Journal of Phonetics (1996) **24,** 423–444

Glottalization of word-initial vowels as a function of prosodic structure

L. Dilley and S. Shattuck-Hufnagel

Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

M. Ostendorf

Boston University, Boston, MA 02215, U.S.A.

Received 2nd March 1995 and in revised form 26th June 1996

Earlier work on the glottalization of word-initial vowels sought an account in terms of the morphosyntactic hierarchy and isolated facts about stress, without accounting for the possible role of phrase-level prosodic structure. More recent work based on prosodic theory (Pierrehumbert & Talkin, 1992; Pierrehumbert, 1995) has shown that prosodic constituents and prominence are important factors in the description of glottalization: word-initial vowels glottalize with higher frequency at the beginnings of intonational phrases, and to a greater degree if the word is pitch-accented. The work reported here extends Pierrehumbert & Talkin's findings using a different method, and reports additional dependencies on prosodic structure. The study is based on a different style of speech and a wider variety of prosodic and segmental contexts, in a speech corpus that was produced with communicative intent (FM radio news style, 3709 word-initial vowels produced by five speakers, three female and two male). Analysis of this corpus shows that glottalization of word-initial vowels is influenced by full *vs.* intermediate intonational phrase boundaries and pitch accent on the target syllable or word, but the effect of the preceding segmental context is small in comparison. These results are robust despite substantial interspeaker differences in both the distribution and nature of glottal onsets, and illustrate the value of revisiting earlier morphosyntactic analyses with prosodic structure in mind.

1. Introduction

A glottal gesture at the onset of a vowel-initial word, such as *issue*, *Arlene* or *about*, has been classified as optional allophonic variation in American English. Until recently, it was unclear in which sentential contexts this is most likely to occur. In fact, initial-vowel glottalization is one aspect of a more general phenomenon: glottalization also occurs at many other locations in the speech signal. It is manifested acoustically in a variety of ways (Batliner, Burger, Johne & Kießling, 1993), typically described as pitch periods that are irregular in timing or shape. In production, the vocal cords are generally thought to be more tightly adducted than

0095-4470/96/040423 + 22 $25.00/0

in the normal (modal) mode of vibration, and the glottal opening is smaller with a relatively longer interval of closure (Stevens, 1988). However, there may be other mechanisms involved. For example, Pierrehumbert & Talkin (1992) suggest a braced vocal fold configuration resulting in irregular vibration without full closure.

Glottalization or laryngealization occurs in a number of contexts in American English, ranging from a single glottal closure accompanying a consonantal segment to a change in voice source characteristics over a region encompassing several segments or even syllables. In phrase-final position, glottalization (sometimes termed ''creak'') has often been described and, although there are few quantitative studies, the results of Henton & Bladon (1988) confirm the common belief that creak occurs in sentence-final position (for Received Pronunciation and Modified Northern Accents of British English). Other suggested loci of glottalization are at syllable-final /t/ and sometimes /p/ (Pierrehumbert, 1994), and where a word boundary occurs between two vowels (Umeda, 1978; Gimson, 1980).

Word-initial contexts may be a particularly likely location for glottalization to occur. For example, in German, Kießling, Kompe, Niemann, Nöth & Batliner (1993) found that word-initial laryngealizations are frequent (58% of tokens studied), and Kohler (1994) reported a high probability of glottal onsets for vowel-initial morphemes internal to polymorphemic words as well. Thus, an understanding of the factors that influence glottalization at word-initial vowels may be important to understanding the phenomenon as a whole. This study focuses on prosodic and segmental context as factors that may influence glottalization of vowels in word-initial position.

Earlier analyses, carried out without the benefit of a theory of prosodic structure, seemed to show that glottalization in this position is unpredictable. Several aspects of utterances were investigated as possible determining factors, including tongue heights of the vowels in a vowel-vowel sequence across a word boundary (Allen, 1970), grammatical function of the target word (Spickenagel, 1974), and frequency of occurrence of the target word (Umeda, 1978), but a large range of examples was not accounted for.

Although these early studies were carried out before explicit grammars of prosody flowered in the 1980's, they sometimes hinted at the role prosodic structure might play. For example, Umeda noted: '' . . . In complicated material, a glottal stop, like a pause, seems to serve as a good boundary marker between two grammatical or semantic units'' (Umeda, 1978, p. 92).[1] Since the boundaries of grammatical and semantic units are likely (although not necessary) locations for intonational phrase boundaries, this remark presages, in morphosyntactic terms, later findings based on prosodic structure. Umeda also hinted at the role of pitch accent in determining whether a syllable becomes glottalized or not. She found that, in general, rare words are more frequently glottalized than common words, but noted that the words *all*, *only*, and *other,* although common, were also glottalized. She noted that this is ''probably because these words are often used rather emphatically'' (Umeda, 1978, p. 90), which we interpret as accented. Hirschberg (1993), Ross, Ostendorf & Shattuck-Hufnagel (1992), and others report that quantifiers are among the most likely of all words to receive pitch accents. Thus Umeda's formulation, although not

[1] Umeda looked at occurrences of full glottal stops only; the present study included other glottal gestures as long as they met the perceptual and acoustic criteria described in Section 2.

expressed in prosodic terms, is compatible with the view that prosodic factors influence the occurrence of glottalization in word-initial vowels. Gimson (1980, p. 169) discusses both prominence and segmental context as determining factors, suggesting that accent plays a role in glottalization: ''a hiatus of vowels belonging to different syllables (especially when the second syllable is accented) may in careful speech be separated by a [ʔ] instead of being joined by a vocalic glide'' and that ''any initial accented vowel may be reinforced by a preceding glottal stop when particular emphasis is placed on the word, whatever the preceding sound''.

More recent work has provided quantitative evidence for a strong relationship between prosodic structure and glottalization of word-initial vowels (Pierrehumbert & Talkin, 1992; Pierrehumbert, 1995), in studies over a limited range of segmental and prosodic contexts using controlled stimuli. Pierrehumbert & Talkin (1992) showed that when a vowel-initial word is situated at the beginning of an intonational phrase, the vowel has a much higher likelihood of being glottalized than in other locations, for both stressed and unstressed (reduced) vowels. Their results also suggested but did not firmly establish that pitch accent was a factor. In later work, Pierrehumbert (1995) found clear differences in the degree of word-initial glottalization between nuclear accented syllables and post-nuclear de-accented syllables.

The study that is reported here confirms and extends these results on a different speech style and in a wider range of prosodic and segmental contexts. The studies by Pierrehumbert and colleagues looked at two vowel-vowel segmental contexts: /ə#ɔ/ and /ə#ə/ contexts (as in . . . *Emma*, *August* . . . and . . . *lima abundance* . . .) in the 1992 study, and /-*ri*#/ preceding different vowel-initial words (as in . . . *watery oatmeal* . . .) in the 1995 study. Earlier work (Gimson, 1980) had suggested that these contexts favor the occurrence of glottal onsets for word-initial vowels. In addition, discourse context was selected to ensure that the relevant portions of all the target utterances were produced with the same prosodic pattern, i.e., with a low pitch accent (L*) on the target syllable in the 1992 study and a high intonation context (H*H-) in the 1995 studies.[2] In contrast, the present analysis included instances of word-initial vowels in a range of segmental environments that occurred in a database of continuous speech which had been produced with communicative intent (FM radio news stories), and which was not restricted to a particular set of segmental or prosodic contexts. The speech was prosodically labeled with pitch accents by syllable and with intonational phrase boundary markers (as part of an ongoing database project), and inspection of these labels showed that the corpus displayed a wide variety of prosodic as well as segmental contexts. In addition to extending earlier findings about the importance of prosodic structure and prominence in determining glottalization patterns to more general contexts, the work reported here also investigates the role of prosodic structure in more detail.

Phonetic context itself has also been suggested as a factor that affects the rate of glottalization. For example, Pierrehumbert (1994) looked at glottalization of word-final stops /p/ and /t/ followed by a variety of sonorant and non-sonorant consonants, and found that a final stop was more likely to be glottalized if it was

[2] In their 1994 study, Pierrehumbert & Frisch note that the low pitch accent is not ideal for determining whether or not glottalization is influenced by phrase onsets and pitch accents, since low f_0 is itself conducive to glottalization.

followed by a sonorant consonant, as for /t/ in *print-maker*. For the phenomenon we are looking at—glottalization of word-initial vowels—earlier work suggested that the presence of a preceding vowel is associated with a higher rate of glottalization of the target onset vowel (Umeda, 1978; Gimson, 1980). In German, Kohler (1994) investigated differences for five classes of preceding contexts in the incidence of glottal stop and/or glottalization of the word-initial vowel. The study described here further explores the effect of preceding segmental context, investigating seven classes of sounds preceding the target syllable, including the effect of a preceding pause and/or glottalization in the preceding segment.

In summary, the questions addressed in this study, which is focused on the glottalization of word-initial vowels, include:

1. Is the finding reported by Pierrehumbert & Talkin (1992), that word-initial vowels are more likely to glottalize when they occur at prosodically significant locations (i.e., at full intonational phrase onsets and in pitch accented words), confirmed in this corpus of continuous communicative speech?
2. Do full intonational phrase boundaries affect the glottalization rate differently from intermediate intonational phrase boundaries?
3. Does a pitch accent on the target vowel affect the glottalization rate differently from a pitch accent later in the word?
4. Does the immediately preceding segmental context and/or preceding pause or glottalization influence the glottalization rate for word-initial vowels?

The rest of the paper is devoted to exploring these questions, as well as addressing new issues which arose in the course of research, such as interspeaker differences in both the rates and acoustic manifestations of glottalization. Section 2 includes a description of the corpus, the prosodic and segmental elements which were analyzed, and the criteria for what was labeled as ''glottalization''. Findings with respect to the influence of the elements of prosody on glottalization, and of preceding segmental contexts, are presented in Section 3. The implications of these findings for speech science and engineering are summarized in Section 4, together with a discussion of the issue of individual speaker differences. The paper concludes in Section 5 by summarizing the results and noting some unresolved questions.

2. Method

2.1. *Corpus*

This study takes advantage of an existing prosodically-labeled database of FM radio news speech collected at Boston University (Ostendorf, Price & Shattuck-Hufnagel, 1995). Five professional radio news broadcasters, three female (f1a, f2b, f3a) and two male (m1b, m2b), were recorded as they read radio news reports on the air or in the laboratory. Four news stories were read by all five speakers; these had been broadcast earlier, transcribed from the broadcast and were read aloud in radio style in the laboratory. Additional stories spoken by f2b and m1b were written ahead of time by the speaker and recorded during a live radio broadcast. The labels ''a'' and ''b'' associated with speaker identifiers indicate whether the announcer's job typically involved reading stories off the wire or composing them ahead of time, respectively. The speakers were adults, 25 to 40 years of age, and with no striking

regional accent in the data we analyzed. The speech was digitized in paragraph-sized files using a 16 kHz sampling rate (16 bits), orthographically transcribed by hand, and automatically aligned to give a phonetic transcription of the text using a speech recognition system constrained to the correct transcription and a multiple-pronunciation lexicon that allowed some allophonic variation.

The digitized and aligned paragraphs were prosodically labeled by either one or two transcribers, using the ToBI (Tones and Break Indices) system of prosodic labeling (Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert & Hirschberg, 1992; Pitrelli, Beckman & Hirschberg, 1994). The ToBI labeling system, which captures both prosodic constituent boundaries and tonal markings of prominences and intonational phrases, is based on the work of Pierrehumbert (1980) and Beckman & Pierrehumbert (1986) on intonation, and of Price, Ostendorf, Shattuck-Hufnagel & Fong (1991) on break indices between pairs of adjacent words.[3] A consistency study was carried out on a subset of the data to determine the degree of cross-transcriber agreement on prosodic labeling for the two sets of transcribers. The study showed that the agreement between labelers was very good: 95% on phrase boundary labels (within ±0.5 out of 5 levels), 90% for presence *vs.* absence of a pitch accent, 81% for type of pitch accent (out of 5 labels, merging the H* and L + H* categories as in (Pitrelli *et al.,* 1994)), and over 95% for type of phrase accent or boundary tone.

When the digitized speech had been prosodically labeled, instances of words which were lexically vowel-initial were identified from the orthography (e.g., *economic*, *associated*, *artist*). Initially, 3915 words were identified in 186 prosodically labeled paragraphs.[4] The results reported here are based on a subset of 3709 target words, which excluded: i) cases of ambiguous glottalization, according to the criteria described in Section 2.2.; and ii) any words with ambiguously-realized lexical stress on the target syllable, as described in Section 2.3. This subset represents 95% of the original total. The distribution for the different speakers is summarized in Table I,

TABLE I. Distribution of data across the different speakers used in this study. The fourth column gives the number of word-initial vowel tokens for each speaker used in the study, and the fifth column gives the number of tokens in the subset that were not preceded by either a pause or a glottalized segment

Speaker	Paragraphs	Words	Tokens analyzed	No preceding pause/glot
f1a	22	1882	468	406
f2b	59	4956	1169	907
f3a	23	2031	527	461
m1b	58	4201	1013	932
m2b	24	2098	532	452
Total	186	15 168	3709	3158

[3] A detailed description of the ToBI system and the transcription software used in this project are available by anonymous ftp from kiwi.nmt.edu; the software is based on the Waves+ software package from Entropic Research Laboratory, Inc.

[4] The total number of word-initial vowels for each speaker before eliminating tokens was 502 (f1a), 1237 (f2b), 541 (f3a), 1079 (m1b) and 556 (m2b).

which also indicates the number of tokens that was not preceded by either a pause or a glottalized segment, since this second subset was used in some of the analyses.

Although the ToBI prosodic labeling effort was conducted independently of this study, much of the data used in this study was ToBI-labeled by the same person who did all of the glottalization labeling. Ideally, one would want different people labeling prosodic structure and glottalization. To safeguard against possible biases in labeling glottalization, a second listener confirmed perceptions of glottalization and examined the waveform of any questionable glottalizations, and tokens that were not clearly + or − glottalized were eliminated. Confidence in the efficacy of this method is increased by the observation that in a separate smaller study, the same global patterns of glottalization at prosodically significant locations are observed for data that were prosodically labeled by one person and labeled for glottalizations by another.

2.2. *Criteria for glottalization*

The criteria used to determine a labeling as glottalized or not glottalized consisted of two ordered requirements. First, a salient perceptual impression of a glottal gesture was required. Second, all those cases perceived as glottalized were subsequently examined more closely, and only those with an irregularity in the speech waveform were labeled as glottalized. This procedure proved more difficult to implement than expected, resulting in the more detailed specification given below; additional discussion appears in Section 5.

Glottalizations were initially noted by one listener, who marked the entire corpus. For cases in which some question arose (5.0% of the full set of vowel-initial tokens), a resolution process was developed that involved a second listener. If perception of glottalization was weak, yet visual evidence of pitch period irregularity was present, or if perception of glottalization was accompanied by only weak visual evidence of irregular pitch periods, then a final decision was reached by consultation between the two listeners. Cases for which it was difficult to determine the glottalization status even after repeated listening by both labelers were categorized as questionable, and were not included in the set of tokens that were analyzed (about half of those initially marked as questionable). The percentage of tokens that were discarded as questionable represented only 2.6% of the candidate tokens (1–5% of total vowel-initial word tokens for any one speaker).

Irregularity in the acoustic waveform could manifest itself in several ways. By far the most common in this speech sample was highly irregular spacing of pitch periods in the acoustic waveform, as seen in Fig. 1. Typically, the irregular pitch period(s) are longer than the regular pitch period of the following vowel. Some (but not the majority) of these cases with irregular pitch periods showed the long uninterrupted decay times characteristic of a glottal stop, as illustrated in Fig. 2. Others showed irregular pitch period durations without these long decay times. For still others, the waveform shapes of successive pitch periods were irregular, but these tokens always showed irregular pitch period durations as well.

Additionally, a relatively rapid dip in fundamental frequency (f_0) appeared to be an adequate cue for the perception of glottalization in tokens where periodicity was

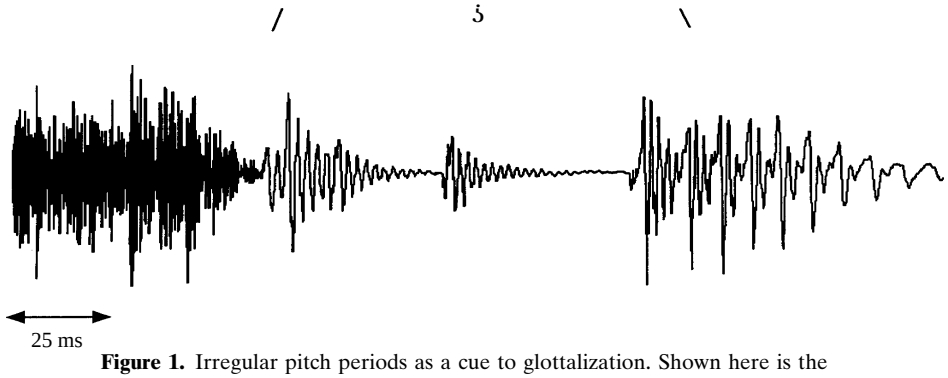

Figure 1. Irregular pitch periods as a cue to glottalization. Shown here is the middle portion of *justice ʔ of* where "ʔ" indicates glottalization.

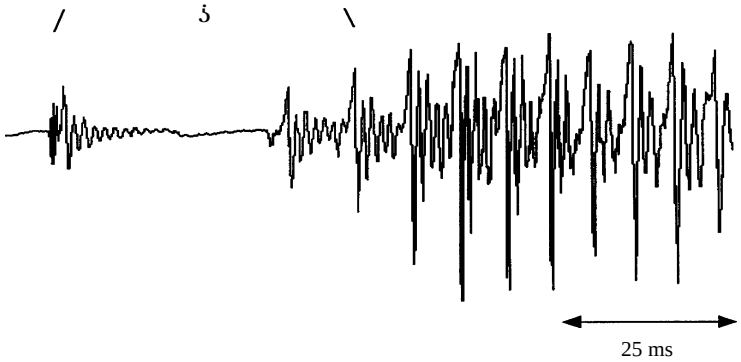

Figure 2. Example of a single glottal pulse followed by irregular pitch periods at the onset of a word-initial vowel. Shown here is *ʔE-* (from *Governor Edward*) where "ʔ" indicates glottalization.

not interrupted, e.g., between two vowels. Evidence from other studies (Houde & Hillenbrand, 1994; Pierrehumbert & Frisch, 1994) indicated that a dip in f_0 is sufficient to signal the presence of a glottal stop in synthesized speech. Fig. 3 shows a lower f_0 in the region perceived as glottalized, in comparison with previous and

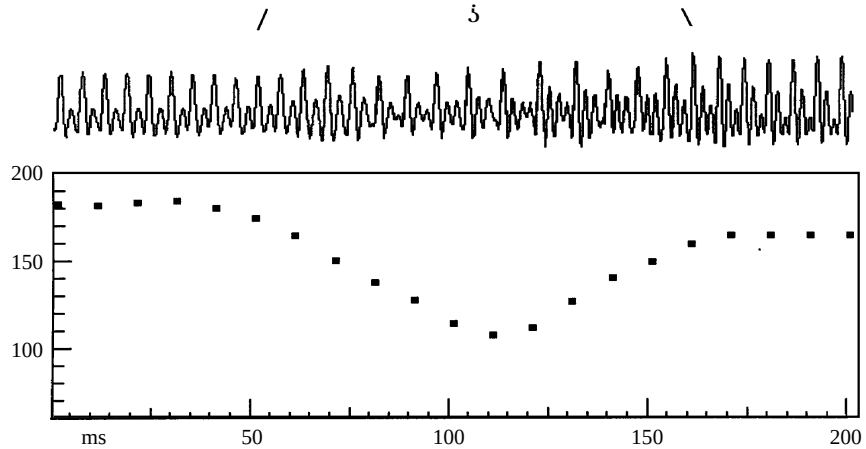

Figure 3. Dip in f_0 between two vowels results in perception of glottalization. Shown here is a portion of policy *ʔ is*, where "ʔ" indicates glottalization.

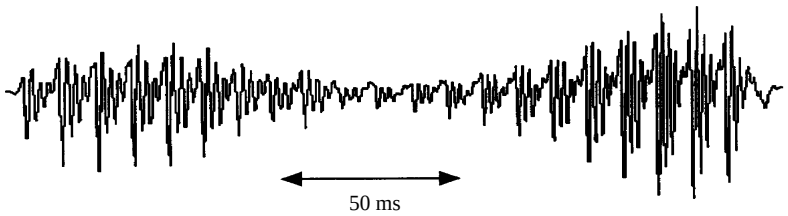

Figure 4. Reduction in amplitude on a word-initial vowel seems to signal glottalization. Shown here is a portion of *leader ɔ Edward* where "ɔ" indicates the "virtual glottalization".

following segments. Such cases were counted as glottalized if they met the perceptual criterion, since they also met the criterion of a perturbation in pitch period duration. A parallel case was observed when the preceding context was not voiced: f_0 was sometimes lower in the region perceived as glottalized than it was in the following portion of the vowel. These tokens, too, were counted as glottalized, because they exhibited the same qualities of perturbed (i.e., lower) f_0 with respect to adjacent context as did vowel-vowel cases like the one illustrated in Fig. 3. Cases where a region of lower f_0 was the only visually apparent cue to glottalization were rare.

One speaker presented a particularly interesting problem with respect to our criteria for glottalization. This speaker, m1b, had a tendency to mark the onsets of initial vowels in a way that was both consistent and perceptually salient, but which showed up in the acoustic waveform as a dip in amplitude, with little if any pitch period irregularity. Houde & Hillenbrand (1994) refer to these events as "virtual glottalizations". One such token is shown in Fig. 4. These tokens gave a strong perceptual impression of some kind of marking, but did not meet the criterion of having irregularly spaced pitch periods. We therefore called them questionable, because there was some acoustic evidence in the form of the amplitude dip. Most of the tokens in the corpus which fell into this class were produced by speaker m1b, and represent 2.7% of the total candidate tokens he produced and over half of the m1b tokens marked as questionable.

2.3. *Analysis factors*

As noted above, the likelihood of a glottal onset for the target syllables was analyzed in terms of both prosodic and segmental contexts.

2.3.1. *Prosodic contexts*

Both syllables labeled as glottalized and those labeled as non-glottalized were categorized according to their prosodic context in the following ways:

Position in the intonational phrase. Syllables were initially marked as +/− phrase-initial at the level of the intermediate intonational phrase. Those syllables which were +phrase-initial were later divided into two additional categories, depending

depending on whether or not the intermediate phrase was the first one in a full intonational phrase (and thus initial in both types of phrase). The distinction between full and intermediate intonational phrases was proposed by Beckman & Pierrehumbert (1986), and corresponds to the presence *vs.* absence of a final boundary tone on the last syllable of the phrase, as well as a deeper *vs.* shallower boundary (4 *vs.* 3 in the ToBI break indices), respectively.

Presence of pitch accent on the target syllable or word. Each syllable in the corpus had been previously labeled as +/− pitch accented. Instances of syllables marked with uncertainty (possible pitch accent *?, which is permitted in the ToBI labeling system and was used for less than 5% of the pitch accents) were included in the category ''+pitch accent on the target syllable.'' For those syllables which were ''−pitch accent on the target syllable'', there was an additional category of ''+/− pitch accent later in the word''. This last category includes word-initial reduced vowels (e.g., the first vowel /ə/ in *attorney*), which by definition cannot be pitch-accented, as well as full vowels (e.g., the first vowel /ɛ/ in *education*) which did not happen to be pitch accented.

Realized lexical stress. Each target syllable was labeled with +/− full vowel, according to its actual pronunciation in the utterance. Some words, e.g., *adult*, can be produced either with a lexically-stressed full vowel in the first syllable [ædʌlt], or with a reduced, non-stressed first syllable [ədʌlt]. In other examples, optional reducibility is not directly related to the placement of lexical main stress, e.g., *environmentally*, where the initial *en-* may be produced with a full or a reduced vowel. Because of this variability, it was necessary to label each target syllable perceptually as to whether it had been realized as a full vowel or reduced vowel; cases of ambiguously-realized lexical stress were omitted from the study. A total of 2.8% of all possible word-initial vowels were discarded because of ambiguous lexical stress, some of which may have also been eliminated because of questionable glottalization.

2.3.2. *Segmental contexts*

The nature of the preceding segmental context was also taken into account. First, tokens were divided according to whether there was a pause (not a glottal stop) of at least 50 ms preceding the target vowel. (Streeter (1978) cited others to argue that pauses of 50 ms or more are used by listeners in syntactic disambiguation, and there were very few silent regions of less than 50 ms in our data.) Preceding pause contexts were separated out to test whether glottalization might be more frequent in this context as a consequence of irregular movement of the vocal folds as they begin to vibrate after a period of silence. For all the tokens not preceded by a pause, the segment immediately preceding the word-initial vowel was categorized according to the orthographic transcription as a stop, fricative, affricate, nasal, liquid, or vowel. (Earlier studies had suggested that vowel-vowel boundaries represent a more favorable context for glottalization.) An additional aspect of the acoustic context noted was glottalization of the preceding segment, in order to ensure that an apparent glottalized vowel was not due to delayed offset of glottalization from a

prior segment. Whether a preceding segment was glottalized was again determined perceptually and checked by a second listener if there was a question, as described earlier. The glottalization rates of the target syllables were tabulated separately for these different contexts to test the effect of the preceding sound on glottalization of a word-initial vowel.

3. Results

In analyzing the set of 3709 word-initial vowels, our first observation was that there were significant differences among the speakers in the rate and character of glottalizations. As noted earlier, speaker m1b differed substantially from the other four speakers in having a very low overall rate of glottalization. For the target word-initial tokens analyzed, speaker m1b had a 13% rate of glottalization, while speakers f1a, f2b, f3a, and m2b had rates of 40%, 44%, 38%, and 24%, respectively.[5] Despite these cross-speaker differences, the general findings of Pierrehumbert & Talkin (1992) and Pierrehumbert (1995), that intonation-phrase-initial position and pitch-accent placement influence glottalization of word-initial vowels, were confirmed and strengthened in the data from each speaker. These results and additional findings with respect to more detailed prosodic structure are discussed below.

3.1. *Phrase-initial position*

All speakers glottalized significantly more often when the syllable occurred at the beginning of an intermediate or full intonational phrase, i.e., at a ToBI break of size 3 or 4. In order to confirm that the increase in glottalization rate was not simply explained by the presence of a pause or glottalization of the preceding segment, in Fig. 5 we give differences in glottalization rates for +/− phrase-initial conditions for the subset of tokens that were not preceded by either a pause or glottalization. Even in this restricted context, we observed a significantly higher rate of glottalization for phrase-initial syllables ($z > 3.78$, $p < 0.001$, $n > 406$ for m1b and stronger results for the other speakers).[6] As discussed later, the rate of glottalization increases when the syllable is preceded by either a pause or glottalization of the preceding segment, which occurs frequently at intonational phrase boundaries. As a consequence, when all tokens are taken into consideration, there is a much higher rate of glottalization in phrase-initial position. Particularly large increases are observed for f2b (48% for restricted context *vs.* 70% for all tokens) and m2b (39% *vs.* 51%).

In the results above, we grouped intermediate and full intonational phrases together to form the phrase-initial category, in contrast to the Pierrehumbert & Talkin (1992) study that considered only full intonational phrases. Thus, it is interesting to ask if there is a difference between full and intermediate intonational

[5] If the questionable glottalizations are counted as glottalizations, the difference between speakers is still substantial: 42%, 45%, 39%, 17%, and 26%, for f1a, f2b, f3a, m1b, and m2b, respectively.

[6] All statistical significance results reported here give numbers for the weakest case of the set of possible speakers and are based on the large sample Gaussian approximation of the binomial distribution for comparing two probabilities. Sample size for the smallest sample case is indicated by *n*.

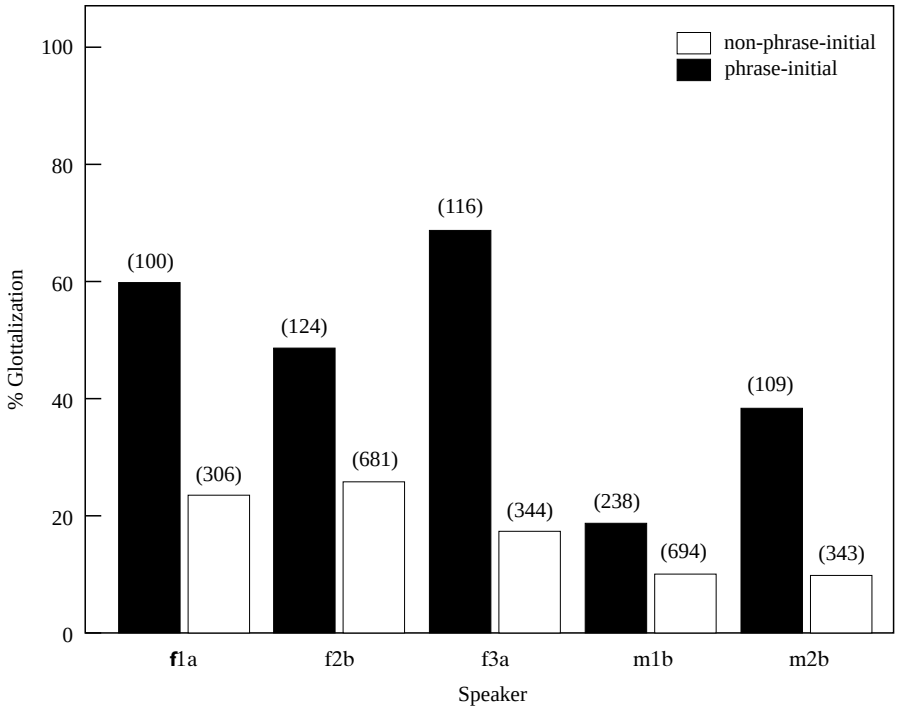

Figure 5. Glottalization rates (%) for each of the five speakers for words that were not preceded by a pause or glottalization, for +/− phrase-initial position. The number in parentheses indicates the total number of candidate tokens for each condition.

phrases and/or whether intermediate intonational phrases should be included in the phrase-initial category. For this reason, we separated three levels of prosodic constituents: non-phrase-initial position (ToBI break levels 0–2), initial in intermediate phrases (3), and initial in full intonational phrases (4). Including all tokens and averaging over all lexical stress conditions, there is a significant difference observed between all three levels: $z > 2.83$, $p < 0.003$, $n > 376$ for a difference between non-phrase-initial and intermediate-phrase-initial position for all speakers, and $z > 2.23$, $p < 0.013$, $n > 159$ for a difference between intermediate and full intonational phrases for all speakers except m1b. However, it appears that the difference between intermediate and full intonational phrases is due almost entirely to the reduced vowel tokens. If we control for full *vs.* reduced vowel status, the difference for reduced vowels is significant for each speaker if all tokens are included ($z > 2.47$, $p < 0.007$, $n > 60$). Similar trends are observed when the tokens with preceding pause or glottalization are omitted, as shown in Fig. 6, but most of the differences are not significant for individual speakers because of the smaller number of full intonational phrase tokens. For full-vowel tokens, with or without the tokens preceded by pause or glottalization, there is no difference between intermediate and full intonational phrases. A possible explanation for the lack of a difference in the full vowel tokens is the confounding effect of pitch accent, which cannot occur on reduced vowels. As shown next, phrase-initial syllables that are also pitch accented are highly likely to be glottalized. The remaining difference, between

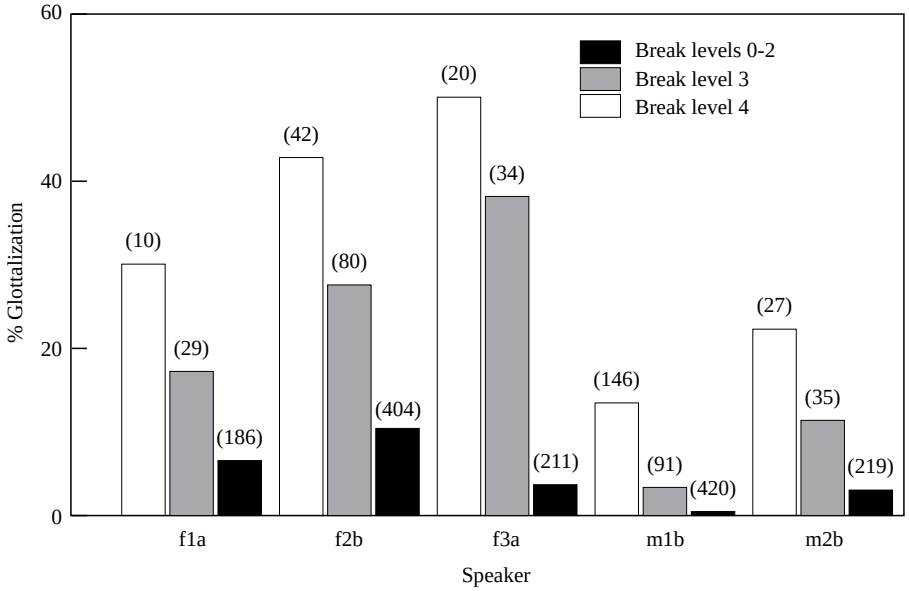

Figure 6. Glottalization rates (%) for each of the five speakers for words that begin with a reduced vowel and are not preceded by either a pause or a glottalized segment, and for three contexts: beginning of a full intonational phrase (4), beginning of an intermediate phrase (3), and non-phrase-initial (0–2), where the numbers indicate ToBI break indices. The number in parentheses indicate the total number of candidate tokens for each condition.

non-phrase-initial and intermediate-phrase-initial position, is observed for both the full and reduced vowel subsets whether or not the tokens preceded by a pause or glottalization are excluded: $z > 1.795$, $p < 0.036$, $n > 215$ for the reduced vowels, and $z > 2.01$, $p < 0.023$, $n > 152$ for the full vowels.

3.2. *Pitch accent*

Pierrehumbert (1995) showed that word-initial vowels in a vowel-vowel context are more glottalized when the syllable receives a nuclear accent. Here, we investigate whether word-initial vowels are more likely to be glottalized in accented words, where segmental context is unconstrained. Since phrase-initial position is such an important factor in predicting glottalization, we separated out that condition when looking at the effect of pitch accent. The results are summarized in Table II, which excludes the tokens preceded by either a pause or glottalization to more clearly indicate the effect of pitch accent in phrase-initial position. (In phrase-initial position, the glottalization rate is high for all contexts when preceded by a pause or prior glottalization.) In addition, we separated out unaccented full vowels from reduced vowels. For all speakers and in all conditions, the presence of a pitch accent increased the likelihood that a word-initial vowel would be glottalized. In phrase-initial position, the differences are significant when averaged over all speakers ($z = 3.60$, $p < 0.0002$, $n = 327$), but not for individual speakers except for f2b. The result was more striking for non-phrase-initial position, where the

TABLE II. Percentages (and counts) of tokens glottalized in different prosodic contexts for each of the speakers used in this study. Stress level is indicated by +/−F for full *vs.* reduced vowel, and +/−A for accented *vs.* unaccented syllable. Results are based on token syllables that were not preceded by a pause or glottalization

	−Phrase-Initial			+Phrase-Initial		
Speaker	−F, −A	+F, −A	+F, +A	−F, −A	+F, −A	+F, +A
f1a	**6%** (128)	**17%** (46)	**80%** (64)	**22%** (36)	**84%** (44)	**100%** (12)
f2b	**10%** (282)	**18%** (60)	**58%** (191)	**29%** (104)	**52%** (61)	**90%** (29)
f3a	**3%** (157)	**12%** (41)	**56%** (80)	**36%** (42)	**90%** (29)	**93%** (27)
m1b	**0%** (314)	**4%** (82)	**34%** (167)	**5%** (137)	**33%** (49)	**55%** (29)
m2b	**5%** (150)	**0%** (26)	**27%** (88)	**14%** (57)	**63%** (27)	**75%** (20)

differences are significant for each individual speaker ($z > 2.99$, $p < 0.0014$, $n > 110$). We also note that accent and phrase position appear to be dependent variables in the sense that the rate of glottalization cannot be predicted by simple multiplication or addition of the independent rates for these two factors, and in that these two factors are shown to be dependent for the glottalized cases using a Chi-squared contingency table test ($df = 3$, $p < 0.005$).

A second question that can be asked of the data in Table II is whether there is a difference between reduced vowel syllables and unaccented full vowel syllables. The answer is again yes for individual speakers, but in this case the evidence is stronger for the phrase-initial contexts ($z > 3.02$, $p < 0.0007$, $n > 71$). In non-phrase-initial position, there is no difference for speaker m2b, but a significant difference for all other speakers ($z > 1.86$, $p < 0.032$, $n > 174$). Although the reduced-vowel tokens have substantially lower glottalization rates than unaccented full-vowel tokens, glottalization of reduced vowels in phrase-initial position is still relatively common. However, even if we include all reduced-vowel tokens, the phrase-initial rates observed here—45%, 52%, 54%, 9%, and 30% for f1a, f2b, f3a, m1b, and m2b, respectively—are much lower than the over 90% rates observed in vowel-vowel contexts by Pierrehumbert & Talkin (1992) for their two speakers. This difference is probably due to speaker variability, the low tone context and/or the use of laboratory recordings, since the results discussed in the next section suggest that phonetic context is not a significant factor in phrase-initial context.

The results in Table II control only for presence of pitch accent on the target syllable, since other studies (e.g., Shattuck-Hufnagel, Ostendorf & Ross, 1994) suggest that pitch-accent patterns by syllable within the word are an important factor in the description of phrase-level intonational prosody. One might also ask, as did Pierrehumbert & Talkin (1992), whether the presence of a pitch accent later in the word increases the likelihood of glottalization, for example on the initial full-vowel

syllable *au-* in *augmenTAtion* or the initial reduced-vowel syllable *a-* in *aBUNdance,* where upper case signifies a pitch accent. The results from separating three classes of tokens—unaccented word, accent later in the word, and accent on the target syllable—show that the three classes differ. Overall (and for individual speakers with at least 10 tokens in a category) there was a consistent trend toward higher rates of glottalization for word-initial vowels where there was an accent later in the word relative to unaccented words, and still higher rates for word-initial vowels where the accent was on the target syllable. Since only a small number of tokens had an accent on the word but not on the target syllable, statistics were computed by combining data from all speakers. In contexts where glottalization is very likely (phrase-initial full vowels) or very unlikely (non-phrase-initial reduced vowels), the differences between the three cases are not significant. However, for other contexts (non-phrase-initial full vowels; phrase-initial reduced vowels) there were significant differences between all three categories: $z > 4.14$, $p < 0.00002$, $n = 669$ and $z > 3.06$, $p > 0.0002$, $n = 334$, respectively, for the subset of tokens excluding those preceded by a pause or a glottalized segment.

3.3. *Phonetic context*

Others have found that segmental context had an effect on the rate of glottalization, so we also examined the effect of the preceding segmental context on vowel-onset glottalization. Recall that we hypothesized that the presence of a pause or preceding glottalization might be associated with a higher frequency of glottalization in the target syllable. In fact, there is a much higher incidence of glottalization when either of these factors preceded the target word. Glottalization rates were 85% for tokens preceded by a glottalized segment but no pause, 64% for tokens preceded by a pause but no glottalization, 87% when preceded by both, and 23% when preceded by neither. Kohler (1994) found a similar increase in rate of glottalization in German for unstressed vowels preceded by a pause. There are two possible accounts for this observation, one mechanical and the other prosodic. It may be that the higher incidence of glottalization is simply due to mechanical constraints of starting a vowel after a pause and offset delay of cessation of preceding glottalization. Or, it could be that the preceding pause and glottalization are coincidentally present because of a phrase boundary (92% occur at phrase boundaries), and the vowel onset glottalization is simply associated with the presence of a phrase boundary. Although both factors could play a role, our data support at least the prosodic account, and the number of tokens in non-phrase-initial position is too small to resolve the issue of whether mechanical factors also account for glottalization of word-initial vowels.

The effect of the preceding pause and/or glottalization (which we interpret as a reflex of the prosodic boundary) is much larger than effects due to segmental context. For tokens preceded by only a pause or by only glottalization, where the number of tokens is small, the differences across segmental class are insignificant. For tokens that are not preceded by either a pause or a glottalization, there are small but significant differences among some segmental classes. This result suggested that segmental context might be a more important conditioning factor at phrase-internal word boundaries, where there are fewer cues (e.g., no pause or boundary tone) to word onset. To check this hypothesis, we computed the frequency

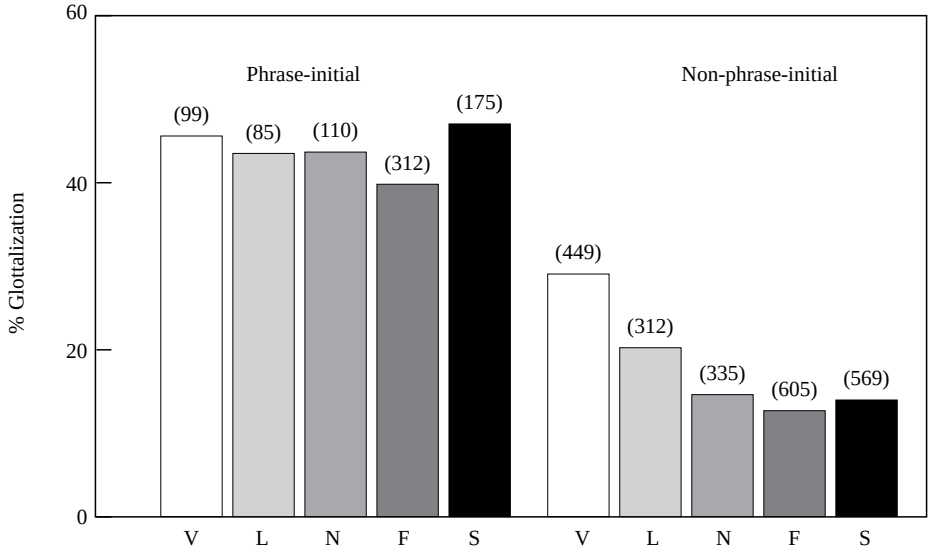

Figure 7. Glottalization rates (%) for different preceding contexts—vowel (V), liquid (L), nasal (N), fricative (F), and stop (S)—averaging over all five speakers, omitting tokens that are preceded by a pause or a glottalized segment, and separating +/− phrase–initial condition.

of word-initial vowel glottalization for different preceding segment classes, omitting all tokens preceded by a pause or glottalization and comparing phrase-internal *vs.* phrase-initial positions. As shown in Fig. 7, segmental context appears to be important primarily at phrase-internal word boundaries and has a smaller effect than does prosodic context. (The affricates are not included because too few were observed to provide meaningful statistics.) With the exception of a marginally significant difference between stops and fricatives ($z = 1.52$, $p < 0.064$, $n = 487$), the differences between phonetic classes in phrase-initial position are insignificant. In phrase-internal position, the only segmental classes that are significantly different are vowels and liquids ($z = 2.73$, $p < 0.0032$, $n = 761$ for vowel *vs.* liquid, $z = 1.87$, $p < 0.031$, $n = 647$ for liquid *vs.* other classes). Thus, our data are not inconsistent with the claim of Gimson (1980) and others that a preceding vowel context is conducive to glottalization, but we find that the differences in glottalization rates due to preceding segmental context are small relative to the differences due to prosodic contexts.

4. Discussion

The results of this analysis confirm and strengthen the observations of Pierrehumbert & Talkin (1992) and Pierrehumbert (1995), showing that glottalization of word-initial vowels is influenced by prosodic structure in continuous speech. Further, by distinguishing between more detailed classes of prosodic and segmental context, we obtained new results that have implications for research in other areas.

One area is the study of theories of intonational prosody, and the implications of these theories for cognitive processing representations. In particular, the difference

in rate of glottalization for reduced vowels at the onset of 4s *vs.* 3s strengthens Beckman & Pierrehumbert's (1986) proposal for a distinction between full and intermediate intonational phrases, and suggests that speakers represent these constituents during the speech production planning process. This finding raises the question of whether the marking of a new phrase onset in this way is an active process, or can simply be attributed to the influence of some other factor associated with phrase boundaries. For example, one might postulate that the higher rate of glottalization for vowels at full intonational phrase boundaries is due to mechanical factors associated with the higher rate of pauses and phrase-final creak at the offset of the preceding phrase. That is, glottalization might be a consequence of a delay in the cessation of phrase-final creak or attributable to the physiological requirements for beginning a vowel after a pause. However, mechanical factors alone cannot explain the results, since phrase-onset glottalizations are often present even in the absence of preceding glottalization or pause, and since differences based on preceding segmental context are minimal. Instead of a mechanical account, these findings support the hypothesis that speakers actively signal the onset of a new prosodic constituent, i.e., that glottalization may be one of several cues to phrase onset. A similar role for pitch accent is suggested by Shattuck-Hufnagel *et al.*'s (1994) analysis of early accent placement within the word, showing that this is more likely when the word carries the first pitch accent of a new intermediate intonational phrase. Taken together, these observations support the view that both full and intermediate intonational phrases play a part in the representations used by speakers to plan and produce utterances.

The glottalization of word-initial vowels can be thought of in terms of other "strengthening" phenomena that occur at word onsets when they occur at prosodically significant locations. For example, Pierrehumbert & Talkin (1992) found that both /h/ and glottalization of word-initial vowels are produced with greater gestural magnitude at prosodically significant locations. For /h/, "accentuation increases gestural magnitude, making . . . consonants more consonantal" (p. 109); "gestural magnitude was greater in phrase-initial position" (p. 111). For glottalization of word-initial vowels, their measure of the degree of pitch period irregularity showed a greater irregularity for phrase-initial tokens and for accented phrase-medial tokens, compared with unaccented phrase-medial tokens. A measure of aspiration duration for word-initial /t/ also revealed phrase-initial and accent-related increases. Similarly, Fougeron & Keating (1995) found increased amounts of tongue-palate contact for /n/ in phrase-initial position (and to successively smaller extents for constituents lower in the prosodic hierarchy). These findings, along with our own, suggest that glottalization of word-initial vowels at prosodically significant locations may represent a strengthening of the articulatory gesture associated with the onset of the prosodic constituent or prominence. The perceptual consequences of these articulatory differences remain to be explored.

The results of this study also have implications for computer speech processing. The connection between glottalization and prosodic structure could be utilized in speech synthesis to more clearly signal prosodic (and thereby some aspects of lexical and syntactic) structure, as shown in initial work by Pierrehumbert & Frisch (1994). Furthermore, the fact that glottalization occurs frequently at the beginning of an intonational constituent as well as at a pitch accent suggests that it could be exploited as a cue for recognizing prosodic patterns. The additional cues that glottalization provides could improve automatic detection of prosodic markers,

both for corpus transcription and for speech understanding applications where prosody is useful for interpreting sentence meaning and speaker intent.

We note that further work will be necessary in order to distinguish several different aspects of the role of low f_0 in eliciting glottalization of onset vowels. Effects of low tone can arise in one of three ways: from a preceding low pitch accent or boundary tone, from a low tone on the target vowel itself, or from a following low tone. With respect to preceding low context, anecdotally, we have seen patterns that support Pierrehumbert & Frisch's (1994) remark that low tone configurations are often realized with final creak. However, when tokens with low-tone-induced creak in the preceding segments were removed, the effects of prosodic structure were still observed. With respect to low tones on the target syllable, which Pierrehumbert & Frisch (1994) note are conducive to glottalization, some of the accented tokens may have been glottalized because of association with an L* or L* + H pitch accent, rather than with pitch accent *per se.* However, we are confident that the prosodic structure constraints on glottalization of word-initial vowels are not dependent on the presence of low target tones, because L* accents were rare in this corpus. With respect to following low-tone context, we sometimes observed that speakers creaked on an unstressed syllable in the last word of a phrase that ended with L- or L%. If this creaky portion of the utterance included the target vowel of a vowel-initial word, it was marked as glottalized. Such cases, though infrequent, may account for some of the glottalized tokens observed in unexpected contexts (e.g., non-phrase-initial reduced vowels). Thus, while the relation between low tonal targets and laryngealizations needs further study, the prosodic structure constraints we report here are unlikely to be accounted for as the consequence of low f_0.

The analyses reported here revealed profound differences among individual speakers in overall rates of glottalization (which ranged from 13% to 44%) and in the magnitude of the effect of different factors that influence the glottalization of word-initial vowels. Despite these differences, the connections between glottalization and several aspects of prosodic structure were significant for all five speakers and are further supported by a similar study of word-initial vowels in spontaneous speech. In that study, we examined glottalization of word-initial vowels in about 10 minutes of spontaneous dialogue between two speakers taken from the Trains corpus (Heeman & Allen, 1995). We found a similar distribution of glottalization at intonational phrase onsets and pitch accented syllables, as well as a striking difference between the two speakers in the rate of glottalization and in the acoustic correlates of these glottal events. For example, one of the speakers showed an associated breathy quality (Klatt & Klatt, 1990) and the other did not.

The considerable cross-speaker differences in the acoustic manifestation of glottalization for the FM radio news corpus are compatible with findings from other studies suggesting that individual speakers can vary substantially in the details of their laryngeal function. For example, Stevens (1994) reported individual speaker differences in the nature of the glottal waveform for different degrees of prosodic prominence. Similarly, Kießling *et al.* (1993) provided evidence for a strong speaker dependence in the acoustic form of what they term laryngealizations: their work on laryngealization detection in a corpus from a single German speaker did not transfer to a second speaker. These results, together with acoustic observations like those described below, underline the need for investigating glottalization patterns in more speakers.

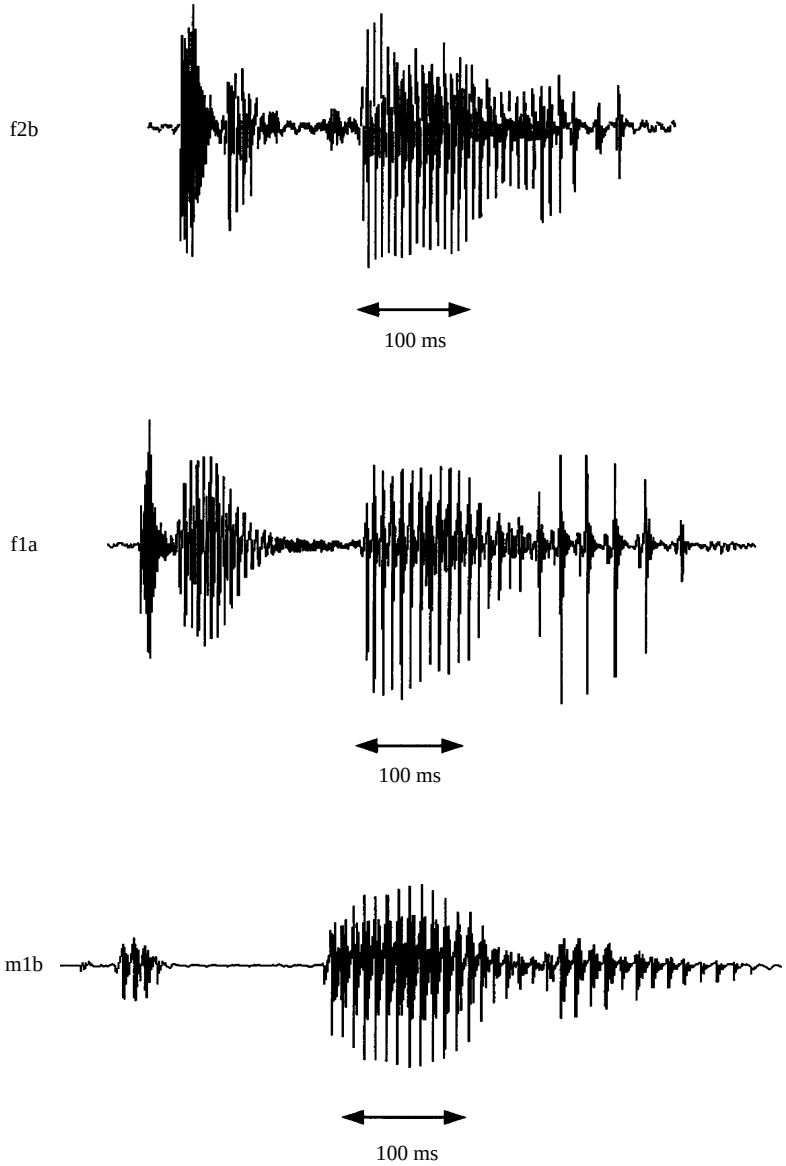

Figure 8. The final portion of three renditions of [*hard act*] *to follow* spoken by announcers f2b (top), f1a (middle), and m1b (bottom). For this paragraph-final phrase, speakers f1a and f2b glottalize the final syllable of *follow*, but m1b does not.

An illustrative example of how speakers from our corpus differed in their use of a related source variation at the ends of phrases is shown in Fig. 8. These examples show three different speakers producing the same phrase in a context in which laryngealized or ''creaky'' voice is commonly found: a phrase-final syllable in an L–L% intonational context (here in a paragraph-final phrase that ends with the words [*a hard act*] *to follow*). The female speakers shown in the top two panels exhibit creaky waveforms; speaker m1b, however, shows little sign of such irregular

pitch periods. As noted earlier, this speaker also showed a low rate of glottalization at vowel onsets, but he often marked a vowel onset with a substantial and extended dip in amplitude compared with the peak amplitude in the preceding and following vowels, as seen in Fig. 4. Speakers f1a and f2b seldom showed this pattern of amplitude dip at vowel onsets. From preliminary inspection of the waveforms, this contrast appeared quite general for these speakers: m1b often showed a marked decrease in amplitude in contexts where other speakers often glottalized.

Thus, as noted in the discussion of criteria for labeling vowel onsets, speaker m1b's manner of marking phrase offsets was consistent and perceptually salient, but was unlike that of other speakers in that it appeared to rely more heavily on changes in amplitude than on perturbations in periodicity. There is some precedent for this observation in the literature. For example, Houde & Hillenbrand (1994) and Hillenbrand & Houde (in press) reported that, for synthetic stimuli, either a dip in f_0 or a dip in energy at the onset of the two syllables in *oh-oh,* phonetically [ʔɔʔo] or [ʔʌʔo], is adequate to produce a rendition that listeners find acceptable. In a preliminary test of this hypothesis, a more detailed analysis was performed on a sample of the word-initial vowels for speaker m1b. These analyses revealed that in most cases one or more formants, usually F_2 and F_3 (though sometimes F_1) were dramatically reduced in amplitude by at least 5 dB for a short time period. It may be that the amplitude dip, with its associated sense of onset marking, arises from the same family of gestures as the irregular pitch periods associated with perceptible glottalization. Interestingly, for speaker m1b, 25 of the 29 tokens that were characterized by an amplitude dip alone occurred at the same prosodically significant locations as the tokens that met our criteria for glottalization, i.e., at intonational phrase onsets and at pitch accents. These results are reminiscent of the observations reported by Manuel, Shattuck-Hufnagel, Huffman, Stevens, Carlson & Hunnicutt (1992), showing that in words like ''support'', even when the first syllable was so reduced that there was no vowel-like periodic signal, there is an indication in the final portion of the /s/ frication that a glottal gesture toward voicing has occurred. Moreover, Manuel *et al.* showed that listeners can make use of this information to distinguish such utterances of *support* from utterances of *sport.* Apparently, even when the signal contains only impoverished cues to this glottal gesture, listeners can interpret them appropriately, raising the possibility that the same thing is true at the onset of vowels. This underlines the importance of further studies to determine precisely what is occurring at the larynx when each of these correlates is seen in the acoustic signal.

At issue are the questions of what occurs at the glottis when glottalization is perceived in different contexts and what are objective acoustic criteria for identifying different types of glottalization. For example, it is not clear whether the perturbations observed for glottalized portions of the vowel are caused by a weaker articulatory gesture of the same type that results in a glottal stop, or whether a different mechanism is involved. Pierrehumbert & Talkin (1992) found that for at least one of their speakers, there was little evidence to support the view that glottalization of word-initial vowels is produced by partial or complete adduction of the vocal folds. Instead, they propose a mechanism of bracing or tensing of partially abducted vocal folds, which results in irregular vibration without full closure during each glottal cycle. Previously proposed quantitative acoustic measurements of glottalization include pitch period jitter and the difference (or ratio) between the

first harmonic and either the second harmonic or the peak frequency of the first formant (Ladefoged, Maddieson & Jackson, 1988; Pierrehumbert & Talkin, 1992; Vayra, 1994), and several qualitative features (Batliner *et al.,* 1993). All researchers point out, however, that simple thresholds on these measures are unreliable when compared across vowel contexts or across speakers. Further research into the physiological mechanism which underlie the production of various types of glottalization should cast light on these questions.

The large variability across speakers in both rate of glottalization and its acoustic manifestation raises questions of how this variability might be accounted for, and there are multiple ways of handling the issue theoretically. At one extreme, the variability might be explained by a collection of rules that are simply undiscovered as yet. Or, it may be that glottalization is one of a collection of strategies that speakers use to mark prosodic events, which together are rule-governed but which may individually vary with importance across speakers. At another extreme, one could hypothesize that degree of articulation forms a continuum, and that different speakers operate at different ranges along this continuum. Variation is less problematic for computational models, which can handle differences in glottalization with probabilistic algorithms, just as pronunciation variations are currently handled.

5. Conclusions

In a larger and more diverse set of utterances and contexts, this study replicates Pierrehumbert & Talkin's (1992) observation that speakers are more likely to glottalize word-initial vowels when those vowels occur at the beginning of a new intonational phrase. Moreover, we showed that glottalization of word-initial vowels is more likely when the target word is marked with a pitch accent, extending the results of Pierrehumbert (1995). In addition, this analysis led to new findings that showed differences in glottalization rates at full *vs.* intermediate intonational phrase onsets, and differences for pitch accents on the target syllable *vs.* later in the word. Even reduced vowels in word-initial position are likely to be produced with a glottal onset if they occur at the beginning of a full intonational phrase, although not as often as full vowels. The presence of a preceding pause or glottalization strengthened the phrase boundary effect, but preceding segmental context was a significant factor only in the absence of a phrase boundary.

Besides supporting the active use of prosodic structure by human speakers in the planning of fluent communicative speech, the work reported here raises a number of questions for further investigation. For example, to what extent do listeners make use of the correlation between prosodic structure and glottal onset of word-initial vowels? What are the differences, if any, between the glottal gestures (and their acoustic correlates) associated with the onset of a word-initial vowel for pitch accents *vs.* prosodic phrase onsets and/or laryngeal gestures in other locations (e.g., the creak associated with the final syllable of an intonational phrase, or the glottal closure associated with a syllable-final voiceless stop, particularly /t/)? Finally, the observation that some speakers appear to mark the onset of a word-initial vowel that begins a new intonational phrase with a perceptible amplitude marker that is not obviously identifiable in the speech waveform with the acoustic shape of the glottal gesture employed by other speakers requires further investigation. A clear understanding of the articulatory mechanisms associated with the glottal marking of

various segmental and prosodic phenomena, and of the acoustic consequences of these mechanisms, would contribute materially to our understanding of glottalization patterns as a source of information about the structure of spoken utterances.

The authors wish to thank D. Talkin, the editor P. S. Beddor, and anonymous reviewers for valuable comments on the manuscripts. This work was supported by two grants from ARPA and NSF (NSF numbers IRI-8905249 and IRI-9314967) and an NIH grant (number DC0-2125).

References

Allen, J. (1970) The glottal stop as a junctural correlate in English *Journal of the Acoustical Society of America,* **47**, 57.

Batliner, A., Burger, S., Johne, B. & Kießling, A. (1993) Müsli: A classification scheme for laryngealizations. In *Proceedings of the ESCA workshop on prosody,* Working Papers 41, 176–179. Lund: Department of Linguistics and Phonetics.

Beckman, M. & Pierrehumbert, J. (1986) Intonational structure in Japanese and English, *Phonology Yearbook III,* 15–70.

Dilley, L. & Shattuck-Hufnagel, S. (1995) Variability in glottalization of word onset vowels in American English, in *Proceedings of the XIIIth international congress of phonetic sciences,* Stockholm, Vol. 4, 586–589.

Fougeron, C. & Keating, P. A. (1995) Demarcating prosodic groups with articulation, UCLA *Working Papers in Phonetics,* **92,** 61–87 and *Journal of the Acoustical Society of America,* **97** (S1), 3384.

Gimson, A. C. (1980) *An introduction to the pronunciation of English,* 3rd edition. London: Edward Arnold.

Heeman, P. A. & Allen, J. (1995) *The Trains* 93 *dialogues.* TRAINS Technical Note 94–2. Rochester, NY.: University of Rochester Computer Science Dept.

Henton, C. & Bladon, A. (1988) Creak as a sociophonetic marker. In *Language, speech and mind: studies in honour of Victoria A. Fromkin* (L. M. Hyman and C. N. Li, editors), pp. 3–29. London: Routledge.

Houde, R. A. & Hillenbrand, J. M. (1994) The role of voice pitch in the perception of glottal stops, *Journal of the Acoustical Society of America,* **95**(S1), 2872.

Hillenbrand, J. M. & Houde, R. A. (in press). The role of F_0 and amplitude in the perception of intervocalic glottal stops, *Journal of Speech and Hearing Research.*

Hirschberg, J. (1993) Pitch accent in context: predicting prominence from text, *Artificial Intelligence,* **63**, 305–340.

Kießling, A., Kompe, R., Niemann, H., Nöth, E. & Batliner, A. (1993) Voice source state as a source of information in speech recognition: detection of laryngealizations, NATO ASI.

Kohler, K. J. (1994) Glottal stops and glottalization in German, *Phonetica,* **51**, 38–51.

Klatt, D. & Klatt, L. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers, *Journal of the Acoustical Society of America,* **87**(2), 820–857.

Ladefoged, P., Maddieson, I. & Jackson, M. (1988) Investigating phonation types in different languages, In *Vocal physiology. Voice production*: *mechanisms and functions* (O. Fujimura, editor), 297–317. New York: Raven Press.

Manuel, S. Y., Shattuck-Hufnagel, S., Huffman, M., Stevens, K. N., Carlson, R. & Hunnicutt, S. (1992) Studies of vowel and consonant reduction, In *Proceedings of the international conference on spoken language processing,* pp. 943–946. Banff University of Alberta.

Ostendorf, M., Price, P., & Shattuck-Hufnagel, S. (1995) The Boston University radio news corpus, *Boston University Technical Report* ECS-95-001.

Pierrehumbert, J. (1980) *The phonology and phonetics of English intonation.* PhD Thesis, Massachusetts Institute of Technology.

Pierrehumbert, J. (1994) Knowledge of variation, invited talk at CLS 30. In *Papers from the 30th regional meeting of the Chicago Linguistics Society,* Chicago: University of Chicago.

Pierrehumbert, J. (1995) Prosodic effects on glottal allophones. In *Vocal fold physiology*: *voice quality control* (O. Fujimura and M. Hirano, editors), 39–60. San Diego: Singular Publishing Group.

Pierrehumbert, J. & Frisch, S. (1994) Source allophony and speech synthesis. In *Proceedings of the ESCA/IEEE workshop on speech synthesis,* 1–4.

Pierrehumbert & Talkin (1992) Lenition of /h/ and glottal stop. In *Papers in laboratory phonology II: gesture segment prosody* (G. Doherty and D. R. Ladd, editors), 90–117. Cambridge: Cambridge University Press.

Pitrelli, J. F., Beckman, M. E. & Hirschberg, J. (1994) Evaluation of prosodic transcription labeling reliability in the ToBI framework, In *Proceedings of the international conference on spoken language processing,* **1**, 123–126.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S. & Fong, C. (1991) The use of prosody in syntactic disambiguation, *Journal of the Acoustical Society of America,* **90**, 2956–2970.
Ross, K., Ostendorf, M. & Shattuck-Hufnagel, S. (1992) Factors affecting pitch accent placement, In *Proceedings of the international conference on spoken language processing,* **1**, 365–368. Banff University of Alberta.
Shattuck-Hufnagel, S., Ostendorf, M. & Ross, K. (1994) Pitch accent placement within lexical items in American English, *Journal of Phonetics*, **22,** 357–388.
Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. (1992) TOBI: A standard for labeling English prosody, In: *Proceedings of the international conference on spoken language processing,* **2**, 867–870. Banff University of Alberta.
Spickenagel, N. J. (1974) Rules for the insertion of glottal stops in synthetic speech, unpublished AT & T memo cited by Umeda (1978).
Stevens, K. (1988) Modes of vocal fold vibration based on a two-section model. In *Vocal physiology: voice production mechanisms and functions* (O. Fujimura, editor). New York: Raven Press.
Stevens, K. (1994) Prosodic influences on glottal waveform: preliminary data, presented at the international symposium on Prosody, Yokohama, Japan, 53–64.
Streeter, L. (1978) Acoustic determinants of phrase boundary perception, *Journal of the Acoustical Society of America,* **64**(6), 1582–1592.
Umeda, N. (1978) Occurrence of glottal stops in fluent speech, *Journal of the Acoustical Society of America,* **64**(1), 88–94.
Vayra, M. (1994) Phonetic explanation in phonology: laryngealization as the case for glottal stops in Italian word-final stressed syllables. In *Phonologica. Proceedings of the 7th international phonology meeting,* (W. Dressler, M. Prinzhorn and J. Rennison, editors), Torino.
.