# Intonation Structure and Disfluency Detection in Stuttering

Timothy Arbisi-Kelm*
* University of California Los Angeles, timrbc@ucla.edu

## Abstract

Despite traditional evidence that lexically-stressed content words are the most common loci of disfluencies in stuttered speech (Natke, et al., 2002; Prins, et al., 1991; Brown 1938), the fact that certain function words—crucially, those directly preceding stressed content words—are also disfluently produced (Au-Yeung, et al., 1998; Howell & Sackin, 2000), suggests that the actual trigger of a disfluency occurs later than the speech perturbation itself. It was hypothesized in this paper that stutterers' disfluencies would be accompanied by more prosodic irregularities prior to the actual disfluency than would non-stutterers' disfluencies, and that the underlying disfluencies would be triggered by metrically prominent material in the phrase. Three stutterers and their age-matched controls were recorded performing a spontaneous narration of a picture book. Results supported the hypothesis that metrically prominent, or pitch-accented, words would attract a higher rate of stuttered disfluencies than would unaccented words. Stutterers also produced a higher percentage of their disfluencies on the nuclei of words than did controls; moreover, these disfluencies attracted pitch accents more often than did those of controls, providing further evidence of anomalous derivative disfluencies surfacing in advance of underlying ones.

## 1. Introduction

According to current models of disfluency production in normal speech, such as those proposed in Levelt (1983) and Shriberg (1999), disfluencies generally signal a speaker's detection, and attempted correction, of an error in language production. However, the fact that errors can be repaired covertly—that is, detected and corrected before their overt articulation—makes the identification of the production error a challenging task (Dell 1985; Levelt 1983). A number of studies have shown that speakers will often produce different types of evidence which signal the origin of the disfluency—i.e., whether it is phonological, semantic, syntactic, etc. In slips-of-the-tongue, for example, speakers may exchange segments across word boundaries, revealing an error made at a post-lexical level (Fromkin 1971; Shattuck-Hufnagel 1979; Dell & Reich, 1980). Errors at a lexical (or pre-lexical) level, meanwhile, are evidenced by TOT (tip-of-the-tongue) phenomena, which are characterized by a speaker's inability to retrieve a fully-specified lexical form. Often a speaker is able to retrieve only such metrical information as a word's number of syllables, suggesting that some aspects of a word's metrical structure are retrieved before the segmental content (Caramazza 1997; Levelt 1989). Production errors such as these, however, are opaque to the listener, since they occur before the problematic word's articulation.

Nonetheless, analyzing disfluencies and their interaction with other linguistic phenomena can provide clues to locating areas of breakdown in both normal and impaired speech. Levelt (1983) breaks down the structure of a speech error repair into

three main components: the reparandum (the error to be repaired), the editing phase, and the repair. Two studies of speaker repair types (Levelt & Cutler 1983; Cutler 1983) revealed that phonological speech error repairs always prosodically resembled the actual errors, while lexical error repairs often had higher F0 values and longer segment duration than their errors. Analyses of repetitions in spontaneous speech have shown that repair F0 values differed depending on whether a repetition represented a recovery from a disfluency or a pragmatic stalling strategy: only in the first case did repairs reset F0 to the same level of the repeated word, while F0 value in the latter "prospective" repetitions was neither reset nor increased (Shriberg 1994; Shriberg 1999). At the same time, reparanda in the prospective cases were consistently longer in duration than their corresponding repairs, suggesting that even in overt repetitions error detection is often covert. A study of vowel quality differences in function word production by Fox Tree & Clark (1997) found similar evidence for covert detection and repair. When the determiner "the" was produced as its longer, tense-voweled variant, speech was interrupted 81% of the time, and only 7% of the time after the standard lax variant. The fact that the vowel quality was changed (rather than simply lengthened) gives further evidence that speakers may anticipate production trouble in advance of its articulation, and make corrective adjustments online.

Evidence of particularly early error anticipation has been found in the speech of stutterers. In Viswanath (1989), stutterers performing an oral reading task produced significantly longer words immediately *before* words which were stuttered, while control groups showed no pre-disfluency duration differences. Similarly, Au-Yeung, Howell & Pilgrim (1998) and Howell, Au-Yeung & Sackin (1999) found a strong tendency of stutterers to produce disfluencies on function words immediately preceding content words (e.g., "a dog"), while rarely on function words following content words (e.g., "took it"). This distributional evidence suggests that it is the relative location of the function word within a larger phrase—crucially, early in the phrase—that is critical to its disfluent production. Howell & Sackin (2000) provided a more detailed analysis of this tendency and concluded that although both function word and content word disfluencies are found in stutterers' speech, each disfluency type represents a different aspect of a production breakdown: content word disfluencies are presumably triggered by factors intrinsic to the word itself (e.g., phonological complexity), while function word disfluencies constitute postponements in executing an incomplete phonological plan. Taking these results together, then, it appears that in stutterers' speech, qualitatively different disfluencies surface in predictable locations within a prosodic phrase, suggesting that the distribution patterns of disfluency types are constrained by prosodic phrase structure.

The prominent role of prosodic structure in predicting stuttering disfluency patterns has been recognized in a number of earlier studies. Lexically-stressed syllables (Brown 1938; Natke, Grosser, Sandrieser & Kalveram, 2002) and word-initial position (Prins, Hubbard & Krause, 1991) have been shown to be highly correlated with stuttered disfluency rate. While such studies successfully accounted for the effects of word-level prosody, they did not analyze the full range of suprasegmental factors at work in speech production, of which lexical stress is only one type. According to the Autosegmental-metrical model of intonational phonology proposed by Pierrehumbert and her colleagues (e.g., Pierrehumbert 1980; Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988), all levels of metrical stress are organized hierarchically: word-internally, stressed

syllables are more prominent than unstressed syllables, while phrasal prominence (pitch accent) marks a degree of prominence higher than the word level. The highest level of metrical stress in this model is the nuclear pitch accent, the final pitch accent of an English intermediate phrase. These prominence relations can be represented by a metrical grid, as shown in Figure 1:

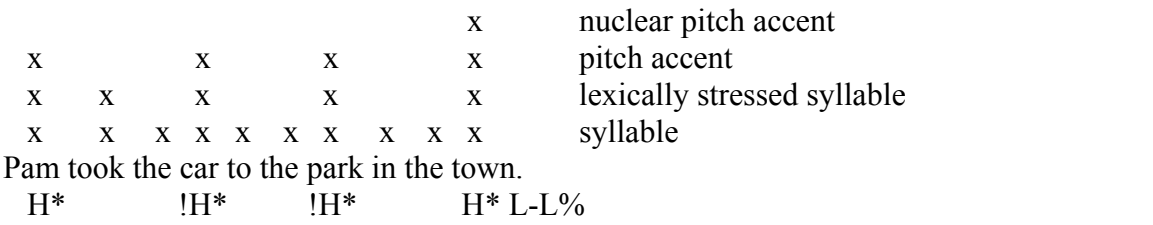


**Figure 1: Levels of stress in an intonational phonology model**

Each grid mark reflects a single degree of prominence, and tones are associated with pitch-accented syllables (e.g., H*), as well as with the intermediate (e.g., L-) and intonation boundaries (e.g., L%) of a phrase. Intermediate phrases minimally consist of a pitch accent and a phrase accent, and are part of a larger structure called an intonation phrase. Assuming this framework, it could therefore be proposed that any link between stuttering and lexical stress is not necessarily an indication of difficulty with word-level stress per se, but rather of a general instability in producing linguistic prosodic prominence.

The possibility that stutterers' disfluencies originate in prosodic breakdowns would have implications for the question of how much prosodic structure is available to speakers in general during early linguistic planning. Unlike incrementalist models of speech production (e.g., Levelt 1989; Levelt, Roelofs & Meyer, 1999), prosody generation models such as those developed by Ferreira (1993) and Keating & Shattuck-Hufnagel (2002) presuppose the accessibility of larger prosodic constituent structures relatively early in the production process—crucially, before phonological encoding. If stutterers' production breakdowns are due to a component critical to construction of prosodic structure, then the effects of such a breakdown should manifest themselves when they are detected in the speech planning process. By the predictions of non-incrementalist models, this would occur pre-lexically, and thus effectively account for the early anticipatory disfluencies commonly observed in stuttering. At the same time, if the distribution of stutterers' disfluencies is constrained by the prosodic environment, then this might constitute further support for a non-incrementalist model of production.

A significant body of literature has investigated this question of the relationship between planning and disfluency detection in both normal and disfluent speech. Various speech monitoring models have been developed to account for a speaker's ability to detect potential disfluencies during speech planning (e.g., Levelt 1983; Blackmer & Mitton 1991; Postma & Kolk 1993). These models share a core assumption that speakers are able to inspect their speech programs before the onset of articulation, with one result being the ability to make corrections of errors detected during this inspection. It is still a source of speculation as to how exactly the inner structure of this monitor should be characterized, but it is generally agreed that the monitoring system operates through one or more feedback loops from a particular production stage (e.g., phonetic encoding) back

to another (e.g., conceptualization), potentially via the speech perception system—though this latter claim is also a matter of debate.

A speech error repair model known as the "Covert Repair Hypothesis" was developed to account for both stutterers' and nonstutterers' disfluencies (Kolk 1991; Postma, Kolk & Povel, 1990; Postma & Kolk 1993). Essentially, this hypothesis claims that a speaker may both detect and correct an error before it encroaches on overt articulation. The authors postulated that while both stutterers and nonstutterers are equipped with the same monitoring architecture, for stutterers the system is particularly active due to the proliferation of speech errors resulting from a phonological encoding deficit. In other words, stutterers have an underlying deficit in selecting phonemes for an utterance plan; what actually surfaces in the output of stutterers' speech, however, is not the improperly organized phonological plan but rather the stops and restarts of the monitor as it covertly detects these errors.

The assertion that stuttering is rooted in a phonological encoding deficit, but surfaces only through subsequent attempts of a speech monitor to correct the encoding errors, is a controversial hypothesis that has been tested in a number of studies with inconclusive results (e.g., Postma & Kolk 1992a, b; Wijnen & Boers 1994; Burger & Wijnen 1999; Melnick & Conture 2000). More recent studies have proposed a different interpretation of how the speech monitoring system interacts with disfluency production in stutterers' speech. Rather than assume a phonological encoding deficit, Vasic & Wijnen (2005) instead hypothesized that the impairment in stuttering is rooted in the monitor itself: an "over-vigilant" monitor falsely identifies aberrations in the speech plan, and interrupts the production process as a response to these hyper-vigilant detections. Specifically, the authors proposed that the monitor identifies any *discontinuous* aspects of the speech stream—i.e., accented words, aspirated segments, temporal variations, or other linguistically prominent elements—as speech errors, and in an attempt to correct the perceived errors will interrupt and restart articulation, producing observable disfluencies. Thus, in the Vasic & Wijnen monitoring model, this latter evaluation process itself is faulty in that it incorrectly identifies normal discontinuities of speech as failing to meet pre-determined prosodic criteria.

Neither previous speech error nor disfluency studies have addressed these questions of speech plan inspection—or, "lookahead"—from within a model of intonation. Such an approach would provide a framework through which disfluencies could be analyzed for their prosodic properties and behavior, as well as their interactions with these larger phrasal structures. As cited above, several studies revealed different behavior of stuttered disfluencies depending on their context within a prosodic phrase (Viswanath 1989; Au-Yeung, et al., 1998). It was therefore hypothesized in the present study that stutterers would produce more anticipatory disfluencies (e.g., pre-pausal vowel lengthening) in their speech than would normal speakers. The second hypothesis of this study was that if disfluencies in stuttering are due to a deficit in formulating prosodic structure (e.g., Brown 1938; Prins, et al., 1991; Natke, et al., 2002), then disfluencies should also surface in metrically prominent locations: specifically, these non-anticipatory (hereafter, "target") disfluencies were predicted to occur most often on the most metrically prominent material in an intermediate phrase (i.e., nuclear pitch accents).

## 2. Method

### *2.1. Structure of experiment*

The experiment was divided into two related tasks. In Task 1, both stutterers and control speakers narrated a wordless picture book with natural, spontaneous speech. Task 2 was similar in that it involved the same narration, although this time only three controls (and no stutterers) participated. The goal of Task 2 was for each control, who was already familiar with the story from Task 1, to read the narrative produced by his age- and gender-matched counterpart in the stuttering group—that is, controls read the stutterers' narratives from Task 1.

The motivation for Task 2 was to provide a reference prosody for each speaker in the stuttering group, by which intended prosodic phrase boundaries and tonal types could be estimated. Previous studies have used similar methods in order to examine characteristics of different speech populations such as alaryngeal speakers (van Rossum 2005) and autistic and stuttering children (Fosnot & Jun 1999). An obvious challenge to this approach is that prosody varies considerably across speakers, both in constituent structure and tone assignment. Nevertheless, a reference prosody was proposed as a an additional means of interpreting the stutterers' raw data, so that some comparisons could be made while keeping lexical and segmental content constant.

### *2.2. Participants*

Six subjects—three adult male stutterers and their age- and gender-matched controls—were selected to participate in a story-telling task. Stuttering severity for stutterers was moderate, as determined by assessments provided by licensed speech-language pathologists. In order to control for age effects, subjects and controls were chosen to represent one of three age groups: 30-39, 50-59, and 70-79 years of age.

### *2.3. Procedure*

Individual subjects were seated in a quiet room, each for a single session of approximately one hour. Instructions were simply to narrate the picture book, "Frog Where Are You?" (Mayer 1969), as if sharing the story with someone for the first time. This procedure was chosen because it allows subjects to produce spontaneous and natural-sounding utterances delivering the same story, while using a similar or same set of lexical items for the characters and objects shown in the pictures. In order to facilitate the creation of a narrative structure, subjects were instructed to peruse the book before the task and form a general idea of the story.

### *2.4. Data analysis*

All recordings were digitized, sampled at a rate of 11025 Hz. Using the PitchWorks signal analysis software program (SciCon R&D), F0 tracks and waveforms were displayed and the prosodic information was coded in accord with the MAE-ToBI (Mainstream American English—Tones and Break Indices) system of transcribing English intonation (Beckman & Ayers 1994; Beckman & Hirschberg 1994).

In order to assess the reliability of prosody and disfluency transcription, a second transcriber was employed to re-analyze 10% of the data (1812 words). Inter-judge agreement scores were obtained by using a formula of percent of judgments agreed: (agreed / agreed + disagreed) x 100. A reliability score of 92.5% was found for pitch accent assignment, while break index assignment (i.e., boundaries and disfluencies) had an inter-agreement score of 86.8%. The total agreement score was 89.2%.

### 2.4.1. Coding disfluencies

Stuttered disfluencies were defined in accord with previous analyses—namely, as the "occurrence of irrelevant sounds, repetition of sound or of syllable, silent blocks" (Bosshardt 1993). However, a formalized classification was necessary in order to code disfluencies within the ToBI intonation model. The principal information to be recorded in marking a disfluency was the location of its production in the phrase. Also important, however, was capturing the type of break that occurred within the utterance as a consequence of the stuttered disfluency. For example, in the previous literature the distinction between disfluencies occurring before the release of an onset and those occurring after its release are often expressed by using the term "block" to refer to the former, and "repetition"—or "prolongation", if there is no restart—to the latter, though generally only as a descriptive reference. Such distinctions were taken as motivations for separate disfluency break index categories, eventually resulting in an extended system of break indices for disfluencies. A list of the single disfluency break indices used in the current study, along with a short description of each type, is displayed below in Table 1:

**Table 1: List of single disfluency break index types**

| Disfluency type | abbrev | Description |
|---|---|---|
| restart | t | restarting of a segment, syllable, word, or entire phrase |
| prolongation | p | abnormal and/ or unplanned prolongation of a segment |
| cut | c | a partially completed word |
| pause | ps | abnormal and/or unplanned pause between or within words |
| filler | f | filler "words" or segments (e.g., "um", "uh") |

While the list of disfluency types in Table 1 captures all the major distinctions among disfluencies produced by both stutterers and control subjects, more complex disfluencies also resulted from combinations of these general types. This was accounted for by simply combining the relevant disfluency break indices. For instance, a disfluency in which a prolongation was followed immediately by a pause was represented by the break

index ‘p.ps’ (‘p’ + ‘ps’); similarly, a cut word (‘c’) followed by a pause (‘ps’) and then a filler (‘f’) was coded as ‘c.ps.f’. In this way it was possible to accurately reflect the multiple events which can occur within a single disfluency.

English ToBI uses a break index scale to code the degree of juncture between two words, with ‘0’ representing a clitic boundary, ‘1’ a phrase-medial word boundary, ‘3’ an intermediate phrase break, and ‘4’ an intonation phrase break. In order to integrate this study’s expanded disfluency system into the English ToBI transcription convention, each disfluency diacritic was matched with the break index number corresponding to the particular disfluency’s degree of juncture. For instance, when a word was cut and prosodically cliticized to a following word, ‘0c’ was used; when a phrase-medial word was cut, ‘1c’ was used; when a word at the end of an intermediate phrase was cut, ‘3c’ was used; and when a word at the end of an Intonational Phrase was cut, ‘4c’ was used.

### 2.4.2. Coding script data

Analyzing the target data from the reference prosody scripts was important for testing the robustness and relevance of patterns found in the natural data, given that essentially all of the conditions—with the exception of word position—require some approximation of the intended (targeted) structure, if their interactions with disfluencies are to be better understood. Any evidence of a planned nuclear pitch accent, for example, may become opaque if a preceding disfluency resulted in a new ip break—thus rendering the pitch accent phrase-initial instead of phrase-final.

Determining the target was formalized by the following definitions:

> Definition 1: *Target*
> In the environment of a non-pause disfluency, the *target* is the word on which the disfluency occurs, provided that no more than the onset of the word was produced; if the nucleus was also produced, then the target was the word following the disfluent form, including any subsequent material separated by a break index of ‘0’. Similarly, for a pause disfluency, the target is the word immediately following the pause, including any subsequent material separated by a break index of ‘0’.

> Definition 2: *Prosodic Word*
> A *prosodic word* is a prosodic unit consisting of a lexical word (pitch-accented word) and its satellite material, the latter of which is separated from the lexical word (or further satellite material) by a ‘0’ break index boundary.

An example demonstrating both definitions is shown in Figure 2 below, where the speaker is attempting to utter the phrase “goes to sleep”:

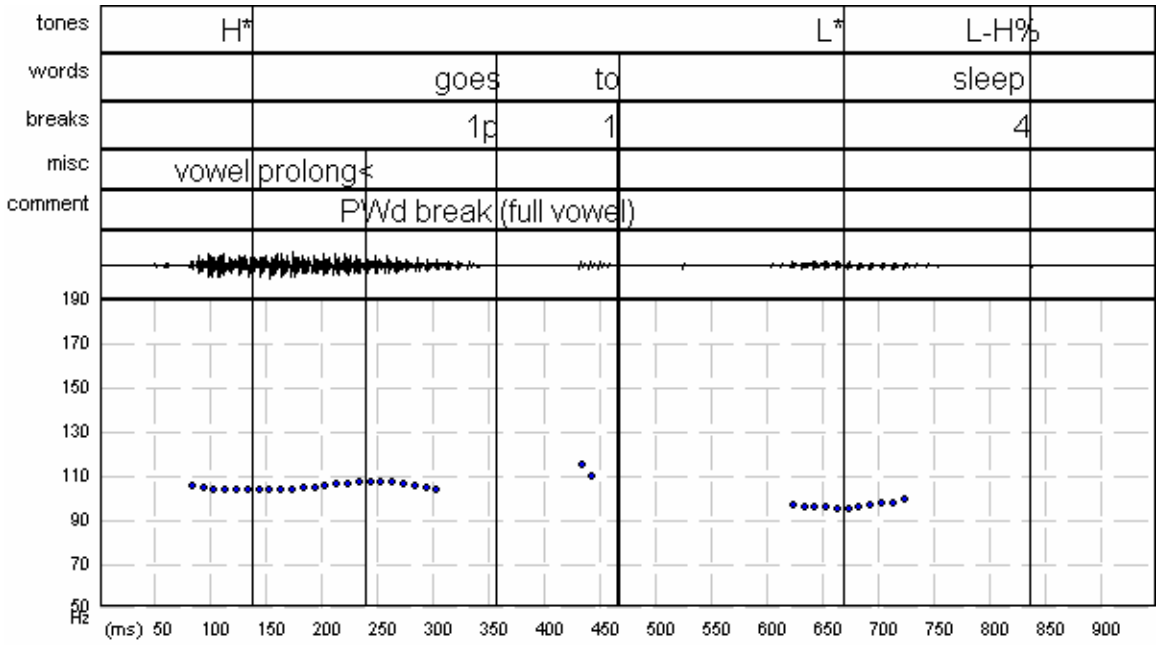


**Figure 2: A post-disfluency target: the vowel of "goes" is prolonged, suggesting an anticipatory delay before the target "to sleep".**

Descriptively speaking, the actual disfluency occurs in the prolongation of the syllable nucleus in the word "goes". However, since the syllable nucleus of the word on which the disfluency occurred was successfully produced, the target for this disfluency is interpreted as the next word. The size of all prosodic constituents and their breaks was determined by referring to the prosody produced by a control speaker. In this way it became possible to hypothesize the word and phrase boundaries intended by the stuttering speaker. In some cases a disfluency would fundamentally alter an intonation structure if a speaker was forced to re-start a word or phrase. In Figure 2 the effect is more subtle; while it appears the prolongation simply extends the duration of the word "goes", it also arguably has the effect of slowing down production of the next words: "to" and "sleep". The reference prosody provided by the control speaker illustrates this interpretation:

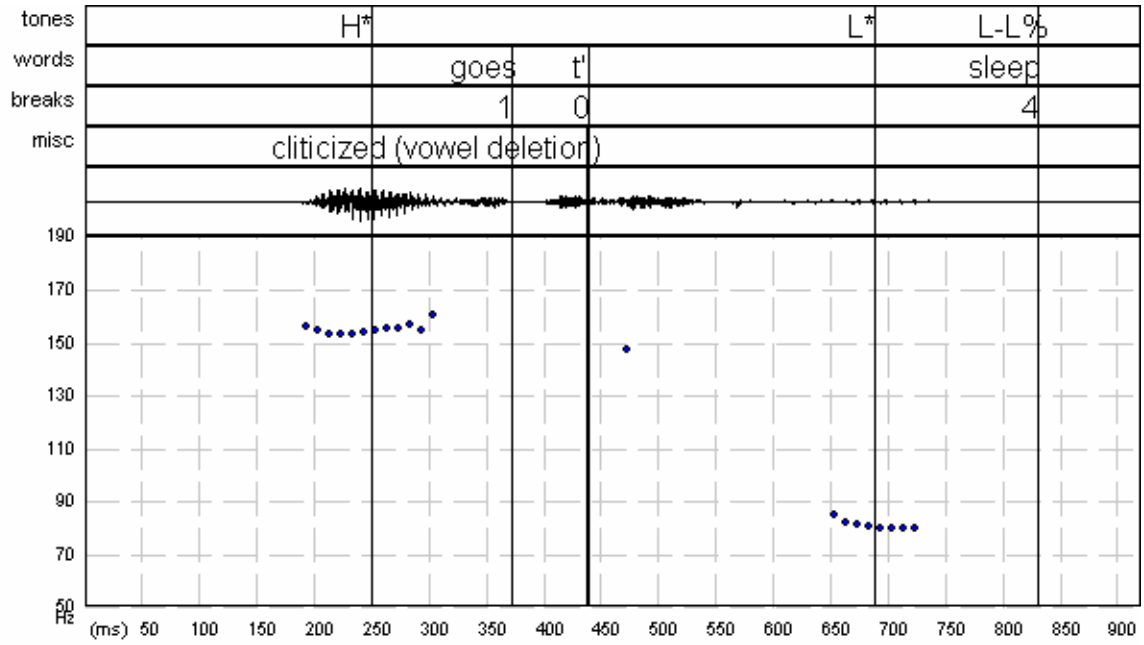


**Figure 3: Evidence that the post-disfluency target is the accented Prosodic Word "to sleep", as the function word "to" is cliticized to the content word "sleep".**

In the example shown in Figure 3, the reference prosody reveals a clitic boundary between the function word "to" and the final word "sleep", which receives a nuclear pitch accent. Following Definition 1, the target includes both the word following the disfluency as well as any further material separated only by a clitic-sized break (BI '0'). Thus, the target must include both "to" and "sleep". Furthermore, following Definition 2, since "to" is prosodically cliticized to "sleep", the two form a single Prosodic Word. This implies that the target should be a Prosodic Word.

The implication of this is that the prolongation disfluency of "goes" can now be said to occur immediately before the nuclear pitch-accented word. That is, despite the actual production by the stutterers, which suggests that only the word immediately following the disfluency and not those followed by a clitic-sized break (i.e., "to") must therefore be the target, the reference prosody instead provides evidence that "to" is indeed followed by break index '0', thus forming a Prosodic Word immediately following the disfluency. Hence, the target of the disfluency is the entire nuclear pitch-accented Prosodic Word "to sleep".

One final important formalization is how to determine a target depending on where in a syllable a disfluency is realized. For instance, in the prolongation example from Figure 2, it was maintained that since the syllable nucleus of the word in which the prolongation surfaced was successfully produced, the word following the prolonged word was the actual underlying target. This is consistent with what is referred to as Wingate's (1976) "fault line" analysis, derived from his findings that stutterers demonstrate difficulty not in production of onsets or nuclei, but in the transition from the former to the latter. Thus, once a speaker has begun production of the nucleus, it may be concluded that the syllable itself has been successfully produced, since the critical transition between onset and nucleus has therefore been achieved[1]. Applying this metric to the evaluation of the reference prosody, disfluencies surfacing on syllable onsets were categorized as target

disfluencies, since the critical juncture between onset and nucleus in such cases was not successfully crossed. When a disfluency occurred on any part of the rime, however, it was categorized as anticipatory, since articulation advanced past the onset-rime juncture:

> Definition 3: *Anticipatory disfluency*
> An *anticipatory disfluency* is any disfluency realized on a syllable nucleus.

## 3. Results

Because of the relatively small group size, and the overall low number of disfluencies in the controls' data, descriptive statistics were used to analyze and present the data for all variables and contrasts.

### *3.1. General prosodic characteristics*

The basic descriptive characteristics of both the natural speech narrations ('Nat') and reference prosody ('Ref') for each subject are presented in Table 2, listing the raw numbers of words, disfluencies, intermediate phrases (ip), and pitch accents (PA).

**Table 2: Prosody characteristics for each subject in both Task 1/ Nat (natural speech narrations) and Task 2/ Ref (reference prosody script-reading): words, ips, disfluencies, and PAs.**

| | S1/C1 | | | S2/C2 | | | S3/C3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Nat S1 | Ref C1 | Nat C1 | Nat S2 | Ref C2 | Nat C2 | Nat S3 | Ref C3 | Nat C3 |
| # words | 1212 | 1005 | 472 | 571 | 508 | 494 | 2814 | 1828 | 712 |
| # ip | 357 | 232 | 127 | 153 | 90 | 117 | 701 | 329 | 151 |
| M wd/ip | 3.53 | 5 | 4.21 | 3.76 | 5.91 | 4.26 | 4.13 | 5.56 | 4.72 |
| SD of wd | 2.19 | 2.04 | 1.60 | 2.16 | 2.85 | 1.88 | 2.33 | 2.90 | 2.26 |
| # disfl | 374 | N/A | 19 | 190 | N/A | 16 | 1140 | N/A | 124 |
| M disfl/ip | 1.09 | N/A | 0.15 | 1.25 | N/A | 0.14 | 1.68 | N/A | 0.82 |
| SD of disfl | 1.12 | N/A | 0.38 | 1.11 | N/A | 0.47 | 1.48 | N/A | 0.95 |
| # PA | 566 | 470 | 256 | 310 | 258 | 298 | 1257 | 824 | 279 |
| M PA/ip | 1.65 | 1.98 | 2.00 | 2.05 | 2.71 | 2.57 | 1.85 | 2.26 | 1.85 |
| SD of PA | 0.87 | 0.98 | 0.94 | 1.11 | 1.35 | 1.04 | 0.92 | 1.18 | 0.93 |

While the total number of words in a narrative differed widely among subjects of both groups, the mean number of words in an intermediate phrase (ip) did not differ considerably among subjects. Stutterers produced slightly shorter ips (average 3.53-4.13 words/ip) than did controls (average 4.21-4.72 words/ip), and these ranges were non-overlapping. At the same time, all stutterers produced many more words overall than their age-matched controls. The principal reason for this effect was the significantly higher production of disfluencies in stutterers' speech, which includes not only overt whole-word repetitions, but any prosodically distinct partial words. Finally, there is some evidence of a possible age effect, as S3 and C3 both produced the highest word/ip ratios in their respective groups, while S1 and C1 produced the lowest ratios.

Similar to intermediate phrases, total number of pitch accents varied more among stuttering subjects than among controls (range for S: 310-1257; range for C: 256-298). The average number of pitch accents per ip, however, was more consistent among subjects of both groups. Furthermore, stutterers produced considerably fewer PAs per ip than did controls.

Differences between a subject's actual narrative ('natural' data) and reference prosody ('script' data) are attributable to the fact that disfluencies added unintended material to a speaker's production, as was seen in the difference between word number in stutterers' and controls' natural data. For instance, any time a speaker restarted a whole word, each restart represented a separate attempt of the *same* target. A script indicating the intended utterance, however, would require removal of the extraneous restart to produce the shorter intended phrase. Thus, in this most common case, a script interpretation resulted in less prosodic material (e.g., fewer words) than was in the original narrative.

In order to determine whether contextual effects, such as the influence of adjacent tones and phrasal position of the disfluency, resulted in different patterns between stutterers and controls, these contexts were analyzed for each disfluency. Disfluencies were thus examined for the following conditions: location of disfluency within the syllable, degree of metrical prominence of the disfluent word, and position of the disfluency within the intermediate phrase.

### *3.2. Disfluencies and syllable location*

Group comparisons revealed an overall difference between stutterers' and controls' disfluency production with respect to location within the syllable. 76.3% of control group disfluencies were realized on syllable onsets, while for the stuttering group only 61.7% of disfluencies occurred on syllable onsets. Compared with the controls, the significantly higher rate of disfluencies in stutterers' syllable nuclei is evidence of a higher anticipatory disfluency rate. As Figure 4 illustrates, however, this difference was evident only for speakers S2 and S3, while S1 patterned more similarly to the controls. The first hypothesis was therefore only partially supported by this evidence.

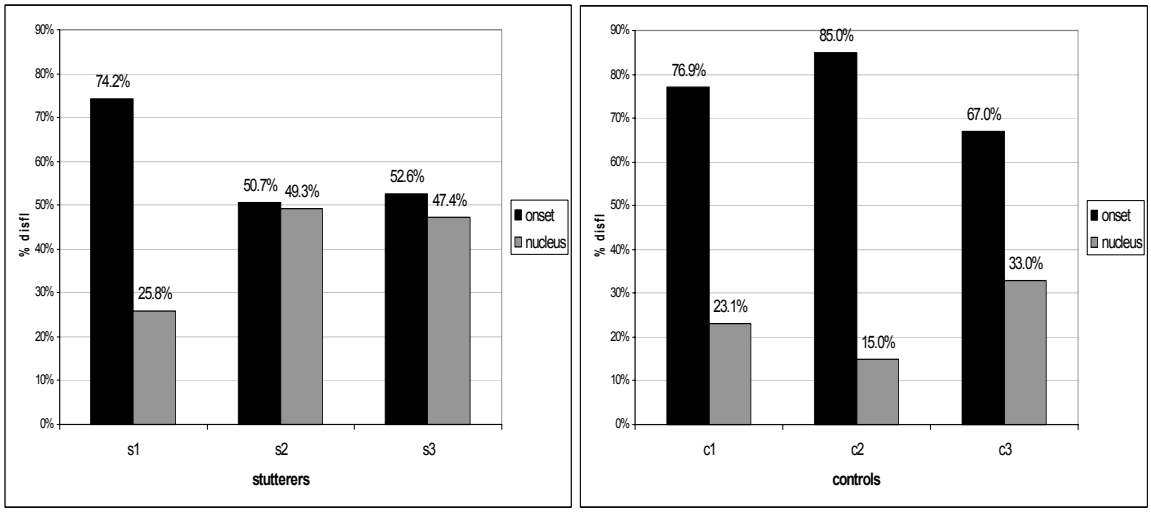


**Figure 4: Disfluency location within the syllable: stutterers vs. controls**

Adding a second condition—prominence level—to the comparisons of disfluency rate and syllable location, uncovers further evidence of anticipatory disfluencies in stutterers' productions. The disfluency rates at onset position for both PA (stutterers: 23.2%; controls: 23.4%) and nuclear pitch accent (NPA) forms (stutterers: 14.8%; controls: 15.2%) were extremely similar for both groups, as shown in Table 2. The largest difference between stutterers and controls was in the generation of disfluent PAs at nucleus position, with stutterers reaching a disfluency rate of 18.9% and controls just 4.4%:

**Table 3: Interaction of syllable position (onset vs. nucleus) and prominence level (PA vs. NPA vs. unaccented): stutterers vs. controls**

| Speaker \ disfl location in syllable | PA | | NPA | | unaccented | |
|---|---|---|---|---|---|---|
| | onset | nucleus | onset | nucleus | onset | nucleus |
| S1 | 83 | 17 | 68 | 28 | 96 | 41 |
| | 24.9% | 5.1% | 20.4% | 8.4% | 28.8% | 12.3% |
| S2 | 42 | 55 | 30 | 22 | 33 | 25 |
| | 20.3% | **26.6%** | 14.5% | **10.6%** | 15.9% | 12.1% |
| S3 | 240 | 225 | 135 | 105 | 168 | 159 |
| | 23.3% | **21.8%** | 13.1% | **10.2%** | 16.3% | 15.4% |
| C1 | 8 | 1 | 10 | 2 | 2 | 3 |
| | 30.8% | 3.8% | 38.5% | 7.7% | 7.7% | 11.5% |
| C2 | 12 | 2 | 3 | 1 | 2 | 0 |
| | 60% | 10% | 15% | 5% | 10% | 0% |
| C3 | 17 | 4 | 11 | 9 | 47 | 24 |
| | 15.2% | 3.6% | 9.8% | 8% | 42% | 21.4% |
| TOTAL Stutterers | 365 | 297 | 233 | 155 | 297 | 225 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 23.2% | **18.9%** | 14.8% | **9.9%** | 18.9% | 14.3% |
| TOTAL Controls | 37 | 7 | 24 | 12 | 51 | 27 |
| | 23.4% | 4.4% | 15.2% | 7.6% | 32.3% | 17.1% |

Once again, speaker S1's production pattern conforms more closely to that of the controls, although controls' disfluencies are quite small in number. Overall, the higher occurrence of disfluencies on vowels of pitch-accented syllables in the stutterers' speech, but not in the controls' speech, provides additional support for the prediction of the first hypothesis. While controls do produce some disfluencies on syllable nuclei, only a fairly small proportion are produced with pitch accents, suggesting that stutterers' anticipatory disfluencies may be qualitatively different from controls'.

### *3.3. Disfluencies and prominence level*

Stutterers and controls differed considerably in the distribution of disfluencies across unaccented, accented, and nuclear-accented words. As Figure 5 shows, all three stutterers consistently produced disfluencies on a higher percentage of both pitch-accented and nuclear pitch-accented words than they did of unaccented words. Overall, 70% of stutterers' pitch-accented words were disfluently produced, compared with 31.5% of NPAs and 23.2% of unaccented words. Controls showed little variation across prominence levels, producing disfluencies on 9.8% of PAs, 8.6% of NPAs, and 11.9% of unaccented words.

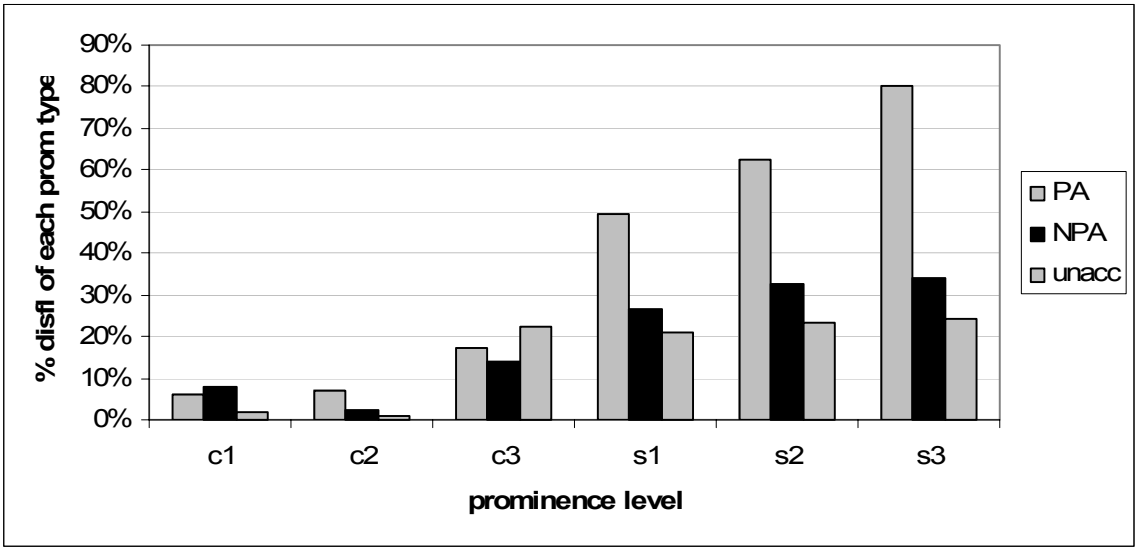


**Figure 5: The percent disfluent of each prominence type: controls vs. stutterers**

These results provide only partial support for the prediction of the second hypothesis: disfluency rate was highest in metrically prominent positions (i.e., in pitch-accented words), though it was expected that nuclear pitch accents would have attracted the highest disfluency rate.

Nevertheless, it is feasible that when unaccented words are produced disfluently, they appear pitch-accented as a result of the disfluency itself, such as in a vowel prolongation. In the same way, words which are designated to receive pitch accent in the early stages of language production may ultimately not receive one as a result of disfluency-induced prosodic reorganization. Nuclear pitch accents, in particular, are vulnerable to phrasal reorganizations, since, by definition, they occur in ip-final position: a disfluency which forces prosodic restructuring, then, could force a re-assigning of nuclear pitch accent.

It was therefore even more important to use a reference prosody to which natural prominence patterns could be compared. Table 4 lists the individual prominence disfluency data for all groups: stutterers (Nat S1-S3), their reference prosody (Ref C1-C3), and controls (Nat C1-C3). For subjects S1 and S3, pitch-accented and nuclear pitch-accented words were disfluency targets in the script-interpreted reference prosody (Ref C1 and Ref C3) more often than in the natural data (Nat S1 and Nat S3), while unaccented words were targets less often. Thus, the key pattern found in the natural data—disfluencies occurring more often on pitch-accented than unaccented words—was again supported by the script interpretations, with the exception of the reference prosody (Ref C2) for S2.

**Table 4: Comparison of prominence disfluency percentages for stutterers (Nat S), reference prosody (Ref C) and controls (Nat C).**

| | S1/C1 | | | S2/C2 | | | S3/C3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Nat S1 | Ref C1 | Nat C1 | Nat S2 | Ref C2 | Nat C2 | Nat S3 | Ref C3 | Nat C3 |
| PA disfl | 103 | 131 | 8 | 98 | 64 | 13 | 444 | 468 | 22 |
| % | 49.3% | **55%** | 6.2% | 62.4% | 38.1% | 7.2% | 80% | **94.5%** | 17.2% |
| NPA disfl | 95 | 77 | 10 | 50 | 30 | 3 | 237 | 240 | 21 |
| % | 26.6% | **33.2%** | 7.9% | 32.7% | **33.3%** | 2.6% | 33.8% | **72.9%** | 13.9% |
| unacc disfl | 136 | 80 | 4 | 61 | 78 | 2 | 375 | 207 | 97 |
| % | 21.1% | 15% | 1.7% | 23.4% | **31.2%** | 1% | 24.1% | 20.6% | 22.4% |

### *3.4. Disfluencies and phrasal position*

In order to analyze the results with respect to phrasal position, all data were organized into three phrasal classes: ip-initial, ip-medial, and ip-final[2]. In the group comparisons, shown in Figure 6, stutterers produced the highest disfluency level on ip-initial position (24.8%), followed by medial (17.9%) and final (10.6%). This trend is drastically reversed in the reference data interpretation, as the target disfluency percentage is highest for medial (46.5%) and final (47%) positions, and significantly lowered for initial

position (19.1%). Once again, controls did not reveal marked disfluency differences among the three phrase locations.

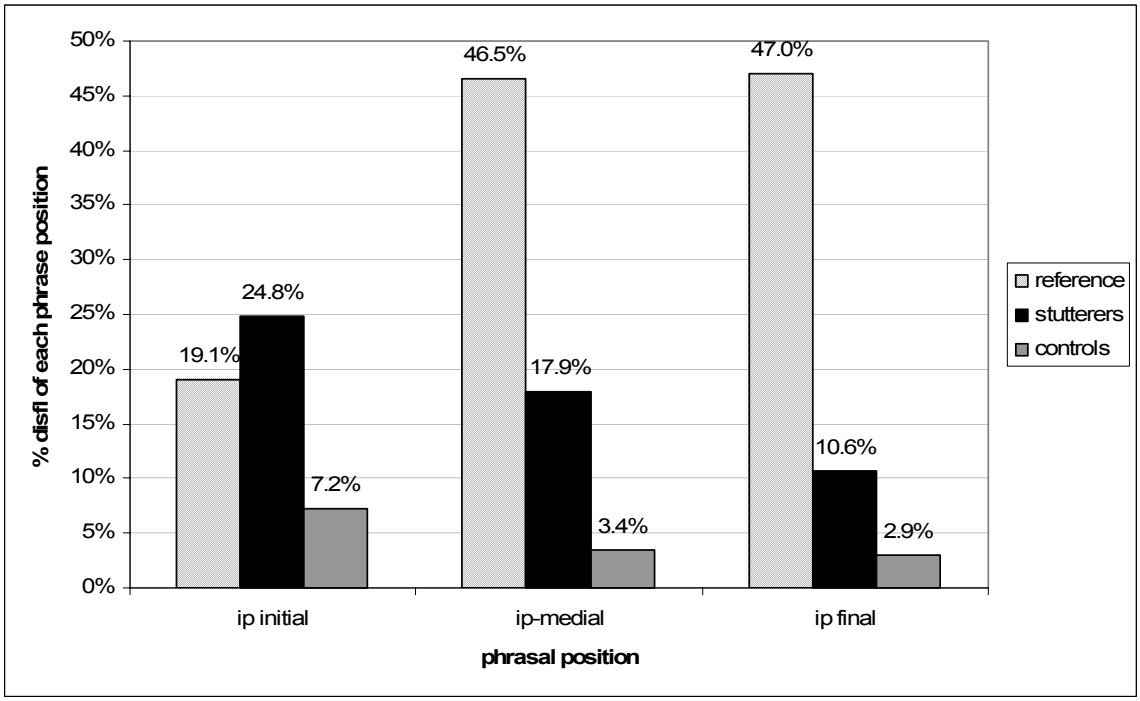


**Figure 6: The percent disfluent of each phrase position: reference vs. stutterers vs. controls.**

In Table 5, the breakdown of the individual data reinforces the overall group trends observed in Figure 6. Although the reference disfluency percentages differ slightly among subjects, the consistent result is that target disfluencies are highest in medial and final positions, contrasting with the higher relative percentage in initial position for the stutterers' natural data. Thus, the prediction from the second hypothesis—that disfluencies would surface more frequently on high-prominence positions—was not supported by the natural data, but the phrasal reinterpretation of the reference prosody suggests that both ip-medial and ip-final positions are in fact underlying triggers of disfluencies.

**Table 5: Comparison of prominence disfluency percentages for stutterers (Nat S), reference prosody (Ref C) and controls (Nat C).**

| | S1/C1 | | | S2/C2 | | | S3/C3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Nat S1 | Ref C1 | Nat C1 | Nat S2 | Ref C2 | Nat C2 | Nat S3 | Ref C3 | Nat C3 |
| ip-initial disfl | 57 | 50 | 2 | 37 | 17 | 1 | 156 | 48 | 24 |
| % | 20.5% | 23.1% | 1.6% | 28.9% | 21.3% | 0.9% | 26% | 15.7% | 17.4% |
| ip-medial disfl | 77 | 169 | 0 | 43 | 131 | 8 | 312 | 666 | 23 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| % | 13% | **30.8%** | 0% | 14.7% | **38.5%** | 3% | 20.4% | 56% | 5.4% |
| ip-final disfl | 31 | 65 | 1 | 16 | 23 | 1 | 60 | 195 | 9 |
| % | 11.2% | 30.1% | 0.8% | 12.5% | 28.8% | 0.9% | 10% | **63.7%** | 6.5% |
| # initial / final wds[3] | 278 | 128 | 601 | 122 | 115 | 138 | 216 | 80 | 306 |
| # medial wds[4] | 591 | 292 | 1532 | 222 | 268 | 423 | 548 | 340 | 1190 |

Analyzing the distribution of disfluencies with respect to three phrasal positions—initial, medial, and final—ultimately showed that, for each subject, significantly fewer disfluencies were generated in a *planned* ip-initial position than they were in ip-initial positions from the actual narrative. In other words, accounting for the effect of disfluencies on the surface ip structure, it was shown that ip-initial position was an infrequent locus of underlying disfluencies. However, analyzing disfluency distribution as a function of only three possible locations may have resulted in an artificially high representation of ip-medial disfluencies. The reason for this is that while initial and final positions were restricted to single words, medial position included all possible positions in between. For instance, while each word of a three-word ip would correspond to exactly one of the three ip positions, a six-word ip would result in an unequal correspondence: one word each would correspond to initial and final positions, but the four words in between would be necessarily assigned to medial position.

In order to decompose this phrase-medial prosodic material into analyzable constituents, Phonological Phrases (PhPs) were identified within each speaker's intermediate phrases as interpreted by the reference prosody. Like Prosodic Words, Phonological Phrases are prosodic structures which are often smaller than an intermediate phrase, and which include a head (pitch accent) and its complementary prosodic material (Selkirk 1986; Nespor & Vogel 1986; Hayes 1989). For the purpose of this analysis, Phonological Phrases were defined operationally as crucially differing from Prosodic Words with respect to constituency: constituents of Phonological Phrases need not be cliticized to the pitch-accented head word. As Figure 7 shows, although the function word "the" is not cliticized to "frog", all three words are interpreted as belonging to a single Phonological Phrase.

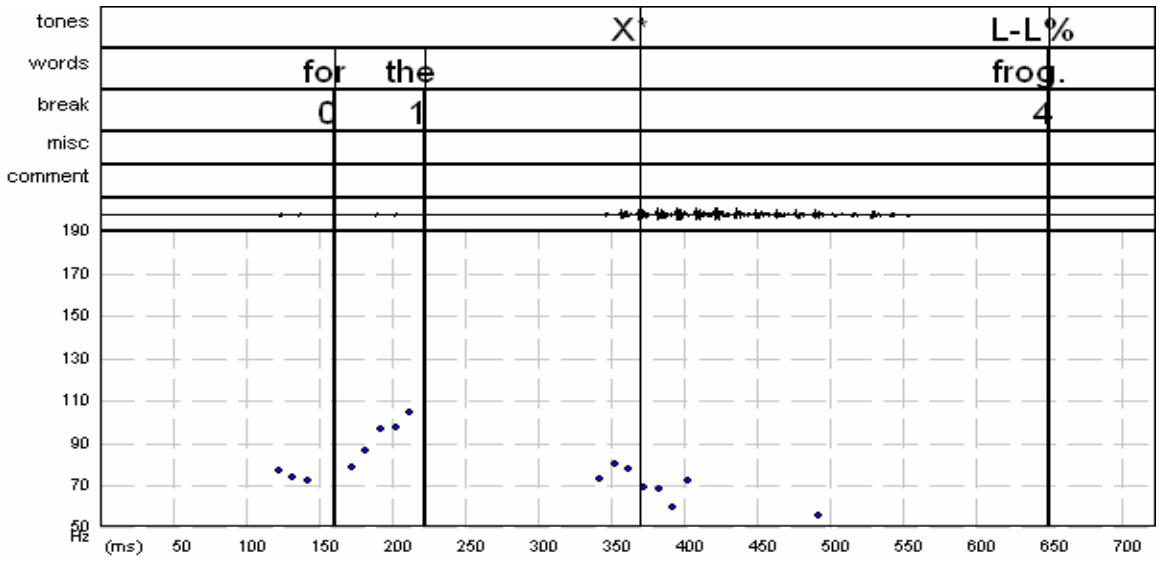


**Figure 7: An example of a Phonological Phrase, where all satellite words ("for", "the") are included with the nucleus ("sleep") as constituents of the same phrase.**

Phrase length ranged from a single PhP (1 pitch accent) to six PhPs (six pitch accents). Figure 8 shows the proportion of total target disfluencies located in each PhP for all three reference scripts combined, beginning with the smallest possible ip where disfluency location could vary (i.e., ips consisting of two pitch accents), and ending with ips containing six pitch accents.

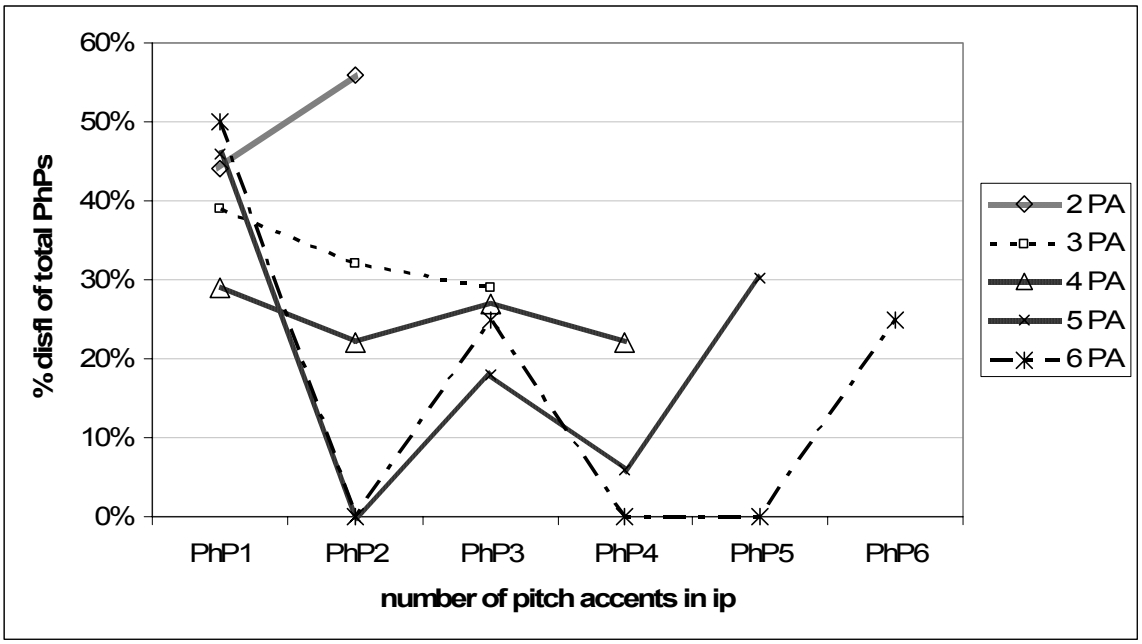


**Figure 8: Target disfluency rate for Phonological Phrases (PhP), produced within intermediate phrases consisting of between 2 and 6 pitch accents (PA).**

As the graph illustrates, with the exception of ips consisting of two pitch accents, for all other ip types, the first Phonological Phrase contained the highest percentage of target disfluencies. Interestingly, beginning with ips consisting of 3 PAs, the difference between the disfluency rates for the first and second PhPs gradually increased as the number of pitch accents per ip increased. Specifically, while the first PhP maintained a steady rate of disfluency through all phrase lengths, target disfluencies were decreasingly found in the second PhP position as the total number of PAs increased. The third and fourth PhPs revealed similar increasing differences of disfluency distribution: while a relatively stable percentage of target disfluencies surfaced in the third PhP regardless of ip length, the relative percentage of target disfluencies in the fourth PhP decreased as ip length increased. The final noteworthy effect was the consistent percentage of final PhPs that contained target disfluencies. For all ips regardless of length, the final PhP was the locus of a target disfluency at least 25% of the time.

In short, an analysis of target disfluency distribution within ips divided into Phonological Phrases reveals a more nuanced picture. While the coarser analysis illustrated in Figure 6 showed that the initial position of the ip, as interpreted by the reference prosody, was the location of few disfluencies overall, Figure 8 shows that a relatively high percentage of target disfluencies occurred in the initial Phonological Phrase for ips of all lengths. A similarly high percentage of target disfluencies surfaced in final PhP position. Finally, with respect to other medial positions, while a steady percentage of disfluencies occurred in the third PhP regardless of ip length, decreasing percentages of target disfluencies surfaced in the second and fourth PhPs as ip length increased.

## 4. Discussion

The results of this experiment in general supported the first hypothesis that stutterers would generate a higher percentage of "anticipatory" disfluencies than would controls. The results of the syllable position analysis provided strong evidence supporting a categorical distinction between target and anticipatory disfluencies: namely, stutterers produced a significantly higher percentage of their disfluencies on the nuclei (38.3%) of words than did controls (23.7%). Moreover, these disfluencies also attracted pitch accents more often (29%) than did those of controls (12%), providing further evidence of anticipatory disfluencies in advance of underlying ones. The overall effect was mitigated, however, by the fact that one of the three stuttering speakers produced a comparatively low number of disfluencies on syllable nuclei.

Prominence comparisons revealed that all stutterers consistently produced more disfluencies on accented than on unaccented words, while controls showed no consistent pattern. However, contrary to the prediction of Hypothesis 2, pitch accents were produced with more disfluencies than were nuclear pitch accents. When the natural data were compared with the script data, the results revealed that high-prominence words (i.e., pitch-accented and nuclear pitch-accented) occurred much more frequently in script-interpreted target position than did unaccented words, with nuclear pitch-accented words receiving the highest frequency of disfluencies for two of the stuttering subjects. This result only partially supported the second hypothesis, which predicted that nuclear pitch-

accented words would attract the highest disfluency rate, followed by pitch-accented and unaccented words, respectively.

Finally, with respect to phrasal position, the natural data analysis did not support the first hypothesis, which predicted that ip-final position—the default location of nuclear pitch accent—would generate the highest rate of disfluency. However, because disfluencies often resulted in premature termination of planned ips, a reference prosody was again used to hypothesize a representation of the underlying prosodic phrase structure. For all three scripts, target disfluencies occurred with highest frequency on accented forms, and on NPAs for two of the scripts.

A finer-grained analysis of Phonological Phrase distribution suggested evidence of stutterers' access to larger prosodic structure in their detection of these disfluencies. While target disfluencies occurred most often in the first and final PhPs of an intermediate phrase, smaller effects were manifested consistently in other PhPs depending on the number of PhPs (i.e., phrases with a PA as their nucleus) that were generated in the ip. Specifically, disfluencies occurred in the second and fourth PhPs less frequently as total PA number increased, while the number of disfluencies produced in the third PhP remained consistent regardless of ip length.

The results shown here, which reveal that stutterers appear to be sensitive to prosodic breakdowns well before articulation of the problematic material has ensued, do suggest that speakers have access to prosodic structures as large as an intermediate phrase before fully retrieving the phonological content. This evidence therefore favors a non-incrementalist prosody generation model (e.g., Ferreira 1993; Keating & Shattuck-Hufnagel 2002) similar to the model outline above. The fact that stutterers frequently produce disfluencies both in prosodically predictable anticipatory and target positions supports the hypothesis that prosody generation is impaired at an early level for stutterers.

In addition, the fact that control subjects demonstrated significantly fewer anticipatory disfluencies than stutterers—as well as little evidence of consistent disfluency distribution patterns with respect to prominence level and phrasal position—implies that the triggering of disfluencies in normal speech originates not in pitch accent assignment or prosodic structuring, but rather in the types of speech errors discussed in Levelt (1983). For example, control subjects overall produced more disfluencies on syllable onsets, suggesting a speech error originating not in the computation of prominence patterns, but rather in the later process of phonological encoding. On the other hand, tip-of-the-tongue errors often signify a semantic error produced in earlier conceptual planning stages. Crucially, however, in errors such as these there is no reliable evidence of prosodically constrained distributional patterns, as the evidence in both anticipatory and target disfluencies in stutterers has shown.

The findings of this paper are also compatible with the predictions of both a covert monitoring hypothesis as proposed by Postma & Kolk (1993), which assumes an impairment rooted in phonological encoding, as well as a hyper-vigilant monitoring hypothesis (Vasic & Wijnen 2005), which locates the dysfunction in the monitor itself. However, a third possibility which combines aspects of both hypotheses might also be considered. On the one hand, the results in this paper support an interpretation of stuttered disfluency production as a prosodic deficit; however, rather than originating in a breakdown in phonological encoding, the results here suggest an intonationally-rooted impairment—specifically, a failure of speakers to properly build a prosodic structure

around metrically prominent events (i.e., pitch accents). At the same time, the role of a speech monitor in detecting these prosodically vulnerable points is strongly motivated by the high percentage of anticipatory disfluencies found at predictable points in the intonational structure. Furthermore, the possibility that the actual stuttering impairment is located ultimately in the activity of the monitor itself—i.e., that the monitor, responding to intonationally prominent phenomena, is improperly interpreting this information as disfluent—is equally compatible with the results presented here.

Given that much of what constitutes disfluency evidence is either indirect or opaque, it is also possible that the processes underlying stuttering are automatic and only tangentially related to higher-level language processing. Models such as the Motor Plan Hypothesis stress the physiological etiology of stuttering, and involves distinct levels of motor plan assembly and muscle plan preparation which may be disrupted in stutterers' speech (van Lieshout, et al. 1996). Although the linguistic prosody central to the present analysis is generally assumed to differ categorically from "non-linguistic" properties of prosody associated with motor timing, there is nevertheless a strong relationship between the two. It is well-documented in previous stuttering literature that activities which ensure rhythmic predictability, such as singing, reading in chorus, and reading when accompanied by a metronome, have ameliorative effects on stuttering (e.g., Stager, et al., 1997). In their account of stuttering and monitoring, Vasic & Wijnen (2005) suggest that these forms of external timing function to distract the overly vigilant speech monitor of stutterers—namely, by requiring the monitor to be attentive to the external rhythm and ensuring input-output alignment, rather than attend to linguistically salient information of a normal speech plan that would otherwise distract it. In lieu of such an external meter, the authors argue, stutterers' speech monitors will inevitably apply excessively strict temporal and rhythmic constraints to the production plan, resulting in self-interruption and thus disfluency.

Postulating a prominent role of rhythmicity constraints on stutterers' speech monitoring is compatible with the results found in this paper. On several occasions, all three speakers produced utterances which, while entirely fluent segmentally, were prosodically anomalous in creating a highly regular disyllabic foot structure. Particularly interesting is the fact that speakers generated these examples spontaneously, as opposed to previous experiments which prompted speakers to follow an externally set meter.

The fact that rhythmically predictable meter is conducive to fluent production in the speech of stutterers may underlie a prosodic breakdown that is fundamentally timing-based. According to an influential model of prosody generation proposed in Ferreira (1993), prosodic constituents—specifically, Prosodic Words (PWds)—are assigned abstract timing intervals crucially before any segmental content has been assigned to the words. Metrical grids are constructed for each PWd, with the prominence level of each determined by its function in the entire prosodic structure: for instance, a contrastively focused PWd would be assigned a higher prominence level than all other PWds; similarly, an utterance-final PWd, by virtue of its occurrence at the end of the most prosodic constituents (i.e., since it terminates a PWd, a Phonological Word, and an utterance) should be assigned the highest prominence level of the utterance. Finally, these individual metrical grids are used by the Prosody Generator to form a metrical grid for the whole utterance. It is possible that this generation of a larger grid is somehow impaired for stutterers, as it requires integration of metrical grids which vary in size of

timing interval. Attempting to produce a predictable rhythmic structure, however, consisting of metrical grids sharing both stress and timing interval size, would presumably simplify the grid-generation process, as well as the metrical structure of the grid itself, thereby avoiding stuttered disfluencies.

The timing component posited by Ferreira (1993) would have important ramifications for a prosodic impairment. As mentioned above, if Prosodic Words are the constituents of timing in a prosodic structure, and a metrical grid is created from the larger prosodic structure formed by these PWd constituents, then presumably a breakdown in any one of these constituents would disturb the execution of global or local timing patterns planned within the utterance.

In summary, a strong case can therefore be made for the following claims: 1) minimally, speakers have access to the entire intermediate phrase, although the internal structures and prominence relationships are planned in incremental chunks spanning two Phonological Phrases; and 2) while normal disfluencies are produced as a result of errors in either conceptual or phonological encoding, stuttered disfluencies are triggered by errors in prosody structure generation—specifically, in building intermediate phrases and, to a lesser degree, building smaller units composed of two PhPs.

## 5. Conclusion

Thus, in conclusion, prosodic prominence relations appear to play an important role in triggering stuttered disfluencies. These results are consistent with both the hypothesis that triggers of disfluency originate in points of high prominence, as well as the related hypothesis that such triggers are a part of larger prosodic phrases whose internal structure is crucial in determining when a stuttering speaker detects an impending breakdown.

## 6. References


Au-Yeung, James; Howell, Peter; and Pilgrim, Lesley (1998). Phonological words and stuttering on function words. *Journal of Speech, Language, and Hearing Research*, 41, 1019-1030.

Beckman, Mary E. & Pierrehumbert, Janet (1986). Intonational Structure in Japanese and English. *Phonology Yearbook* 3, 255-309.

Beckman, Mary E. & Ayers, Gayle (1994). Guidelines for ToBI labeling. *Language and Cognitive Processes*, 11(1/2), 17-67.

Beckman, Mary E. & Hirschberg, Julia (1994). *The ToBI Annotation Conventions.* Unpublished manuscript, Ohio State University and AT&T Bell Telephone Laboratories.

Blackmer, Elizabeth & Mitton, Janet (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39, 173-194.

Bosshardt, Hans-Georg (1993). Differences between stutterers' and nonstutterers' short-term recall and recognition performance. *Journal of Speech and Hearing Research*, 36, 2 286-293.

Brown, Spencer F. (1938). Stuttering with relation to word accent and word position. *Journal of Abnormal & Social Psychology*. Volume 33, 112-120.
Burger, Remca & Wijnen, Frank (1999). Phonological Encoding and Word Stress in Stuttering and Nonstuttering Subjects. *Journal of Fluency Disorders*, 24, 91-106.
Caramazza, Alfonso. 1997. How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14, 177–208.
Cutler, Anne (1983). Speaker's Conceptions of the Functions of Prosody. In: A. Cutler and D.R. Ladd, (Ed.), *Prosody: Models and measurements*. Springer, Heidelberg.
Dell, Gary S. & Reich, Peter A. (1980). Toward a unified theory of slips of tongue. In V.A. Fromkin (Ed.). *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand.* New York: Academic.
Dell, Gary S. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science*, 9, 3-23.
Ferreira, Fernanda (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 2, 233-253.
Fosnot, Susan & Jun, Sun-Ah. (1999) Prosodic Characteristics in Children with Stuttering or Autism during Reading and Imitation. In *Proceedings of 14th International Congress of Phonetic Sciences (IcPhS '99)*, San Francisco, USA, 1925-1928
Fromkin, Victoria A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27-52.
Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20, 253-306.
Howell, Peter; Au-Yeung, James; and Sackin, Stevie (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language, and Hearing Research*, 42, 2, 345.
Howell, Peter & Sackin, Stevie (2000). Speech rate manipulation and its effects on fluency reversal in children who stutter. *Journal of Developmental and Physical Disabilities*, 12, 291-315.
Keating, Patricia & Shattuck-Hufnagel, Stephanie (2002). A Prosodic View of Word Form Encoding for Speech Production. *UCLA Working Papers in Phonetics*, 101, 112-156.
Kolk, Herman (1991). Is stuttering a symptom of adaptation or of impairment? In: Peters, H.F.M., Hulstijn, W. & Starkweather, C.W. (Eds.), *Speech motor control and stuttering*. Amsterdam: Elsevier/ Excerpta Medica.
Levelt, Willem J.M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.
Levelt, Willem J.M. & Cutler, Anne (1983). Prosodic marking in speech repair. *Journal of Semantics*, 2(2), 205–217.
Levelt, Willem J.M. (1989). *Speaking: From intention to articulation*. Cambridge, Mass: MIT Press.
Levelt, Willem J.M.; Roelofs, Ardi; and Meyer, Antje S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
van Lieshout, Pascal H.H.M.; Hulstijn, Wouter; and Peters, Herman F.M. (1996). Speech production in people who stutter: testing the Motor Plan Hypothesis. *Journal of Speech and Hearing Research*, Volume 39, 76-92.
Mayer, Mercer (1969). *Frog, Where are You?* New York: Dial Books.

Melnick, Kenneth & Conture, Edward (2000). Relationship of length and grammatical complexity to the systematic and nonsystematic speech errors and stuttering of children who stutter. *Journal of Fluency Disorders*, Volume 25, Issue 1, 21–45.
Natke, Ulrich; Grosser, Juliane; Sandrieser, Patricia; and Kalveram, Karl T. (2002). The duration component of the stress effect in stuttering. *Journal of Fluency Disorders*, Volume 27, Issue 4, 305-318.
Nespor, Marina & Vogel, Irene (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.
Pitchworks Software, version 6.4.n. (1999-2003). Sciconrd.com.
Pierrehumbert, Janet (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT, published 1988 by IULC.
Pierrehumbert, Janet & Beckman, Mary E. (1988). *Japanese Tone Structure*. The MIT Press, Cambridge, MA.
Postma, Albert; Kolk, Herman; and Povel, Dirk-Jan (1990). Speech planning and execution in stutterers. *Journal of Fluency Disorders*, 15, 49–59.
Postma, Albert & Kolk, Herman (1992a). The effects of noise masking and required accuracy on speech errors, disfluencies and self-repairs. *Journal of Speech and Hearing Research*, 35, 537-544.
Postma, Albert & Kolk, Herman (1992b). Error monitoring in people who stutter. Evidence against auditory feedback defect theories. *Journal of Speech and Hearing Research*, 35, 1024-1032.
Postma, Albert & Kolk, Herman (1993). The Covert Repair Hypothesis: Prearticulatory repair Processes in Normal and Stuttered Disfluencies. *Journal of Speech and Hearing Research*, 36, 472-487.
Prins, David; Hubbard, Carol P.; and Krause, Michelle (1991). Syllabic Stress and the Occurrence of Stuttering. *Journal of Speech, Language, and Hearing Research*, 34, 5, 1011-1016.
van Rossum, Maya (2005). *Prosody in Alaryngeal Speech*. Utrecht: Utrecht Institute of Linguistics.
Selkirk, Elisabeth (1984). *Phonology and Syntax: The Relation between Sound and Structure*, The MIT Press, Cambridge, MA.
Selkirk, Elisabeth (1986). On derived domains in sentence phonology. *Phonology*, 3, 371-405.
Shattuck-Hufnagel, Stephanie (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum.
Shriberg, Elizabeth (1994). *Preliminaries to a Theory of Speech Disfluencies*. PhD thesis, University of California at Berkeley.
Shriberg, Elizabeth (1999). Phonetic consequences of speech disfluency. In *Proceedings of the International Congress of Phonetic Sciences* (ICPhS-99), San Francisco, Vol. I, 619-622.
Stager, Sheila V.; Denman, Daniel W.; and Ludlow, Christy 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.

Vasic, Nada & Wijnen, Frank (2005). Stuttering as a monitoring deficit. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, and F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech*. Hove (East Sussex): Psychology Press.

Viswanath, Nagalapura (1991). Temporal structure is reorganized when an utterance contains a stuttering event. In Peters, H.M., Hulstijn, W., and Starkweather, C.W. (Eds.), *Speech Motor Control and Stuttering*. Elsevier Publishers: Amsterdam, 341-346.

Wijnen, Frank & Boers, Inge (1994). Phonological Priming Effects in stutterers. *Journal of Fluency Disorders*, 19, 1-20.

Wingate, Marcel E. (1976). *Stuttering: Theory and Treatment*. New York: Irvington-Wiley, (Chap. 9).

[1] In onsetless syllables, this would be determined by successfully initiating production of the nucleus.

[2] This comparison did not include pauses, since ip-initial vs. ip-final location is not a possible distinction for pauses.

[3] The total number of ip-initial and -final words was determined by summing the number of ips consisting of at least two words.

[4] The total number of ip-medial words was determined by summing the number of medial words for all ips consisting of at least three words.