# An Analysis of the Intonational Structure of Stuttered Speech

*Timothy Arbisi-Kelm*

Department of Linguistics
University of California, Los Angeles, CA, USA
timrbc@ucla.edu

## Abstract

While previous studies have successfully revealed areas vulnerable to disfluency at the word level in stuttering, identifying the specific factors responsible for this instability has proved difficult. Analyzing the effects of phrasal prosody, which governs such word-level factors as lexical stress [1], is critical in order to account for the relations between word-level and phrase-level effects, and how they affect patterns of disfluency in stuttered speech.

In a story-telling task performed by two stutterers and two control subjects, it was found that stutterers' disfluencies were accompanied by more prosodic irregularities prior to the actual cause of the disfluency. In particular, changes in f0 and duration affect the realization of cues in the disfluent environment, resulting in fundamental alterations of intonational phrase structure. Anticipatory and target-realized disfluencies contribute different acoustic cues to their immediate environments, the results of which often create conflicting prominence relationships among prosodic constituents, thereby losing information key for conveying meaning.

## 1. Introduction

Stuttered speech is characterized by disfluencies which arrest the production of intonational phrase structures. While linguistic factors such as syllable onset position, semantic content, word length, word onset position, and utterance onset position have all been shown to correlate with high disfluency rates [2], prosody—i.e., prominence (pitch accents) and grouping (intonation phrase boundaries)—can account for most of these factors. Furthermore, subsequent studies have shown that stressed syllables are particularly vulnerable to disfluency [3][4]. But while previous work has provided evidence of f0 variability in stuttered speech [5], that the underlying deficit does not lie in the physical generation of airflow is evidenced by the lack of disfluencies generated when stutterers sing, read in chorus, read accompanied by a metronome, and produce other speech in which rhythmic predictability is guaranteed [6].

On the other hand, evidence from an experiment manipulating intonation structure of various sentence types showed that with the exception of focused nuclear pitch accent (NPA), disfluency patterns follow the hierarchical metrical scale of prominence: that is, stuttered disfluencies are a function of metrical prominence, predictable by properties of intonation [7]. Since pitch accents bear the greatest level of prominence, they attract a high rate of disfluency. That focused NPAs in this experiment were unexpectedly more fluent than non-PA'd syllables, however, revealed that what is faulty is not the production of PAs per se, but rather the production of grouped pitch accents—i.e, intonational phrase structures.

It was thus hypothesized that the disfluencies in stuttering will not only occur at a higher rate than in normal speech, but that their effects will surface much earlier in a given utterance, and also result in intonational anomalies that preclude a speaker from producing intended prosodic structures.

## 2. Method

### 2.1. Subjects

Four subjects—two stuttering males (age 72 and 51) and two controls (male 55, female 63)—were recruited from the Los Angeles area to participate in this experiment.

Subjects were seated in a quiet room, each for a single session of approximately one hour. Subjects wore a head-mounted SM10A Shure microphone, with the signal passed to a Marantz portable casette recorder (PMD222). Instructions were simply to narrate a frog picture book [8] as if sharing the story with someone for the first time. This procedure is particularly suitable to an intonation study because it allows subjects to create spontaneous-sounding sentences.

To ensure naturalness of speech, each subject was instructed to peruse the book before beginning in order to form a general idea of the story, thus facilitating the creation of a spontaneous-sounding narrative structure.

### 2.2. Data Analysis

All data samples were audiotape recorded, sampled at a rate of 11025 Hz, and stored digitally. Using the *PitchWorks* signal analysis software program (SCICON R&D), data files were coded in accord with the ToBI (Tones and Break Indices) model of transcribing English intonation [9]. A total inventory of five pitch accents (H*, L*, L+H*, L*+H, H+!H*) plus downstepped H tones (e.g., !H*, L+!H*, L*+!H) were used as well as phrase accents (H-, L-, !H-) and boundary tones (H%, L%). Tones, words, and break index markers were recorded on individual tiers above the pitch track of each sentence stimulus.

A modified version of the English ToBI transcription model was used to code word break data. In addition to the original inventory of normal breaks indicating word ('1'), intermediate phrase (ip) ('3') and intonational phrase (IP) ('4') boundaries, a revised set of disfluency categories was proposed to account for the variety of data occurring in stuttered speech. The total disfluency inventory contained the following types: prolongation ('p'), pause ('ps'), cut (word interruption) ('c'), and repetition ('r'), as well as all attested combinations of these. For example, a combination of a prolonged and cut word followed immediately by an ip boundary was designated by a break index '3pc'.

All word, tone, and break index information were labeled on tiers above the f0 data. A miscellaneous tier was used to indicate the exact location of a given disfluency, since disfluency boundaries often did not coincide with word and phrase boundaries.

The figure below shows an example of a prolongation disfluency occurring in a single intermediate phrase:

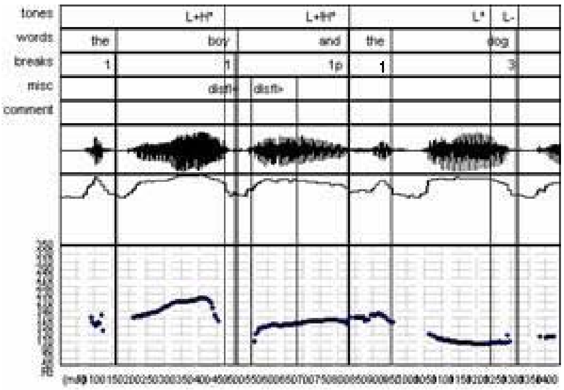


*Figure 1: Pitch track of prolongation disfluency.*

The tone tier indicates the location of pitch accents (e.g., L+H*), intermediate phrase accents (e.g., L-), and intonational phrase accents (e.g., H%). The break index tier reveals the size and type of boundary separating words: in Figure 1 above, each word except for the final one is followed by a word-sized break (break index '1'), while the occurrence of a disfluency is marked by the diacritic 'p' (prolongation) following the prolonged word; the disfluency index in the miscellaneous tier shows the exact location of the disfluency—in this case, a prolongation of the vowel in "and".

Labels for pauses and cuts were also placed at the right edge of the word, while repetitions were labeled immediately before the second iteration of the repeated segment.

## 3. Results and Discussion

The following comparisons are divided into two sections: 1) descriptions of the major prosodic category differences and 2) more detailed analysis of prosodic behavior in the disfluent environment.

### 3.1. Descriptive comparisons

Figure 2 below shows the raw numbers for words, intonation phrases (IP), intermediate phrases (ip), disfluencies, and pitch accents (PA), as well as the average number of disfluencies and PAs within each phrase type:

| Types\subject | S1 | S2 | C1 | C2 |
|---|---|---|---|---|
| Total # of words | 1148 | 535 | 492 | 489 |
| Total # of IPs | 214 | 81 | 84 | 91 |
| Total # of ips | 341 | 150 | 109 | 126 |
| Total # of disfl. | 314 | 168 | 8 | 9 |
| Avg. disfl./IP | 1.47 | 2.05 | 0.10 | 0.10 |
| Avg. disfl./ip | 0.92 | 1.14 | 0.07 | 0.07 |
| StDev disfl/ip | 0.97 | 1.01 | 0.29 | 0.26 |
| Total # of Pas | 535 | 288 | 272 | 232 |
| Avg. PA/IP | 2.50 | 2.05 | 3.24 | 2.55 |
| Avg. PA/ip | 1.56 | 1.92 | 2.39 | 1.83 |
| StDev PA/ip | 0.88 | 1.07 | 1.13 | 0.93 |

*Figure 2: Phrases, disfluencies, and pitch accents.*

#### 3.1.1. *Intonational Phrases*

Raw number of intonation (IP) and intermediate phrases (ip) did not differ consistently among subjects. While S1 produced a comparatively high number of words, IPs, and ips, S2 behaved similar to both control subjects in this area.

#### 3.1.2. *Pitch accents*

More surprising were the average rates of PA for both phrase types. It was expected that frequent disfluencies would result in smaller phrases overall for stuttering subjects, though this was only the case for S1. S2 again showed results more similar to those of the controls.

#### 3.1.3. *Disfluencies*

As anticipated, stuttering subjects produced a higher rate of disfluencies for both phrase types, and with a higher degree of variation than that of control subjects.

Comparisons across disfluency types were less revealing across subjects, due to the wide disparity of disfluency rates between stutterers and controls. However, as seen in Figure 3 below, the distribution of disfluency types showed some general similarities among all groups:

| Types\subject | S1 | S2 | C1 | C2 |
|---|---|---|---|---|
| prolong | 0.21 | 0.45 | 0.02 | 0.04 |
| SD | 0.46 | 0.56 | 0.13 | 0.19 |
| pause | 0.34 | 0.27 | 0.04 | 0.02 |
| SD | 0.58 | 0.50 | 0.18 | 0.15 |
| cut | 0.13 | 0.15 | 0.01 | 0.01 |
| SD | 0.41 | 0.42 | 0.09 | 0.09 |
| repetition | 0.03 | 0.03 | 0 | 0 |
| SD | 0.18 | 0.16 | 0 | 0 |
| combination | 0.20 | 0.24 | 0.02 | 0 |
| SD | 0.43 | 0.50 | 0.19 | 0 |

*Figure 3: Distribution of disfluency types*

Numbers represent the average number of each category type per ip. Disfluencies were separated into the four major types (prolongations, pauses, cuts, and repetitions), and a category of combinations (e.g., a prolonged and cut word). Overall, repetitions were the least common disfluency type across both groups. Disfluency rates were too low among the controls to reveal any within-category differences, but stutterers showed much higher rates for prolongations (.21/.45) and pauses (.34/.27) than for cuts (.13/.15) and repetitions (.03/.03).

### 3.2. Contextual comparisons

#### *3.2.1. Prolongations*

Prolongation disfluencies fell into two major classes: those realized before pitch accented words (*anticipatory*) and those realized on pitch accented words (*target*). Anticipatory prolongations were recognizable generally by their lengthened vowel duration (as opposed to onset prolongation on targets), as well as by their tendency to occur in early phrasal positions. In Figure 4 below, an anticipatory prolongation surfaces on both the second utterance of "while" (3p) and "in" (1p):

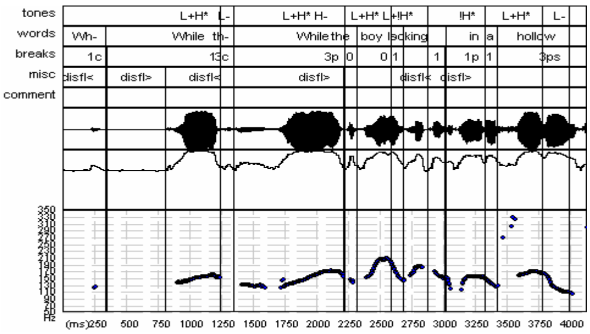


*Figure 4: Anticipatory prolongation*

The crucial evidence supporting the interpretation of these prolongations as anticipatory are the two fluent productions of "while", since even after the first successful attempt the speaker nevertheless restarts the phrase. Similarly, the prolongation after "in" would also appear to serve as an anticipatory delay since disfluencies on function words are less common than on content words in adult stuttered speech [10]. While not produced with overt disfluencies, the targeted PAs "boy" and "hollow" were nevertheless presumably anticipated as potentially problematic words.

As a direct result of extending the duration of affected segments, prolongations also triggered further prosodic irregularities. Since pitch accents consist of pitch, duration, and amplitude cues, increased duration on an unstressed word (e.g., function word) attracted prominence and thus altered the overall prosodic structure of an intonation phrase. This is illustrated in examples like Figure 4, where the function word "in" has triggered a pitch accent, despite the fact that this syllable is neither lexically stressed nor contrastively focused.

A second noteworthy irregularity appearing in prolonged environments was also a component of pitch accent: f0 movement. As segments are prolonged, the pitch value aligned with that segment does not necessarily hold steady: to the contrary, f0 interpolation between pitch targets results in rising and falling tone. As a result, a number of prolonged disfluencies were also realized with extended tones; for instance, a L* with eventual H* target began rising toward that target even while the segment stalled, thus resulting in a PA more similar to a L+H* (such as on "while" above). Similarly, final boundary cues also may rise (H-H%) or fall (L-L%), and thus many prolongation disfluencies carrying these boundary cues were confounded with the originally planned accents, thus presenting obstacles to realizing an intonation phrase.

#### *3.2.2. Pauses*

Like prolongations, pauses also were encountered in both anticipatory and target contexts. The feature distinguishing the latter from the former was the evidence of onset closure from the upcoming target word: commonly called a block in much of the current literature, this event was manifested acoustically as an abrupt cessation of phonation upon a word's offset. The following silence was a result of a delay in both releasing the attempted onset and beginning the word's phonation. This contrasts with the anticipatory pause, whose preceding word terminates with normal boundary cues.

Once again, as with prolongations there were also tonal consequences for postponed or blocked phonation. IP and ip breaks were particularly confusable with word-internal pause disfluencies, since any slight rising or falling of f0 appeared much like a boundary tone when followed immediately by a pause. And in the opposite case where there was a planned ip/IP boundary, pauses on targets (i.e., onset blocking) often created conflicting cues by obliterating any pitch excursion that would have signaled the boundary's edge, yet nevertheless created a clear post-boundary break cue in the manifestation of the disfluent pause.

#### *3.2.3. Cuts and repetitions*

While rarer in the data than prolongations and pauses, cut words and repeated segments resulted in similar irregularities in that cues resulting from disfluency competed—or conspired—with originally planned cues to significantly alter intonation phrase structure.

As with other disfluency types, F0 movement also played a crucial factor in determining the extent of influence the disfluency would have on its environment. In particular, resetting f0 had the effect of either providing a post-boundary ip-external cue—by resetting f0 higher—or an ip-internal cue (such as downstep)—by resetting f0 lower. Thus, in addition to triggering unplanned boundaries, disfluencies also could cause ips to extend beyond normal boundary breaks:

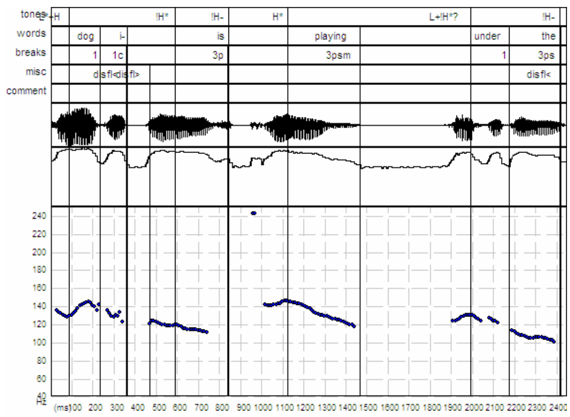


*Figure 5: A cut triggers downstep, a prolongation triggers an ip break, and a pause fails to trigger an ip break.*

In the example above, a cut (1c) triggered falling f0 which, coupled with the anticipatory prolongation (3p), attracted pitch accent on the function word "in". Aligned with the continued duration of this syllable, falling pitch excursion cued an ip boundary which was solidified by the post-boundary tone reset. Finally, despite the boundary cue naturally resulting from the anticipatory pause (3psm), f0 failed to reset in the middle of a syntactic environment where a boundary was likely not originally intended.

Although while once again comparisons across groups is difficult given the low number of disfluencies for control speakers, it is nevertheless revealing to compare the specific behaviors of these disfluent environments between subjects:

| Disfl | Behavior | S1 | S2 | C1 | C2 |
|---|---|---|---|---|---|
| prolong | PA same word | 38 | 62 | 0 | 0 |
| | ip break | 64 | 23 | 0 | 1 |
| | downstep | 19 | 18 | 1 | 0 |
| | f0 rise/fall | 32 | 48 | 0 | 1 |
| | onset prolonged | 44 | 2 | 2 | 3 |
| | no effect on ip | 6 | 9 | 3 | 0 |
| pause | PA next word | 2 | 19 | 0 | 1 |
| | ip break | 118 | 27 | 3 | 1 |
| | downstep | 8 | 11 | 1 | 0 |
| | f0 reset | 6 | 3 | 0 | 0 |
| | block | 46 | 1 | 0 | 0 |
| | no effect on ip | 23 | 10 | 3 | 2 |
| cut | PA next word | 6 | 7 | 0 | 0 |
| | ip break | 9 | 5 | 0 | 0 |
| | downstep | 1 | 2 | 0 | 0 |
| | f0 rise/fall | 18 | 3 | 0 | 0 |
| | no effect on ip | 29 | 9 | 2 | 1 |
| repeat | ip break | 3 | 1 | 0 | 0 |
| | f0 rise/fall | 4 | 0 | 0 | 0 |
| | no effect on ip | 13 | 7 | 0 | 0 |
| other | PA function words | 27 | 62 | 1 | 1 |
| | Target | 206 | 67 | 8 | 8 |
| | Anticipatory | 108 | 101 | 0 | 1 |

*Figure 6: Prosodic behavior in disfluent environments*

## 4. Conclusions

In summary, it was shown that the principal difference between stuttered and control disfluencies, aside from total number, is the context: stuttered disfluencies are often realized in context of prosodic irregularities prior to the primary disfluency, while normal disfluencies are almost exclusively realized on the target. The result for stuttered speech is a combination of alterations of tone (f0 rise or fall) and duration (lengthened or interrupted), with these variations themselves resulting in further irregularities such as degraded or false cues, and otherwise unorthodox intonational patterns that jeopardize a speaker's intended meaning.

## 5. References

[1] Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. PhD thesis, MIT, published 1988 by IULC.

[2] Brown, S.F., 1945. The loci of stutterings in the speech sequence. Journal of Speech Disorders 10, pp. 181–192.

[3] Natke, U., Grosser, J., Sandrieser, P., Kalveram, K.T. (2002). The duration component of the stress effect in stuttering. Journal of Fluency Disorders Volume 27, Issue 4 , 305-318.

[4] Prins, D., Hubbard, C., Krause, M. (1991). Syllabic Stress and the Occurrence of Stuttering. Journal of Speech and Hearing Research, 34, 5, 1011-1016.

[5] Healey, E.C., & Ramig, P.R. (1986). Acoustic measures of stutterers' and nonstutterers' fluency in two speech contexts. Journal of Speech and Hearing Research, 29, 325-331.

[6] Stager, S.V., Denman, D.W., & Ludlow, C.L. (1997). Modifications in aerodynamic variables by persons who stutter under fluency-evoking conditions. Journal of Speech, Language, and Hearing Research, 40(4), 832-847.

[7] Arbisi-Kelm, T. (2004). An Intonational Analysis of Disfluency Patterns in Chronically Stuttered Speech. (unpublished manuscript).

[8] Mayer, M. (1969). Frog, Where are You? New York: Dial Books.

[9] Beckman, M., and Ayers, G. (1994). Guidelines for ToBI labeling. Language and Cognitive Processes, 11(1/2), 17-67.

[10] Dayalu, V.N., Kalinowski, J., Stuart, A., Holbert, D, Rastatter, M. (2002). Stuttering frequency on content and function words in adults who stutter: a concept revisited. Journal of Speech, Language, and Hearing Research, 45, 871–878.

[11] Beckman, M., and Ayers, G. (1994). Guidelines for ToBI labeling. Language and Cognitive Processes, 11(1/2), 17-67.