonation. The predictions for an alternative proposal, the Phonetic Feature Enhancement Account (e.g., Cho, 2005; Cho, Lee, & Kim, 2014; Cho & McQueen, 2005; de Jong, 1995), depend on how the phonetic features for voiced and voiceless stops in English are specified. If the specification for voiced stops is \{+voiced\}, then it might be predicted that there should either be no difference between domain-initial and medial positions for the duration of phonation, or that phonation should be increased in initial position in order to maximally distinguish it from voiceless stops. However, if the phonetic feature for phonologically voiced stops is instead \{-voiced, -spread glottis\}, then it might be expected that there should be less full voicing in stronger prosodic positions to enhance the \{-voiced\} feature. Thus, the results of the current study of English are compatible with both those of Kuzla and Ernestus’s (2011) examination of German and with the Phonetic Feature
Enhancement Account if the proper specification is \{-voiced, -spread glottis\} (or perhaps no specification at all for \{spread glottis\}; see below regarding comparisons with German): there is a smaller proportion of stops containing any voicing at all in absolute utterance initial position (25%), than there is in phrase-medial, word-initial stops (65%), which is smaller than word-medial or final position (85%-90%). On the other hand, the implementation of phonation may not be one of the main cues to distinguishing between voiced and voiceless stops in English in prosodically prominent positions (cf. Cole et al. 2007, who do not even look at closure phonation as a cue), but rather a cue which is more helpful for determining word position. This would be compatible with the previous speculation that the intermediate rates of phonation for stops in word-initial, phrase-medial position may help in word segmentation.

*Effects of stress placement.* Two other factors examined in this study that affect the relative proportions of partial, full, and no voicing are preceding and following stress specification. As shown in Figure 2, when the vowel preceding the obstruent is specified as stressed, the target obstruent is almost twice as likely to be fully voiced compared to when it is specified for unstressed. The opposite effect is seen for the following vowel: when it is unstressed, the target obstruent is fully voiced nearly twice as often. These findings are compatible with Lisker’s (1970) findings that voiced stops preceded by a stressed syllable have demonstrably lower peak oral pressure than voiceless stops in that environment, and also than voiced stops that are followed by a stressed syllable, which is more compatible with vocal fold vibration. Stress has less of an effect on complete devoicing, suggesting that stress mainly influences whether an obstruent is going to be partially or fully voiced. The effect of stress seen in this data is consistent with previous work on consonant reduction in the American English flapping position (Bouavichith & Davidson, 2013; Warner & Tucker, 2011), which has shown
that voiced consonants tend to weaken—e.g. are realized as shorter and fully voiced, or as an approximant—in post-stress environments.

*Effects of preceding context.* The influence of the preceding segment differs depending on whether the target obstruent is a stop or a fricative. When the preceding sound is a sonorant (here referring to vowels and approximants), stops and fricatives are similar and are only fully devoiced 11-13% of the time, as indicated in Figure 3. Likewise, a similar situation holds for voiced stops preceding both obstruents, which condition full voicing in about a quarter of the cases and full devoicing in 30% of the cases. More differences emerge when the preceding segment is a voiceless stop, voiced fricative, or nasal. Stops preceded by voiceless obstruents are fully devoiced in >75% of cases, which is to be expected, since reinitiating voicing after a voiceless closure interval is similar to the situation of being preceded by a pause, as described above (Shih, et al., 1999; Westbury, 1979). The comparison case for fricatives preceded by voiceless stops (with only 12% devoicing and 53% full voicing) is not reliable, since there is a total of only 17 cases (compared to 274 for voiceless stops before stops). Stops preceded by voiced fricatives are often fully devoiced (66%), whereas fricatives preceded by voiced fricatives remain fully or partially voiced in 90% of tokens. While these results should be interpreted in light of the fact that the voicing of adjacent obstruents were coded to reflect their underlying specification and not necessarily their phonetic implementation, there is still an interesting interaction between preceding voiced stops and voiced fricatives, and their differential effects on the target voiced stops and fricatives. Finally, while nasals condition a high proportion of full voicing for stops, this is not the case for fricatives.

The results demonstrate that overall, fricatives in phrase medial position are more consistent in maintaining partial or full voicing during the constriction than stops are. For all
preceding segment types, complete devoicing ranges from 10-30%. In contrast, full devoicing in phrase-medial stops ranges from 11%-78%. One generalization that can be made is that voiced stops preceded by any kind of obstruent, regardless of voicing, tend to be realized with more complete devoicing than voiced fricatives in the same environment (this is true even for stops preceded by voiced stops, though the overall proportion of devoicing is lower than for other preceding obstruents). This pattern is consistent with previous observations that as the period of obstruction lengthens (e.g. obstruent-stop sequences, as opposed to sonorant-stop sequences), the less likely it is for the closure in the second part of the sequence to remain voiced (e.g., Westbury & Keating, 1986). As already noted for fricatives, they show a similar pattern as stops when preceded by a voiced stop, but this is not the case when preceded by a voiced fricative. Here, complete or partial voicing of the fricative is much more likely (90%) than it is for a stop (34%). The proportion of voiced fricative-voiced fricative tokens containing voicing is a little higher than that reported in Stevens et al. (1992), at 68%.

One possibility to account for this difference between stops and fricatives is that preceding context is interacting with word position. An examination of the proportions of stops and fricatives preceded by voiced obstruents shows that 97% of the stops in this category are also in word-initial position. Yet, a comparison with the fricatives that are also in initial position mirror the same pattern as the results for word position as a whole: 95% of initial fricatives preceded by a voiced obstruent (e.g., across a word boundary) are partially or fully voiced, whereas 44% of stops in the same position are partially or fully voiced. Thus, the difference between fricatives and stops cannot be attributable to an interaction between preceding segment and word position. An alternative possibility is that the fricatives in second position in the consonant sequences are better able to maintain voicing because speakers produce them with
weaker frication, perhaps intentionally. Though tokens with no evidence of an aperiodic component were eliminated from the analysis, it is possible that speakers weakened frication in fricatives preceded by obstruents, since doing so would assist in the maintenance of voicing (e.g., Haggard, 1978; Ohala, 1983; Stevens, et al., 1992).

Another issue that may be related to an asymmetry between stops and fricatives concerns the phonological specification of voicing in Germanic languages. Though a comprehensive account of the relationship between phonetic implementation and phonological voicing categories is beyond the scope of this paper, some work on German—again, which has voicing patterns similar to those of English—may bear on these findings. Based on acoustic data from German, Jessen and Ringen (2002) and Beckman et al. (2009) have argued that the acoustic implementation of phonation in German is consistent with a featural specification of [+spread glottis] for voiceless stops and no specification at all for voiced stops. In contrast, fricatives have a contrast for [voice], which explains why voicing is maintained with much more consistency than voicing in stops. A similar conclusion was reached by Tsuchida et al. (2000) for English following the examination of transillumination data for voiceless stops and fricatives, but it should be noted that study lacked any investigation of the phonetic implementation of voiced obstruents. Nevertheless, the overall greater proportion of voicing in fricatives as compared to stops in the current study is consistent with a difference in the operative featural specifications for stops and fricatives in English.

Regardless of the possible phonological difference between stops and fricatives, the opposing effect of a preceding nasal on a voiced stop versus a voiced fricative may be closely tied to the differing effect of nasalization on these two obstruents. The need to maintain a high oral pressure in fricatives is already compromised by the pressure requirements for phonation,
and Solé (2009) argues that if the velopharyngeal port is open during the production of a voiced fricative (as might be expected under a coproduction account of adjacent gestures, e.g., Fowler & Saltzman, 1993), strong frication is even further impaired and voiced fricatives are more likely to be realized as frictionless continuants. By devoicing, the fricatives can better retain their frication noise without having to modify their gestural coordination with the adjacent nasal. Moreover, one can speculate that when following nasals, fricatives could devoice to enhance the frication noise rather than weaken to retain the voicing in order to remain maximally acoustically distinct from the preceding sonorant/nasal. In contrast, when stops are preceded by nasals, the same nasal leakage allows phonation to be prolonged temporarily as the velum is closing (Ohala & Ohala, 1993; Rothenberg, 1968; Solé, 2009; Westbury, 1983). Indeed, the preceding nasal is so effective in maintaining voicing on a following stop that it has the highest proportion of fully voiced stops (61%), even compared to a preceding sonorant (34%). The English pattern shown here is presumably phonetically related to the cross-linguistically common phonological process that requires voiced stops after nasals (Hayes, 1999; Pater, 1999).

4.2. Voicing shape: how phonation is implemented in the obstruent constriction

The aerodynamic environments provided by adjacent contexts sets the background for a discussion of the time-course of phonation throughout the obstruent constriction. The results of this study indicate that when a word-medial obstruent is partially voiced in American English, it does not have the same acoustic characteristics for both stops and fricatives. Instead, it is evident that the location of voicing in the closure portion depends on the manner of the obstruent. As illustrated in Figure 6, whereas stops show a bleed pattern in >80% of all tokens (referring to the carryover of voicing from the preceding vowel or sonorant that dies out part of the way into the
stop), fricatives more frequently display a *trough* pattern (the carryover of voicing that dies out, followed by the initiation of voicing again after a period of voicelessness). For fricatives, the majority pattern for the shape of partial voicing also interacts with word position, and there is a greater proportion of bleed cases in word-final position than in word-initial or word-medial position. Figure 7, which shows the average proportion of voicing in three intervals of equal duration, confirms these patterns.

It is also evident from Figure 6 that two other logical possibilities for voicing shape—*hump* and *negative VOT*—are rarely implemented. *Hump* refers to voicing that starts after the stop closure or onset of frication, and ends before the offset of the consonant. The paucity of this shape is not surprising, since such a pattern would not be taking advantage of the voicing in the adjacent sonorants that are very common in this data. *Negative VOT*, in contrast, refers to a pattern of voicing that begins partway through the obstruent and continues past the release of the obstruent. It might be expected that this pattern would be more common, for two reasons. First, this is the pattern that can be present in ‘pre-voiced’ stops in phrase-initial position after a pause. Though pre-voicing in such stops is relatively rare, it is present, and the variable duration of such voicing reported in previous studies (e.g., Flege, 1982; Keating, 1984; Lisker & Abramson, 1964; Lisker & Abramson, 1967; Scobbie, 2009; Westbury, 1979) suggests that the onset of phonation does not always co-occur with the onset of the stop closure. Thus, the negative VOT pattern seems to be available phrase-initially, and might be expected to occur in other positions as well. On the other hand, it is least likely that speakers will initiate phonation precisely when oral pressure is highest, as in the case of stops, if they do not have to; instead speakers can opt for the bleed pattern in order to ensure that voicing is produced in the closure. Though negative VOT alone is rarely implemented in these results, the trough pattern—which essentially
incorporates negative VOT as part of the pattern—does occur in 14% of the word-medial stops. While trough occurs relatively consistently across word-position for phrase-medial stops, the rates are still a small proportion of the total and are 10% lower than utterance-initial position. The considerably higher incidence of the trough pattern in fricatives will be discussed further below.

The second reason for assuming that this pattern would be more prevalent is more terminological in nature; by definition, negative VOT refers to the amount of time before the stop release that voicing begins (e.g., Ashby & Maidment, 2005; Ladefoged & Johnson, 2014; Reetz & Jongman, 2009). If partial voicing during a stop closure is actually characterized by the bleed pattern, then measuring the duration of that voicing as a negative number that takes the offset of the burst as a reference point is ill-defined. Ultimately, the results of this study suggest that referring to partial voicing in the closure of a stop that is phonologically specified for voice as ‘negative VOT’ is misleading.

The majority patterns, bleed (for stops) and trough (for fricatives), are not surprising, because they capitalize on the phonation from the surrounding sonorants that are the most frequent environments in this sample of American English. Once voicing has been initiated, it will continue during an obstruent closure until the oral pressure becomes high enough that voicing ceases (without extra maneuvers to increase the size of the pharyngeal cavity, e.g., Ohala, 1983; Proctor, Shadle, & Iskarous, 2010; Westbury, 1983). In the production of a voiced obstruent, speakers can take advantage of the ‘carryover’ phonation that is provided from a preceding sonorant to prolong the voicing produced during the obstruent constriction. Articulatory data confirms that speakers often implement various maneuvers to expand the volume of the supralaryngeal cavity and control supralaryngeal pressure in order to maintain
voicing in voiced stops as compared to voiceless ones (Bell-Berti, 1975; Koenig & Lucero, 2008; Proctor, et al., 2010; Rothenberg, 1968; Westbury, 1983). In comparison, studies have shown that while there is typically a little carryover (bleed) voicing at the beginning of the closure for voiceless stops, it is much shorter than the duration of voicing found for voiced stops (Docherty, 1992; Flege & Brown, 1982; Keating, 1984; Suomi, 1980). Whether speakers are both actively expanding the supralaryngeal volume for voiced stops and truncating the carryover voicing in voiceless stops is uncertain, but it is clear that American English, which does not obligatorily maintain a fully voiced closure in the implementation of voiced stops even in intervocalic position (cf. Abdelli-Beruh, 2004 for French; Burton & Robblee, 1997 and Samokhina, 2010 for Russian; Keating, 1984 for Polish), nevertheless distinguishes the amount of voicing produced in the closures of voiced and voiceless stops.

While the bleed pattern was most common for stops and relatively common for fricatives, the trough pattern was observed about 14% of the time for stops (regardless of word position), but 68% of the time for fricatives in initial and medial word position and 34% of the time in final word position. This finding suggests that the overlap of fricative and vowel gestures, and consequently the re-initiation of voicing before the fricative is completed, is probably less difficult than the comparable onset of closure-medial voicing for stops. Similar to the earlier discussion of the strength of frication noise, Docherty (1992) speculates that when there is a transition from a critical constriction to an open vocal tract, there is first a gradual attenuation of frication noise that allows for the initiation of vocal fold vibration even before all aperiodic energy disappears. A similar gradual weakening of the constriction while maintaining the primary cue to place is not available for stops, which can explain why the trough pattern is much more prevalent for fricatives than for stops. Cursory inspection of the few trough cases that exist
for stops demonstrate that some of the tokens are lacking a burst, suggesting that the supraglottal pressure during the closure was already low and possibly more amenable to phonation.

5. Conclusion

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. While several of these observations have been made in previous literature, this study used a large corpus of connected (read) speech to investigate how patterns found in isolated words or phrases relate to more naturalistic conditions. For example, when adjacent to another obstruent, voiced stops and fricatives are less likely to retain voicing since the obstruction is long. On the other hand, preceding nasals are shown to enhance voicing in stops and inhibit it in fricatives. Phrase-initial stops are only produced with voicing in about a quarter of the tokens because initiating voicing when oral pressure is high is difficult, whereas final obstruents do generally retain at least some partial voicing, because voicing is usually already initiated when the constriction for the obstruent is formed. While English allows for these phonetic considerations to be maximally prominent, they can also be overruled in languages that consistently produce voiced obstruents with fully voiced constrictions, as in French or Polish. Even some English dialects may have more stringent requirements for producing fully voiced obstruents, as demonstrated by Jacewicz et al.’s (2009) data for speakers from North Carolina. The results from the current study also suggest that even when voicing does have more aerodynamic support, as in the case of the majority of phrase-medial, word-initial stops, it may not necessarily take optimal advantage of
the surrounding phonation. A potential explanation for this finding is that English speakers use partially voiced or fully devoiced stops as an acoustic cue to being a word onset, which may help in speech segmentation.

The analysis of partial voicing in this study shows that the implementation of phonation is not a uniform phenomenon. Contrary to what is implied by the oft-used term ‘negative VOT’, the most common implementation of partial voicing is bleed, which takes advantage of the carryover voicing from the preceding segment. The second most common shape for partial voicing is the trough, which can be considered a combination of bleed and negative VOT, except that the trough is mainly prevalent for fricatives since the re-initiation of voicing would be aerodynamically disfavored in stops. The results for partial voicing provide more detail beyond categorical distinctions between partial and full voicing and complete devoicing, and confirm the phonetic tendencies that are to be expected when a language does not require fully phonated closures for voiced obstruents.

Acknowledgments

I would like to thank James Whang, Steven Foley, and Benjamin Gittelson for their help in analyzing the data for this study. I also thank Gillian Gallagher, Suzy Ahn and Sang-Im Lee-Kim for their helpful comments on this work and the audience at the Acoustical Society of America meeting in Providence and at ICPhS 2015 in Glasgow for their input. This research was supported by faculty research funds from New York University.

Arppe, A. (2013). polytomous: Polytomous logistic regression for fixed and mixed effects (Version 0.1.6, R package).


Figure 1. Proportions of phrase-medial unvoiced, partially voiced, and fully voiced obstruents by word position.
Figure 2. Proportions of phrase-medial unvoiced, partially voiced, and fully voiced obstruents by (a) preceding vowel’s specification for stress and (b) following vowel’s specification for stress.
Figure 3. Proportions of phrase-medial unvoiced, partially voiced, and fully voiced obstruents by preceding segment.
Figure 4. Proportion of overall voicing during the constriction for fricatives and stops in phrase-medial position. Error bars show standard error.
Figure 5. Examples of the partial voicing shapes (not including fully voiced or completely devoiced obstruents)

(a) Example of *bleed* in the phrase ‘a boiling’
(b) Example of *trough* in the phrase ‘was usually’
(c) Example of negative VOT in the phrase ‘was aghast’
(d) Example of *hump* in the phrase ‘a dodo’
Figure 6. Proportion of voicing shapes for fricatives and stops in phrase-medial position. See text for explanation of ‘voicing shapes’.
Figure 7. Proportion of voicing for phrase-medial stops and fricatives by interval. ‘Initial, medial, final’ refer to word position.
Table 1. Proportion of voiced (vs. unvoiced) obstruents in initial, medial and final phrase position

<table>
<thead>
<tr>
<th></th>
<th>Fricative</th>
<th>Stop</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>--</td>
<td>0.25</td>
</tr>
<tr>
<td>medial</td>
<td>0.86</td>
<td>0.80</td>
</tr>
<tr>
<td>final</td>
<td>0.35</td>
<td>0.57</td>
</tr>
</tbody>
</table>
Table 2. Multinomial mixed effects regression coefficients in log-odds for voicing categories for phrase-medial obstruents. * indicates significance $p < .05$, ** indicates $p < .001$

<table>
<thead>
<tr>
<th></th>
<th>voiced</th>
<th>partial</th>
<th>unvoiced</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.129</td>
<td>-1.655**</td>
<td>-2.746**</td>
</tr>
<tr>
<td>Manner (stop)</td>
<td>-0.171**</td>
<td>0.06</td>
<td>0.203*</td>
</tr>
<tr>
<td>Word Position (final)</td>
<td>-0.231**</td>
<td>0.057</td>
<td>0.607**</td>
</tr>
<tr>
<td>Word Position (initial)</td>
<td>-0.452**</td>
<td>0.066</td>
<td>0.473**</td>
</tr>
<tr>
<td>Preceding Stress (stressed)</td>
<td>0.422**</td>
<td>-0.035</td>
<td>-0.03</td>
</tr>
<tr>
<td>Following Stress (stressed)</td>
<td>-0.212**</td>
<td>0.236**</td>
<td>0.414**</td>
</tr>
<tr>
<td>Preceding Segment (nasal)</td>
<td>-0.602**</td>
<td>-0.208</td>
<td>0.655**</td>
</tr>
<tr>
<td>Preceding Segment (vcd fric)</td>
<td>-0.017</td>
<td>-0.003</td>
<td>-0.202</td>
</tr>
<tr>
<td>Preceding Segment (vcls fric)</td>
<td>-3.883**</td>
<td>-1.151**</td>
<td>1.445**</td>
</tr>
<tr>
<td>Preceding Segment (vcd stop)</td>
<td>-0.347*</td>
<td>-0.278*</td>
<td>0.69**</td>
</tr>
<tr>
<td>Preceding Segment (vcls stop)</td>
<td>0.371*</td>
<td>-0.379</td>
<td>-0.258</td>
</tr>
<tr>
<td>Obstruent Duration</td>
<td>-20.546**</td>
<td>14.174**</td>
<td>-1.777</td>
</tr>
<tr>
<td>Manner (stop)*Prec Seg (nasal)</td>
<td>0.841**</td>
<td>-0.306*</td>
<td>-0.926**</td>
</tr>
<tr>
<td>Manner (stop)*Prec Seg (vcd fric)</td>
<td>-2.146**</td>
<td>-0.584*</td>
<td>1.685**</td>
</tr>
<tr>
<td>Manner (stop)*Prec Seg (vcd stop)</td>
<td>0.476*</td>
<td>-0.039</td>
<td>-0.161</td>
</tr>
<tr>
<td>Manner (stop)*Prec Seg (vcls stop)</td>
<td>-2.284**</td>
<td>-0.723</td>
<td>1.622*</td>
</tr>
</tbody>
</table>
Table 3. Linear mixed effects regression model for proportion of voicing for phrase-medial obstruents. Significant t-values, indicated with *, are those with an absolute value greater than 2.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.005</td>
<td>0.028</td>
<td>36.14*</td>
</tr>
<tr>
<td>Manner (fricative)</td>
<td>0.066</td>
<td>0.060</td>
<td>1.09</td>
</tr>
<tr>
<td>Interval (2)</td>
<td>-0.479</td>
<td>0.026</td>
<td>-18.40*</td>
</tr>
<tr>
<td>Interval (3)</td>
<td>-0.722</td>
<td>0.026</td>
<td>-27.73*</td>
</tr>
<tr>
<td>Word Position (medial)</td>
<td>0.045</td>
<td>0.029</td>
<td>1.56</td>
</tr>
<tr>
<td>Word Position (final)</td>
<td>0.018</td>
<td>0.035</td>
<td>0.51</td>
</tr>
<tr>
<td>Duration</td>
<td>-1.705</td>
<td>0.339</td>
<td>-5.03*</td>
</tr>
<tr>
<td>Manner (fric)*Interval (2)</td>
<td>-0.115</td>
<td>0.057</td>
<td>-2.02*</td>
</tr>
<tr>
<td>Manner (fric)*Interval (3)</td>
<td>0.506</td>
<td>0.057</td>
<td>8.91*</td>
</tr>
<tr>
<td>Manner (fric)*Word Pos (medial)</td>
<td>-0.039</td>
<td>0.054</td>
<td>-0.72</td>
</tr>
<tr>
<td>Manner (fric)*Word Pos (final)</td>
<td>-0.122</td>
<td>0.055</td>
<td>-2.20*</td>
</tr>
<tr>
<td>Interval (2)*Word Pos (medial)</td>
<td>0.013</td>
<td>0.018</td>
<td>0.72</td>
</tr>
<tr>
<td>Interval (3)*Word Pos (medial)</td>
<td>0.039</td>
<td>0.018</td>
<td>2.13*</td>
</tr>
<tr>
<td>Interval (2)*Word Pos (final)</td>
<td>0.004</td>
<td>0.027</td>
<td>0.14</td>
</tr>
<tr>
<td>Interval (3)*Word Pos (final)</td>
<td>-0.010</td>
<td>0.027</td>
<td>-0.39</td>
</tr>
<tr>
<td>Duration*Interval (2)</td>
<td>0.528</td>
<td>0.344</td>
<td>1.54</td>
</tr>
<tr>
<td>Duration*Interval (3)</td>
<td>0.011</td>
<td>0.344</td>
<td>0.03</td>
</tr>
<tr>
<td>Duration*Manner (fric)</td>
<td>-1.834</td>
<td>0.432</td>
<td>-4.25*</td>
</tr>
<tr>
<td>Duration*Word Pos (medial)</td>
<td>-0.327</td>
<td>0.404</td>
<td>-0.81</td>
</tr>
<tr>
<td>Duration*Word Pos (final)</td>
<td>0.347</td>
<td>0.479</td>
<td>0.72</td>
</tr>
<tr>
<td>Man (fric)*Interval (2)*Word Pos (med)</td>
<td>-0.028</td>
<td>0.062</td>
<td>-0.46</td>
</tr>
<tr>
<td>Man (fric)*Interval (3)*Word Pos (med)</td>
<td>-0.146</td>
<td>0.062</td>
<td>-2.35*</td>
</tr>
<tr>
<td>Man (fric)*Interval (2)*Word Pos (fin)</td>
<td>-0.031</td>
<td>0.063</td>
<td>-0.49</td>
</tr>
<tr>
<td>Man (fric)*Interval (3)*Word Pos (fin)</td>
<td>-0.281</td>
<td>0.063</td>
<td>-4.48*</td>
</tr>
</tbody>
</table>
Table 4. Proportion of voicing in the first, second and third interval for individual consonants, divided by word position. All tokens are in phrase-medial position.

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>Second</th>
<th>Third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>0.89</td>
<td>0.43</td>
<td>0.20</td>
</tr>
<tr>
<td>d</td>
<td>0.93</td>
<td>0.50</td>
<td>0.14</td>
</tr>
<tr>
<td>g</td>
<td>0.88</td>
<td>0.41</td>
<td>0.18</td>
</tr>
<tr>
<td>dʒ</td>
<td>0.94</td>
<td>0.54</td>
<td>0.16</td>
</tr>
<tr>
<td>v</td>
<td>0.82</td>
<td>0.23</td>
<td>0.58</td>
</tr>
<tr>
<td>z</td>
<td>0.61</td>
<td>0.18</td>
<td>0.50</td>
</tr>
<tr>
<td>Medial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>0.95</td>
<td>0.49</td>
<td>0.29</td>
</tr>
<tr>
<td>d</td>
<td>0.93</td>
<td>0.56</td>
<td>0.22</td>
</tr>
<tr>
<td>g</td>
<td>0.94</td>
<td>0.42</td>
<td>0.32</td>
</tr>
<tr>
<td>dʒ</td>
<td>0.93</td>
<td>0.49</td>
<td>0.08</td>
</tr>
<tr>
<td>v</td>
<td>0.84</td>
<td>0.29</td>
<td>0.60</td>
</tr>
<tr>
<td>ð</td>
<td>0.89</td>
<td>0.53</td>
<td>0.36</td>
</tr>
<tr>
<td>z</td>
<td>0.76</td>
<td>0.17</td>
<td>0.41</td>
</tr>
<tr>
<td>ʒ</td>
<td>0.87</td>
<td>0.18</td>
<td>0.56</td>
</tr>
<tr>
<td>Final</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>0.91</td>
<td>0.40</td>
<td>0.19</td>
</tr>
<tr>
<td>d</td>
<td>0.95</td>
<td>0.52</td>
<td>0.22</td>
</tr>
<tr>
<td>g</td>
<td>1.00</td>
<td>0.57</td>
<td>0.27</td>
</tr>
<tr>
<td>dʒ</td>
<td>0.87</td>
<td>0.29</td>
<td>0.32</td>
</tr>
<tr>
<td>z</td>
<td>0.88</td>
<td>0.45</td>
<td>0.26</td>
</tr>
</tbody>
</table>