VARIABILITY IN GLOTTALIZATION OF WORD ONSET VOWELS IN AMERICAN ENGLISH

Laura C. Dilley and Stefanie Shattuck-Hufnagel
Research Laboratory of Electronics, MIT, Cambridge, MA 02139, USA

ABSTRACT

American English glottalization of word-onset vowels at the beginning of a new intonational phrase or at a pitch accent holds across speakers and speaking styles, although individuals differ in glottalization rates and in sensitivity to prosodic elements, and manifest glottalization in a number of ways.

1. INTRODUCTION

Speakers of American English often glottalize in certain locations, e.g. at the ends of phrases ("final creak"), at a final voiceless stop (glottalized /t/) and at a word-onset vowel, as in "able" or "alone". For word-onset vowels, the likelihood of glottalization is increased significantly in certain prosodic contexts, e.g. when the vowel-initial word begins a new intonational phrase or is pitch accented [7]. For reduced vowels, glottalization is more likely when the new phrase is a full rather than an intermediate intonational phrase; for full vowels, it is more likely when the pitch accent occurs on the target syllable rather than later in the word [2].

Although prosodic boundaries and prominences constrain the glottalization of word-initial vowels in similar ways across speakers, there is also substantial variation among speakers [3]. For example, for 4 radio news broadcasters, overall glottalization rates were 41%, 37%, 22% and 7%.

This striking rate variation raises several questions. Do the differences arise in part from the use of different texts? Do the findings hold across speaking styles? Can individual differences in onset-vowel glottalization be related to other aspects of glottalization behavior? This paper compares glottalization rates for word-initial vowels for FM radio news speakers producing the same texts, for non-professional speakers reading isolated sentences, and for speakers producing spontaneous speech.

2. FM RADIO NEWS SPEAKERS

Since the use of different texts may partially account for the range of glottalization rates in Dilley et al. [2], we analysed a corpus in which all speakers produce the same texts, for evidence that a) speakers differ in their prosodic interpretation of those texts, and b) these differences might contribute to contrasts in overall glottalization rate. Four news stories were originally broadcast by one speaker and later read in the lab by all the speakers in FM news style. Renditions of all 4 stories were available for 5 speakers (3 female, 2 male) including the original 4.

2.1 Database and Analysis. Details of the BU FM Radio News corpus are described in Ostendorf et al. [6]. Briefly, the speech was recorded in the studio during broadcast, orthographically transcribed, phonetically aligned and labelled for part of speech and for prosody, using the ToBI transcription system [10], [8]. The ToBI (TOnes and Break Indices) system marks 5 levels of prosodic constituent boundaries, and, for each intonational phrase, location (by syllable) and type of pitch accents, location and type of phrase accent, location and type of boundary tone, and highest prominence-related F0. For this study, additional labels were added by hand to each vowel that began a word: phonetic lexical stress (Full Vowel vs. Reduced Vowel), and glottalization (+Glot vs. -Glot). The criteria for glottalization were two-fold: a perceptual impression of glottalization, and a marker in the wave form, usually an irregularity in pitch period duration (or a dip in F0), but occasionally in wave

form shape (e.g. diplophonia). Tokens which did not satisfy both of these criteria were labelled as -Glot, or as Questionable Glot e.g. if the boundary between the target vowel and a preceding glottalized segment was unclear, or only one of the two criteria was met. The four stories contained 573 vowel-initial words in about 5 minutes of speech per speaker. We noted the overall glottalization rate for word-initial vowels for each speaker, and analysed the effect of several types of prosodic environment. Results for tokens preceded by a syllable that contained a glottalized segment (often the result of phrase-final creak in the preceding phrase), and/or by a pause were analysed separately. Removed from the analysis were tokens with questionable phonetic stress (2% of total) and tokens with questionable glottalization (2%).

2.2 Results and Discussion. Rates of glottalization for word-initial vowels preceded by a pause > 50 ms or earlier glottalization were 94-100%, indicating that these factors are strongly associated with high glottalization rates. The number of tokens with such preceding contexts varied from a high of 88 (speaker f2) to a low of 19 (speaker m1), illustrating the fact that different speakers produced different distributions of pause and/or word-final glottalization for the same text.

Table 1 shows results for all 5 speakers for vowel tokens not preceded by a pause or creaky segment. Rates are shown separately for all tokens, for pitch accent and phrase-intial contexts, and for the remaining contexts.

Table 1. Glottalization rates overall (a), for pitch accent and phrase-onset contexts (b), and for the remaining contexts (c) for 5 FM radio news speakers.

%	f1	f2	f3	m1	m2
(a)	32	34	29	13	16
(b)	67	58	62	30	33
(c)	8	15	6	2	2

Individual overall glottalization rates for word-onset vowels range from 34% to 13%. In addition, glottalization is more likely at prosodic phrase boundaries and pitch accents, even for speakers m1 and m2, who have low rates of glottalization. ($p \leq .0001$)

Additional examples of speaker differences hint at a gender effect. For example, the 3 female speakers glottalized non-phrase-initial Full Vowels in pitch accented words significantly more often than in non-pitch-accented words ($p \leq .01$), but the 2 male speakers did not. Moreover, the 3 female speakers in Table 1 have higher rates (34%, 32%, 29%) than the 2 males (16% and 13%). To determine whether similar effects arise for other speakers, we turned to a speech database of isolated sentences read by non-professional speakers in the lab.

3. NON–PROFESSIONAL SPEECH

3.1 Database and Analysis. The LEX database consists of isolated sentences read aloud in the lab by 4 non-professional speakers (2 female, 2 male) and digitized. For the present study, we selected the sentences that contained word-onset vowels, and labelled the prosody and glottalization as above. The overall rates and the effect of prosodically significant contexts are shown in Table 2.

Table 2. Glottalization rates overall (a), for pitch accent and phrase onset contexts (b), and for the remaining contexts (c), for 4 non-professional speakers.

%	f-j	f-s	m-k	m-m
(a)	29	5	<1	13
(b)	67	12	3	43
(c)	6	**0**	**0**	**0**

3.3 Results and Discussion. Like the FM newscasters, these non-professional speakers show a range of overall glottalization rates, from less than 1% to 29%, and a tendency for Phrase Initial and Pitch Accent contexts to elicit more glottalization. Speaker m-k was the exception: his overall glottalization rate was so low that distinctions between prosodic contexts did not emerge. The pattern of lower glottalization rates for male speakers observed in the radio news speech is called into question: female speaker f-s showed one of the

lowest rates of glottalization, both overall (5%) and for prosodically significant contexts (12%), and both her rates were lower than those for male speaker m-m (13% and 43%), indicating that individual variation makes it difficult to assess gender-related differences with a small number of speakers.

4. SPONTANEOUS SPEECH

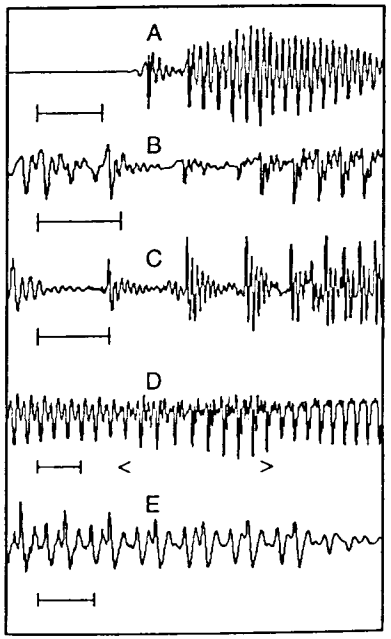
4.1 Database and Analysis. To determine whether the influence of prosodic context on glottalization rates is also found in spontaneous speech, we analyzed a set of utterances from the ATIS corpus, collected from travel agents enacting the task of making airline reservations using spoken language to interact with a computer [12]. We examined a subset of 155 utterances spoken by many different speakers, each containing a disfluency. The digitized utterances were aligned and labelled as above. Since each speaker produced just a few utterances, only descriptive results are given here. Tokens with a previous pause or glottalized syllable were removed, leaving 138 word-onset vowels for analysis.

4.2 Results and Discussion. The overall rate of glottalization for this corpus was 43%, with a greater likelihood of glottalization for pitch accented and/or phrase initial vowels (64% of 78 tokens) vs. other prosodic contexts (17% of 60 tokens). An interesting aspect of these utterances was the finding that, for reduced vowels, tokens with a pitch accent later in the word were more likely to be glottalized than tokens with no pitch accent on the word (43% vs. 5%). The effect of a pitch accent later in the word was also observed for a single FM radio news speaker by Dilley et al. [3], and suggests that the effects of an accent may extend beyond the boundaries of its immediate syllable (see alsoTurk [11]).

5. WAVE FORM SHAPES

A critical aspect of any study of glottalization rates is the decision about what to categorize as glottalization. Our two-part criterion, perceptual salience and pitch period irregularity, labelled a variety of waveform shapes as +Glot; some examples are shown in Fig. 1.

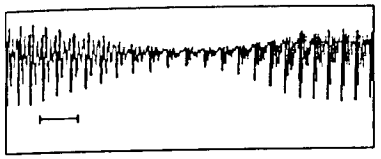
Figure 1. Examples of word-onset vowels heard and labelled as glottalized (FM radio speakers). A: glottal stop, B: general irregularity, C: ecponentially-decaying pulses, D: F0 dip and amplitude change (< > indicates region of lower F0), E: diplophonia. Indicator bar = 20 msec.



Excluded from this category was a kind of vowel onset which did not sound glottalized, yet gave the impression of a boundary or onset marker. This was most common for FM radio news speaker m1, where it was often associated with a dip in amplitude, without other noticeable irregularity in the wave form, e.g. Fig. 2. Houde and Hillenbrand [4] have reported that a dip in amplitude is a sufficient cue to elicit the perception of glottal onsets for the vowels in a synthesized version of the exclamation "oh-oh".

Another speaker, m-k, a non-professional, marked onsets in a perceptually salient way such that they sounded "breathy". We did not observe

Figure 2. Example of a dip in amplitude, heard as onset-like but not as glottalized (speaker m2)



irregularity in the period or shape of the waveform in these cases, but noise was present in the signal; moreover, these tokens had an /h/-like distribution of energy across frequencies. Tokens which were "breathy" were labelled -Glot. The fact that m-k appears to employ this marking strategy quite often instead of glottalizing contributes to his low overall rate of glottalization.

These observations suggest that a wide variety of signal characteristics and perceptually salient phenomena can occur at word-onset vowels, underlining the importance of a better understanding of the glottal mechanisms involved.

5. CONCLUSIONS

Analysis of word-onset vowel glottalization rates in American English, and their variation across individual speakers, prosodic contexts and speaking styles, shows that a) individuals vary substantially in overall glottalization rate as well as in use of prosodic structure, and b) a variety of signal shapes are perceived as glottalized in these vowels. Nevertheless, these preliminary analyses support the view that prosodic structure plays an important role in determining where vowel glottalization will occur. Somewhat similar findings have been reported for German by Kohler (1994). Further work will determine whether glottalization can provide useful cues to prosodic structure, not only as phrase-final creak but also as a marker for phrase onset and pitch accent, and whether appropriate use of glottalization can increase the naturalness and comprehensibility of synthesized speech.

ACKNOWLEDGEMENTS

This research was supported by NSF, NIH and ARPA, and by the MIT Undergraduate Research Opportunity Program. The original analysis of glottalization in FM radio news speech was carried out with Mari Ostendorf; the BU Radio News Corpus was constructed in collaboration with Mari Ostendorf and Patti Price; and the LEX database with Ken Stevens and Sharon Manuel. We gratefully acknowledge help provided by Melanie Matthies.

REFERENCES

[1] Beckman, M. and Pierrehumbert, J. (1986) Intonational structure in Japanese and English, *Phonology Yearbook* 3, 15-70

[2] Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. (1994) Prosodic constraints on glottalization of vowel-initial syllables in American English, *JASA* 94 (Pt 2), 2978

[3] Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. (submitted) Glottalization of vowel-initial syllables as a function of prosodic structure

[4] Houde, R.A. and Hillenbrand, J. (1994) The role of voice pitch in the perception of glottal stops, *JASA* 95 (S1), 2872

[5] Kohler, K. J. (1994) Glottal stops and glottalization in German, *Phonetica* 51, 38-51

[6] Ostendorf, M., Price, P. and Shattuck-Hufnagel, S. (1995) *The Boston University Radio News Corpus*, Boston University ECS Enginerring Report No. ECS-95-001

[7] Pierrehumbert, J. and Talkin, D. (1992) Lenition of /h/ and glottal stop, in *Papers in Laboratory Phonology II* (G. Doherty and D.R. Ladd, eds.), 90-117. Cambridge University Press.

[8] Pitrelli, J. F., Beckman, M.E. and Hirschberg, J. (1994) Evaluation of prosodic transcription labelling reliability in the ToBI framework, *Proc. Int. Conf. on Spoken Language Processing*,

[9] Shattuck-Hufnagel, S., Ostendorf, M. and Ross, K. (1995) Stress shift and early pitch accent placement in lexical items in American English, *J. Phonetics* 22, 357-388

[10] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. (1992), TOBI: A standard for labelling English prosody, *Proc. Int. Conf. Spoken Lang. Processing* 867-870

[11] Turk, A. and Sawusch, J.R. (forthcoming) The domain of the durational effects of accent, MIT Speech Group Working Papers 10

[12] Multi-Site Data Collection for a Spoken Language Corpus, MADCOW, Proc. DARPA Speech and Nat. Lang. Workshop Feb 1992 p. 7