# Is irregular phonation a reliable cue towards the segmentation of continuous speech in American English?

*Kushan Surana & Janet Slifka*

Massachusetts Institute of Technology, Cambridge, MA
Speech Communication Group, Research Laboratory of Electronics
{kushan@mit.edu; slifka@speech.mit.edu}

## Abstract

This paper analyzes the potential use of irregular phonation as a cue for the segmentation of continuous speech. The analysis is conducted on two dialect regions of the TIMIT database which consists of read, isolated utterances. The data set encompasses 114 speakers resulting in 1331 hand-labeled irregular tokens. The study shows that 78% of the irregular tokens occur at word boundaries and 5% occur at syllable boundaries. Of the irregular tokens at syllable boundaries, 72% are either at the junction of a compound-word (e.g "outcast") or at the junction of a base word and a suffix. Of the irregular tokens which do not occur at word or syllable boundaries, 70% occur adjacent to voiceless consonants mostly in utterance-final location. These observations support irregular phonation as an acoustic cue for syntactic boundaries in connected speech. Detection of regions of irregular phonation could improve speech recognition and lexical access models. [Work supported by NIH # DC02978.]

## 1. Introduction

A large body of research exists regarding the range of acoustic cues used to mark boundaries in the speech stream. These cues serve a segmentation purpose for various types of units — including syllables, words, phrases, utterances and dialogs. In American English, these cues include the aspiration of voiceless stop consonants in syllable-initial position, segmental lengthening prior to a major prosodic boundary such as the utterance, and signal amplitude changes in the vicinity of a silent pause as the speaker suspends the sound source. In particular, prior work has focused on specifying the factors which determine the likelihood that a word boundary will be marked with irregular phonation. In general, these factors may arise from a segmental context and/or a prosodic environment. For example, irregular phonation tends to occur at word boundaries between vowels [9], and at syllable final /t/ and sometimes /p/ [4]. The occurrence of irregular phonation at word-initial vowels and its relationship with the prosodic structure of the utterance has also been explored [1, 5]. Their studies show that irregular phonation at word-initial vowels occurs more often at the beginnings of intonational phrases, and to a greater degree if the word is pitch-accented.

As stated, these studies focus on determining the factors that influence the likelihood that a word boundary will be marked with irregular phonation. In this paper, we address a related question with a slightly different focus — given the presence of irregular phonation, what is the likelihood of a word boundary at that location? Similarly, if irregular phonation does not occur at a word boundary, in what context does it occur? The results directly support the use of automatically detected regions of irregular phonation in spoken language systems. First, these irregular regions can help determine the probability of a word-boundary location. Also, with limited additional context, the probability estimate for a word-boundary can be strengthened. Specifically, we examine two cases in more detail — voiceless stop consonants and vowel-vowel junctions.

We examine the occurrence of irregular phonation in relation to the likelihood that it will occur at word and syllable boundaries in connected speech for American English in a speaker-independent analysis on the TIMIT database, which is composed of isolated utterances. The ends of utterances (and phrases) have been observed to be marked with irregular phonation [2, 3]. Given the structure of the database as isolated utterances, we examine utterance-initial and utterance-final irregular phonation as well as syntactic level phrase-initial and phrase-final irregular phonation. The primary focus of the results in this study is in relation to syntactic boundaries. Other studies have examined the influence of prosodic boundaries on irregular phonation [6]. Their study shows a higher rate of irregular phonation at the ends of utterances than at the ends of utterance-medial intonational phrases. Although the relationship between prosody and syntax is not yet fully modeled, there should be considerable overlap between the two via the constraints syntax imposes on the choices that the speaker makes among the possible prosodic realizations for a given utterance.

## 2. Irregular phonation

Studies of irregular phonation span the fields of signal processing, linguistics and speech production. There are varying interpretations of what it means for phonation to be irregular and the term is, in general, used differently between researchers. The next few paragraphs clarify the use of the term in this paper.

Normal, voiced speech is characterized by quasi-regular vibration of the vocal folds. Although the vocal folds oscillate regularly in general, when the variables transglottal pressure, vocal fold tension, and vocal fold adduction — among others — are in particular ranges, irregularities in vocal fold vibration are observed for certain combinations of the values of these variables. These irregularities in vocal fold vibration are visible in the speech waveform and are more pronounced than the small cycle-to-cycle variations associated with the quasi-periodic quality of regular phonation.

Prior research regarding voice quality and phonation often use the terms *"modal"* and *"periodic"* interchangeably with *"regular"* phonation. Similarly, *"nonmodal"* and *"aperiodic"* are often used to denote *"irregular"* phonation. This paper avoids the use of these terms as they are not synonymous with regular or irregular phonation. For example, nonmodal phonation includes irregular, aperiodic phonation such as vocal fry as well as regular, periodic phonation such as breathy voice. Regions in the speech waveform with very low frequency, periodic glottal pulses are also not typical of the quasi-periodic pulses in the normal range of phonation for a given speaker and are classified as irregular in this study, in spite of being periodic.

Specifically, irregular phonation is defined as: **"A region of phonation is an example of irregular phonation if the speech waveform displays either an unusual difference in time or amplitude over adjacent pitch periods that exceeds the small-scale jitter and shimmer differences, or an unusually wide-spacing of the glottal pulses compared to their spacing in the local environment, indicating an anomaly with respect to the usual, quasi-periodic behavior of the vocal folds."** In general, jitter differences $< 1\%$ and shimmer values $< 0.5$ dB are considered normal.

Table 1: Boundary labels for irregular token occurrence

| Word level | Phrasal level | Stops | Other |
|---|---|---|---|
| Word-final | Utt-final | p | Vowel-vowel |
| Word-initial | Utt-initial | t | |
| Syll-final | Phrase-final | k | |
| Syll-initial | Phrase-initial | | |
| | Last phonation in utt. | | |
| | First phonation in utt. | | |

## 3. Data set

The irregular tokens were extracted from a subset of the TIMIT corpus (1990), a phonetically-labeled database of isolated utterances, recorded with a 16 kHz sampling rate. The database includes time-aligned orthographic, word, and phone transcriptions. In this study, a subset of the database is used — those utterances produced by speakers from the dialect regions "Northern" (dr1) and "New England" (dr2).

In the TIMIT database the phone label 'q' or glottal stop is used to label an allophone of /t/ or to mark an initial vowel or vowel-vowel boundary. The criteria for applying this label 'q' is not tied to the acoustic realization, as is the case in this study, and is not used to label all possible cases of irregular phonation. For these reasons, the irregular tokens were hand-labeled and extracted by analyzing the waveform in both the temporal and frequency domains to find regions which corresponded to the stated definition of irregular phonation. The resulting data set consists of utterances from 114 different speakers.

The word transcription of the TIMIT database was used to determine word and utterance boundaries. Regions of irregular phonation were classified in relation to the syntactic boundaries of syllable, word, phrase and utterance. Phone-related instances for /p/, /t/, /k/ and vowel-vowel sequences were classified using the TIMIT phonetic transcription. A summary of the classification categories is given in Table 1 for four categories — word level, phrasal level, voiceless stop consonants and vowel-vowel boundaries. Within the word-level and phrasal-level categories, the labeling is mutually exclusive. For example, a word-initial occurrence of irregular phonation is not counted as syllable-initial. Similarly, an utterance-initial occurence of irregular phonation is not marked as phrase-initial.

## 4. Results

Figure 1 shows the percentage, as well as the absolute number, of irregular tokens that occur at word and syllable boundaries. 78% of the irregular tokens occur at word boundaries — 45% at word-final locations and 33% at word-initial locations. An additional 5% of the irregular tokens occur at syllable boundaries. These tokens were re-analyzed, leading to two main observations. First, of the 69 irregular tokens occurring at syllable boundaries, 50 occurred (72%) either at the junction of a compound word (e.g. "outcast") or at the junction of a base word and a suffix. For example, irregular phonation was noted at the end of 'equip' in 'equipment'. Secondly, 52 of the 69 irregular tokens (75%) at syllable boundaries conincided with a voiceless stop location (either /p/, /t/ or /k/).

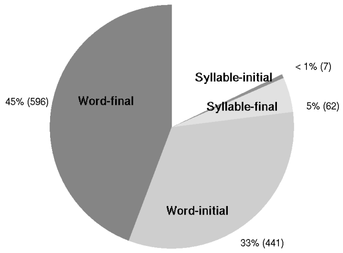

Figure 1: *Breakdown of irregular phonation at word and syllable boundaries. The absolute number is shown next to the percentage within brackets. (Based on 1331 tokens)*

Figure 2 shows the percentage, as well as the absolute number, of irregular tokens at phrasal boundaries. Combined, 48% of the irregular tokens occur at phrasal boundaries — 27% at utterance boundaries while another 21% at syntactic phrase boundaries. In Figure 2, the irregular tokens occurring at the last place of phonation within the utterance are combined with the utterance-final tokens. Similarly, the irregular tokens at the first place of phonation within the utterance are combined with the utterance-initial tokens.

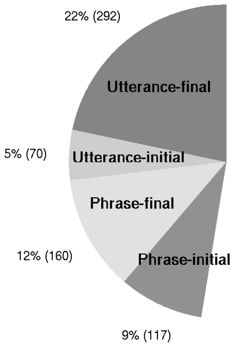

Figure 2: *Breakdown of irregular phonation at syntactic phrase and utterance boundaries. The absolute number is shown next to the percentage within brackets. (Based on 1331 tokens)*

Figure 3 shows the percentage, as well as the absolute number, of the irregular tokens that occur at voiceless stop /p/, /t/ or /k/ and at vowel-vowel junctions. 24% of the irregular tokens occur at voiceless stop consonants and 10% occur at vowel-vowel junctions.

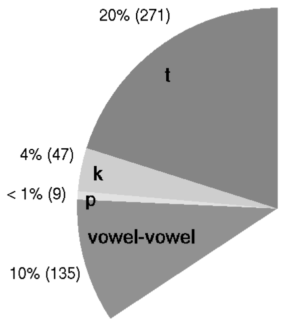

Figure 3: *Breakdown of irregular phonation at voiceless stops and vowel-vowel boundaries, The absolute number is shown next to the percentage within brackets. (Based on 1331 tokens)*

A further study of the irregular tokens at voiceless stop locations and vowel-vowel junctions was conducted in relation to word-boundaries (Figure 4). All the irregular tokens at vowel-vowel junctions occur at word-boundaries, i.e. either in word-initial or word-final position. For the irregular tokens at voiceless stops, 268 of the 326 occur at word-final position while another 44 occur at syllable-final position. All 44 of the syllable-final irregular tokens for voiceless stops occur either at the junction of a compound word or at the junction of a base word and a suffix.

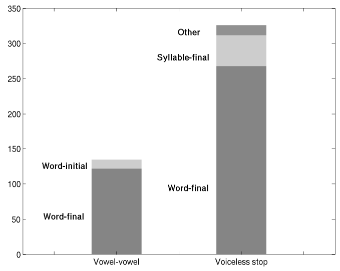

Figure 4: *Breakdown of irregular phonation at word level boundaries for vowel-vowel junctions and voiceless stops*

Additional analysis was conducted on cases of irregular phonation which do not coincide with either a word or syllable boundary in order to determine the context in which the irregular phonation occurs. Table 2 lists the five broad contexts in which these irregular tokens occur.

Table 2: Contexts in which irregular phonation at word-medial position occur

| |
|---|
| Before or after a voiceless consonant |
| Before or after a voiced consonant |
| Before or after a sonorant consonant |
| Function word ’a’ |
| Other |

A total of 225 irregular tokens occur at neither word nor syllable boundaries. Figure 5 shows their distribution among the five categories listed in Table 2. Of the 225 irregular tokens not at word-boundaries, 158 occur adjacent to a voiceless consonant. Analyzing these tokens showed that 130 of these occur either in utterance-final location or before a pause in the utterance. In Figure 5, utterance-final voiced consonants (stops & fricatives) are grouped with the voiceless consonants since such realizations are largely devoiced. One such example is shown in Figure 6 (a) where the word “subject” occurs in utterance-final location and the irregular token precedes the voiceless stop /k/.

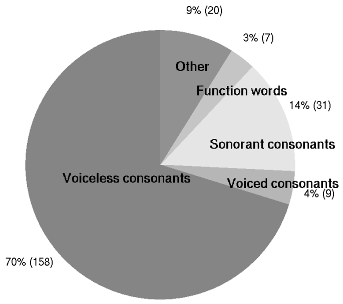

Figure 5: *Breakdown of irregular phonation which does not occur at word or syllable boundaries*

Of the 9 irregular tokens which occur adjacent to a voiced consonant, 6 are at utterance-initial or phrase-initial position. 16 of the 31 irregular tokens next to sonorant consonants occur either at the last word of the utterance or at pre-pausal locations. The 7 irregular tokens at function words encompass the entire word and hence are classified as neither word-initial nor word-final. The remaining 20 irregular tokens classified under the “Other” category include 10 tokens which show irregular phonation in vowel-medial position. This particular behavior is observed across multiple speakers. Figure 6 (b) shows one particular example where the irregular token occurs within the vowel /ae/ in the word “packing”.

## 5. Discussion

This paper addresses the question of whether or not all detected instances of irregular phonation in American English are associated with a boundary location. The results are collected in a speaker-independent analysis across 114 different speakers and show that 78% of the irregular tokens occur at a word boundary. Batliner *et al.* (1993) examined instances of irregular phonation for German speech. One-third of the database consisted of real spontaneous utterances, while the rest consisted of the same utterances read by the same speakers nine months afterwards. From a total of 1191 irregular portions of speech, 58% occurred in word-initial position and 18% occurred at the end of a word. The results of the present study for American English are highly consistent with the results of Batliner *et al.* (1999), and support the conclusion that irregular phonation is a strong acoustic cue for the detection of word boundaries.

The sentences in our database have a total of 10994 word boundaries, and of these, 1037 are marked with irregular phonation. In other words, 10% of the word boundaries are marked with irregular phonation. This percentage could increase with the identification of other strong acoustic cues. The detection of a subset of the word boundaries in a speech stream (based on robust acoustic cues such as irregular phonation and regions of silence) can provide segmentation of the speech stream into limited regions for proposing a cohort of word candidates in spoken language systems. Appropriately limiting the search region prevents the cohort from growing unmanageably large. The results of the present study are in conjunction with an effort towards the development of a system for automatic classification of regions of phonation as either regular or irregular. This *classification system* is a first and fundamental step towards an eventual *detection system*. With the current system, in a test set, 292 of 320 irregular tokens (recognition rate of 91.25%), and 4105 of 4320 regular tokens (recognition rate of 95.02%) are correctly identified [8].

A secondary question examined the 22% of the tokens of irregular phonation which did not occur at a word boundary and asked if there are still consistent observable trends in relation to other types of boundaries. Of these 22% of the tokens (294 in total), 50 were found to occur at syllable boundaries located at the junction of a compound word or between a base word and a suffix (such as -ment, -ly, or -en). An additional 130 tokens, which do not occur at a syllable boundary, occur in the vicinity of a voiceless (or devoiced) consonant at the end of an utterance, and 6 tokens, occur following a nominally voiced stop consonant at the start of an utterance (/b/, /d/ or /g/).

Recently, physiological correlates to irregular phonation, in utterance-final location, for utterances ending with a vowel, have been quantitatively studied [7]. The results show that when the end of the utterance coincides with the speaker taking a breath, the conditions associated with the respiratory actions to finish one breath and prepare for the next inhalation tend to give rise to a particular type of irregular phonation - one that is produced with relatively widely abducted vocal folds or produced as the vocal folds are in the process of continuing to abduct. This configuration yields irregular phonation which is highly damped and is in contrast to definitions of glottalization associated with tightly adducted vocal folds. In the present data, 58% of the tokens not occurring at a word or syllable boundary occurred in the vicinity of the end of an utterance. For example, in an utterance ending in the word ‘subject,’ the utterance ends with a voiceless consonant production, but the last instance of phonation in the utterance is irregular. In such cases, a physiological basis similar to that in [7], may create conditions conducive to irregular phonation.

Overall, for the 22% of irregular tokens which do not occur at a word boundary, 63% of them do occur in a boundary-related environment (such as syllable or utterance). These results further support the conclusion that, if in a spoken language system, an instance of irregular phonation is detected, a speech boundary should be hypothesized. The type of the boundary will depend on additional analysis which might include acoustic cues related to the specific nature of the irregular phonation, other acoustic cues related to the prosodic structure (such as duration and intonation), or the information regarding the segmental context.

Future work could entail expanding this study from read, isolated utterances to spontaneous speech. It is likely that irregular phonation at word boundaries would occur more often in spontaneous, than read speech. The study also focuses on the role irregular phonation can play in automatic speech recognition and lexical access models using data from a large set of speakers. The tradeoff to this approach is that it ignores speaker specific characteristics of irregular phonation that could prove highly useful for some speakers.

## 6. Conclusion

This study conducts a speaker-independent analysis using multiple speakers from two dialect regions in the TIMIT database to analyze if irregular phonation is a useful cue to segment continuous speech. Given that regions of phonation can be classified as regular or irregular with a high degree of consistency [8], the present results confirm that regions of irregular phonation can reliably serve as a segmentation cue for speech recognition and speech parsing. Irregular phonation is one of a range of acoustic cues to speech boundaries, and further studies offer the possibility of combining several cues to build a prosodic structure in a spoken language system.

## 7. Acknowledgements

This work was supported by NIH/NIDCD # DC02978. The first author would like to thank Stefanie Shattuck-Hufnagel for her advice and support.

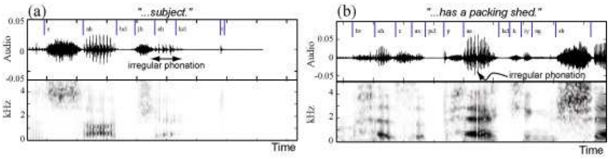

Figure 6: *Two examples of irregular phonation which do not occur at word boundaries. (a) is an example of an irregular token adjacent to a voiceless consonant in utterance-final location while (b) shows an irregular token in vowel-medial position*

## 8. References

[1] Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M., 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24, 423-444.

[2] Kreiman, J., 1982. Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, 10(2), 163-175.

[3] Lehiste, I., 1979. Sentence boundaries and paragraph boundaries — perceptual evidence. In *The Elements: a parasession on linguistic units and levels*, Chicago Linguistics Society, 15, 99-109.

[4] 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.

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

[6] Redi, L. & Shattuck-Hufnagel, S., 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29, 407-429.

[7] Slifka, J., 2005 (in press). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice*.

[8] Surana, K., 2005 (expected). Classification of vocal fold vibration as regular or irregular in normal, voiced speech. *M.Eng. thesis*, Massachusetts Institute of Technology.

[9] Umeda, N., 1978. Occurrence of glottal stops in fluent speech. *Journal of the Acoustical Society of America*, 64, 1, 88-94.