UNIVERSITY OF CALIFORNIA

Los Angeles

An Intonational Analysis of Disfluency Patterns in Stuttering

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Linguistics

by

Timothy Richard Arbisi-Kelm

2006

The dissertation of Timothy Richard Arbisi-Kelm is approved.

Susan Curtiss

Mario Mendez

Carson Schütze

Sun-Ah Jun, Committee Chair

University of California, Los Angeles

2006

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It goes without saying that there are many people who have helped me reach my academic goals during my time at UCLA. First of all, I would like to express deep gratitude to my committee chair, Sun-Ah Jun, for her guidance, support, and faith in my abilities. It was through her inspiration that I stumbled upon the theoretical questions that eventually grew into this dissertation. I will always appreciate the confidence and patience she has maintained in my development as a researcher. She made it all possible.

Also, I would like to thank each of my other committee members—Susie Curtiss, Carson Schütze, and Mario Mendez—for their unique contributions to my research, and for their tireless support and enthusiasm. Their expertise and training have resulted in the formation of my research abilities, and motivated me to look at difficult problems from new perspectives. I owe many thanks to the other faculty members of the Phonetics Lab—Pat Keating, Bruce Hayes, Kie Zuraw, and the late Peter Ladefoged—for listening to my ideas and offering their brilliant insights.

I wish to also thank all the other faculty members whom I worked with, or who helped me navigate the challenges of my graduate education: Anoop Mahajan, Tim Stowell, Nina Hyams, Ed Stabler, Philippe Schlenker, and Colin Wilson, among others. And finally, I thank stuttering researcher Dr. Gerald Maguire, whose personal journey and groundbreaking discoveries in stuttering gave me a hopeful model and inspiration for my own work in this field.

To all of my good friends and colleagues, I thank you for your camaraderie and insights. To Kuniko Nielsen, for sharing coffee and ideas with me; to Christina Esposito, for sharing Sicilian recipes and ideas with me. Thanks also to other Lab members Andy Martin, Molly Shilman, Sameer Khan, Lauren Varner, Ingvar Lofstedt, and Katya Pertsova, for your support and interaction, and for putting up with my constant presence dissertating in the "cool room".

When I did make it out of the Phonetics Lab, blinking my eyes in the bright sunlight, I was fortunate to be surrounded by my many thoughtful and inspiring fellow students. Thanks to Alex Hurst, Shabnam Shademan, Greg Kobele, Marcus Smith, Christina Kim, David Schueler, Jeff Heinz, Ben Keil, and Julia Berger-Morales, among others, for always being open to discussing random ideas and linguistic questions. And to Jason Kandybowicz, for his energy and smiles, and our inspired David Lynch talks.

In my years prior to the dissertation marathon, I have fond memories of the many UCLA students who have moved on, motivating me by their great successes. Thanks to Ying Lin, who was always there to give support and advice, and who had the most healthful and delicious-looking lunches; to Mariame Sy, who always greeted me with a smile; to Heriberto Avelino, whose practical jokes kept me on my toes. Also to Sahyang Kim, who was so fun to work with on the dyslexia project and who taught me how to use Praat; and to Rebecca Scarborough, for the exciting research discussions we had in the P-Lab. I wish to thank Leston Buell, for his great sense of humor, and Harold Torrence, who was always waiting for his bagel to be toasted when I arrived each morning in the P-Lab kitchen. Thanks to Stephen Wilson for sharing with me his expertise in so many areas, particularly his knowledge of brain imaging. Also to Adam Albright and Melissa Epstein, fellow P-Lab members who were superb and motivating role models. And finally, to John Foreman, the first to meet me when I arrived in L.A. for the first time, and gave me such a warm welcome to the department.

I owe so much to UCLA staff members Kathryn Roberts, Michael Serdenia, and Natasha Levy, for always being so generous with their time, and helping me meet the challenges of everyday graduate student life.

For hours of her time devoted to coding F0 files, and patiently learning my complicated labeling system, I thank my research assistant, Rachel Schwartz. And of course, because of the enthusiastic participation of the subjects in this study, I was given the opportunity to finally test the hypotheses which motivated this dissertation. In particular, I wish to thank deeply the individual participants who, as lifelong stutterers, know intimately the frustrating and fascinating experience of what it means to lose control of language—and yet have the strength and courage to continue to speak.

I also thank Cambridge University Press, The MIT Press, and Penguin Group (USA) Inc., for their permission in using some of the figures in the text of my dissertation.

As a part-time Angeleno, it feels obligatory to thank the Academy, though I haven't made any movies yet. There's still time.

And in my non-linguistic world, I of course owe everything to my family. My eternal thanks to my parents, Richard and Anna Marie Arbisi, for surrounding me with love, and teaching me to live with enthusiasm and joy and hope. To my sister and brother, with whom I share my oldest and most beautiful memories. To my nieces and nephew, who remind me that life must be lived in a perpetual state of wonder and awe. And to my in-laws, who always give me support and love.

Most of all—yes, the best for last—I owe my deepest gratitude to my beloved wife, Heidi. To Heidi, for the long nights she spent coding data, for her incredible organizational talents, for working as a full-time professional, a full-time student, and still devoting all her time to helping me travel through this long journey. For her sacrifice. For her love. For all that she has helped me to become. For always being at my side.

VITA

March 14, 1971	Born, Rockford, Illinois
1993	B.A., Philosophy St. John’s University Collegeville, Minnesota
1993-1994	Volunteer Translator Christian Foundation for Children and Aging Senahú, Guatemala
2000	Eugene Cota Robles Fellowship University of California, Los Angeles Los Angeles, California
2002	Research Assistant Department of Linguistics University of California, Los Angeles
2003	M.A. Linguistics University of California, Los Angeles Los Angeles, California
2003-2004	Teaching Assistant Department of Linguistics University of California, Los Angles
2005	Research Assistant Department of Linguistics University of California, Los Angeles
2005	Dissertation Year Fellowship Award University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Arbisi-Kelm, T. (2005). An analysis of the intonational structure of stuttered speech. In *Proceedings of Interspeech 2005 - Eurospeech—9th European Conference on Speech Communication and Technology*, Lisboa, Portugal.

Arbisi-Kelm, and Jun, S. (2005). A comparison of disfluency patterns in normal and stuttered speech. In *Proceedings of Disfluency in Spontaneous Speech Workshop*, Aix-en Provence, France.

Arbisi-Kelm, T. (2004). An Analysis of Intonation Patterns in Stuttered Speech. Poster presented at the *Meeting of the Acoustical Society of America*, San Diego, CA, USA.

ABSTRACT OF THE DISSERTATION

An Intonational Analysis of Disfluency Patterns in Stuttering

by

Timothy Richard Arbisi-Kelm

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2006

Professor Sun-Ah Jun, Chair

While previous studies have revealed areas vulnerable to disfluency at the word level in stuttering, identifying the factors responsible for this instability has proved difficult; moreover, the inconsistent results are complicated by a failure to control for the effects of phrasal prosody, which govern such word-level factors as lexical stress. It was hypothesized in this dissertation that stuttered disfluencies are caused by a deficit in building prosodic structure. Specifically, it was predicted that words bearing a higher level of metrical prominence would be more prone to disfluent production in stutterers' speech than words of lower prominence. The second prediction was that, for stutterers,

advance detection of underlying breakdowns in prosodic structure creation (‘target disfluencies’) would result in a categorically different type of disfluency (‘anticipatory disfluencies’), with the latter surfacing in predictable locations of an intonation phrase—crucially, prior to the target disfluency. For control subjects, it was hypothesized that while some anticipatory disfluencies would surface, they would do so less frequently than for stutterers, and in prosodically unpredictable locations.

In a story-telling task performed by four stutterers and four control subjects, it was found that, for stutterers, metrically prominent words attracted a higher rate of disfluencies than did unaccented words. In addition, analyses of the prosodic context of each disfluency revealed that stutterers produced significantly more anticipatory disfluencies than did controls.

In order to determine more accurately the location of stutterers’ target disfluencies, a reference prosody was collected by instructing control subjects to read scripts of stutterers’ narrations. Comparisons of script results with the stutterers’ natural data further confirmed that target disfluencies occurred most often on pitch-accented words—particularly the first and final pitch-accented word of an intermediate phrase. The consistent location of disfluencies on roughly every other pitch-accented word, meanwhile, suggested the constraining influence of smaller prosodic structures on disfluency distribution patterns.

It was concluded that while normal disfluencies are produced as a result of errors in either conceptual or phonological encoding, stuttered disfluencies are triggered by an impairment in prosodic structure generation—specifically, in building intermediate phrases and, to a lesser degree, smaller units within the ip.

CHAPTER 1

INTRODUCTION

1.1 Stuttering as a linguistic disorder

Even today, after the remarkable advances in brain imaging technology over the past few decades, it is still not a universally accepted claim that developmental stuttering is a properly linguistic disorder. That is to say, while imaging techniques such as PET and fMRI continue to improve in spatial resonance, how to interpret the abnormal bihemispheric activation patterns identified during stuttered speech is still an unresolved question in stuttering research (Fox, et al., 2000; Ingham, et al., 1996, 2000; Braun, et al., 1997). For instance, whether these activation patterns reflect the underlying breakdown itself or rather a motoric derivative of stuttering cannot be demonstrated given the temporal limitations of fMRI (Blomgren, et al., 2003). Even if we can conclude that the disproportionate bilateral activation is pre-articulatory, however, the debate shifts to whether the root problem originates in lower-level malfunctioning of motoric timing or in higher-level cognitive processes, such as phonological encoding or lexical retrieval. In short, even in tracing the etiology of stuttering to a clearly neurogenic abnormality, we are still quite far from determining the nature of stuttering.

Such persistent limitations with brain imaging data make it more imperative that other methods of exploring stuttering be availed of. While strictly neurological studies of language disorders are still in their relative infancy, properly analyzing the linguistic evidence that accompanies processing breakdowns dates back to at least the early part of the 20th century with S.F. Brown's groundbreaking analyses of linguistic factors such as segmental quality and word length (Brown 1938). Since then, the majority of research concerning potential linguistic factors of stuttering has converged on two key areas: sentence planning and word-level planning. Factors involved in sentence planning have borne conflicting results with respect to whether syntactic complexity—as defined, for example, by the number of embedded clauses within an utterance (Kleinow & Smith 2000, Kadi-Hanifi & Howell 1992)—or utterance length (Yaruss 1999; Maner, et al., 2000; Tornick & Bloodstein 1976; Melnick & Conture 2000) is the more crucial factor in determining a form's susceptibility to disfluent production.

Similarly, while there exists clear evidence of areas of processing vulnerable to disfluency at the word level, identifying the specific factors responsible for this instability has proved difficult. For example, many previous studies have revealed a direct association between higher disfluency rate and the presence of word onsets (Brown 1945; Hahn 1942; Quarrington, et al., 1962), yet other studies found no such effect (Soderberg 1962; Dworzynski, et al., 2003). Analyzing subsegmental evidence has yielded similarly inconclusive results: correlations between disfluency rate and voice onset times (VOTs) are well-founded in the literature, as several experiments have shown that stutterers appear to produce longer VOTs in fluent speech than do nonstutterers (Agnello 1975; Adams & Hayden 1976; Hillman & Gilbert 1977; Healey & Ramig 1986). However, this result was not reproduced in a study where subjects uttered nonsense syllables (Watson &

Alfonso 1982), while Boutsen (1995) found this VOT lengthening difference was significant only in onsets of stressed syllables.

Finally, while studies of lexical stress have been more successful in uncovering significant correlations between stressed syllables and disfluently produced syllables (Natke, et al., 2002; Prins, et al., 1991; Brown 1938), it is nevertheless uncertain precisely how a stressed syllable might elicit a disfluent production (see 2.2.2.3.2).

However, these previous studies did not control phrase-level factors, and it is possible that their word-level effects are influenced by phrasal prosody. 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 word-level. The highest level of metrical stress in this model is the nuclear pitch accent, located in the rightmost position of an English intermediate phrase. Assuming this framework we could therefore hypothesize that any link between stuttering and factors such as lexical stress and longer VOTs is not an indication of difficulty with these word-internal factors per se, but rather of a general instability in producing prosodic prominence; that is, disfluency is triggered not solely by physical factors such as the amount of airflow but also by the semantic and syntactic factors represented in intonational structure. Furthermore, exploring stuttering from a prosodic perspective makes it possible to account for both word-level and sentence-level factors, as well as the interactions among them.

A recent experiment conducted by the author explicitly tested this claim that disfluency is directly correlated with prominence-lending intonational events (Arbisi-Kelm 2004). In a sentence-reading task, stutterers produced significantly more disfluencies on metrically prominent words than on non-prominent words. It was shown that, with the exception of contrastively 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 (Beckman & Pierrehumbert 1986; Beckman & Hirschberg 1994). Since nuclear pitch accents bear the greatest level of prominence, they therefore attract the highest rate of disfluency. However, contrastively focused NPA'd syllables in this experiment were found to be more fluent than non-pitch-accented syllables. This result was surprising since non-focused default NPAs (i.e., the final pitch accent in an ip) were indeed produced with an expected high rate of disfluency. It is possible, however, that the nature of the task—i.e., read speech, rather than spontaneously-produced speech—was largely responsible for this effect. Since speakers were only required to read simple sentences, they had arguably already completed early stages of linguistic processing before initiating speech, thus facilitating production. Additionally, given that lexical content and syntactic structure were already provided to speakers in the form of written sentences, it is not uncertain whether the task was truly testing prosodic planning, since at least some elements of prosodic structure may be accessed pre-lexically (Levelt 1989; Levelt, et al., 1999; Keating & Shattuck-Hufnagel 2002).

Thus, while the results confirmed that both word-level and utterance-level effects can, at least in part, be accounted for by evidence of prosodic impairment at the processing level, determining from a processing perspective what exactly constitutes the deficit is the over-arching question which is yet to be answered. Nevertheless, the results represent further evidence for the still-controversial claim that the disfluencies characteristic of stuttering are indeed ultimately rooted in the linguistic domain.

1.2 Goal of the thesis

The main premise of this thesis is that disfluencies in stuttering are a product of a speaker's impairment in the planning, and subsequent production, of prosodically prominent material. Specifically, I will argue that intonationally prominent events—i.e., pitch accents—are particularly prone to triggering disfluencies, since they constitute the metrically salient events which must be planned in advance in order to form prosodic phrases. Furthermore, since nuclear pitch accents bear the highest prominence level within the metrical prominence hierarchy, I predict that they will trigger the greatest number of disfluencies relative to all other prosodic positions.

The second premise is that disfluencies in stuttering will not only occur at a higher rate than in nonstutterers' speech, but that categorically different types of disfluencies will surface in predictable locations within an intonation phrase. In particular, I will argue that, for stutterers, disfluencies are either underlying (‘target’) or derived (‘anticipatory’), with the latter occurring as a result of the detection of the former. I will attempt to show that ‘target’ disfluencies are associated most often with the most metrically prominent material in an intermediate phrase (i.e., nuclear pitch accents), and that ‘anticipatory’ disfluencies are produced predictably before—and in close proximity to—these ‘target’ disfluencies. For non-stutterers, I predict that while anticipatory disfluencies will surface, they will do so at a lower rate than for stutterers, and in prosodically unpredictable locations.

The investigations of this thesis will bear on two principal questions. The more immediate question is how disfluencies are detected and triggered in both stuttered and normal speech. Although previous stuttering research has produced robust findings linking stuttered disfluencies to breakdowns in linguistic processing, consensus has not been reached regarding which specific components of linguistic processing are impaired. However, three contemporary hypotheses have been particularly influential in motivating recent stuttering literature: 1) stuttering results from a deficit in phonological encoding (Wingate 1989); 2) stuttering results from an impairment in speech production monitoring (Vasic & Wijnen 2005); 3) stuttering results from a breakdown in post-linguistic motor processing. While the first hypothesis posits a deficit in a speaker’s accessing the sounds and/or stress of a given word, the second hypothesis proposes that the defect lies in an “overvigilant” speech monitor’s falsely detecting disfluencies that are not present. It is expected that the results of the experiments in this thesis will expose disfluency type and distribution patterns key to determining at what production stage a breakdown occurs, and what the nature of this breakdown is. The third hypothesis, which ultimately predicts no linguistic role in determining disfluency production, will also be evaluated in terms of the types of disfluency patterns that emerge in the data.

The second principal question to be addressed concerns the planning and generation of prosodic structure in both stutterers' and nonstutterers' speech. In short: how much prosodic structure is constructed before phonological encoding ensues? Analyzing the disfluencies produced in spontaneous speech provides a number of advantages for investigating this question. On the one hand, previous studies have shown that speakers often detect disfluencies in advance of their production (Levelt 1983; Shriberg 1999). Thus, the type and patterns of disfluencies which surface in speech can function as a partial window—or, "lookahead"—to how much of a given utterance a speaker has planned and prepared for articulation. For instance, if a speaker produces the onset of a particular syllable, it can be assumed that the semantic content—and at least some of the phonological form—of the word to which that syllable belonged had been planned.

At the same time, while a wealth of speech error evidence has been utilized to explore this question of lookahead range (e.g., Fromkin 1971; Selkirk 1984; Levelt 1989), similar analyses of speech *disfluency* data—particularly in stuttering—are lacking. Moreover, neither previous speech error nor disfluency studies have analyzed lookahead data from within a model of intonation. The importance of the latter lies in providing a framework through which disfluencies can be analyzed for their prosodic properties and behavior. In this thesis I will use the model of Intonational Phonology (Pierrehumbert 1980; Beckman

& Pierrehumbert 1986; Ladd 1996) to guide the disfluency analysis, in an attempt to provide satisfactory answers to the questions outlined above.

1.3 Organization of the dissertation

The dissertation is organized as follows:

Chapter 2 introduces an overview of developmental stuttering, a literature review of previous stuttering research, and a summary of the Intonational Phonology framework assumed in the present analysis.

Chapter 3 outlines the design structure of two intonation experiments in which stuttering subjects and controls both performed a narration task, and controls completed a related script-reading task in order to provide a reference prosody for closer analysis of the spontaneous narrations by stutterers. A classification and coding system for disfluency types is introduced.

Chapter 4 presents and examines the results of the narrative task for both stutterers and controls. Individual and group comparisons are made with respect to a set of prosodically motivated conditions.

Chapter 5 looks at the results of comparisons made between the natural data analyzed in Chapter 4 and the reference prosody created by controls' readings of stutterers' transcripts.

Chapter 6 summarizes and discusses the work presented in preceding chapters, as well as presents suggestions for future work.

CHAPTER 2

BACKGROUND

2.1 Modeling language production

Analyzing the speech disfluencies proper to disorders such as stuttering as a function of breakdown in word production derives from a rich literature spanning the past several decades. While the assumptions underlying models of language production are diverse—ranging from strictly derivational frameworks (e.g., Levelt 1989; Levelt, et al., 1999) to connectionist activation-spreading networks (e.g., Dell 1997)—the models nevertheless share a principal concern with lexical access and word-form encoding. Suprasegmental and phrase-level phenomena such as prosody and intonation, being considered post-lexical processes, are applied incrementally at word-form spell-out (Levelt 1989).

A Leveltian model of language production (Levelt 1989; Levelt et al. 1999) proposes an order of discrete stages beginning with conceptual preparation and terminating in articulation. Figure 1 shows the order of these stages:

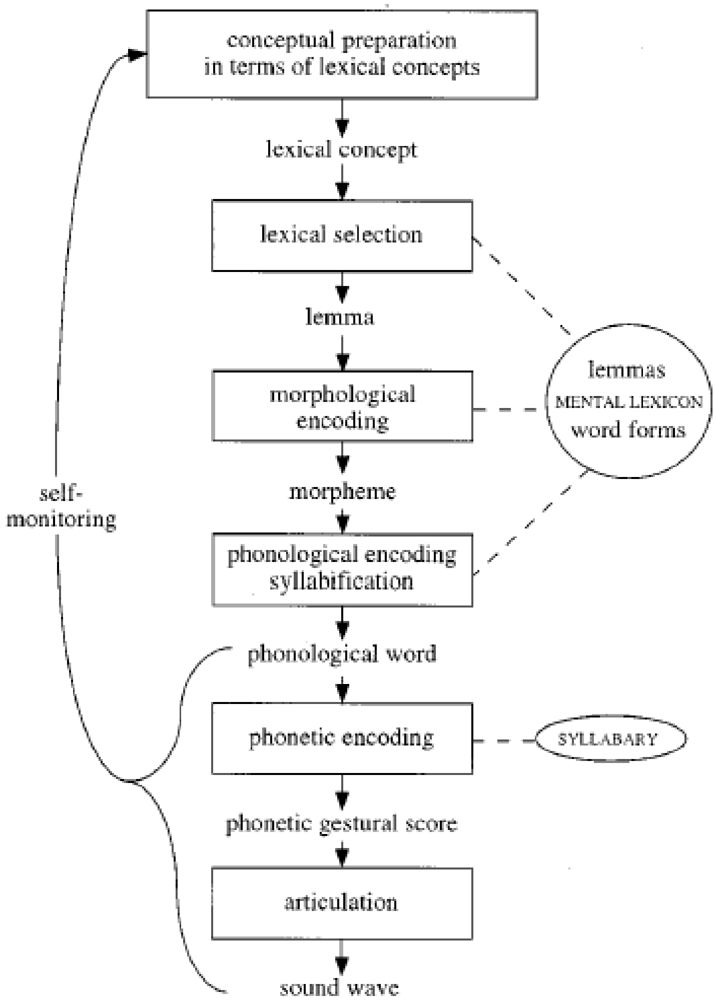


Figure 1: Leveltian stages of language production (Levelt et al. 1999)

As seen in Figure 1, production stages begin with the formulation of a lexical concept, which is then specified for word-level syntactic and semantic information (e.g., gender, number). This abstract 'lemma' then receives its morphological content from separately stored roots and affixes, which are filled in through parallel retrieval of the word form's metrical and segmental information. Finally, this sequence of segments is addressed to a syllabic gestural score, which, depending on the score's frequency, is either stored or generated online; once specified, the gestures are then articulated into speech.

Levelt (1989) models prosodification in greater detail by elaborating the phonological encoding stage. Figure 2 shows the various levels of input and output filtering through what Levelt calls the Prosody Generator, whose role is to incrementally produce a sentence's metrical and intonational structure:

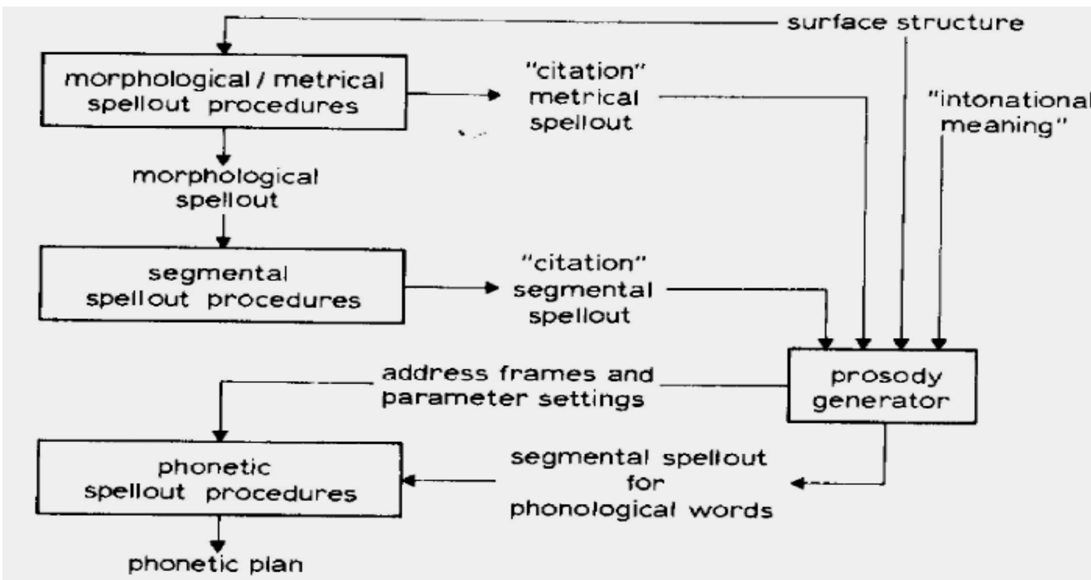


Figure 2: Phonological encoding model (Levelt 1989)

The Prosody Generator receives information via morphological and segmental spellout as well as directly from the surface structure and intonational meaning of the intended utterance. This neatly captures the metrical and intonational aspects of prosodic planning: while segmental spellout provides a phonological syllable plan, phrase-level intonational information is simultaneously integrated, with the output resulting in prosodic words ready for phonetic spellout.

As observed in Keating & Shattuck-Hufnagel (KSH 2002), however, proposing word-form encoding and phrase-level intonation planning as independent processes is not without complications. In particular, Levelt's model assumes an incremental generation of each representation, from the lexical word to phonetic spell-out. Metrical structure, for example, is therefore constructed frame by frame in a left-to-right fashion, and thus surface structure does not need to be accessed in constituents larger than the Prosodic Word. A number of specific cases nevertheless challenge this proposed limited access: for example, it has been shown that the majority of speech errors involving movement of segmental material actually cross Prosodic Word boundaries (e.g., ***k****ey* ***ch****ain*→ ***ch****ee* ***k****ain*) (Fromkin 1971). Levelt (1989) accounts for this by claiming that some speaking contexts--such as in formal or slow speech--allow "lookahead" into the next Prosodic Word. A more challenging case would involve lookahead stretching over more material and boundaries, such as in the example: *the Knicks* [nIks] *and the Celtics* [sEltIks]→ *the* ***S****icks* [**s**Iks] *and the* ***Kn****eltics* [**n**EltIks] (KSH 2002). Not only would an incremental method be further taxed in allowing an even larger buffer to encompass a separate phrasal boundary, but, more interestingly, it could not predict that either misplaced segment would fail to attach to the onset-free position in *and*; in other words, even with expanded lookahead, incrementalism cannot easily account for what appears to be a speaker's access to a larger planning structure (KSH 2002).

Yet another example that challenges incrementalism is found in the process known as Beat Movement (Selkirk 1984), where a speaker must resolve a stress clash by altering the stress pattern of the earlier form: e.g., *sixTEEN DOLLars→ SIXteen DOLLars* (Levelt 1989). Once again, slight relaxation of the buffer limit accounts for a speaker's ability to access the metrical structure of another Prosodic Word, but iteratively applying Beat Movement puts greater demands on the buffer to increase indefinitely: e.g., *sixTEEN JapaNESE INstitutes→ sixTEEN JAPanese INstitutes→ SIXteen JAPanese INstitutes.*

In summary, the representation of prosody in the Leveltian model as a component fed by the output of both word-level phonological encoding and phrase-level surface structure allows lexical and grammatical information to be seamlessly integrated into the model's prosodic structure. Nevertheless, as KSH (2002) show, the assumption that these word-level and phrase-level processes are necessarily independent causes problems in accounting for phenomena that suggest speakers do have access to larger prosodic structures when processing word-forms. An alternative model, proposed in KSH (2002), is one in which articulation is governed not only by word-level phonological representations, but also by phrase-level prosodic structures. In other words, the claim is that higher levels of prosodic constituent structure are available to a speaker even during encoding of the Prosodic Word, thus entailing a "lookahead" feature that can reach up into these larger constituents, as demonstrated in the previous examples.

The question of how much prosodic structure is available before encoding of the Prosodic Word is directly relevant to the question of determining points of breakdown in disfluent speech. If disfluencies proper to stuttering are ultimately triggered prosodically—rather than strictly determined by a word's lexical features—then we would expect more disruptions to occur at prosodically predictable positions in sentence production, such as on metrically prominent words.

2.2 Stuttering background

Developmental stuttering is a language disorder characterized by chronic disruptions in speech manifested as sound blocks, repetitions, and prolongations (DSM-IV-R, 1994). Identifying the precise nature of the impairment has proved a challenging task, as reflected by the wide range of methodologies with which stuttering has been investigated over the past century (Wingate 1988). A long-standing debate in the field has concerned the degree to which either linguistic factors or motoric factors constitute the underlying causes of stuttering. The advent of more sophisticated tools for charting language production have shown increasing evidence for the former (Blomgren, et al., 2003; Ingham 2001). The following section provides a short summary of this research.

2.2.1 Neurological substrate

Experiments exploring neurological factors associated with stuttering have increased dramatically in recent years, given the rapid advancements in PET and fMRI neuroimaging. Brain-scanning experiments over the past ten years have repeatedly demonstrated that neural activity in stuttering differs markedly from that in normal language production (Fox, et al., 2000; Ingham, et al., 1996, 2000; Kalinowski, et al., 2000; Braun, et al., 1997). In general, results show increased activation in key pre-motor regions of the right hemisphere during stutterers' production of *fluent* speech, coupled with decreased activation in left hemisphere auditory areas such as the superior temporal gyrus. Crucially, these activation patterns have been shown to occur in both silent (imagined) and overt speech, which supports the recently revived argument that stuttering involves linguistic deficiencies independent of strictly motor behavior (Ingham, et al., 2000). Results such as these reflect a higher-level language impairment underlying chronic disfluency.

2.2.2 Linguistic factors

2.2.2.1 Word-level semantic and syntactic factors

A number of studies have tested for higher-order linguistic deficits in stuttered speech production. A fairly robust finding reproduced in several stuttering experiments is that content words are consistently less fluent than function words (Dayalu, et al., 2002; Howell, et al., 1999); one interpretation of this is that higher-level processes such as lexical retrieval and semantic encoding may be a locus of breakdown for stutterers. In the Dayalu, et al. (2002) study, participants read aloud a list of individually presented content and function words, exhibiting significantly more disfluencies in reciting the content words. The authors explained this difference as resulting from the differences in lexical frequency between content and function words, with the latter's higher relative frequency facilitating more fluent production. The Howell, et al. (1999) study found a similar pattern in the spontaneous speech of adults.

However, experiments measuring response time have shown mixed results: in a two-part picture naming task, Prins, et al. (1997) revealed a statistically significant delay in stutterers' production of verbs compared with that of controls; when asked to name either the animal (e.g., "dog") or action (e.g., "run") or both in a series of pictures, subjects showed no statistically significant differences in noun production, while both verb and noun+verb productions were significantly longer for stutterers. Notably, the verb latency difference fully accounts for the two-word difference between groups.

The second picture-naming experiment in this study compared subject responses varying by length (1- vs. 2-syllable) and lexical frequency (frequent vs. infrequent). Subjects were asked to memorize a small number of words differing by these factors, and then respond with the word that corresponded to the picture presented (e.g., "jar"= 1-syllable, frequent). Results showed no length effect, but a significantly greater latency was found for the infrequent items in stutterers' speech as compared with controls' utterances.

Measuring lexical effects was also the object of a study by Packman, et al. (2001), though their focus was to compare stutterers' performance on reading both words and non-words. The assumption was that non-word reading bypassed lexical selection, and thus any differences between the two conditions would reflect an effect of lexical retrieval. The results showed no significant fluency difference between words and non-words, which the authors interpreted as evidence that lexical retrieval problems are not a factor in stuttering—i.e., high rates of disfluency and latency were found in the production of both words and non-words. It is unclear, however, if the reading task in the latter experiment truly controlled for lexical retrieval effects, since reading phonotactically legal non-words may still require lexical search. In other words, when subjects encountered non-words in reading passages, they may have still treated them as potential words, wittingly or not. In fact, the authors' finding that longer latencies occur in stutterers' reading of non-words than actual words supports this possibility.

In summary, it appears from the content vs. function word studies that semantic and syntactic factors exert some influence over stutterers' disfluency patterns, though characterizing the extent of this effect would require careful control of syntactic (e.g., phrase structure), and intonational (e.g., prominence relationships) factors in order to present more conclusive evidence. For instance, producing words in a sentence context invokes a number of phrasal effects that could confound any attempts to measure just lexical-semantic factors.

2.2.2.2 Sentence planning factors

Factors involved in sentence planning have shown conflicting results with respect to whether syntactic complexity—as defined by the number of embedded clauses within an utterance—or utterance length is the more crucial factor in determining a form's susceptibility to disfluent production. In a phrase-reading task comparing simple, compound, and complex sentences, Maner, et al. (2000) found a weak effect of utterance length and no effect of syntactic complexity on lip movement variability in stutterers. Subjects read a 6-syllable phrase both in isolation (e.g., *Buy Bobby a puppy*) and also within two "lower" and "higher" complexity sentences: the "lower" complexity sentences were compound sentences (e.g., *You buy Bobby a puppy, and I buy Matt toys*) and "higher" complexity were complex sentences, though with the key phrase in the matrix clause (e.g., *You buy Bobby a puppy now if he wants one.*). That only an effect of length was found motivated the authors to conclude that syntactic complexity was not a relevant factor in stuttering, although the failure to elicit sentences with the target phrase in the embedded clause may have been responsible for this lack of effect; in other words, comparing *You buy Bobby a puppy* when it occurs in a matrix clause for both longer sentence types arguably does not represent a significant production difference because the embedded clause has not been accessed or fully planned at this point. Indeed, an almost identical study conducted by Kleinow and Smith (2000) found the opposite tendency—i.e., an effect of syntactic complexity but not utterance length—when stimuli included sentences that crucially located the target phrase within an infinitival clause.

Thus, the sentence *He wants to* *buy Bobby a puppy* *at my store*, where the key phrase occurs in the infinitival clause, resulted in greater lip movement variability in stutterers than did the sentence *You* *buy Bobby a puppy* *now if he wants one*. The remaining sentence types—single matrix clauses and compound sentences—correlated with less lip movement variability than both the embedded and infinitival clause examples.

Similar studies have produced inconsistent results, with some evidence supporting the claim that disfluencies are more a function of utterance length (Tornick & Bloodstein 1976; Yaruss 1999), other evidence favoring the argument that utterance complexity is more influential (Kadi-Hanifi & Howell 1992), and still others indicating that both may play an equal role (Melnick & Conture 2000).

The Tornick & Bloodstein (1976) study provided evidence of an intriguing interaction between utterance length and utterance position, as more disfluencies were triggered utterance-initially in longer sentences than in shorter sentences, though the proportion of disfluencies was nevertheless equal in both sentence types. In other words, the rate of disfluencies was not simply a correlate of utterance length, though utterance length did appear to exert influence on the locus of disfluency. Despite the authors' tentative conclusion that motor planning may have resulted in this tendency, the fact that the number of disfluencies did not differ significantly as a function of utterance length implies that other structural effects are influencing the disfluency patterns.

Other studies have shown similar evidence of sentence-initial vulnerability to disfluent production. In a story-reading task (Prins, et al., 1991), stutterers produced a higher rate of disfluencies per syllable within the first three words of an utterance (45%)

than in later words (33%). This result was found to be independent from the second main effect of lexical stress, which the authors found correlated with stressed content words most often found in later clausal positions. Likewise, two studies recording spontaneous speech samples (Koopmans, et al., 1992; Au-Yeung, et al., 1998) found higher rates of disfluency on the first two words of an utterance. However, both analyses made the further observation that initial disfluencies were confounded with a word's grammatical function: specifically, disfluent utterance-initial words were function words more often than content words, while later disfluent words were more frequently content words. Au-Yeung, et al. (1998) found that the tendency to stutter on utterance-initial function words was more pronounced in children, while for adults disfluencies surfaced more on content words. However, for all subjects, when disfluencies were produced on function words, the function words almost invariably preceded content words (e.g., "*the* school" vs. "looked *after*"), regardless of their position within the utterance. The authors argued that this evidence supported an analysis in which units smaller than the utterance—"phonological words", consisting of one or more function words followed by a content word—should be considered the key predictors of disfluency occurrence.

A study of Norwegian (Kaasin & Bjerkan 1982) provides counter-evidence to the utterance-initial disfluency findings, as it was found that what the authors called "critical words" were stuttered more often than "noncritical words", and the largest proportion of critical words appeared in the final position of sentences (e.g., "You shall meet in the *library.*"). Subjects were given instructions to convey a verbal message, and consistently produced words critical to the message's content at utterance-final position. While the

Tornick & Bloodstein (1976) and Prins et al. (1991) studies—as well as Brown (1938)—have shown the opposite result, i.e., that stuttered words tend to occur towards the beginning of an utterance, Kaasin & Bjerkan (1982) posit that in natural speech speakers tend to place "critical words" at the ends of utterances; however, speakers in Tornick & Bloodstein (1976), Prins, et al., (1991) and Brown (1938) all engaged in oral reading tasks, which may have resulted in less natural sentence production patterns. On the other hand, since a high number (52%) of sentence-initial words in Brown (1938) were also content words, it is plausible that they parallel the "critical words" in KB82 in functioning as disfluency triggers—that is, they induce stuttering because of their informational content.

It is therefore apparent that semantic factors such as lexical frequency, and syntactic factors such as grammatical class and sentence complexity, contribute to predicting the probability of disfluency occurrence, though inconsistently. On the one hand, the fact that content words are produced disfluently more often than function words is consistent with the claim that semantic complexity is at least one trigger of disfluency in stuttering. At the same time, this semantic effect interacts with a word's position within the utterance; that is, while content words on the whole have been shown to be more disfluent than function words, function words produced in utterance-initial position are spoken with a greater disfluency rate. However, there is also evidence that words critical to an utterance's overall message attract a high disfluency rate; yet rather than surfacing utterance-initially, they tend to occur in final position. At the very least, determining the relationship between content words and these “critical words” is necessary in order to make more definitive predictions regarding semantic and sentence planning factors.

The next section will look at the results of studies that have focused on how stuttering is related to word-level factors such as segmental features and lexical stress, and how these intersect with semantic and sentence-level factors.

2.2.2.3 Word-level prosodic factors

2.2.2.3.1 Segmental

Since stuttering is descriptively an articulatory disorder, it is not surprising that segmental factors have received considerably greater attention than other linguistic factors in past stuttering research. Nevertheless, the effects of phonological and phonetic factors on stuttering have proven no easier to isolate. While there exists clear evidence of areas of vulnerability to disfluency at the segmental level, identifying the specific factors responsible for this instability is less clear. For instance, many previous studies have revealed an association between higher disfluency rate and onset type (Brown 1938; Brown 1945; Hahn 1942; Taylor 1966; Quarrington, et al., 1962). Brown’s seminal work was vital in focusing research attention on strictly linguistic factors during a period when stuttering causes were largely assumed to be psychological in nature. Influenced by contemporary theory, the hypothesis was that stuttering would not be predictable on either phonetic or physiological grounds. He conducted an experiment in which 32 young adult stutterers read five passages from a variety of publications. The total number of disfluent words constituted 9.7% of all words (collapsed across subjects), and 91% of these occurred on the word-initial segment. Furthermore, in a ranking of each segment's disfluency rate, all consonants except [h] were stuttered more often than were any vowels.

That the disfluent segments appeared word-initially proved to be a particularly significant finding. Hahn (1942) duplicated these results in a similar experiment in which 43 subjects read four versions of a 550-word passage, with each reading differing by social context—i.e., from low "social pressure" (e.g., reading in isolation) to high (e.g., reading to an audience). Results again showed that word-initial position was most vulnerable (98.13% of total disfluencies), and that, again with the exception of [h], all consonants triggered more disfluencies than did vowels. Taylor (1966) recorded 12 adult male stutterers reading the first 500 words of a sixth-grade reader and found that 88.5% of word-initial disfluencies occurred on consonant-initial forms, versus 11.5% on vowel-initial. However, no significant fluency difference between consonants and vowels was reported in a similar reading experiment by Soderberg (1962), where subjects read three lists of 15 5-syllable meaningful but unrelated phrases. Though word onsets were varied by type (voiced consonant, voiceless consonant, and vowel), there were no significant differences in either mean frequency or mean duration of stuttered segment types. The author did note that disfluent consonants overall tended to have longer duration than vowels, which may have reflected a slight onset effect.

Whether disfluency rate is also affected by varying consonant types was explored by Quarrington, et al. (1962), by comparing subjects' productions of utterance-initial words in 64 6-word sentences. Word-initial onsets were varied by their correspondence to the "segment disfluency ranking" from Brown (1938) (mentioned above). Quarrington, et al. (1962) compared segments that Brown (1938) found had a relatively high mean stuttering rate ([g], [d], [l], and [p]) with those having a low mean stuttering rate ([w], [s], [f], and [h]). The authors found no significant difference between these segment categories, though it is clear that the groupings are confounded by differences in manner, aspiration, and place. For example, both categories contain voiced and voiceless segments, and the category with high mean stuttering rate contains both aspirated (e.g., [p]) and unaspirated (e.g., [g], [d]) stop segments.

Thus, while it is clear from these studies that segmental quality has some effect on disfluency rate—specifically, disfluencies generally occur more frequently on initial consonants than on initial vowels—the more consistent pattern is that stuttered disfluencies occur word-initially. The implications of this will be explored in greater detail in section 2.2.2.3.2.

Previous studies have also investigated more fine-grained phonetic properties for evidence of stuttering disfluency patterns. In the results from a sentence-reading experiment, Healey & Ramig (1986) revealed that in comparison with the speech of nonstutterers, chronic stuttering produces significantly longer voice onset times (VOTs) for voiceless stops[1]; vowel and consonant duration, meanwhile, were longer for stutterers

[1] Comparisons were made only between fluently produced segments.

only in real-word phrases ("take the shape") as opposed to nonsense strings ("ipi saw ipi"). Using a comparable reading task, Hillman and Gilbert (1977) had also found that stutterers produced longer VOTs for voiceless stops than did nonstutterers, though subjects instead read sentence stimuli contextualized within a larger passage. Adams (1987) elicited naturally produced speech samples by presenting a series of line drawings, and prompting child subjects to name and describe the depicted object. This more open-ended task was chosen to avoid interfering with normal intrasubject response variability. Unlike the previous two studies, whose results suggested that stutterers may have a difficulty with voiceless—and therefore aspirated—segment production (i.e., voiceless stops), Adams (1987) indicated that both voiced and voiceless stops had significantly longer VOTs for stuttering than for nonstuttering children. The author attributes this difference to the greater variability of speech-motor control found in conversational speech; such a reason might explain why no significant VOT differences were found at all in Watson and Alfonso (1982), a study which compared speakers' production of nonsense syllable strings varying by voicing and vowel type (e.g., /əput/ /ətǽt/ /əkit/)[2]. However, two more studies examining voiceless segments found evidence of anomalous production patterns in stutterers: Boutsen (1995) found protracted VOTs for [t] when it preceded a stressed vowel in stutterers' speech (e.g., "<u>ST</u>AT stat stat"), while in Jayaram (1983), monolingual and bilingual stuttering speakers of English and Kannada[3] read word

[2] Subjects were instructed how to pronounce the strings, and were allowed rehearsal time to practice them aloud.
[3] Kannada has a 4-way contrasting stop series, including aspirated and unaspirated voiceless stops (Schiffman 1983)

lists and produced significantly more stuttered disfluencies on voiceless stops (i.e., aspirated segments) than voiced stops (i.e., unaspirated segments).

The results of the above studies suggest that sub-segmental phonetic features such as voicing and aspiration play some role in triggering disfluencies, though they are confounded with higher-level factors like location in the word. Analyzing formant transition data is another way of measuring the potential influence of sub-segmental features on disfluencies. A number of studies have examined consonant-vowel transitions in stuttered speech for irregularities in formant transitions. A study by Yaruss and Conture (1993) analyzed child stutterers' disfluent productions elicited from conversational speech with a parent, and found no anomalous F2 transitions. Looking at stutterers' *fluent* speech, however, Robb, et al. (1998) revealed larger than normal F2 fluctuations after stutterers' production of both voiced and voiceless stops. In this study, monosyllabic target words (C+V+[t]; e.g., "pot") were embedded in short carrier phrases though it is unclear how this would effect greater variation in subjects' responses. A different result was found in Chang, et al. (2002), in which child stutterers 3-5 years of age participated in a picture-naming task. While stuttering and control groups did not differ with respect to their overall degree of CV coarticulation, non-stuttering children produced greater differentiation of F2 transition rates between bilabial and alveolar place of articulation than did stuttering children. The authors interpreted this result as possible evidence that stuttering children have less organized phonological encoding abilities at the subsegmental level. Similarly, in Subramanian, et al. (2003), 20 child stutterers produced smaller second formant (F2) frequency changes near the onsets of CV strings than did their controls, though only voiceless stops were tested.

In sum, while the disfluencies encountered in stuttered speech do reveal articulatory deficits, it is clear from previous research that no one phonetic or phonological feature can be isolated as systematically responsible for articulatory disfluencies. The one pattern these studies do consistently reveal, however, is the occurrence of disfluencies in word-initial position; thus, factors higher than the level of the segment must be considered.

2.2.2.3.2 Suprasegmental

Studies looking at suprasegmental information have uncovered significant correlations between stressed syllables and disfluency (Brown 1938; Bergmann 1986; Prins, et al., 1991; Natke, et al., 2002). Brown (1938) interpreted the strong correlations of disfluencies with stressed words as resulting from a speaker's awareness of psychologically "prominent" or communicatively important words. Wingate (1988), however, pointed out that the four major loci of stuttered disfluencies observed in Brown (1938)—1) on content words; 2) utterance-initially; 3) in longer words; 4) on word-initial consonants—could all be accounted for by the fact that they also represent points of lexical stress. For example, relatively few function words bear stress, while content words and longer words often bear primary and even secondary stress. Furthermore, as stated previously, 52% of the utterance-initial words in the Brown (1938) study were also content words. Subsequent studies have in large part supported these interpretations: subjects performing an oral sentence-reading task presented in Bergmann (1986) stuttered significantly more on stressed than unstressed words, regardless of the location of the word in the utterance. Speakers in the Prins, et al. (1991) study produced utterances that showed a correlation between stuttering and stressed syllable peaks when reading from a text—especially in polysyllabic words: 45% of the total number of stressed syllables were disfluent, compared with just 25% of unstressed syllables. As stated in the previous section, the only exception to this pattern in the Prins, et al (1991) study was the occasional tendency for disfluencies to occur on the first three words of a matrix clause, independent of stress.

Contrary to the results from Bergmann (1986), experiments by Weiner (1984) and Hubbard (1998) both showed a stronger effect of word-initial position in determining disfluency patterns. In the Weiner (1984) study, adult speakers were instructed to read sentences containing words whose stress pattern was determined by the context: e.g., "Give me your ADdress" vs. "AdDRESS the crowd." 91% of stuttered disfluencies occurred at both stressed and unstressed word beginnings. As apparent from the example cited, however, the failure to control the position of the word within the utterance likely confounded the results. Using more uniformly controlled stimuli, Hubbard (1998) nevertheless found the same tendency for subjects to stutter more on word-onsets irrespective of stress location within the word, although none of the target words exceeded two syllables in length as shown below:

“Royal **_a_dults** **_de_mand** methods that stick” (iambic)

“Ready **_ar_mies** **_da_mage** nations that stall.” (trochaic)

Thus, as seen above, disfluencies (indicated by italics) consistently occurred on the first syllables of the bold-faced target words, regardless of stress position (indicated by underline). As the author of the study notes, disyllabic words may represent a special case in which the tendency for stressed syllables to attract disfluencies is unable to surface—presumably because the stressed syllable is too near the word onset to trump the word-initial effect. Indeed, Natke, et al., (2002) includes sentence stimuli with multi-syllabic words and successfully teases out both effects: stressed initial syllables were stuttered 34.3% of the time, compared with 24.3% when unstressed. Non-initial syllables, meanwhile, were stuttered 22.6% of the time when stressed, and only 3.5% of the time when unstressed.

An additional concern raised in both Hubbard (1998) and Natke, et al., (2002) was the failure to control for the effects of metrical structure beyond that of the word. Although Hubbard (1998) improved on previous studies by expanding stimuli to include sentences rather than just single words, the author notes that since the first four words of each sentence were content words, this may have inadvertently resulted in a “newspaper title” effect; thus, each word onset was potentially confounded with the onset of a larger prosodic structure, such as an intermediate or phonological phrase, therefore rendering it impossible to determine which prosodic structure was responsible for the effect. Similarly, Natke, et al., (2002) found an additional effect of stressed syllable length, though surprisingly one in the opposite direction than anticipated—i.e., shorter stressed syllables were disfluent more often than longer stressed syllables both in initial (35.2% vs. 32.6%) and non-initial (28.7% vs. 20.1%) location. Once again, the uncontrolled-for effects of larger prosodic phrases may have been responsible for this effect: for example, while the authors divided syllables into stressed vs. unstressed categories, it is nevertheless possible for a short-stressed syllable to have surfaced as more prosodically prominent than a long-stressed syllable with respect to a larger prosodic phrase. Once again, therefore, failing to isolate the different levels of metrical structure puts into question whether the observed effects were properly associated with word boundaries or boundaries of a larger prosodic phrase.

To summarize, the results from previous studies strongly support the claim that lexical stress is highly predictive of disfluency occurrence in stuttered speech. It is unclear, however, precisely why a stressed syllable might elicit a disfluent production in the first place. Moreover, it is not obvious how to represent the relationship between the factors of stress and word-initial position; the interaction between these word-level factors and phrase-level factors such as utterance position requires deeper explanation. As mentioned in the Prins, et al., (1991) study, in addition to occurring on stressed words, disfluencies also reliably surface toward the beginning of an utterance. But as Kaasin and Bjerkan (1982) showed, stuttering frequency is not simply a function of utterance position but also of sentence stress prominence: in other words, the "critical words" which subjects in KB82 produced utterance-finally presumably triggered disfluency because of their phrasal prominence—not because they occurred utterance-finally. Thus, at this point the question necessitates an understanding of how these prominence relationships, determined by larger prosodic patterns, interact with other linguistic factors to trigger disfluencies.

2.2.2.4 Motivation for considering intonational factors

As observed in the summary of word-level effects in the literature, the extent to which segmental and sub-segmental factors are correlated with disfluent production is difficult to determine. More informative have been the results of syllabic stress studies, which more consistently correlate with disfluency patterns in stuttered speech. Absent from the literature, however, are studies which effectively account for the full range of suprasegmental factors at work in speech production, of which lexical stress is only one type.

A relatively small number of studies have nevertheless expanded their prosodic analyses to include at least such sentence-level features as contrastive stress (Klouda & Cooper 1988; Burger & Wijnen 1998). Klouda & Cooper (1988) found a significant positive correlation between contrastive stress and disfluency in all except sentence-final position: 53% of non-final focused words were disfluent, compared to 28% of sentence-final words. A similar result was observed in Burger & Wijnen (1998), where disfluencies most often occurred on contrastively focused words and in sentence-initial position. However, by just manipulating contrastive focus, these studies failed to test further gradations of metrical stress. For instance, while contrastively focused sentence stress is more prominent than lexical stress, pitch accent represents a level of prominence located between the two—i.e., more prominent than lexical stress, but less prominent than sentence stress (Pierrehumbert 1980). A further distinction not made in either the Klouda & Cooper (1988) or Burger & Wijnen (1998) studies is that between phrasal domains. In other words, while subjects were instructed in both studies to read single sentences, it is not clear whether post-hoc analyses were performed to ensure that speakers produced only one intermediate phrase per sentence. The effect of generating more than one intermediate phrase would be crucial, since intermediate phrases constitute the domain of a nuclear pitch accent (sentence stress), as well as entail a re-setting of fundamental frequency.

Thus, since prominence is marked by intonation, it is essential to analyze prosodic structure in order to understand the relations among its intonational components. Intonational structure determines not only the degree of prominence associated with individual elements of a particular phrase, but also the hierarchical organization of these elements within their phrasal domain. Controlling and manipulating intonational structure is therefore crucial for ascertaining to what extent word-level factors such as longer VOTs and lexical stress actually account for disfluencies; the possible alternative, of course, is that intonational factors themselves trigger disfluencies.

Table 1 is a summary of the linguistic studies of stuttering outlined thusfar, with a short description of linguistic factors and relevant findings:

Linguistic Factor		Finding for stutterers
Word-level	Content/ function	• Content words more disfluent

semantic/ syntactic	words; word class	(Dayalu, et al., 2002) • Function words preceding content words more disfluent than function words following content words (Au-Yeung, et al., 1998) • Noun + verb productions longer latency (Prins, et al., 1997)
	Lexical frequency	• Infrequent words longer latency; (Prins, et al., 1997) • No difference between words and non-words (Packman, et al., 2001)
Sentence planning	Utterance length	• Longer utterances more disfluent (Tornick & Bloodstein 1976; Yaruss 1999; Maner, et al., 2000; Melnick & Conture 2000) • No effect (Kleinow & Smith 2000)
	Syntactic complexity	• More complex utterances more disfluent (Kadi-Hanifi & Howell 1992; Kleinow & Smith 2000; Melnick & Conture 2000) • No effect (Maner, et al., 2000)
	Position within the utterance	• First three words most disfluent (Prins, et al., 1991) • First two words most disfluent (Koopmans, et al., 1992; Au-Yeung, et al., 1998) • First word most disfluent (Brown 1938; Tornick & Bloodstein 1976; Burger & Wijnen 1998) • Final word most disfluent (Kaasin & Bjerkan 1982)
Segmental	Position within the word	• Word-initial most disfluent (Brown 1938, 1945; Hahn 1942; Soderberg 1962; Taylor 1966; Weiner 1984; Hubbard 1998; Natke, et al., 2002)
	Segment type (consonant vs. vowel)	• Consonants more disfluent than vowels (Brown 1938, 1945; Hahn 1942; Taylor 1966) • No effect (Soderberg 1962)
	Consonant type	• Segments [g, d, l, p] more

		disfluent than [w, s, f, h] (Brown 1938) • No effect (Quarrington, et al., 1962)
	VOT	• Longer voiceless stop VOT than controls' (Hillman & Gilbert 1977; Healey & Ramig 1986) • All stops longer than controls' (Adams 1987) • Longer voiceless stop (Jayaram 1983); • Longer for [t] when preceded stressed vowel (Boutsen 1995) • No effect (Watson & Alfonso 1982)
	Segment duration	• Longer consonants and vowels in real-word phrases than controls' (Healey & Ramig 1986)
	Second formant transition (from C to V)	• Larger than controls' (Robb, et al., 1998); No effect (Yaruss & Conture 1993) • Smaller difference between bilabial and alveolar than controls' (Chang, et al., 2002) • Smaller than controls' for voiceless consonants (Subramanian, et al., 2003)
Suprasegmental	Lexical stress	• Stressed more disfluent (Brown 1938; Bergmann 1986; Weiner 1984; Wingate 1988; Prins, et al., 1991; Natke, et al., 2002; Arbisi-Kelm 2004) • No effect (Hubbard 1998)
	Contrastive focus (sentence stress)	• Focused more disfluent (Klouda & Cooper 1988; Burger & Wijnen 1998) • Default nuclear pitch-accented words more disfluent (Arbisi-Kelm 2004)

Table 1: Summary of studies: linguistic factors and relevant findings

In addition to metrical prominence, which can account for the stress effects found in the literature, a second intonational phenomenon can likewise explain the utterance-initial effects observed in many of the same studies. As found in numerous studies on prosody, more extreme lingual articulations—known as "articulatory strengthening"—are produced at the edges of prosodic domains (Pierrehumbert & Talkin 1992; Fougeron & Keating 1997). Thus, utterance-initial location is prominent in that, all else being equal, it requires the greatest amount of linguopalatal contact within a given prosodic domain, and therefore it is not surprising that this position—like metrically prominent "critical words"—is particularly vulnerable to disfluent production.

While intonation directly reflects the properties of prosodic structure, it also interacts with syntactic structure in crucial ways. A sentence's information structure, which contains categories (e.g., focus, topic) derived from the syntax, influences the shape of the prosodic structure (Büring 2003). The relationship between syntactic and prosodic structure is not one-to-one, since the same syntactic structure can be assigned different intonation contours—and vice versa (Beckman & Pierrehumbert 1986; Pierrehumbert & Hirschberg 1990). Some indirect mapping is nevertheless necessary for information structure to be realized in the prosodic structure. As was discussed in 2.2.2.2, while previous studies showed contradictory effects of sentence length and syntactic structure on stuttering, the lack of any formal analysis of prosodic factors potentially prevented a clearer picture of disfluency patterns from surfacing. For instance, in the Kaasin & Bjerkan study, the tendency for stutterers to disfluently produce "critical words" in utterance-final position would be predicted by an intonational phonology model, since in default declaratives the word bearing greatest prominence is often produced at the end of an intonational phrase (Ladd 1996). Likewise, the results of the Tornick & Bloodstein study are less surprising if it is understood that the points of prominence in an utterance are determined not by fixed positions, but rather by the prosodic relations that obtain among these positions. For example, the fact that more disfluencies occurred at the beginning of longer sentences than of shorter sentences, but not necessarily with greater frequency than in other positions in longer sentences, implies that utterance-initial position interacts with other phrasal factors such as metrical prominence: final positions of the immediate phrase, as stated previously, receive greatest prominence in an utterance, and thus may have worked in tandem with utterance-initial strengthening effects to produce disfluencies at similar rates in both early and later positions in the longer sentences. In shorter sentences, on the other hand, the independent effects of utterance-initial articulatory strengthening and utterance-final prominence may have been masked by their close mutual proximity.

Finally, in addition to prosody and syntactic relations, it is clear that semantic factors play a role within intonation structure, since intonation itself conveys information proper to the meaning of the entire utterance. Thus, as seen in section 2.2.2.1, it is not enough to compare semantic information at the level of lexical form: determining a sentence's entire information structure is equally important if the relationships between disfluency patterns and meaning are to be accurately represented.

The shape of the larger picture is at this point quite plain: previous stuttering experiments have effectively revealed a plethora of linguistic factors that alone are not sufficient indicators of the patterns of disfluency encountered in stuttered speech. In the case of word-level effects such as VOT and lexical stress, the lack of reliable evidence for their association with disfluency patterns might be due to intonational factors such as prominence relations among words, as well as the prosodic structure overall. Similarly, the fact that higher-level syntactic and semantic factors interact crucially with intonational factors calls for a more comprehensive analysis of intonation in investigating speech disorders. It is hoped that the results from an intonational analysis can be useful in charting new methods of modeling impaired speech processing and determining the production patterns that characterize normal and disfluent speech alike.

The next section describes the model of intonation adopted in the current study.

2.2.2.5 Intonational Phonology Model

The intonation model of English used in this study is the intonational phonology model, derived from the Autosegmental-Metrical model proposed by Pierrehumbert (1980), Beckman & Pierrehumbert (1986), and Ladd (1996). Intonation is understood as the modulation of pitch over an utterance. Intonational phonology assumes that pitch is represented categorically by high (H) and low (L) tones (or a combination of H and L), and these tones form the basis of a hierarchically organized prosodic structure. Tones are associated either with the stressed syllable of a word or with the boundaries of the phrase. The idea of this mapping is derived from Autosegmental Phonology (Liberman 1975), and the prominence relations among the words can be represented by a metrical grid as shown in Figure 3:

```
                                   x      nuclear pitch accent
 x           x       x             x      pitch accent
 x     x     x       x             x      stressed syllable (content word)
 x     x   x x   x x x    x   x    x      syllable
Pam took the car to the park in the town.
 H*          !H*     !H*           H* L-L%
```

Figure 3: Levels of metrical stress

Each grid mark represents a degree of prominence, with every syllable carrying a minimum of one degree. At the next level are the stressed syllables of each content word, which mark the most prominent syllable within a word. Pitch accent (indicated by an asterisk) is the measure of prominence among stressed syllables. Unlike lexical stress, pitch accents are determined by phrasal rather than lexical properties. Finally, the pitch accent in the rightmost position of an intermediate phrase is referred to as the “nuclear pitch accent”, which bears the greatest prominence of a phrase.

Following the MAE-ToBI (Mainstream American English—Tones and Break Indices) transcription conventions (Beckman & Ayers 1994; Beckman & Hirschberg

1994), English pitch accents can be single (H*, L*, or !H*) or bitones (L+H*, L*+H, L+!H*, L*+!H, or H+!H*), with H tones labeled with the diacritic '!' representing a categorically lowered H tone, or "downstep." The pitch accent is realized on a stressed syllable of a prominent word, within the prosodic phrase in question.

There are two types of tones which function to signal phrasal boundaries: phrase accents and boundary tones. Phrase accents, which can be a low (L-) or high (H-) tone, mark the end of an "intermediate phrase". Intermediate phrases are minimally made up of a pitch accent and phrase accent, and are part of a larger structure called the Intonation Phrase. The boundary tone, meanwhile, is what marks the end of this Intonation Phrase, and can also be low (L%) or high (H%)[4]. These intonational tones are labeled on the 'tones' tier of the ToBI model (see Figure 4).

The second half of the ToBi model name—break indices—indicates a measure of juncture degree between words. The standard break index value between words is '1', while weakened word boundaries—often involving function words, such as in the flap-producing environment "*at a*" in "John threw the ball at a car"—are denoted by a '0' break index value. Break index '3' and '4' signify intermediate and Intonation Phrase boundaries, respectively, while break index '2' is reserved for marking mismatches between tonal cues and juncture cues (e.g., when a break of exceptionally long duration separates two words despite the absence of a tone marking a phrase boundary). For all coding labels, it is standard practice in the ToBI transcription model to label uncertain phenomena with the diacritic '?'. Figure 4 shows examples of how these tone and break

[4] Boundary tones on the left edge of an Intonation Phrase are explicitly marked in English only in the non-default cases (i.e., when indicating a high—%H—initial boundary tone).

indices are labeled on the pitch track, each recorded on separate tiers. The symbol ‘*?’ over the verb “took” indicates that the cues in the pitch track are insufficient for determining with certainty whether a pitch accent is present in this location.

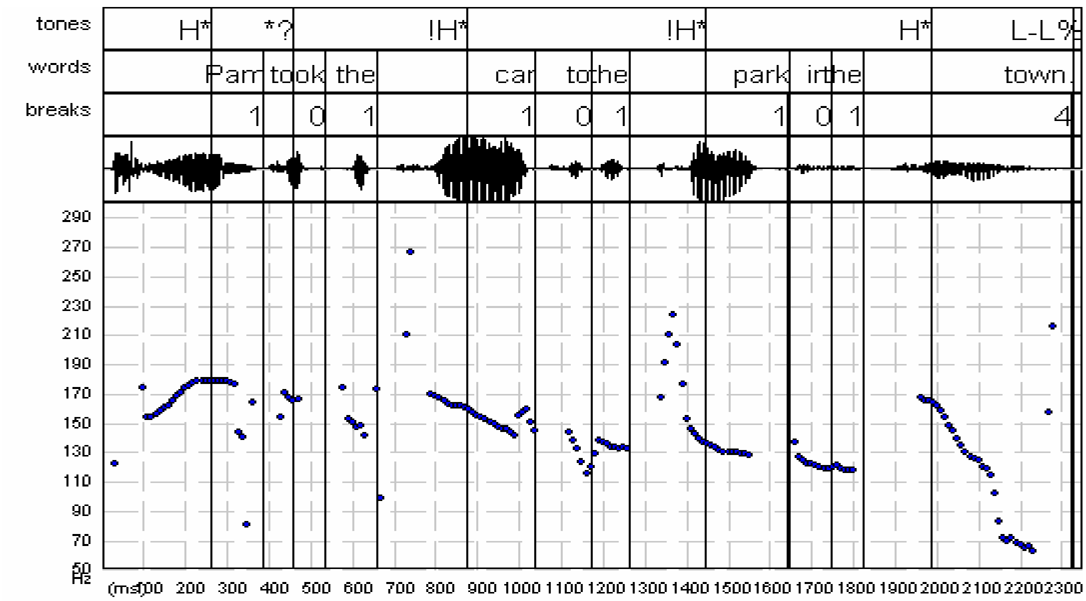


Figure 4: Pitch track of "Pam took the car to the park in the town."

Pitch track diagrams like in Figure 4 are read by aligning the right edge of word labels and break indices to the right edge of the word’s acoustic waveform. For instance, in the ‘words’ tier, the right edge of the word label “Pam” is aligned with the word’s waveform’s right edge; in the ‘breaks’ tier, the break index ‘1’ is also labeled at the actual acoustic boundary of the word, indicating a phrase-internal word break. Similarly, in the ‘tones’ tier, phrase and boundary accents (e.g., ‘L-L%) are right-aligned with the boundary of the phrase. Pitch accent labels (e.g., H*), on the other hand, are aligned with a point on the F0 track that can be associated with the accent; for example, the first H* label in Figure 4 is aligned directly above the peak of the accented vowel in the word "Pam".[5]

In English, a focused word carries nuclear pitch accent, and by manipulating the location of the focused word, the location of the nuclear pitch accent can be changed. For example, in Figure 4 default nuclear pitch accent falls on the final pitch-accented word "town". Placing emphasis on the word "park", however, effectively moves nuclear pitch accent to this focused word, resulting in a higher F0 value over the focused word:

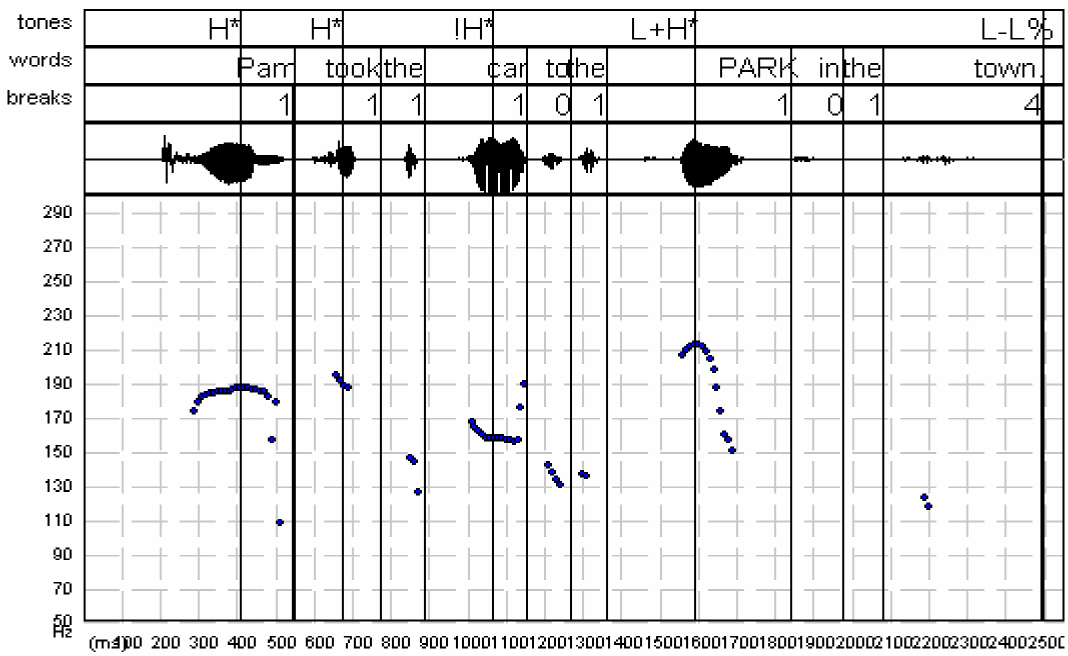


Figure 5: Pitch track of "Pam took the car to the PARK in the town."

[5] Although it is the case in this example, pitch accent labels will not always align directly with F0 peaks and troughs, since these F0 values are not always reached during production of the accented vowel's steady state.

One other interesting effect of focus in English is that all material following a focused element in the intermediate phrase will be "de-accented" (i.e., lose its pitch accent). Thus, as shown in Figure 5, the word "town" does not bear any pitch accent. In a de-accented environment, the lack of pitch accent results in the phrase accent being realized immediately after the focused word. In the example from Figure 5, the realization of the L- phrase accent is clearly illustrated by the immediate F0 fall after the focused word "park", rendering the post-focus "in the town" deaccented.

Intermediate phrases (ips) and Intonation Phrases (IPs) are acoustically marked and distinguished by the presence or absence of crucial boundary cues involving one or more tones and junctures. An ip boundary is marked by a single disjuncture cue before the boundary—i.e., pre-boundary lengthening—and two tonal cues: the pre-boundary F0 cue mentioned above (H- or L-), and a post-boundary re-setting of F0. An Intonation Phrase includes all of these cues plus an additional optional disjuncture cue realized as a post-boundary pause. All four boundary cues are illustrated in Figure 6 below:

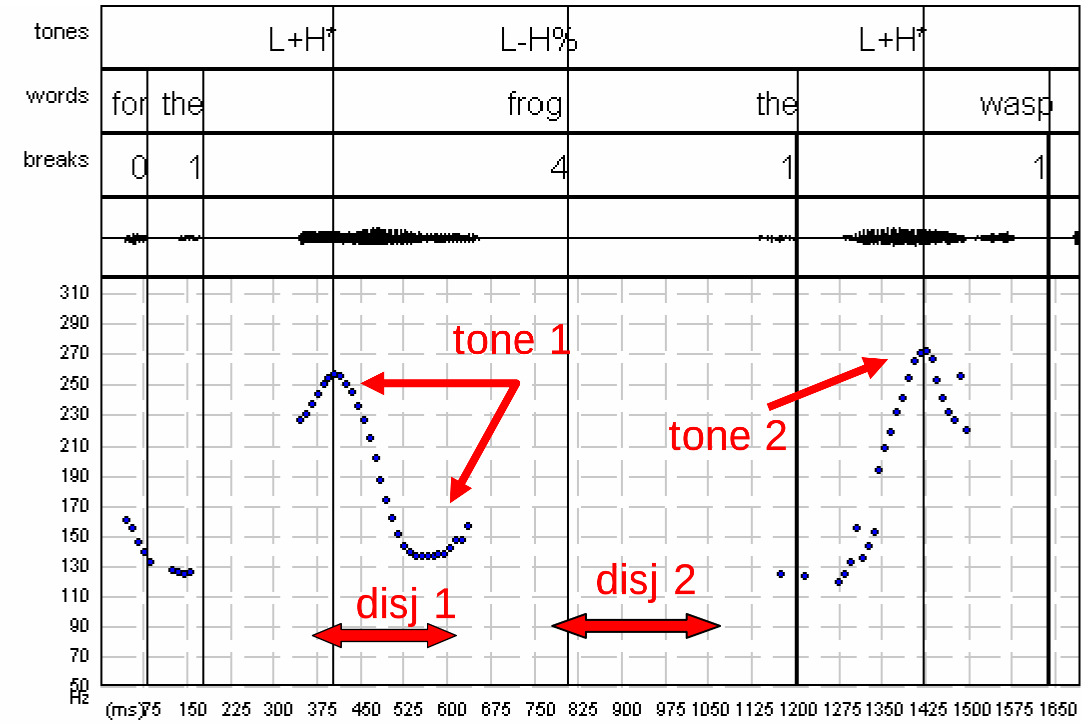


Figure 6: Intonation Phrase break cues example

Since ip and IP boundaries share three of the four possible boundary cues, the cue belonging solely to IP boundaries—post-boundary pauses—is often crucial for distinguishing the two boundary types. However, pre-boundary tones may also assist in identifying the boundary type, provided that the tone marking the ip boundary contrasts with that of the IP boundary. For instance, as Figure 6 exemplifies, the fall and subsequent slight rise on the word "frog" signal the presence of two tones, and thus an IP boundary since no bi-tonal phrase accents are proposed in the intonational phonology model. When pre-boundary ip and IP tones coincide, as they do at an L-L% or H-H% IP boundary, then distinguishing this boundary type from a simple L- or H- phrase accent will more than likely require careful consideration of the remaining cues.

Disjuncture cue 1:	pre-boundary lengthening
Disjuncture cue 2:	post-boundary pause
Tonal cue 1:	pre-boundary F0 fall/ rise (L-H%)
Tonal cue 2:	post-boundary F0 reset

2.2.3 A recent study

An experiment was recently conducted by the author in order to test the hypothesis that disfluency is directly correlated with prominence-related intonational events. Three stuttering subjects participated in an oral sentence-reading task testing a variety of sentence types. One group of stimuli was presented in order to expand on previous prosodic studies, testing from an intonational phonological perspective whether particular word-internal prosodic effects would occur when controlling for intonational factors. The second stimulus group tested the predictions regarding strictly intonational phenomena: namely, whether disfluent production is directly proportional to the prominence-level of a given production. It was anticipated that nuclear pitch-accented syllables, representing the highest degree of stress in an intonation phrase, would be most prone to triggering disfluency, since they bear the greatest level of prominence in the utterance. Accordingly, pitch-accented syllables, bearing the second-highest degree of prominence, were anticipated to be the second-most vulnerable position to disfluent production, followed by stressed syllables and unstressed syllables, respectively.

As was mentioned previously, the results of the study confirmed some of the major hypotheses: in all of the comparisons between pitch-accented and non-pitch-accented positions of stress, the former attracted the higher rate of disfluent speech productions. This supports the principal hypothesis that intonationally prominent words—not simply lexically stressed syllables—are a better indicator of disfluencies in stuttered speech. However, when NPA was triggered by focus—i.e., contrastively focused nuclear pitch accents—it did not attract more disfluencies than the less prominent pitch-accented and stressed-only positions. One possible reason for this result is that it reflects a difference between default nuclear pitch accent—that is, the rightmost pitch accent in declarative sentences, when all words represent new information in the phrase (i.e., broad focus)—and nuclear pitch accent resulting from contrastive focus. In both cases, the nuclear pitch accent denotes the most prominent word of the utterance, but while a default nuclear pitch accent is phonetically indistinguishable from a non-nuclear pitch accent (Ladd 1996; Silverman & Pierrehumbert 1990), focused nuclear pitch accents bear particularly salient acoustic cues. This acoustic difference may therefore have a confounding effect on stutterers' speech which actually facilitates more fluent production.

A second possible explanation for this result was the unnaturalness of the tightly controlled sentence stimuli: in order to control for factors such as segment type (manner, place, voicing), stress position, word length, word frequency, and argument number, all longer sentences contained monosyllabic content words varying only by the place of articulation of their word-onsets. It is thus proposed that the follow-up contain more naturally elicited data. The current study employs a narrative task in which each subject constructs a story based on stimuli from a wordless picture book. In this way, more spontaneous production can be elicited, while still controlling for the content words most likely to be produced by the speakers. More importantly, a narrative task requires that subjects complete all stages of language production, including conceptual preparation and lexical selection.

2.3 Disfluencies in language production

Current models of disfluency production in normal speakers, such as that proposed in Levelt (1983), break down the structure of a speech error repair into several components. Levelt's model assumes three phases: the reparandum, the editing phase, and the repair. The reparandum constitutes the error to be repaired, which may be detected either before (i.e., covertly), during, or after its actual articulation[6]. The point at which a speaker detects the error results in an interruption of the production process, thereby beginning the editing phase. This phase includes the period between the detection of the error and the beginning of the disfluency's repair, and commonly consists of either filled or unfilled pauses. The final stage is the actual repair of the disfluency, when normal fluency resumes.

An example of how these three stages are combined in actual speech is illustrated in Figure 7, where a normal speaker attempts to produce the utterance, "with the dog on the

[6] A production "problem" would not be limited to speech errors per se (e.g., segment substitutions), but rather include lexical search delays, syntactic restructurings, improper word choice, etc.

bed." The moment of interruption occurs after fluent production of the phrase "with the dog on his—", which is then followed by a pause ('3ps') representing the editing phase. The speaker begins fluent speech again after the short pause, making an overt repair in substituting "the" for "his". At this point it becomes clear that the item to be repaired—or, reparandum—was the word "his".[7]

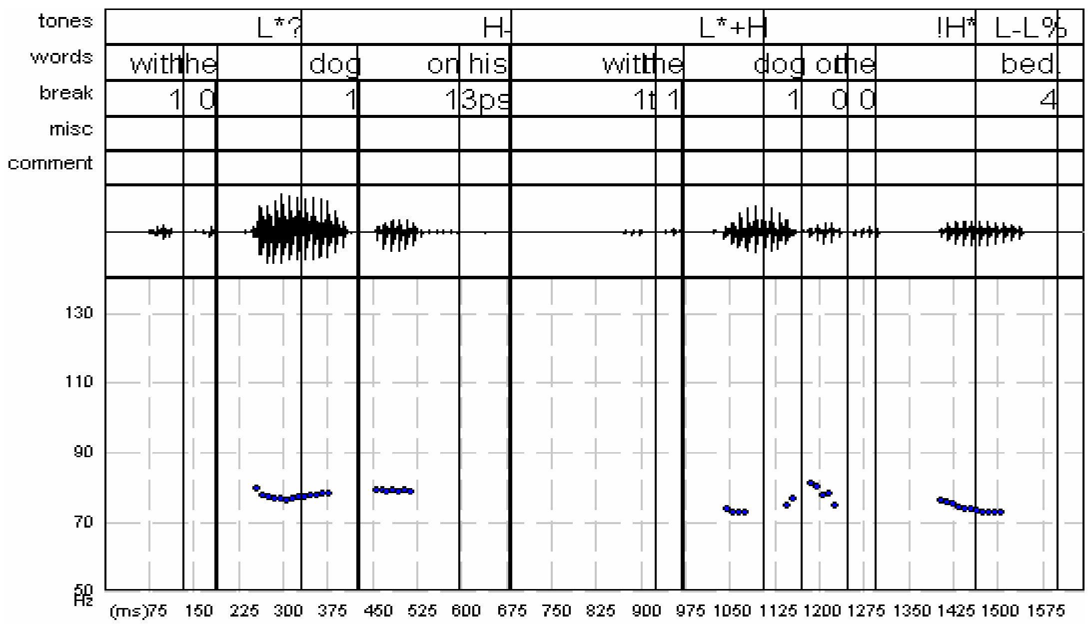


Figure 7: Example of a repaired disfluency: "with the dog on his, with the dog on the bed"

Depending on the cause of the disfluency—i.e., whether it is phonological, semantic, syntactic, etc.—a speaker will generally produce different types of evidence in both the reparandum and editing phase which signal the origin of the disfluency. For instance, in a study of disfluencies found in normal speakers' spontaneous productions, Shriberg (1999) shows that the vast majority of repeated disfluencies (88%) had no clearly

[7] It is also possible, of course, that an error or breakdown was detected in advance of its production (e.g., delay in lexical access of "bed"), and thus the interruption and subsequent pause instead functioned as a covert repair of a word before (i.e., "bed") rather than after (i.e., "his") its articulation.

detectable acoustic transitions between the coda of the word preceding the disfluency and the onset of the disfluent word; that is, interruptions were followed immediately by unfilled pauses before the word repetition. This would imply that in most disfluencies—at least in spontaneously produced single word repetitions—speakers do not show overt evidence of phonological problems, since in these cases no acoustic properties of the following word are manifested.[8]

2.4 Anticipatory vs. target disfluencies

Notwithstanding the inaccessibility of data which could conclusively indicate the stage of processing in which a disfluency is triggered, previous studies have found examples of particularly early disfluency detection in the speech of stutterers. In Viswananth (1989), stuttering and control groups performed five oral readings of two short stories. In an analysis of word duration, the author found that while groups did not differ significantly in their second through fifth readings of the two stories, in the first reading stutterers produced significantly longer words immediately *before* words which were stuttered. Similarly, studies such as Au-Yeung, et al., (1998) found a strong tendency of stutterers to produce disfluencies on function words immediately preceding content words, while rarely on function words following content words.

---

[8] This of course does not rule out a phonological error in favor of a lexical search error, since it may be that the former had indeed occurred but was detected well before its attempted onset.

But despite the robust findings of the aforementioned studies, a number of previous experiments have found higher rates of function word disfluencies only in child stutterers, compared with much higher disfluency rates on content words for adults (Dayalu, et al., 2002; Howell, et al., 1999; Dworzynski, et al., 2003). Howell, et al. (1999) interpreted these results to indicate a different strategy of disfluency repair for children as compared with adult stutterers: while 2-6-year-old child stutterers and fluent speakers of all ages both utilize repetitions to postpone production of following content words, adult stutterers will forego such a delay, thus resulting in a disfluency surfacing directly on the content word. Like Au-Yeung, et al. (1998), however, Howell & Sackin (2000) provided a more detailed analysis of this tendency and concluded that although both delaying (i.e., function word disfluencies) and non-delaying (i.e., content word disfluencies) behaviors are found in stutterers' speech, each disfluency type represents a different aspect of the 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. The observation by Au-Yeung, et al. (1998) that stutterers' function word disfluencies occurred almost exclusively before the content words with which they formed prosodic units ("phonological words") suggests that it is the relative location of the function word within a larger phrase that is critical to its disfluent production. Moreover, results of studies such as Howell, et al. (1999) imply that the distribution pattern of the disfluencies

–i.e., whether they are manifested in a function or content word—indicate whether a disfluency is triggered by detection and postponement of an upcoming production breakdown or the underlying breakdown itself. Henceforth, this supposed distinction will be referred to as that between *anticipatory* (i.e., detected in advance) and *target* (i.e., underlying breakdowns) disfluencies.

2.5 Monitoring and lookahead

Clearly, if production breakdowns in both normal and stuttered speech are often manifested by postponement of problematic material, then speakers must have some planning window within which the breakdown can be detected; in effect, speakers will not always begin overt production of the word whose planning is faulty. One difficult question that follows, however, is to what degree can it be established where in the production process a breakdown has occurred. As studies such as Viswanath (1989) showed, the presence of segmental and tonal information amid multiple disfluencies can provide clues to determining the locus of a production breakdown, as well as how far in advance a speaker may have detected the breakdown.

A significant body of literature has investigated this question of the relationship between planning and disfluency detection in both normal and disfluent speech. In more recent studies, various speech monitoring models have been developed to account for a speaker's ability to detect potential disfluencies during speech planning. The version that will be explored here derives principally from Levelt (1983), Blackmer & Mitton (1991), and 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 debate and 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.

In section 2.3, the anatomy of a speech error repair was outlined for normal speech (Levelt 1983; Shriberg 1999). A similar design known as the "Covert Repair Hypothesis" was developed to account for both stutterers' and nonstutterers' disfluencies (Kolk 1991; Postma, et al., 1990; Postma & Kolk 1993). Essentially, this hypothesis claims that a speaker may both detect and correct an error before it encroaches on overt articulation—hence the phrase, "covert repair". 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. One prediction following from this hypothesis is that if the activity of the monitor is somehow mitigated or distracted through a secondary task, a simultaneous increase in speech errors should result—i.e., undetected and uncorrected by a distracted monitor, phonological errors should cleanly surface as segments that are mis-ordered, substituted, or otherwise ill-formed. In a study implementing a secondary distraction task, it was found that while stuttered disfluencies (i.e., cuts, repetitions, pauses, prolongations) were reduced during performance of the secondary task, there was not in fact a resultant increase in phonological errors (Postma & Kolk 1992a, b).

Burger & Wijnen (1999) tested the phonological encoding hypothesis from a different angle by comparing stutterer and nonstutterer responses to phonological priming. While an earlier similar study (Wijnen & Boers 1994) found that stutterers only benefited from priming when both a word's initial onset and nucleus—compared to just the onset, as was the case for nonstutterers—were primed, these results were not consistently duplicated in Burger & Wijnen (1999).

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, thus producing observable disfluencies. This interpretation is consistent with Levelt's proposal that the monitor detects problems via two methods: 1) by comparing both overtly and covertly parsed output with an intended message; and 2) by applying standard prosodic criteria to the output, such as loudness and rate (Levelt 1983). In other words, not only would a monitor identify errors by comparing specific output segments and word forms with their input forms, but by evaluating the output with respect to various types of production criteria. 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.

As the mixed evidence of these monitoring studies has shown, the question of whether a phonological encoding impairment underlies stuttering disfluencies is difficult to determine, especially given the inevitably covert quality of such encoding breakdowns. Nevertheless, as reviewed in section 2.2.2.3, various studies have shown inconsistent evidence of segmental instability underlying stutterers' disfluencies. An influential series of studies by Wingate (1976) found that stuttering episodes instead correlated highly with stressed syllables: in particular, stuttered disfluencies resulted from a defect in transitioning from the onset of a syllable to its nucleus—a transition Wingate dubbed the syllable's "fault line". The critical feature of this observation that onsets and rimes (i.e., nucleus + optional coda) are the core syllable constituents is supported by psycholinguistic evidence (Levelt 1989; Dell 1986). When applied to the evaluation of stuttered disfluencies, this criterion can serve the fundamental purpose of distinguishing between the anticipatory and target disfluencies defined in the previous section. Simply put, any disfluency involving the onset of a syllable can be considered an underlying (target) disfluency, since the critical juncture between onset and rime has not been successfully produced. Conversely, a disfluency surfacing on any part of the rime can be categorized as anticipatory, as articulation has progressed beyond the onset-rime juncture.

2.6 Hypotheses and motivations

The following hypotheses are proposed based on the novel considerations outlined in this chapter: namely, that the higher-level semantic and syntactic factors traditionally investigated in the literature interact with respect to their functions in a larger intonational structure. Similarly, with respect to word level factors, while the stress studies described above (Brown 1938; Prins, et al., 1991; Natke, et al., 2002) in general support the claim that lexical stress is indeed a significant factor in determining where disfluencies will occur, lexically stressed words are situated within a larger pattern of prominence relations that often override the effects of lexical stress. Thus, in sum, it is hypothesized that stuttered disfluencies are a function of prominence relations, and that therefore their distribution patterns will follow the hierarchical metrical scale of prominence identified above.

The second hypothesis is predominantly motivated by studies that revealed different behavior of stuttered disfluencies depending on their context within a prosodic phrase

(Viswananth 1989; Au-Yeung, et al., 1998). It is predicted here that stutterers will show greater prosodic evidence of detecting underlying (i.e., target) disfluencies than will controls, and this advance detection will result in a higher percentage of anticipatory disfluencies in their speech production. Furthermore, in accord with the previous hypothesis, it is predicted that these target disfluencies will crucially originate most often in the most metrically prominent material in an intermediate phrase (i.e., nuclear pitch accents), and that anticipatory disfluencies will surface predictably before—and in close proximity to—these 'target' disfluencies. This result will not only support the more general claim that a speech monitoring system plays an important role in detecting disfluencies, but also that how the monitor detects disfluencies will be influenced by properties of intonational structure.

Hypothesis 1: It is predicted that words bearing a higher level of prominence will be more prone to disfluent production. Specifically, nuclear pitch-accented words—which bear the highest level of prominence—should attract the highest rate of disfluency, and pitch-accented words should attract a higher rate of disfluency than non-pitch-accented words.

Hypothesis 2: Disfluencies are either anticipatory or target-realized, predictable by properties of prosodic structure. Specifically, stutterers' disfluencies will be accompanied by more prosodic irregularities prior to the actual cause of the disfluency than will nonstutterers' disfluencies.

CHAPTER 3

METHOD

3.1 Structure of experiments

The experiments were divided into two related tasks. Task 1 was a narration task performed by four subjects who stutter and four control subjects. The goal of Task 1 was to narrate 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 to read the transcript of one of three stuttering speakers; that is, each control, who was already familiar with the story from Task 1, read a transcript produced by their age- and gender-matched counterpart in the stuttering group. 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.

3.2 Participants

Eight subjects—four adult male persons who stutter (hereafter, “stutterers”) and their age- and gender-matched controls—were selected to participate in a story-telling task. Subjects for the experiment were recruited by placing flyers around the UCLA campus, by contacting local speech-language pathologists, and by communicating with local stuttering support groups. Individuals who responded to the flyers were screened to ensure that (i) they were native speakers of Standard American English; and (ii) they had a history of developmental stuttering. Stuttering severity for stutterers was moderate, as determined by assessments provided by licensed speech-language pathologists. Age-matched control subjects were chosen for each stuttering speaker, also sharing SAE as their native language. Table 2 below lists the basic data for each subject who participated in the experiment:

	Subjects	Age	Education level	Occupation
Stutterers	S1	71	College degree	Retired real estate agent
	S2	53	College degree	Accountant
	S3	51	College degree (Ph.D.)	Researcher
	S4	35	College degree	Business owner
Controls	C1	72	College degree (Ph.D.)	College professor
	C2	55	College degree	Management consultant
	C3	52	College degree (Ph.D.)	University administrator
	C4	35	College degree	Graduate student

Table 2: Subject information

For Task 1, each of the eight subjects individually performed a spontaneous narration of the picture book. Task 2 involved only the participation of the three control subjects as readers of the stutterers' Task 1 scripts. In order to maintain an even age range of subjects, one control subject from each age bracket—30+, 50+, and 70+ years—was used to read a script in Task 2. Any potential effects of age were controlled by matching a script reader with a stutterer of the same age. Hence, for Task 2, C1 read for S1, C2 read for S2, and C4 read for S4.

3.3 Procedures

Individual 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 cassette recorder (PMD222). 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. Upon completing the experiment, individuals were paid $10 for their participation.

A sample page from the picture book is shown below:

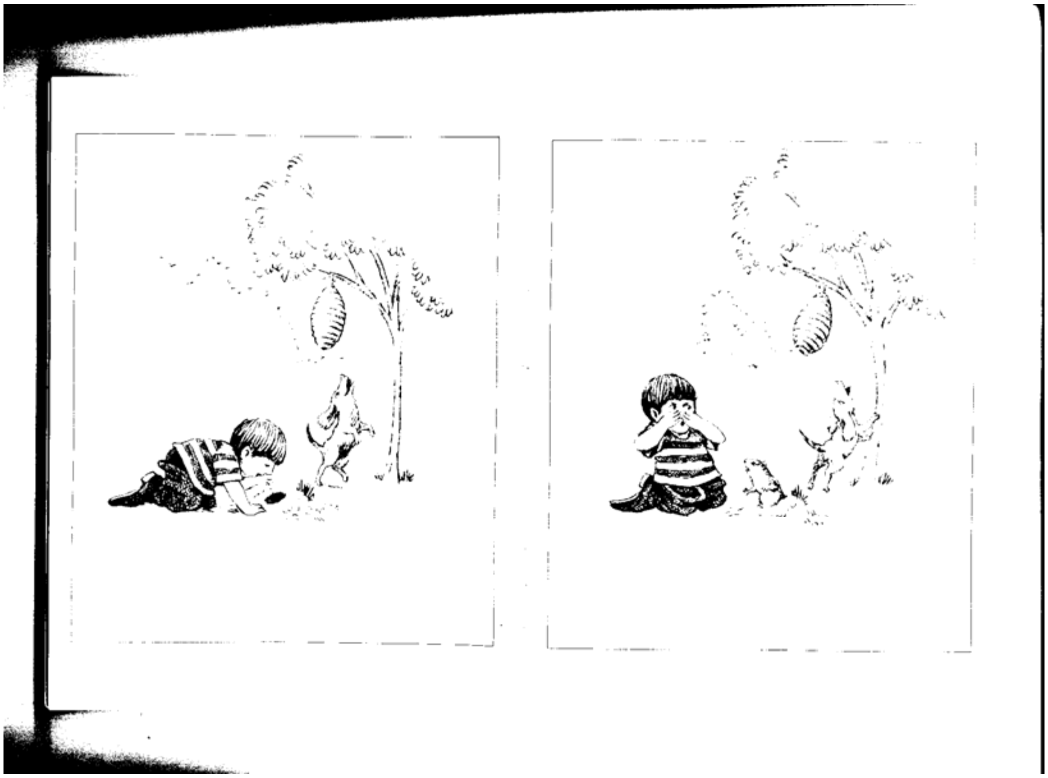


Figure 8: Sample page from the picture book "Frog, Where Are You?" (Mayer 1969)

3.4 Data analysis

3.4.1 Coding disfluencies

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 ToBI system of transcribing English intonation (summarized in section 2.2.2.5). Tones, words, and break index markers were coded on individual tiers above the pitch track of each sentence stimulus (see Figure 10).

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. In the original ToBI model, lengthening resulting from normal disfluencies is indicated by attaching a diacritic 'p' to the numbered break index. For instance, if in the sentence "I said deed again", the production of the word "deed" is arrested, then the break index after the cut would be labeled '1p'—provided that no prosodic phrase break resulted from the cut. Similarly, in the original ToBI transcription convention, disfluencies resulting in hesitations or segment prolongation are labeled '2p', but if the rupture precipitates a phrasal accent and subsequent intermediate phrase break, then the disfluency is labeled '3p'. If pitch is reset, the symbol '%r' is recorded in the tonal tier to indicate this occurrence.

This disfluency nomenclature served as a starting point in elaborating a more structurally detailed system which can accurately represent the wide variety of disfluency types occurring in both stuttered and normal speech, while also capturing the subtle distinctions among them. 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 3:

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”)

Table 3: List of single disfluency break index types

While the list of disfluency types in Table 3 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. The full range of combination disfluency types is outlined in Table 4, with unattested combinations in bold-face:

# of events	Disfluency types	abbrev
2	Restart + prolongation	t.p
	Restart+cut	t.c
	Restart+pause	t.ps
	Restart+filler	t.f
	Prolongation+cut	p.c
	Prolongation+pause	p.ps
	Prolongation+filler	p.f
	Cut+pause	c.ps
	Cut+filler	c.f
	Pause+filler	ps.f
3	Restart+prolongation+cut	t.p.c
	Restart+prolongation+pause	t.p.ps
	Restart+prolongation+filler	t.p.f
	Restart+cut+pause	t.c.ps
	Restart+cut+filler	**t.c.f**
	Restart+pause+filler	t.ps.f
	Prolongation+cut+pause	p.c.ps
	Prolongation+cut+filler	**p.c.f**
	Prolongation+pause+filler	p.ps.f
	Cut+pause+filler	c.ps.f
4	Restart+prolongation+cut+pause	t.p.c.ps
	Restart+prolongation+cut+filler	**t.p.c.f**
	Restart+prolongation+pause+filler	t.p.ps.f
	Restart+cut+pause+filler	**t.c.ps.f**
	Prolongation+cut+pause+filler	p.c.ps.f
5	**Restart+prolongation+cut+pause+filler**	**t.p.c.ps.f**

Table 4: Combination disfluencies

In order to integrate this 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. Table 5 presents a description of all single disfluency break index types:

Disfl type	Break Index	Description
Restart	0t	Restarted word followed by ip-internal tone in a clitic environment
	1t	Restarted word followed by ip-internal tone
	3t	Restarted word followed by post-ip boundary tone
	4t	Restarted word followed by post-IP boundary tone
Prolongation	0p	Prolonged word followed by ip-internal tone in a clitic environment
	1p	Prolonged word followed by ip-internal tone
	3p	Prolonged word followed by post-ip boundary tone
	4p	Prolonged word followed by post-IP boundary tone
Cut	0c	Cut word followed by ip-internal tone in a clitic environment
	1c	Cut word followed by ip-internal tone
	3c	Cut word followed by post-ip boundary tone
	4c	Cut word followed by post-IP boundary tone
Pause	0ps	Pause followed by ip-internal tone in a clitic environment
	1ps	Pause followed by ip-internal tone
	3ps	Pause followed by post-ip boundary tone
	4ps	Pause followed by post-IP boundary tone
Filler	0f	Prosodically cliticized form followed by filler
	1f	ip-internal word boundary followed by filler
	3f	ip boundary followed by filler
	4f	IP boundary followed by filler

Table 5: Break indices for single disfluencies

For the single disfluency types listed in Table 3, each juncture degree and corresponding break index was categorized in accord with the pattern in Table 5: for instance, when a word was prolonged and followed by an ip-internal tone in a clitic environment, the break index after this word was labeled '0p', and so forth. The only disfluency coded differently was the 'filler', which was labeled at the juncture immediately *preceding* the filler. As Figure 9 shows, the filler "um" is coded with the break index '1f', since the filler is preceded by a phrase-internal word boundary. The break index is labeled immediately before the filler, aligned with the word boundary separating the filler from the preceding word "the":

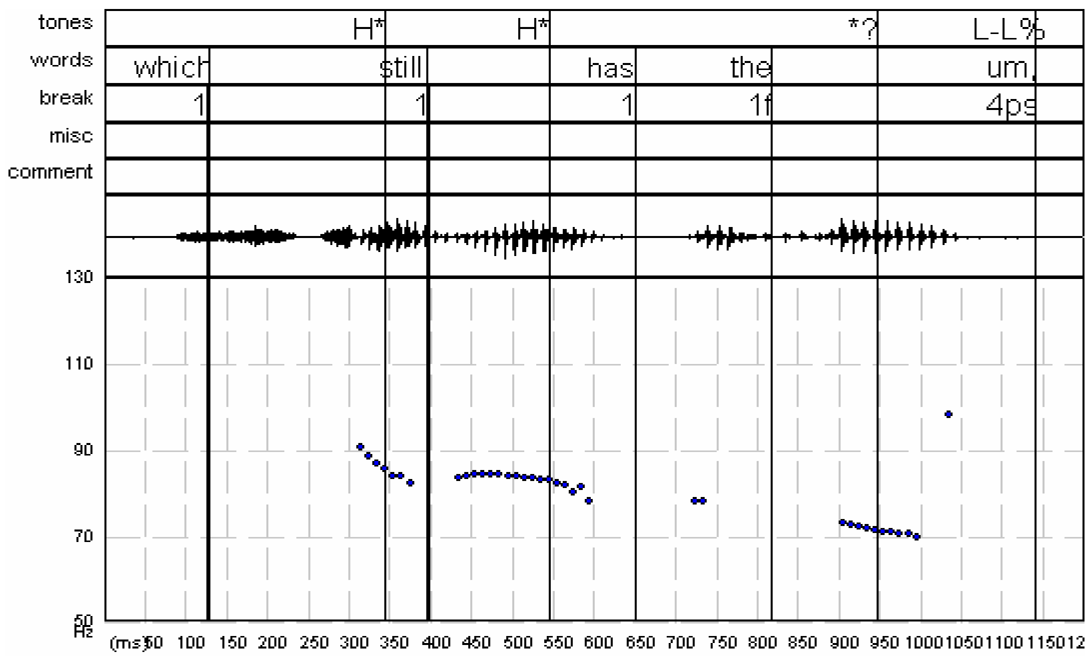


Figure 9: Pitch track illustrating a filler disfluency ('1f' disfluent break)

Combination disfluencies were assigned break index diacritics in the same way as were single disfluencies. Examples of both types will be looked at in greater detail in the following section.

3.4.2 Examples of coding disfluencies

An example of a canonical break index pattern in fluent speech can be seen below in Figure 10. Each word is separated by the break index '1' indicating a normal phrase-internal word boundary, while the phrase terminates in the Intonation Phrase break index marker '4':

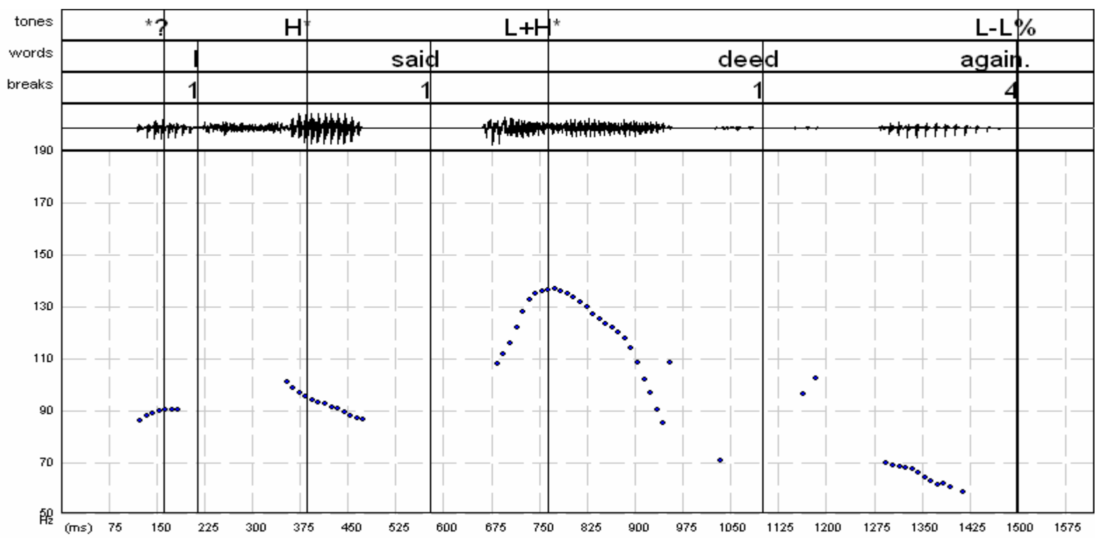


Figure 10: Pitch track illustrating canonical break indices '1' (word break) and '4' (IP break)

The effect of a stuttered disfluency—in this case, a prolongation of the first segment in the word "falter"—on the break index pattern can be seen in the pitch track for the similarly structured sentence below:

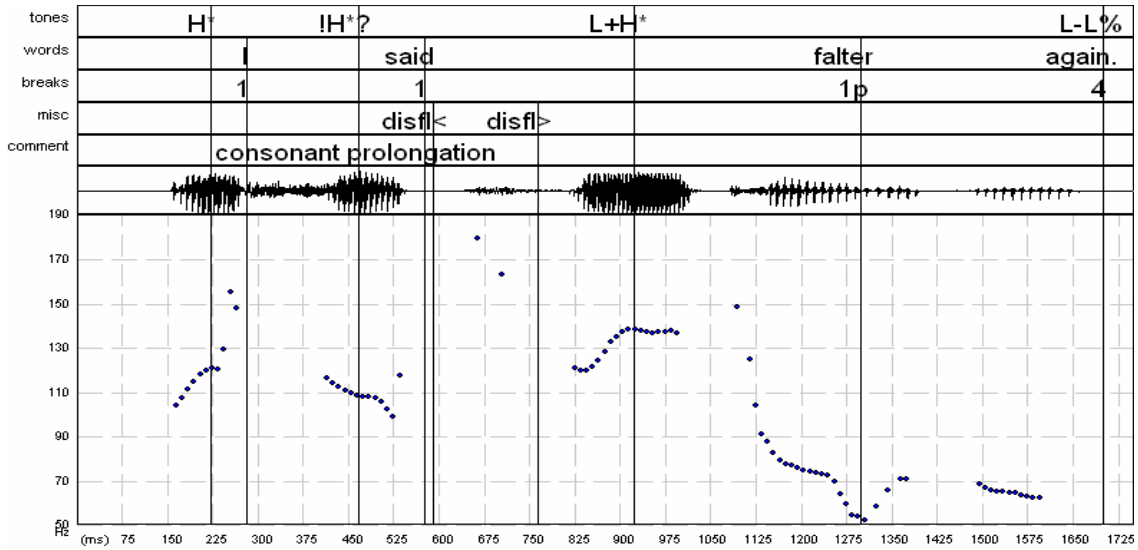


Figure 11: Pitch track illustrating a prolongation (‘1p’ disfluent break)

As observed by comparing Figures 10 and 11, the disfluency marked by the ‘1p’ in the latter figure does not precipitate an intermediate phrase break; rather, phonation has already been initiated for the onset of the following word “falter”, observed by the slight frication energy in the waveform after the offset of the word “said” (marked by the ‘disfl’ labels in the ‘misc’ tier[9]). In accord with the descriptions from Table 5, the break index ‘1p’ is marked immediately following the word in which the prolongation occurs, although the actual disfluency surfaces earlier in the word. In all other respects, the pitch patterns of “said” and the following words in both figures are identical.

The example of a pause below, in contrast, shows a clear absence of phonation following the break and, crucially, preceding the onset of the problematic word:

[9] The misc tier is used to specify a disfluency’s location, marking both the beginning and the end of the event by the diacritics ‘<’ and ‘>’, respectively.

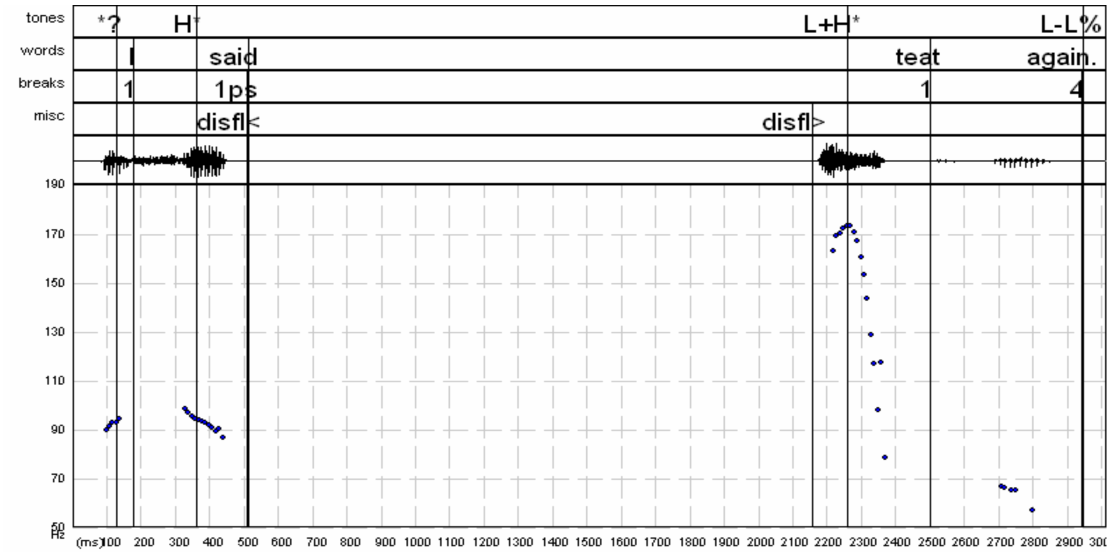


Figure 12: Pitch track illustrating a pause ('1ps' disfluent break)

Unlike the previous example, the break occurs before the onset of the word "teat", and thus earns the label '1ps' after the word "said". The disfluent pause does not provoke a phrase break, which is observable from both the lack of lengthening (i.e., pre-boundary disjuncture cue) and the absence of a phrasal tone (i.e., pre-boundary tonal cue) in the word "said". The L+H* on nuclear pitch-accented "teat" may suggest the re-setting of tone found after ip breaks (i.e., post-boundary tonal cue), but the lack of boundary cues on the preceding word, "said", suggests that the L+H* on "teat" is not the beginning of a new ip.

An example where a stuttered pause induces an intermediate phrase break is displayed in Figure 13:

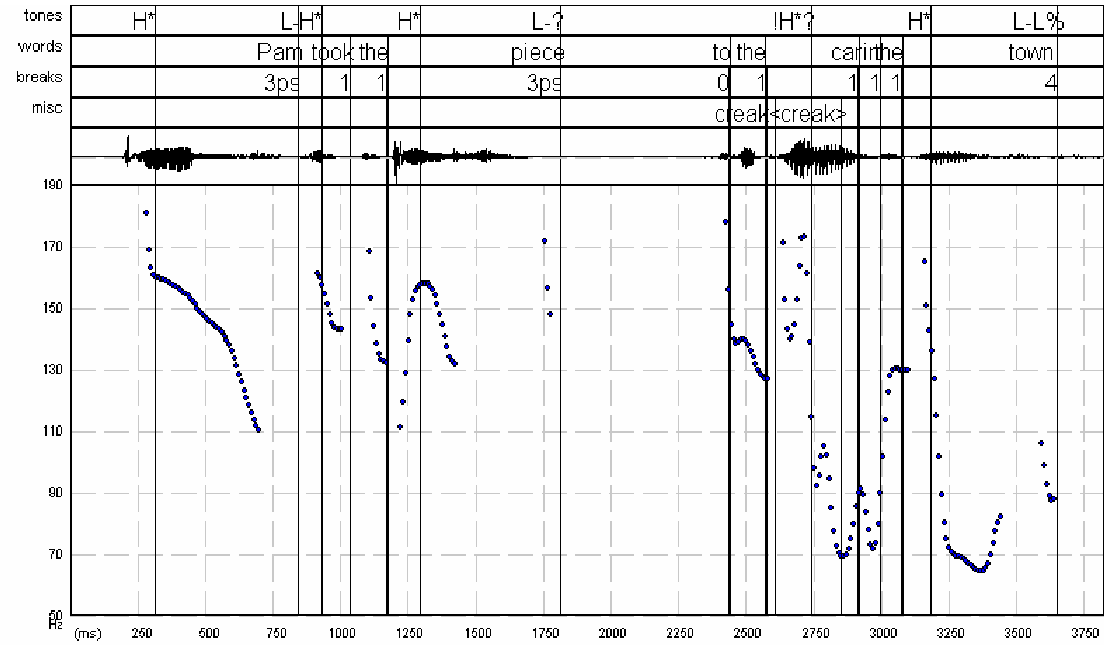


Figure 13: Pitch track illustrating a 3ps: "Pam took the piece to the car in the town."

During the production of the initial pitch-accented word "Pam", a stuttered pause—i.e., prolongation of material crucially preceding the onset of the following word—triggers an intermediate phrase break. The evidence of this phrase break is visible in both the prolonged material before the verb "took", as well as in the unorthodox pitch-accenting of the verb itself. Rather than receive downstepped pitch accent as would occur in a canonical declarative that produces nuclear pitch accent in phrase-terminal position (or no pitch accent at all, as is also common for verbs in longer declaratives), the F0 value during the production of the verb "took" is as high as that of phrase-initial "Pam". Similarly, a second break later in the utterance (i.e., after "piece") signals another pause-level disfluency. While the pause appears relatively long in duration, the question mark after the '3ps' break index indicates an ambiguity in the pitch track cues:

on the one hand, while the pause itself is strong evidence for a phrasal juncture, the downstepped pitch accent over the following word, “car”, would be an indication that an intermediate phrase break did not occur, since downstep is an intermediate phrase-internal phenomenon. It is also possible, however, that the relatively lower pitch accent reflects declination—a gradual and non-categorical lowering of pitch throughout the utterance (Ladd 1996)—and thus the break indeed constitutes a phrase juncture.

In Figure 14, a number of disfluency types are exemplified, including a cut (‘1c’), a restart (‘1t’), a prolongation (‘1p’), and a pause (‘4ps’). The F0 information before the L+H* pitch accent on “frog” is unclear, but since no boundary cues are present after the cut word fragment “f-”, the boundary is labeled with break index ‘1c’. Likewise, the breaks following both restarts (i.e., ‘1t’) mark simple word boundaries:

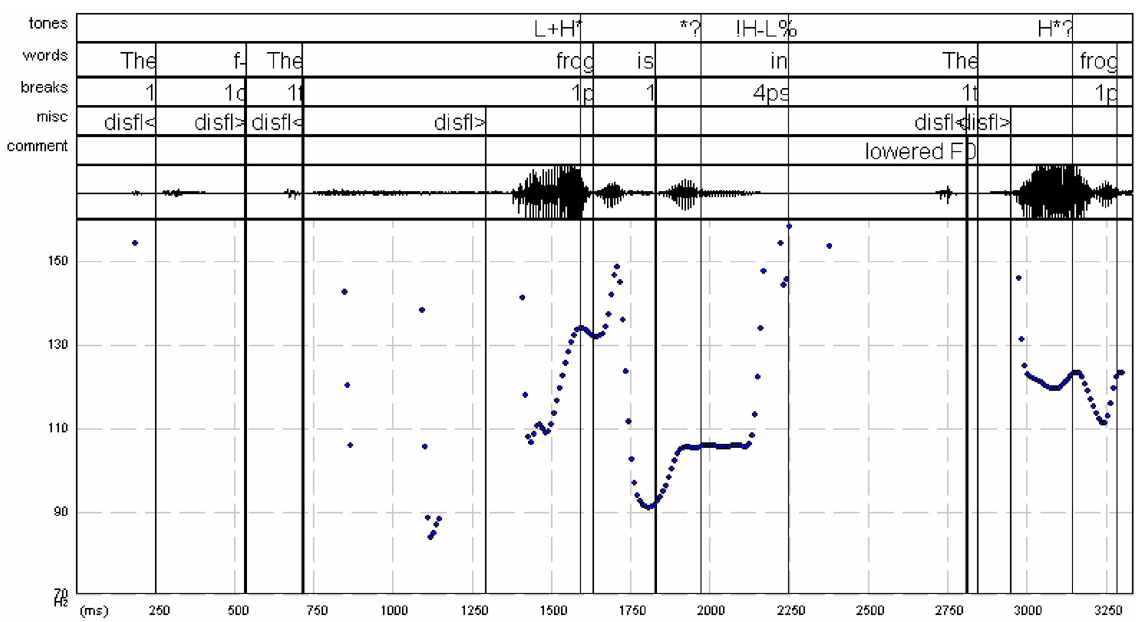


Figure 14: Pitch track illustrating a cut (1c), restart (1t), a prolongation (1p), and a pause (4ps): “The f— the frog is in— the frog”

The only disfluency in Figure 14 resulting in a phrase break is the pause after “in”. The downstepped mid-tones realized on “in” are evidence of a !H-L% bitone[10], and thus a pre-boundary tonal cue for an IP break. Likewise, the pause after “in” functions as a post-IP-boundary disjuncture cue, while the re-setting of F0 on the next pitch accented word “frog” functions as a post-boundary cue. Interestingly, the only missing cue is pre-boundary lengthening, whose absence is indicative of an abrupt—and therefore, presumably, unplanned—break. The fact that disfluencies often result in either the subtraction or addition of boundary cues, and that these cue patterns can also serve as clues to the underlying cause of the disfluency—are important topics that will be investigated in later chapters.

Figure 15 demonstrates a combination disfluency—i.e., one in which more than one of the basic disfluency types occurs in the same prosodic environment. As indicated by the ‘disfl<’ label on the ‘misc’ tier, a disfluency surfaces during the production of the vowel in the word “out”. The disfluency continues after the word boundary, evidenced by the ill-placed pause before the preposition “of”. Because the disfluency contains these two components—a prolongation and an adjacent pause—the boundary is labeled with disfluency break index ‘1p.ps’. Evidence for break index ‘1’ is found in the following pitch accent on the word “of”, which is downstepped and therefore ip-internal.

[10] It is also possible that this mid-tone plateau constitutes simply an ip boundary tonal cue (i.e., !H-), although the fall from the L+H* peak suggests downstepping.

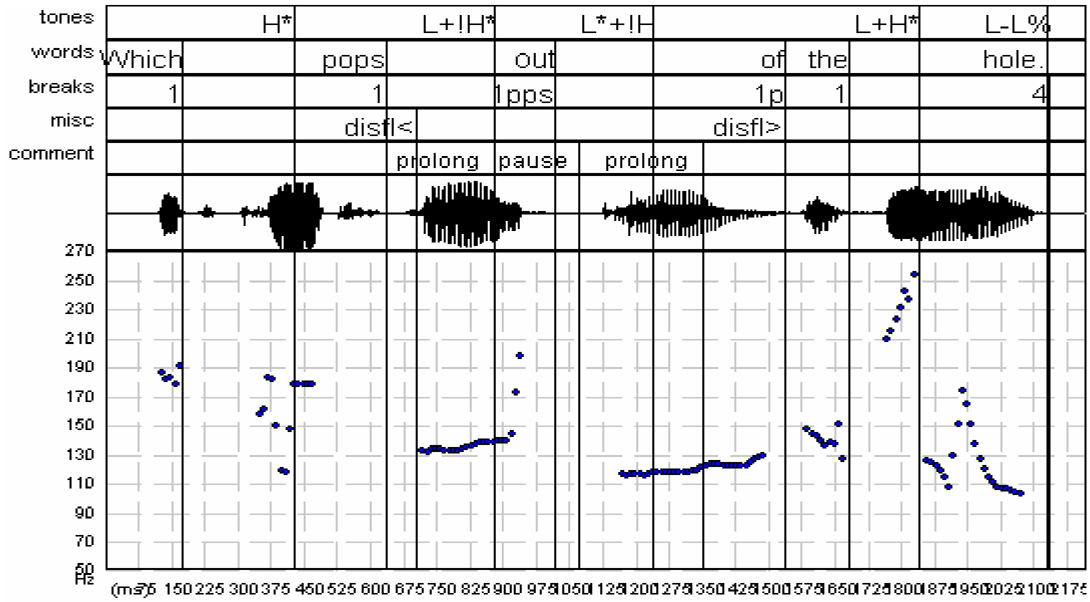


Figure 15: Pitch track illustrating a combination (1pps) and prolongation (1p): "Which pops out of the hole"

A second prolongation+pause disfluency is illustrated in Figure 16, although in this case a re-setting of F0 on the next pitch-accented word after the disfluency provides evidence of an IP break. Unlike the '1p.ps' example in Figure 15, where the post-disfluency pitch accent was downstepped, the H* pitch accent realized on the word "cliff" in Figure 16 is clearly re-set relative to the falling F0 produced during the 'p.ps' disfluency. Together with the presence of the three other boundary cues—i.e., pre-boundary lengthening, pre-boundary F0 movement, and a post-boundary pause—this prosodic information motivates an interpretation of the post-disfluency break as an IP boundary ('4p.ps').

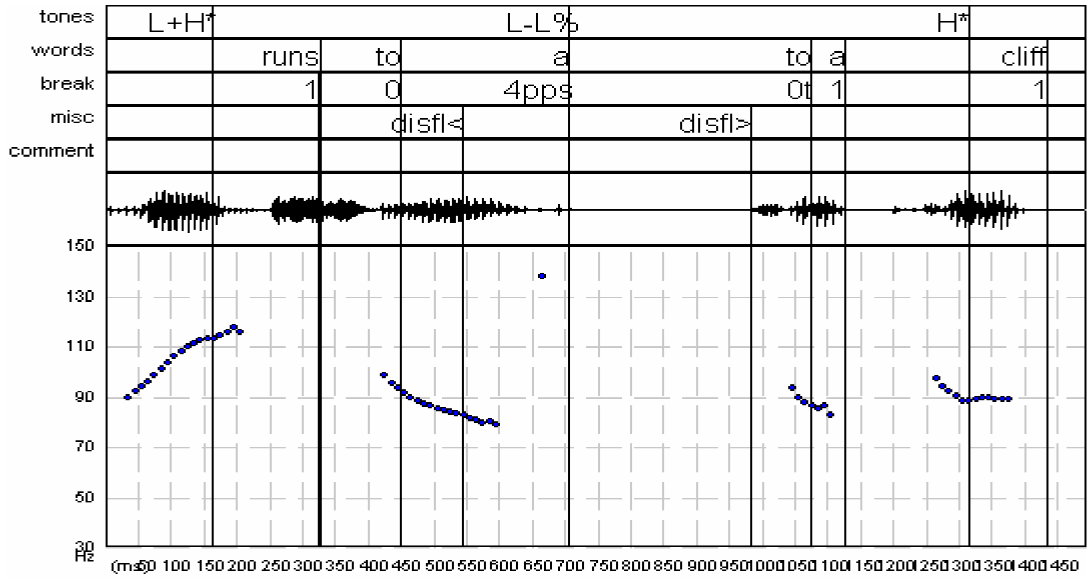


Figure 16: Pitch track illustrating a combination (4pps) and restart (1t): "runs to a—to a cliff"

The final example in Figure 17 shows a somewhat more complex combination disfluency. A 'p.c.ps' combination—which consists of a prolongation, followed by a cut of the attempted word, and finally a subsequent pause—surfaces on the segment "l-". The speaker successfully generates a word's initial segment [l], but a disfluency prevents continued production into the next segment. Instead, the onset is prolonged until the word is abruptly cut off and followed by a pause. These three elements together—the prolonged segment, cut word, and pause—combine to form a p.c.ps disfluency:

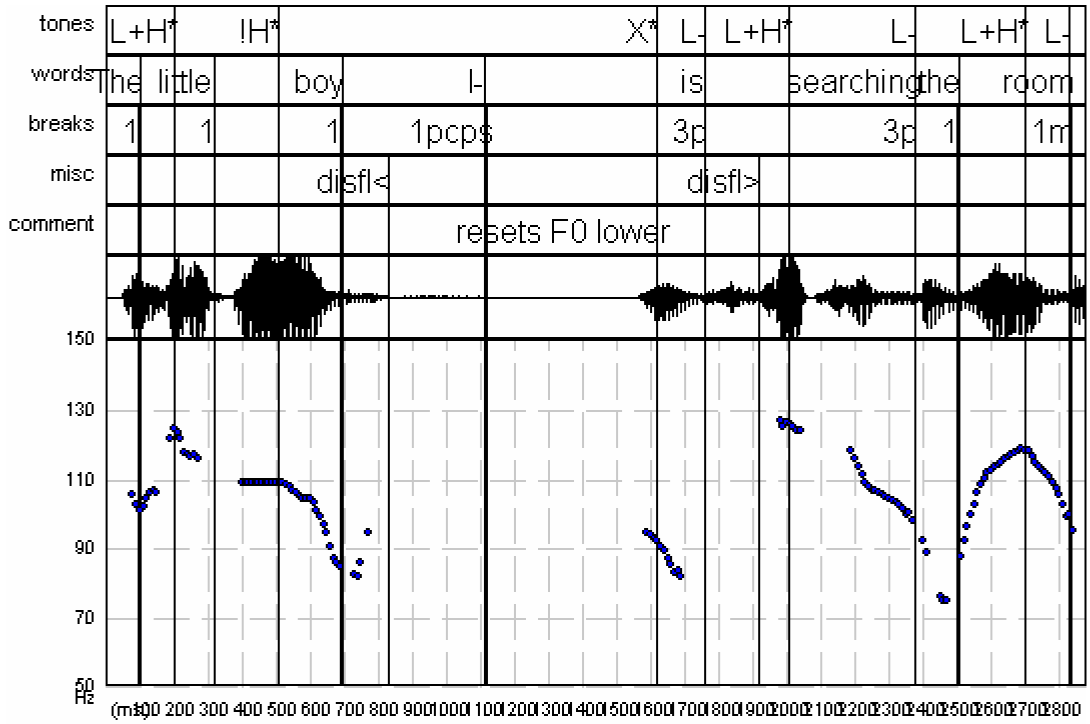


Figure 17: Pitch track illustrating a combination disfluency (prolongation+cut+pause): "The little boy l— is searching the room"

When production is resumed, the speaker produces another disfluency in prolonging the nucleus of the word "is", and a final disfluency in prolonging the onset of the word "searching". Comparing the tonal behavior *following* each disfluency, it is apparent that the latter two prolongations occur at intermediate phrase boundaries, since F0 is reset both in the L+H* pitch accent produced during the word "searching", and in the L+H* realized on the word "room". Thus, disfluency break index '3p' is assigned to both of these prolongations. The first disfluency (i.e., 1p.c.ps), however, is more ambiguous. Although the F0 fall at the end of the word "boy" might constitute a pre-boundary tonal cue, the F0 signal realized on the prolonged vowel of the next fully produced word "in" appears prominent, and thus pitch-accented. Moreover, since the F0 of this pitch accent is not fully re-set, then it is possible that the pitch accent is downstepped and thus ip-internal, therefore motivating a classification of the previous word boundary (i.e., after the cut word ‘l-’) as disfluency break index ‘1p.c.ps’. The ambiguity results from the difficulty in determining what type of pitch accent is manifested on the vowel of the word “is”, since this pitch accent—having surfaced during a disfluency—was ostensibly unintended. Therefore, the pitch accent is labeled ‘X*’.

3.4.3 Mismatches

The final break index category, indicated by the diacritic ‘m’, is used to code cue mismatches which result from a speaker’s intentional variation of speaking rate or style. As shown previously in Figure 6, pre- and post-boundary disjuncture and tonal cues indicate when an intermediate or Intonational Phrase boundary has been produced. Frequently, however, one or more cues may be added or removed, such as when speaking rate is increased or decreased. The example in Figure 18 illustrates two mismatches occurring at the right edges of the words “boy” and “dog”. While pre-boundary lengthening at the ends of both words provides evidence of an ip boundary, neither word is followed by a post-boundary resetting of F0. Thus, the break juncture is a ‘3’, but since it is mismatched with an ip-*internal* post-boundary cue (i.e., lack of F0 resetting), the diacritic ‘m’ is added to indicate that the boundary does not signal a new intermediate phrase.

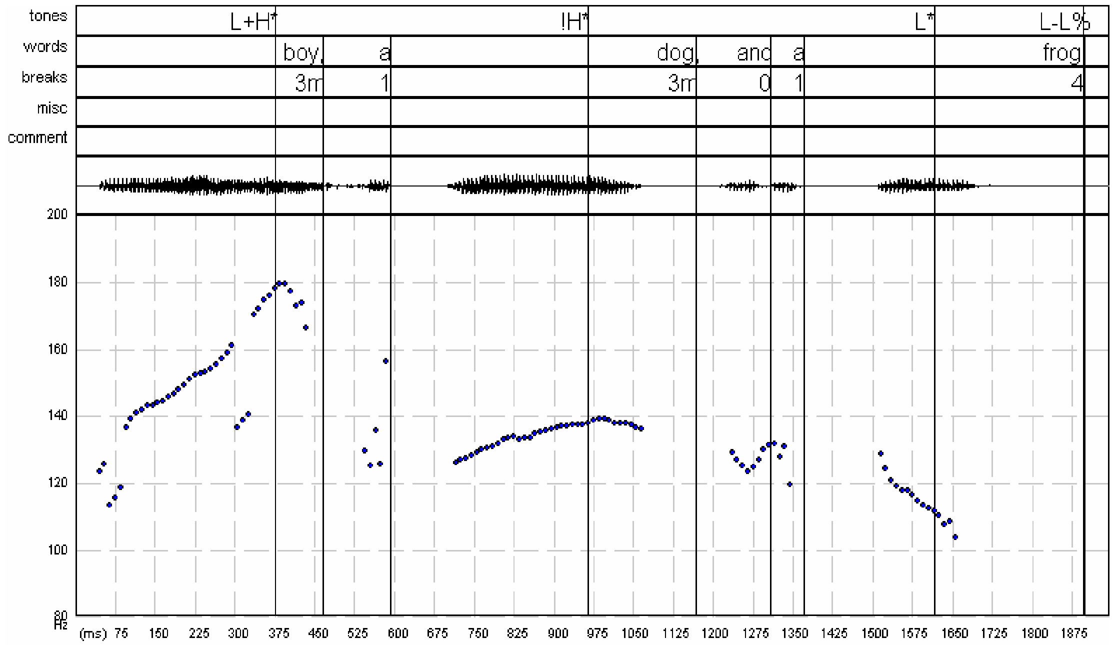


Figure 18: Pitch track illustrating a 3m mismatch: "boy, a dog, and a frog"

As was the case with disfluency break indices, the inventory of mismatched break indices was also expanded to accommodate the full range of mismatch combinations encountered in the data. In each type, the break index number represents the degree of disjuncture occurring at that boundary, followed by the 'm' diacritic to mark the presence of conflicting tonal cues. The complete list of mismatch categories used in this analysis is presented in Table 6.

Mismatch	Break Indices	Description
0m	0 vs. 3	juncture is like reduced clitic boundary; though accompanied by *ip* boundary tone (i.e., juncture size a 0 but tonal behavior like a 3)
0m+	0 vs. 4	juncture is like reduced clitic boundary; though accompanied by *IP* boundary tone (i.e., juncture size a 1 but tonal behavior like a 4)

1m	1 vs. 3	juncture is like word boundary; though accompanied by *ip* boundary tone (i.e., juncture size a 1 but tonal behavior like a 3)
1m+	1 vs. 4	juncture is like word boundary; though accompanied by *IP* boundary tone (i.e., juncture size a 1 but tonal behavior like a 4)
3m	3 vs. 1	juncture is like *ip* boundary, though not accompanied by tonal change (i.e., juncture size a 3 but tonal behavior like a 1)
3m+	3 vs. 4	juncture is like *ip* boundary, while tonal cues indicate an IP boundary (i.e., juncture size a 3 but tonal behavior like a 4)
4m	4 vs. 1	juncture is like *IP* boundary, though not accompanied by tonal change (i.e., juncture size a 4 but tonal behavior like a 1)
4m+	4 vs. 3	juncture is like *IP* boundary, while tonal cues indicate an ip boundary (i.e., juncture size a 4 but tonal behavior like a 3)

Table 6: Break indices for mismatches

CHAPTER 4

RESULTS 1: SPONTANEOUS NARRATION DATA

This chapter will look at the results of the storybook narration task, in which subjects were recorded while giving spontaneous narrations of a picture book. The chapter will be divided into two main sections: 1) descriptions and comparisons of major prosodic categories and disfluency types; and 2) more detailed analyses of prosodic behavior in the disfluent environment. All data were analyzed with Fisher's exact test, unless otherwise noted. Differences of all statistical comparisons were considered significant at $p < 0.05$.

4.1 Description of narration data

Table 7 below shows the descriptive statistics of the narrative in terms of the raw numbers for words, intonation phrases (IP), intermediate phrases (ip), disfluencies, and pitch accents (PA), as well as the average number of words, disfluencies, and PAs within each ip:

	Stutter1	Stutter2	Stutter3	Stutter4		Control1	Control2	Control3	Control4
# words	1212	571	759	2814		472	494	1022	712
# ip	357	153	183	701		127	117	197	151

Avg. wd/ip	**3.53**	**3.76**	**4.15**	**4.13**	4.21	4.26	7.51	4.72
# disfl.	374	190	190	1140	19	16	44	124
Avg. disfl./ip	**1.09**	**1.25**	**1.03**	**1.68**	0.15	0.14	0.22	0.82
sd of disfl	1.12	1.11	1.04	1.48	0.38	0.47	0.56	0.95
# PA	566	310	323	1257	256	298	424	279
Avg. PA/ip	**1.65**	2.05	**1.75**	1.85	2.00	2.57	2.15	1.85
sd of PA	0.87	1.11	0.93	1.01	0.94	1.04	0.99	0.93

Table 7: Phrases, disfluencies, and pitch accents.

4.1.1 The length and prosodic structure of the narrative

As observed in Table 7, total number of words in a narrative differed widely among subjects of both groups. However, the average number of words in an intermediate phrase (ip) did not differ considerably among subjects. But data do show some differences between groups. While stuttering subjects displayed a relatively large degree of variation in total words produced per ip, variation among controls was less pronounced. In addition, stutterers produced slightly shorter ips (average 3.53-4.15 words/ip) than did controls (average 4.21-7.51 words/ip), and these ranges were non-overlapping. In a paired two sample t-test for means, a group comparison of word/ ip rate fell short of significance (p= 0.0811, one-tailed).

4.1.2 Pitch accents

One slight trend appeared with pitch accent rates. 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: 279-424). Rates of pitch accent per ip, however, were more consistent among subjects of both groups. Furthermore, stutterers produced significantly fewer PAs per ip than did controls (p= 0.033, one-tailed).

4.1.3 Disfluencies

As anticipated, stuttering subjects produced a higher rate of disfluency and with a greater 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. Table 8 shows the distribution of disfluency types for each subject in the six categories defined in Chapter 3: restarts ('t'), prolongations ('p'), cuts ('c'), pauses ('ps'), fillers ('f'), and combinations ('comb'):

Types\subject	S1	S2	S3	S4	C1	C2	C3	C4
restarts	56	17	18	169	1	4	3	30
avg/ip	0.15	0.11	0.10	0.25	0.01	0.03	0.02	0.20
SD	0.39	0.34	0.31	0.49	0.09	0.18	0.12	0.42
% of total disfl	14%	9%	9%	15%	5%	24%	7%	24%
prolongations	74	57	68	182	2	4	11	6

avg/ip	0.22	0.38	0.37	0.27	0.02	0.03	0.06	0.04
SD	0.45	0.52	0.59	0.50	0.12	0.22	0.23	0.20
% of total disfl	**20%**	**30%**	**36%**	**16%**	10%	25%	25%	5%
cuts	33	20	14	116	1	2	1	9
avg/ip	0.10	0.13	0.08	0.17	0.01	0.02	0.01	0.06
SD	0.31	0.38	0.28	0.42	0.09	0.13	0.07	0.24
% of total disfl	9%	11%	7%	10%	5%	13%	2%	7%
pauses	109	38	28	93	14	3	17	38
avg/ip	0.32	0.25	0.15	0.14	0.10	0.03	0.09	0.25
SD	0.54	0.48	0.36	0.38	0.33	0.16	0.30	0.48
% of total disfl	30%	20%	15%	8%	**70%**	**19%**	**39%**	**31%**
fillers	8	4	14	132	0	0	1	14
avg/ip	0.02	0.03	0.08	0.19	0	0	0.01	0.09
SD	0.15	0.16	0.27	0.43	0	0	0.07	0.29
% of total disfl	2%	2%	7%	12%	0%	0%	2%	11%
combinations	94	54	48	447	2	3	11	27
avg/ip	0.27	0.36	0.26	0.66	0.02	0.03	0.06	0.18
SD	0.56	0.59	0.53	0.90	0.12	0.16	0.25	0.49
% of total disfl	**25%**	**28%**	**25%**	**39%**	10%	19%	25%	22%

Table 8: Distribution of disfluency types

Disfluency data were collapsed among subjects for each group in order to compare overall disfluency distribution patterns, as shown in Figure 19:

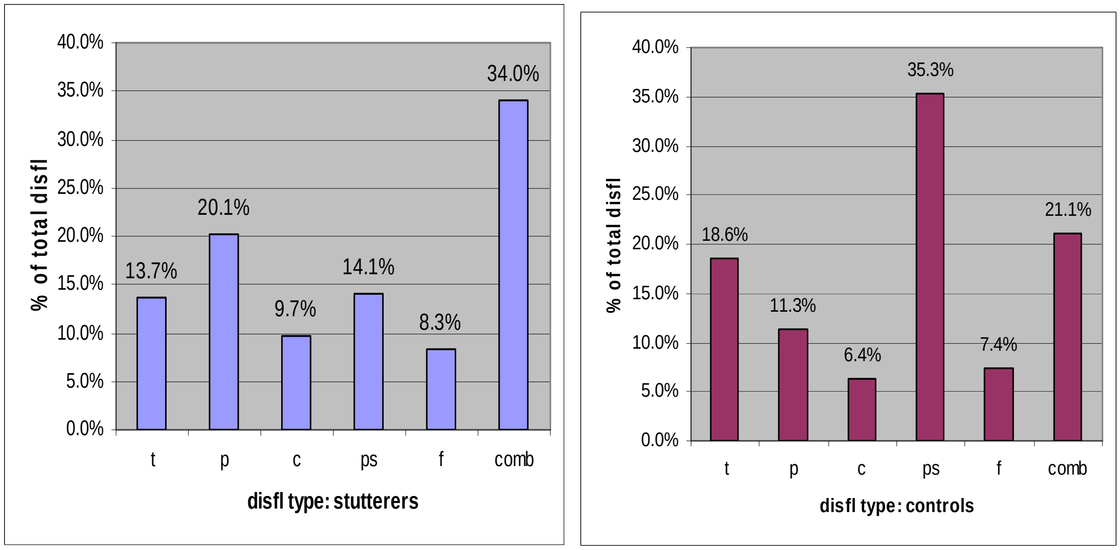


Figure 19: Disfluency type distribution—all stutterers and all controls

Overall, a number of disfluency types were represented in similar proportions between the two groups: restarts (t), prolongations (p), and fillers (f), all ranged between 8.3% and 13.7% for stutterers, and between 6.4% and 18.6% for controls. However, while combinations (34%) and prolongations (20.1%) were the most common type among stutterers, control subjects produced pauses much more frequently than all other types (35.3%). Comparing across groups, stutterers produced significantly more prolongations relative to other disfluency types than did controls ($p=0.002$), while controls produced a significantly greater number of pauses relative to other types than did stutterers ($p<0.0001$). Finally, cuts and fillers each constituted fewer than 10% of the overall disfluencies for both groups.

4.2 Disfluency correlations

As observed in the previous section, while stutterers clearly differed from controls in rate of disfluency, the proportions of disfluency distribution were more similar. However, in order to determine whether contextual effects, such as adjacent tonal behavior and phrasal position of the disfluency, resulted in different patterns between stutterers and controls, these environmental behaviors were examined for each disfluency. Disfluencies were thus examined for the following conditions: degree of prominence of the disfluent word, position of the disfluency within the ip, tonal type in the disfluent environment, and whether the disfluent word was a function or content word.

4.2.1 Disfluencies and prominence

The first condition analyzed was the metrical prominence level of the disfluent environment. For all disfluency types except pauses, this was found by locating the word on which a disfluency was realized: for example, with a prolongation, the prominence level of the prolonged word was recorded. With pauses, however, since the manifestation of the disfluency is non-segmental—i.e., a pause is a period when phonation is absent—the prominence of the following word was recorded instead. This decision was also motivated by the fact that pause disfluencies generally signal production problems with upcoming material—whether lexical or phonetic/phonological.

Combinations were coded in the same manner as were other single non-pause disfluencies, with the exception of combinations that included pauses. Again, since the locus of a pause disfluency is the material following the pause, the prominence level of the word immediately after the pause was recorded as part of the disfluency. Thus, a combination disfluency such as a 'p.ps' (prolongation + pause) was coded for both prominence level of the prolonged material as well as of the post-pause material. This decision ensured that the multiple disfluent events manifested in some combination disfluencies would be accurately recorded and analyzed.

Disfluent words were classified as belonging to one of three levels of metrical prominence: pitch-accented (PA), nuclear pitch-accented (NPA), and unaccented (unacc)[11]. Figure 20 shows a comparison of prominence level of the disfluent words for both groups, merging all subjects:

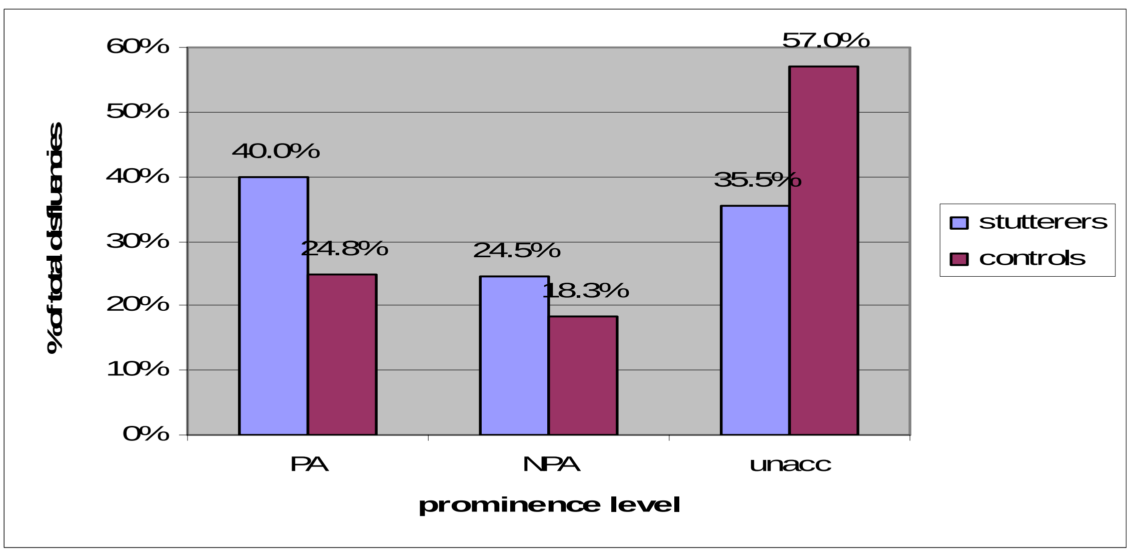


Figure 20: Prominence level of disfluent word: stutterers vs. controls

---

[11] In the model of intonational phonology used here, "unaccented" denotes prosodic material that is without pitch accent; it may be stressed or unstressed.

While the difference in rate of disfluencies occurring with NPAs was relatively small in a group comparison (S=24.5%; C=18.3%), stutterers and controls showed opposite tendencies for pitch-accented vs. unaccented forms. Stutterers produced slightly more disfluencies with pitch accented words (40%) than they did without any accent at all (35.5%), while controls produced more unaccented than pitch-accented disfluencies (57% vs. 24.8%). Furthermore, while over half of control disfluencies (57%) co-occurred with unaccented forms, nearly two-thirds (64.5%) of stuttered disfluencies were on pitch-accented—i.e., either pitch-accented or nuclear pitch-accented—forms.

Breakdown of individual data for stutterers in Figure 21 shows the same trend: over half of each speaker's disfluencies occurred on pitch-accented or nuclear pitch-accented words. Individual controls behaved less consistently, as seen in Figure 22. Speakers C1 and C2 produced patterns similar to those of the stutterers: i.e., a greater percentage of disfluencies on words carrying some type of pitch accent. C3 produced more disfluencies on unaccented words than on either PA'd or NPA'd words, with the latter comparison reaching significance ($p=0.006$). Similarly, C4 also produced significantly more disfluencies on unaccented words than on PA'd ($p<0.0001$) and NPA'd ($p<0.0001$) words.

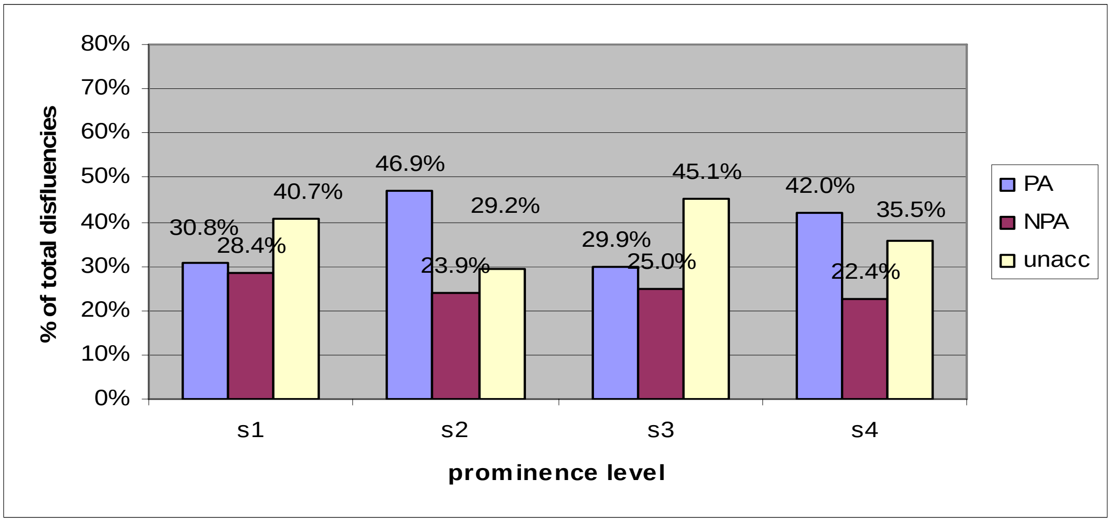


Figure 21: Prominence level of disfluent word: all stutterers

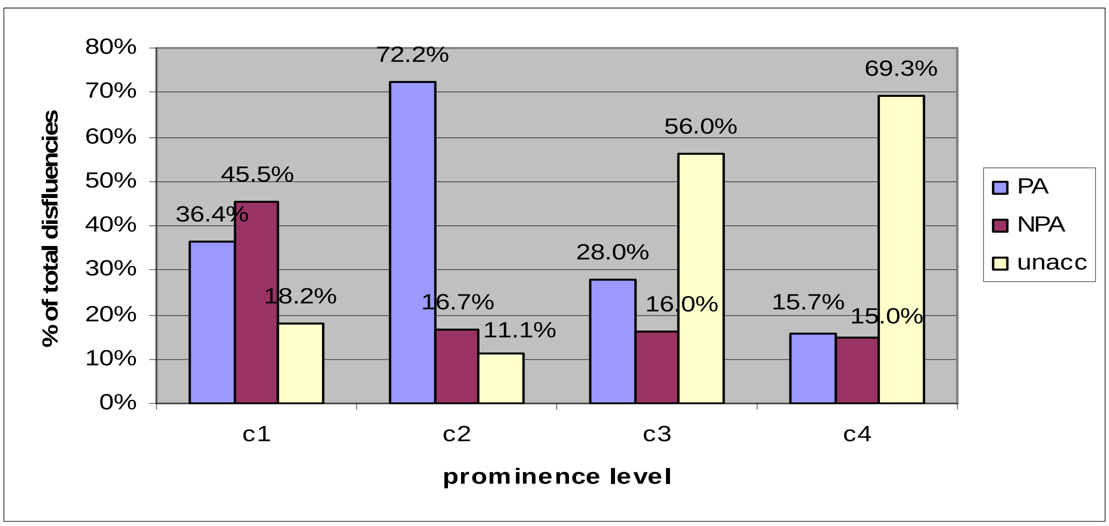


Figure 22: Prominence level of disfluent word: all controls

Since pitch-accented and nuclear pitch-accented words were produced at different rates for each speaker, the disfluency rates for each prominence level were also calculated separately. Collapsing data for all speakers of each group in Figure 23 shows that a substantial percentage of stutterers' pitch-accented words were disfluently produced

(66.5%) while nuclear pitch-accented words contained disfluencies about half as often (31.1%). Only 22.9% of stutterers' unaccented words were disfluent. Within-group comparisons of each prominence level using Fisher's exact test revealed highly significant differences for stutterers ($p < 0.0001$).

As Figure 24 illustrates, within-group individual speaker comparisons support the overall group patterns. Each stutterer generated the highest disfluency rates on pitch-accented words (range: 43.6%-80%), followed by nuclear pitch-accented (26.6%-33.8%) and unaccented (21.1%-24.1%) words.

Controls, meanwhile, while obviously producing much lower disfluency rates than did stutterers, seem to show little difference in rates of disfluency for each prominence type, as shown in Figure 23. Analysis of individual data in Figure 25 confirms this observation.

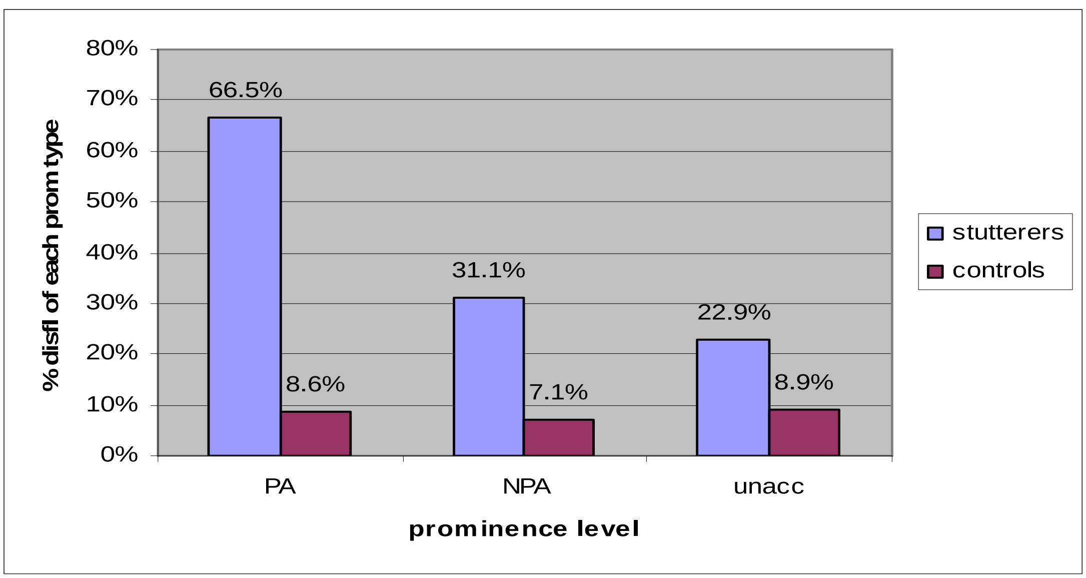


Figure 23: Disfluency rate for each prominence level

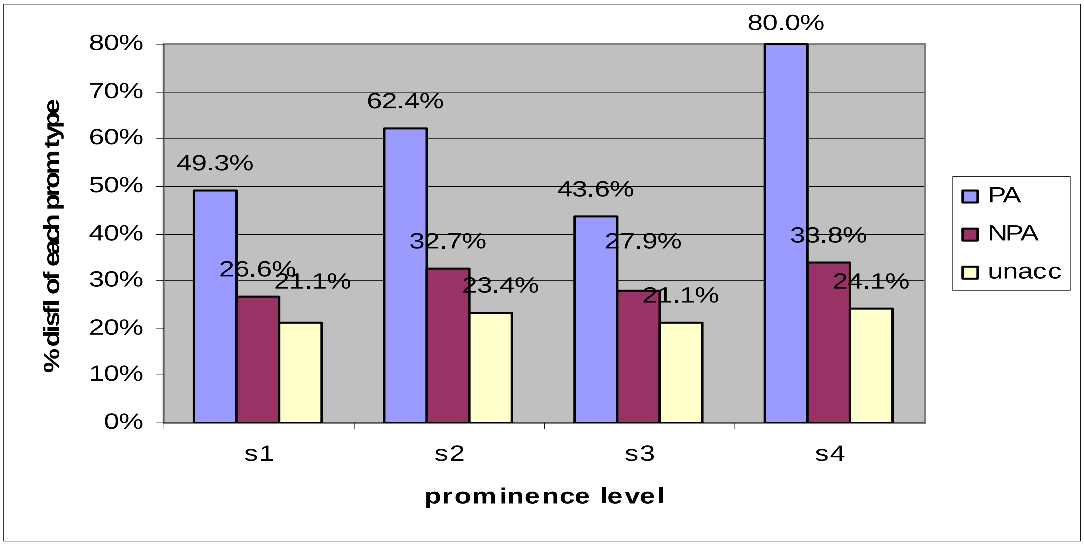


Figure 24: Disfluency rate for each prominence level: all stutterers

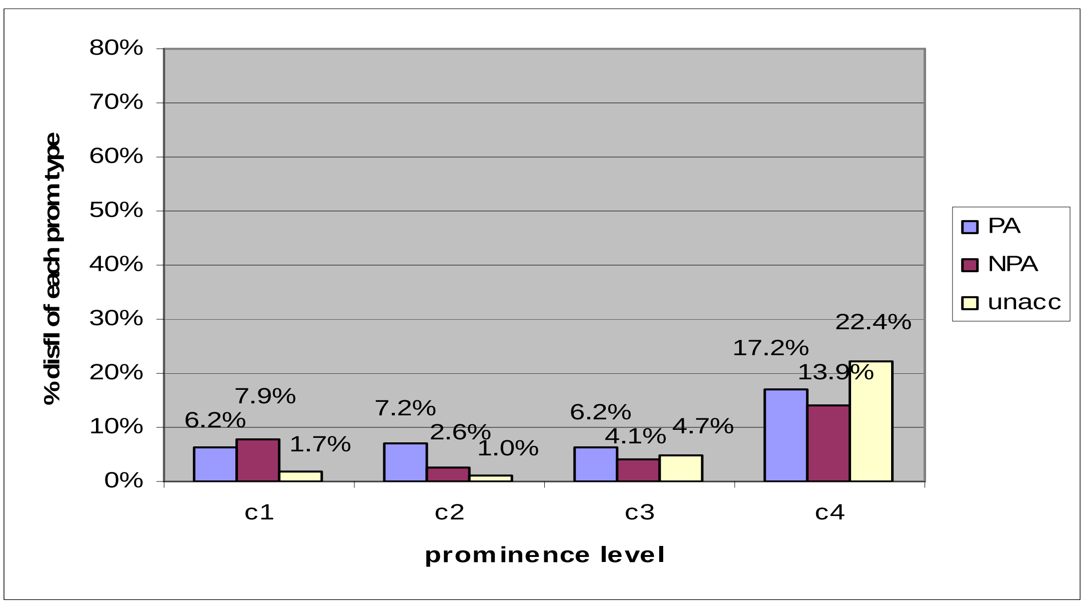


Figure 25: Disfluency rate for each prominence level: all controls

4.2.2 Disfluencies and phrasal position

The locations of disfluencies within prosodic phrases were categorized into three positions: intermediate phrase (ip)-initial, ip-medial, and ip-final position. Pause disfluencies once again required slightly different treatment than other disfluency categories, since pauses can constitute either word boundaries or phrase boundaries themselves—in other words, ip-initial vs. ip-final location is not a possible distinction for pauses. Therefore, pause disfluencies were classified simply as occurring in either ip-medial position or as an ip-boundary.

Combining all subjects revealed some apparent group differences, as stutterers showed a tendency to begin new phrases after producing pauses. As Figure 26 illustrates, 61% of stutterers' pauses constituted ip boundaries, while only 39% were produced ip-medially. Control subjects, however, demonstrated the opposite pattern of maintaining intermediate phrases post-pausally 70% of the time.

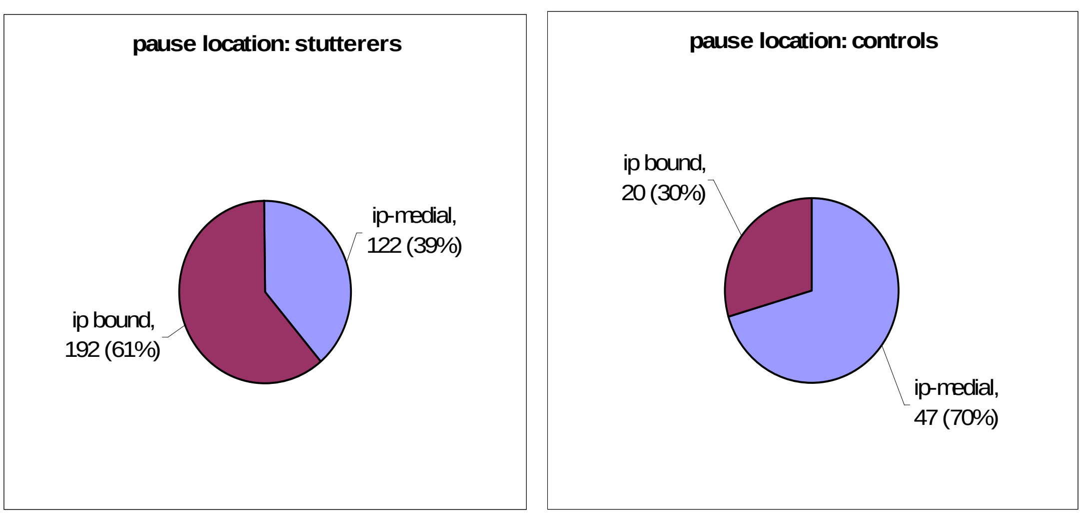


Figure 26: Location of pauses within the intermediate phrase: stutterers vs. controls

Breaking these results down by subject supports the overall trend for stutterers, as all subjects were more likely to begin a new intermediate phrase after a pause than they were to maintain the same phrase. Figure 27 shows that subjects did exhibit some intra-group variation, however, as pause location for S1 differed significantly from S2 (p=0.016) and from S4 (p=0.0007).

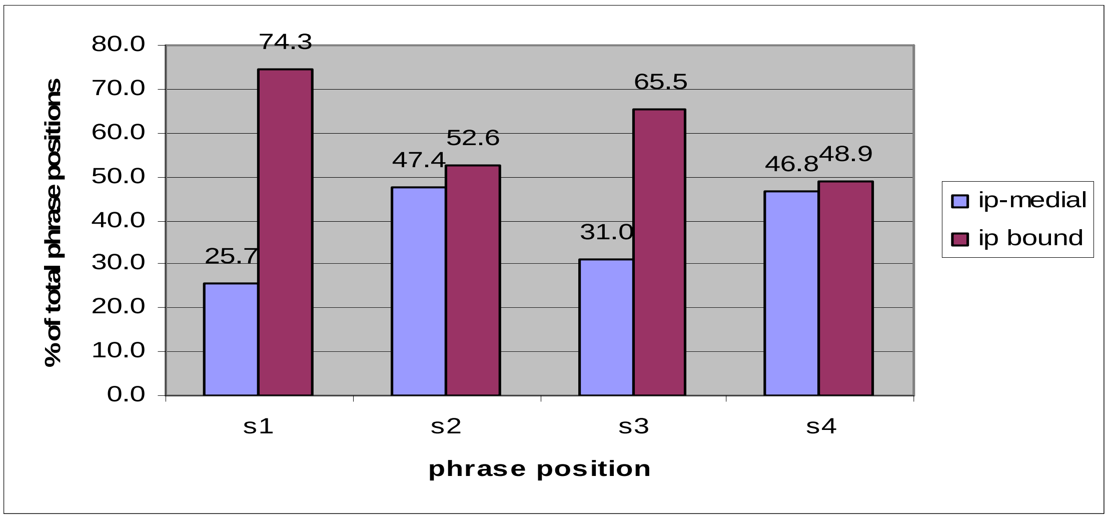


Figure 27: Pause location within the intermediate phrase: individual stutterers

Analyzing the results of individual control subjects, however, revealed a much more inconsistent pattern, as Figure 28 shows. Two subjects (C1 and C3) maintained intermediate phrases after all pauses, though their total number of pauses was quite low (14 and 12, respectively). On the other hand, C2 showed the completely opposite effect, though he produced even fewer pauses (3 pauses total). The low number of control subject pause disfluencies overall (avg. 17.3 pauses/ subject) relative to that of stutterers (avg. 78.5 pauses/ subject) is likely one major reason for such a wide range of pause distribution and lack of a consistent pattern.

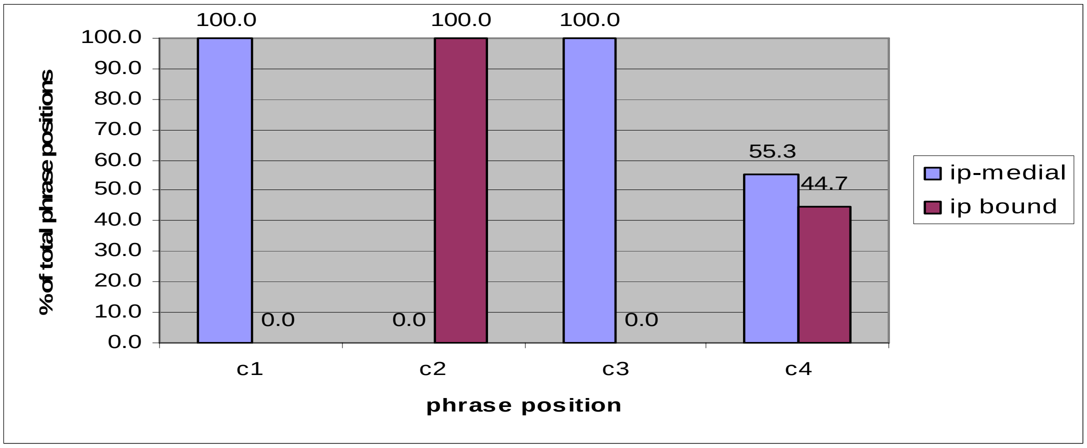


Figure 28: Pause location within the intermediate phrase: individual control subjects

Both groups behaved more similarly in their production of prolongations, as Figure 29 illustrates, locating them most often ip-medially, and least often at the end of an intermediate phrase. However, while stutterers' prolongations occurred in medial position slightly less than half the time (48%), controls produced 65% of their prolongations ip-medially.

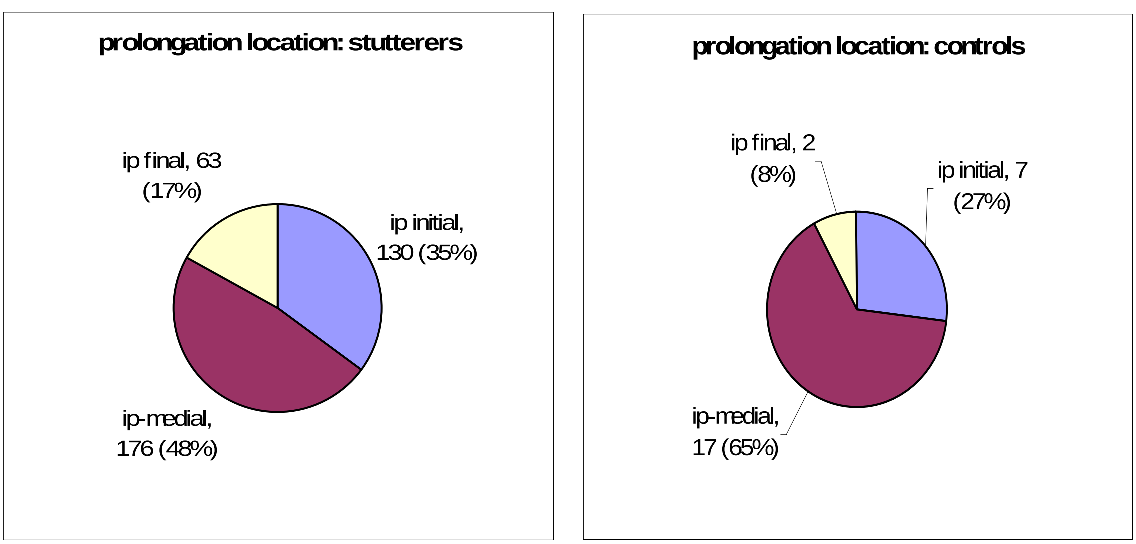


Figure 29: Location of prolongations within the intermediate phrase: stutterers vs. controls

For individual comparisons of stutterers, Speaker S4 behaved differently than the other three speakers in producing significantly fewer ip-final prolongations relative to other positions ($p<0.0001$). S4 also generated significantly more ip-medial prolongations than S3, as Figure 30 reveals. However, S4 still reflected the same general group pattern from Figure 29 by producing the majority of disfluencies in ip-medial position, followed by ip-initial and, lastly, ip-final position. S3, on the other hand, did not replicate the same pattern, as he instead produced ip-medial disfluencies less frequently than in the other two positions.

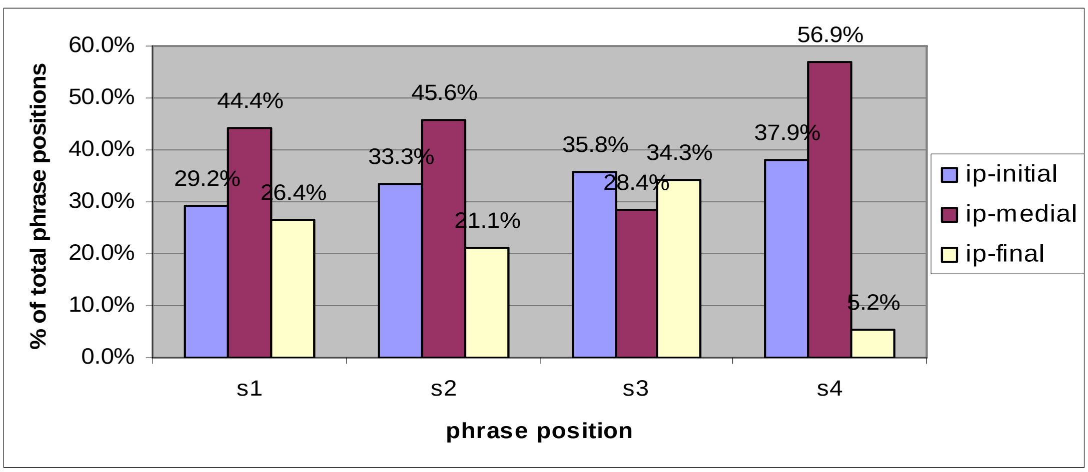


Figure 30: Prolongation location within the intermediate phrase: individual stutterers

Control subjects performed roughly the same in that all subjects but C1 produced ip-medial disfluencies most often, as shown in Figure 31. However, statistical tests were not performed in cases such as these, where individual control subjects' disfluencies were particularly few in number.

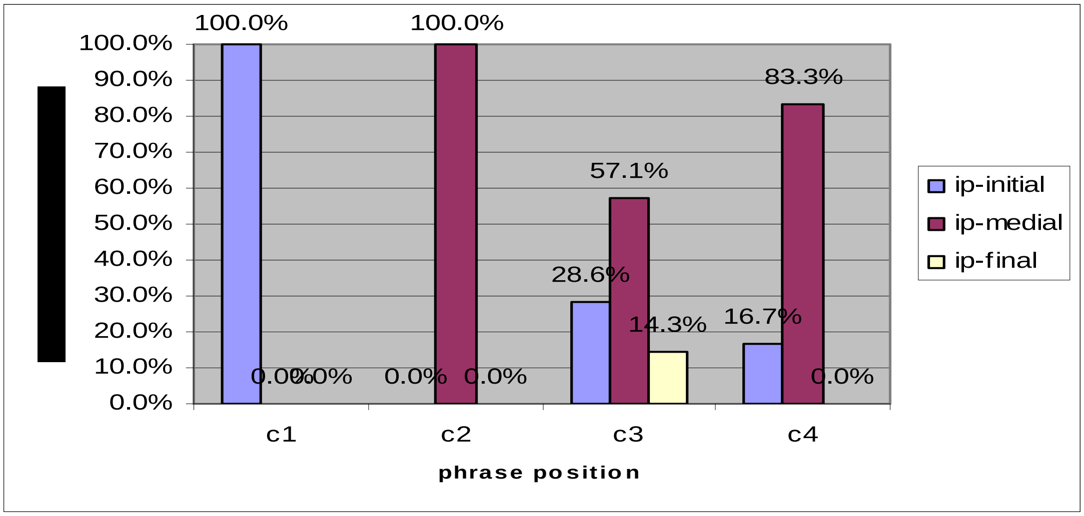
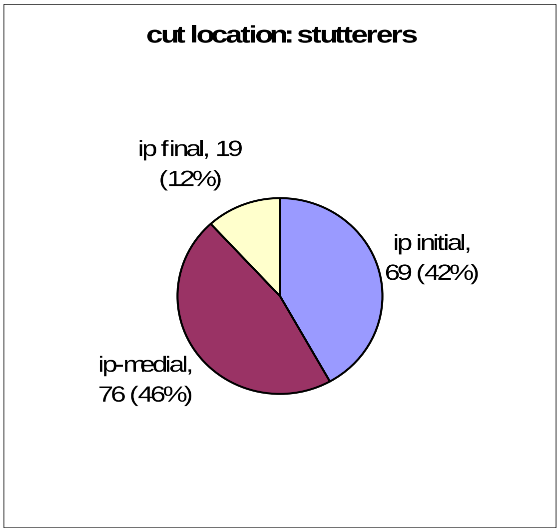


Figure 31: Prolongation location within the intermediate phrase: individual control subjects

Cut disfluencies also occurred most frequently in ip-medial position for both groups, as Figure 32 illustrates. But while stutterers produced the lowest percentage of cuts ip-finally (12%)—as they also did in their prolongation production—controls' cuts were least frequently generated in ip-initial position (19%). The total number of cuts produced by controls was, however, quite low overall (11 tokens). In Figure 33, comparisons of individual stutterers revealed few differences in cut distribution. Individuals in the control group showed more variation across subjects, but this difference may not be meaningful due to the small number of cut-word data.



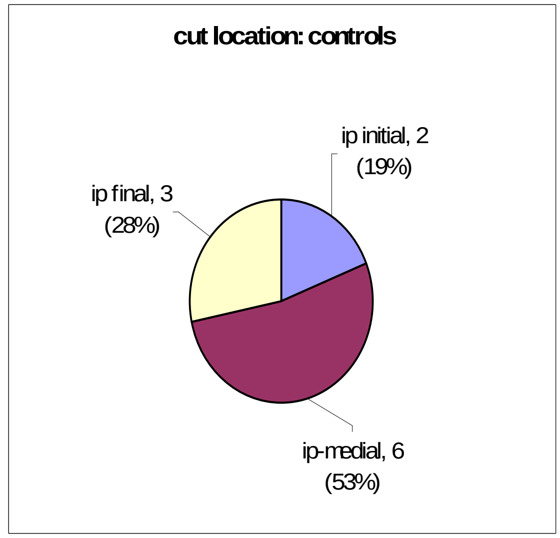


Figure 32: Location of cuts within the intermediate phrase: stutterers vs. controls

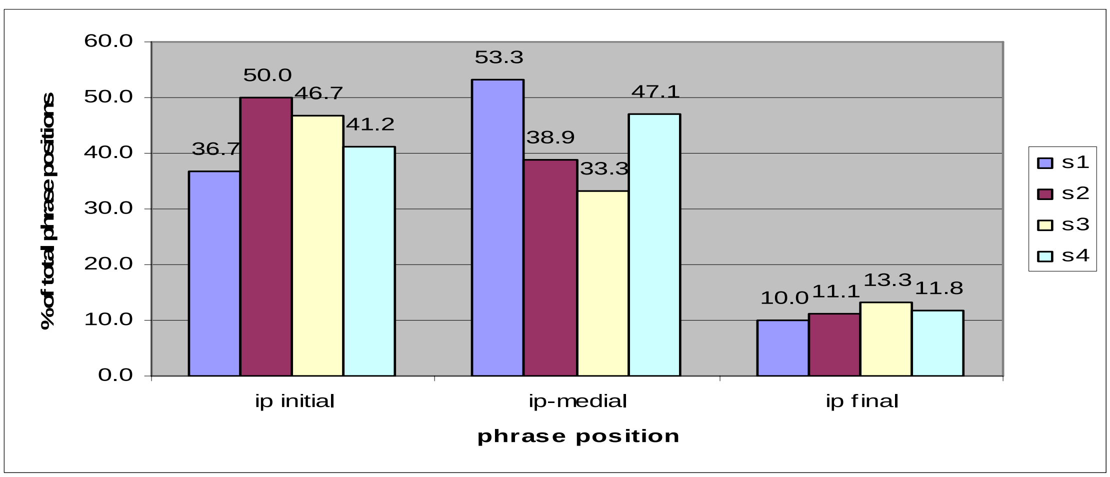


Figure 33: Cut location within the intermediate phrase: individual stutterers

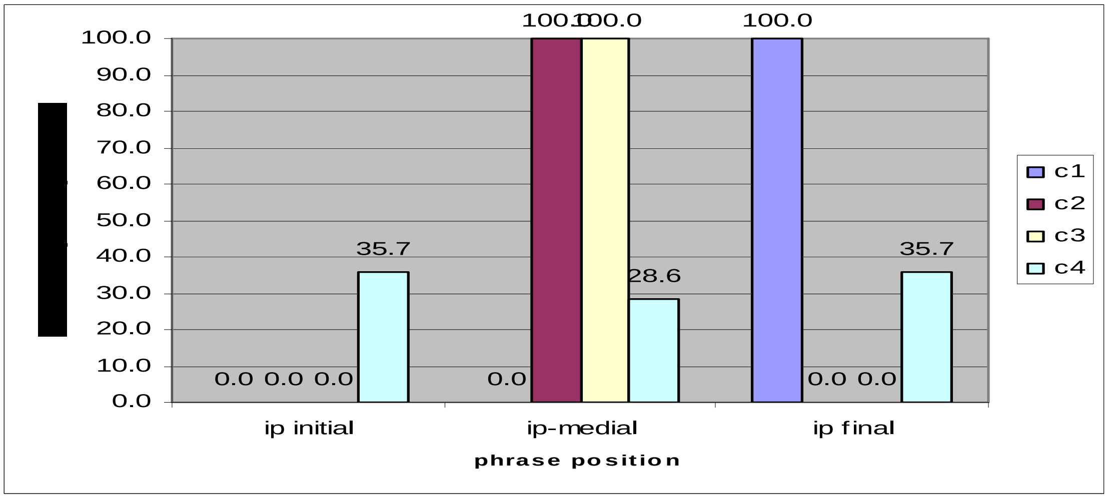


Figure 34: Cut location within the intermediate phrase: individual control subjects

Stutterers overall produced filler disfluencies most often in ip-final position, differing considerably from production patterns of prolongations and cuts. On the other hand, control subjects generated disfluencies more often in ip-initial and ip-final positions than in ip-medial position. Figure 35 shows the distribution of fillers for both groups:

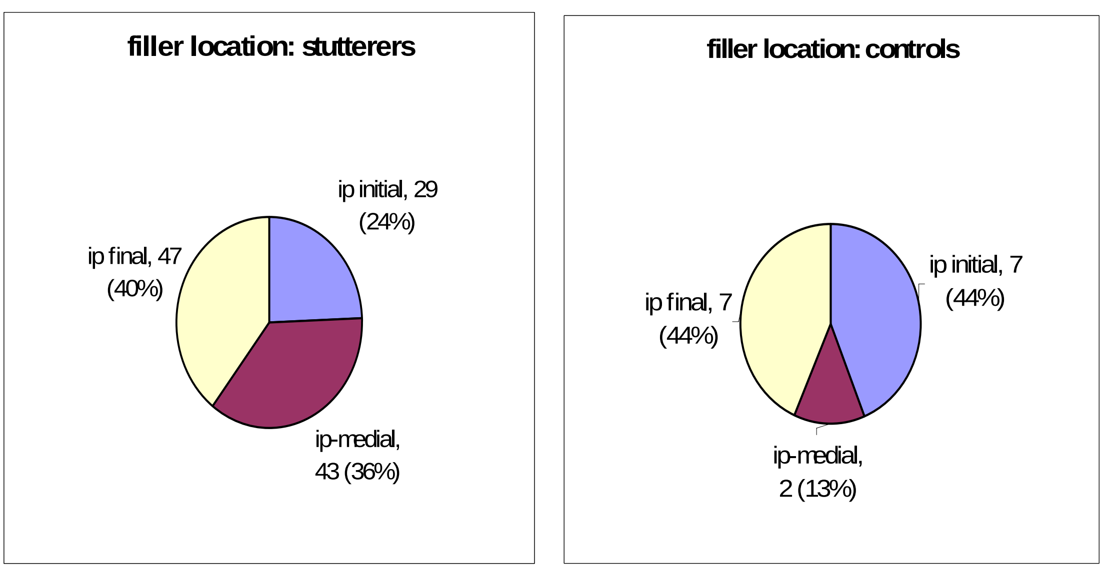


Figure 35: Location of fillers within the intermediate phrase: stutterers vs. controls

Once again, as illustrated in Figure 36 and Figure 37, pairwise comparisons between subjects revealed no significant differences in location of fillers, though overall numbers were low for both groups. Although control subject C3 produced all of his filler disfluencies in ip-initial position, he only produced two tokens total.

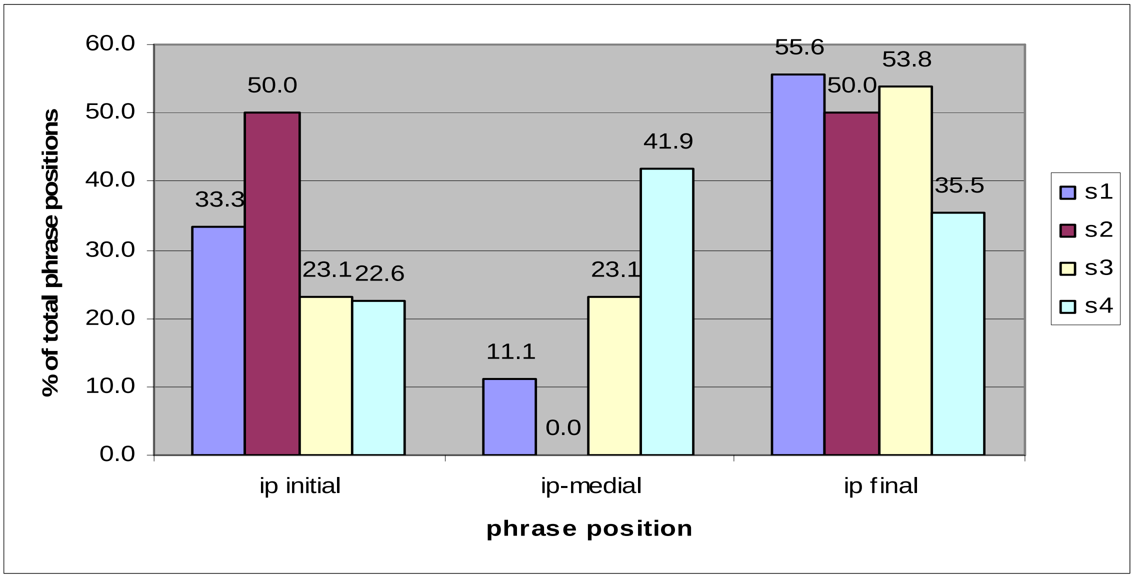


Figure 36: Filler location within the intermediate phrase: individual stutterers

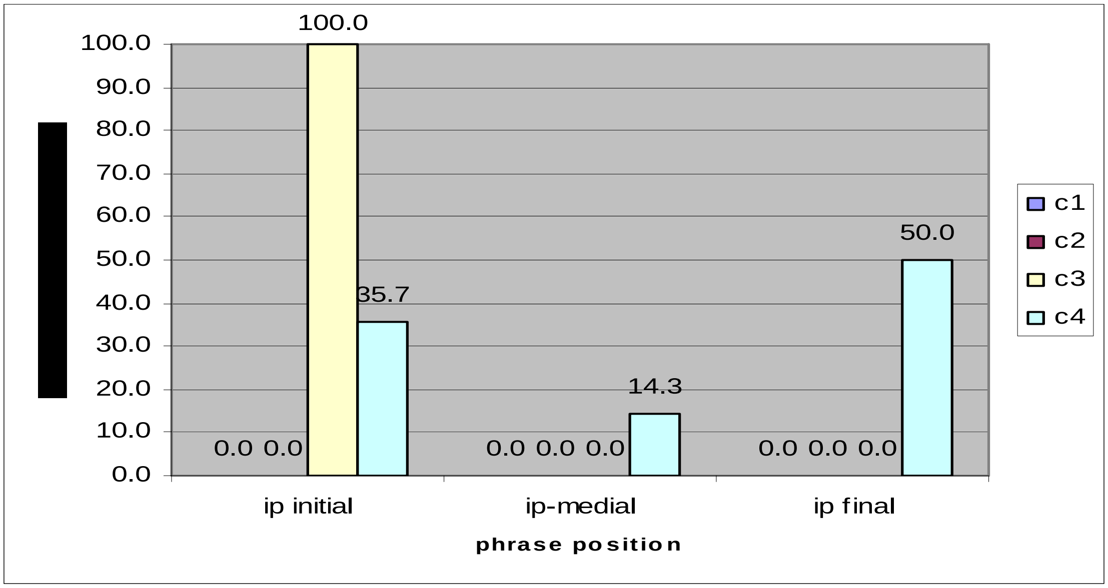


Figure 37: Filler location within the intermediate phrase: individual control subjects

For restarts, however, ip-final position was again the least common disfluency location, representing only 3% of all restarts for both groups. As Figure 38 shows, controls generated restarts in ip-initial (50%) and ip-medial (47%) positions at nearly similar rates, while for stutterers restarts occurred much more often ip-medially (71%).

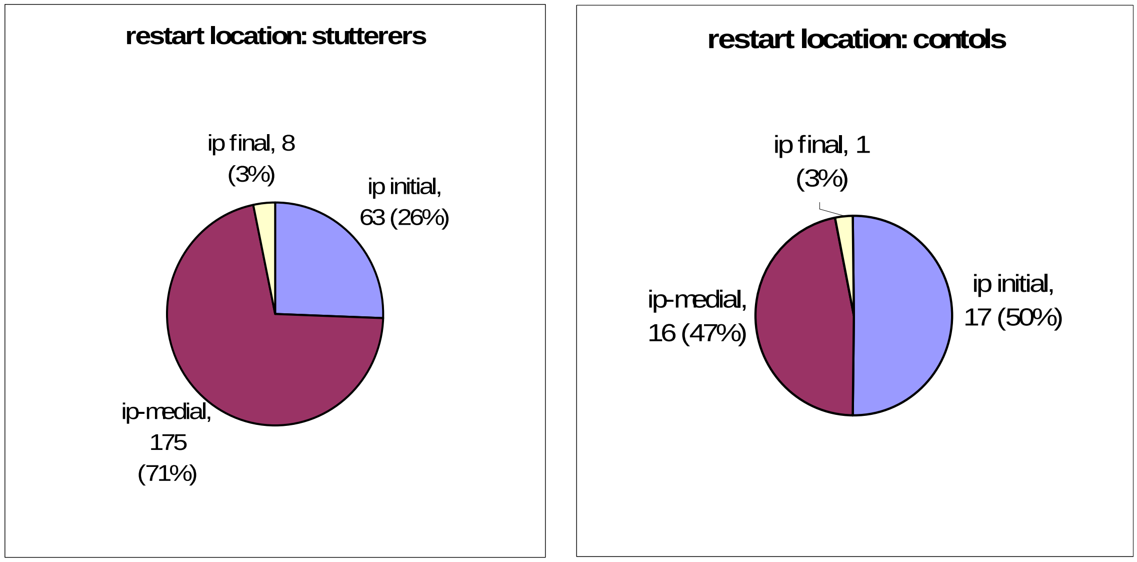


Figure 38: Location of restarts within the intermediate phrase: stutterers vs. controls

All subjects from the stuttering group behaved uniformly with the exception of S4 who produced significantly fewer ip-initial restarts than S1 (p=0.0003), S2 (p=0.0023), and S3 (p=0.0496), and significantly more ip-medial restarts than S1 (p=0.0013).

Two of the four control subjects, meanwhile, did not produce filler disfluencies in their narrative data; of the two who did, C4 behaved similarly to stuttering subjects in producing fewest disfluencies ip-finally (0%). Unlike stutterers, C4 produced slightly more ip-initial than ip-medial disfluencies. Lastly, C3 demonstrated greater variation, but produced only 4 disfluent filler tokens overall.

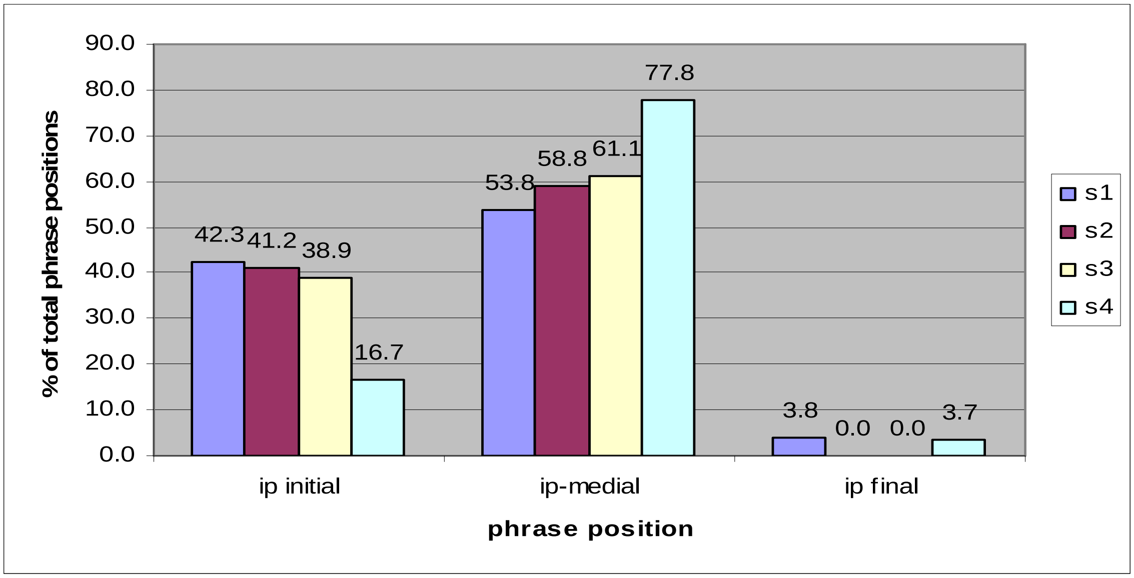


Figure 39: Restart location within the intermediate phrase: individual stutterers

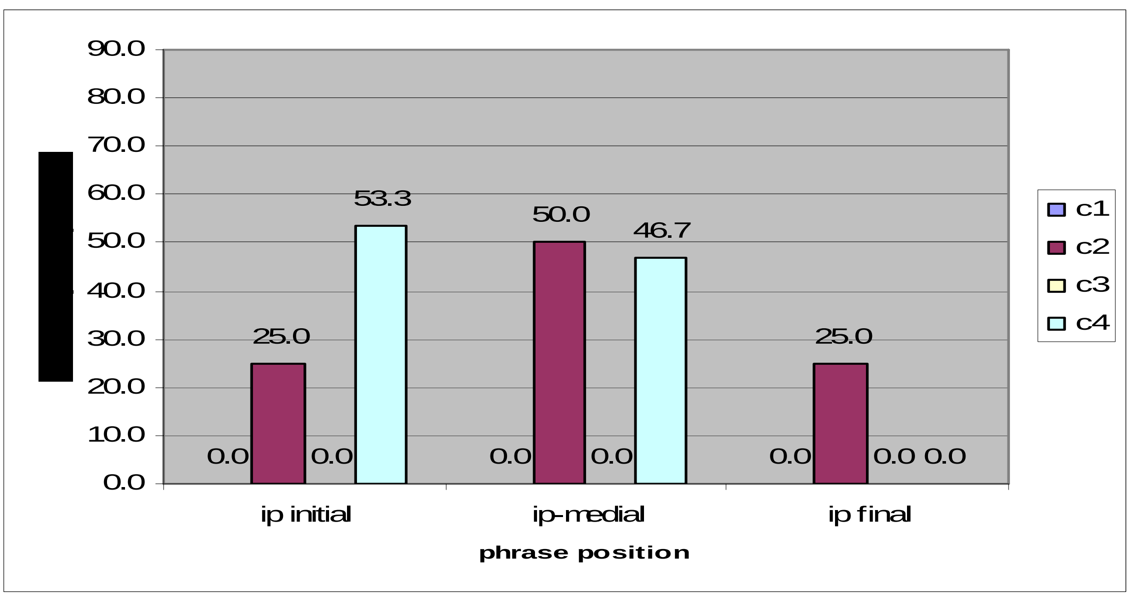
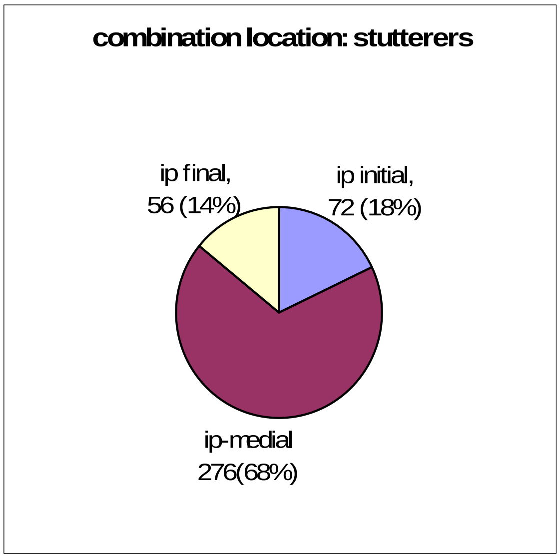


Figure 40: Restart location within the intermediate phrase: individual control subjects

Finally, for both stutterers and controls, approximately two-thirds of combinations surfaced ip-medially. As shown in Figure 41, while stutterers produced slightly more combinations ip-initially (18%) than finally (14%), controls had the opposite result:[12]



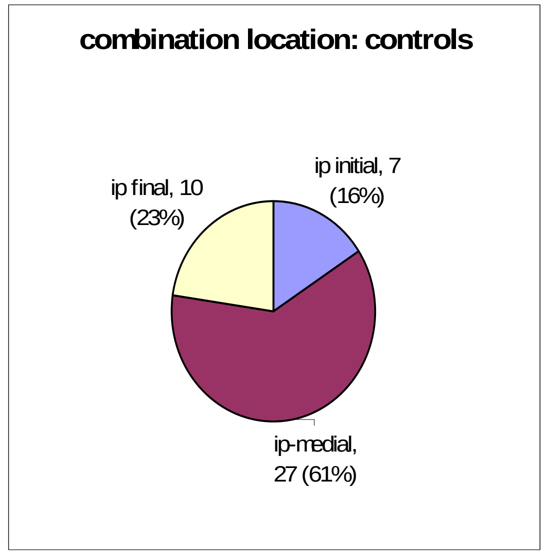


Figure 41: Location of combinations within the intermediate phrase: stutterers vs. controls

4.2.3 Disfluencies and tonal type

Tonal type was recorded for both disfluent words and the material immediately preceding disfluencies. This was done in order to uncover any possible tonal patterns associated with a certain type of disfluency. For pauses, the last tone preceding the pause itself was classified as the pre-disfluency tone, while the tone on the word immediately following the pause was classified as the tone of the disfluency itself. For all other single

[12] Since combination disfluencies were not separated into pause and non-pause types, individual comparisons were not made for this disfluency type.

disfluencies—i.e., restarts, prolongations, cuts, and fillers—the last tone occurring before the disfluent material comprised the pre-disfluency tone, while the tone falling directly on the disfluent portion was labeled the disfluency’s tone.

Figure 42 below shows an example of both a pause and non-pause disfluency:

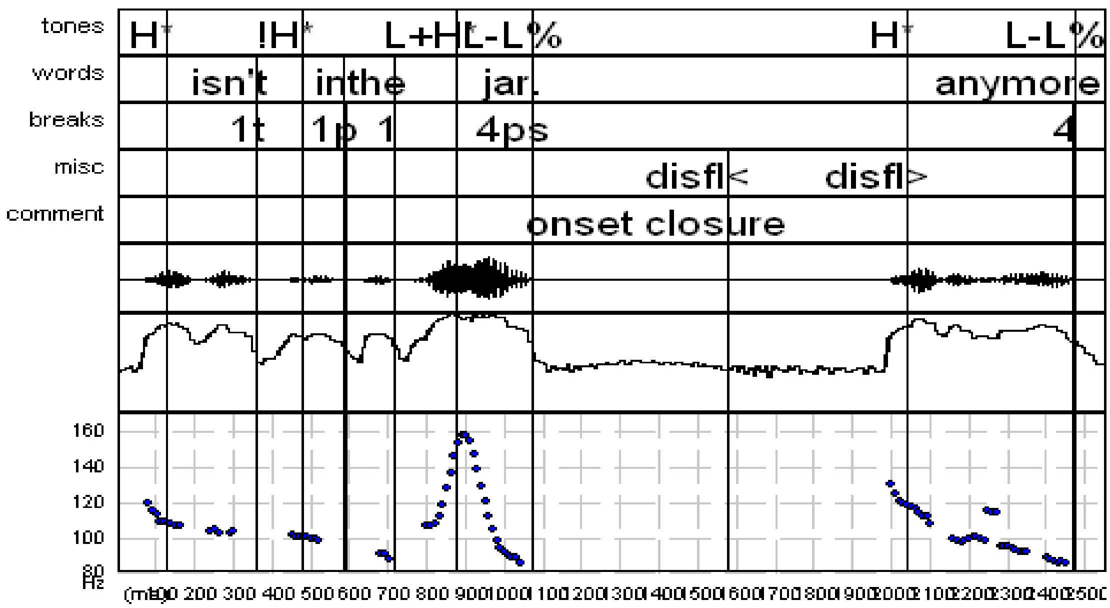


Figure 42: Pre-disfluency and disfluency tonal coding example

In Figure 42, for the prolongation disfluency ‘(1p’), the pre-disfluency tone is the H* occurring on “isn’t”, while the disfluency tone is the !H* on “in” (i.e., the prolonged word). If there had surfaced intervening non-pitch-accented words between the pre-disfluency H* and the disfluent !H*, the H* would nevertheless be counted as the pre-disfluency tone, since it would still represent the final *tone* before the disfluency.

For the pause (‘4ps’), the pre-disfluency tone is L-L%, since it immediately precedes the pause, and the disfluency tone is on the first word after the pause—i.e., the H* on “anymore”. However, unlike pre-disfluency tones for non-pauses—where the tone preceding the disfluency regardless of intervening words is what matters—the disfluency tone for pauses was only counted for the first post-pausal *word*. In other words, only the material immediately following a pause was considered part of the disfluency, and therefore if no tone occurred on this word then the disfluency tone was coded as absent (i.e., toneless).

Combinations including both a pause and non-pause were simply coded as both types, thus combining the guidelines presented above. For example, the contextual environment of a combination prolongation + pause (‘p.ps’) was coded to include the tone immediately preceding the earliest part of the disfluency—i.e., the prolongation—as the pre-disfluency tone, and the last part of the disfluency—i.e., the word immediately following the pause—as the disfluency tone.

Figure 43 shows that when the pre-disfluency tone was a pitch accent, the largest group differences with respect to tonal type were in the production of L+H* ($p<0.0001$) and H* ($p=0.051$). While 39% of control subjects’ disfluencies occurred before L+H*, only 18.2% of stutterers’ disfluencies preceded L+H*. The opposite tendency occurred for H*, which followed 43.5% of stutterers’ disfluencies and just 33% of control subjects’ disfluencies. Before disfluencies, stutterers produced H* more often than other pitch accents, followed by !H* (22.6%) and then L+H* (18.2%). Control subjects, meanwhile, produced L+H* most often before disfluencies, and H* slightly less often (33%).

However, in order to determine if these patterns were unique to disfluencies, pre-disfluency tonal types were compared with tonal types for the whole data set for each group (labeled 'whole' in Figure 43). As shown in the four-way comparison of Figure 43, pre-disfluency pitch accents of the control subjects did not behave differently from their pitch accents overall for any of the categories—a result that was statistically verified, using Fisher's exact test. Stutterers, on the other hand, produced statistically more H* (p=0.0001, 2-tail) and fewer L+H* (p<0.0001, 2-tail) pitch accents in pre-disfluency position than they did in the data as a whole. Differences reached significance also for L* (p=0.047), H+!H* (p=0.0001), and L+!H* (p=0.0002), though these pitch accents were produced with a much lower frequency for all speakers.

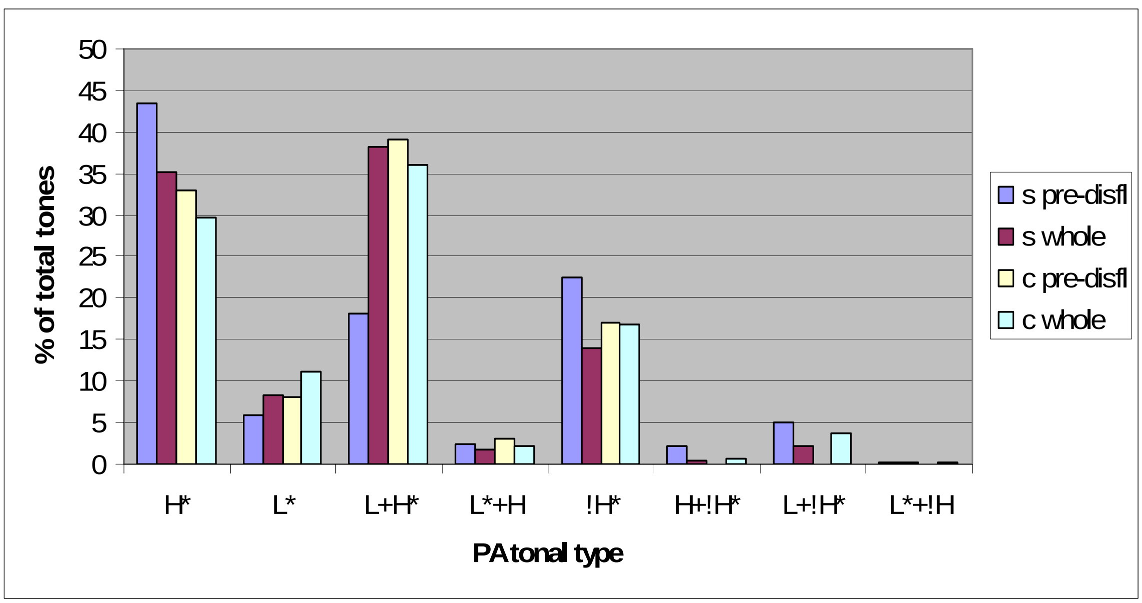


Figure 43: pitch accent tone distribution, all subjects—total and pre-disfluency

Pre-disfluency ip boundary tones (phrase accents) also did not differ appreciably between the two groups, as reflected in Figure 44:

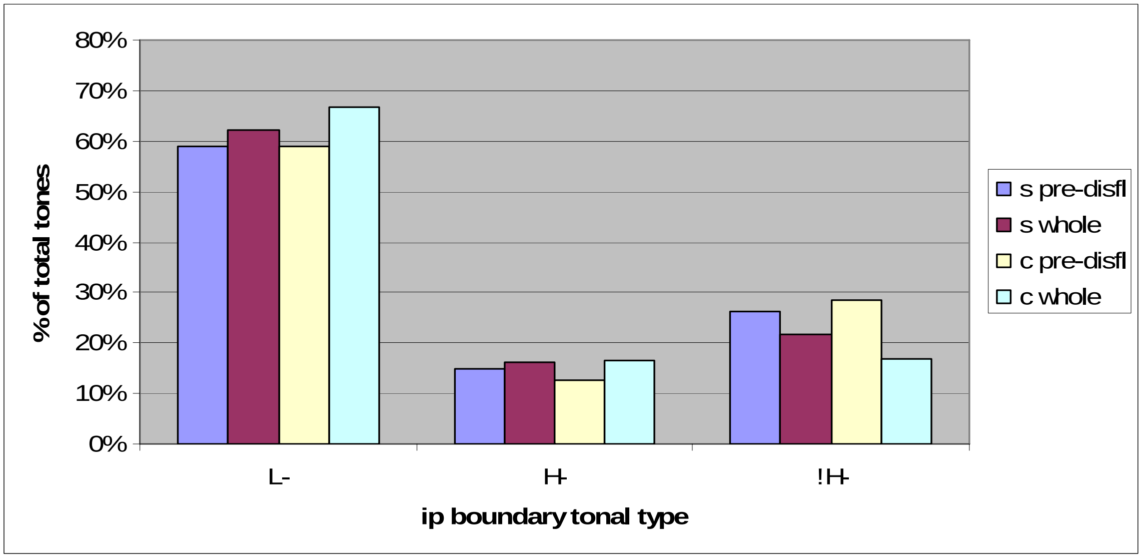


Figure 44: ip boundary tone distribution, all subjects—total and pre-disfluency

As with pitch accents, pre-disfluency phrase accents were compared with overall phrase accent distribution for both groups, and statistical differences were found for stutterers' production of !H-, and controls' production of L- and !H-. In the former case, stutterers produced statistically more !H- phrase accents ($p=0.012$) in pre-disfluency position than compared with their !H- production overall, while controls also produced statistically more !H- accents ($p=0.029$) but fewer L- accents ($p=0.046$) in pre-disfluency position.

Finally, as illustrated in Figure 45, comparisons were made between stutterers and controls with respect to the type of IP boundary tone contours that occurred before disfluencies. Both groups produced statistically different patterns in pre-disfluency environments as compared with their normal boundary tone distributions. While stutterers produced statistically fewer H-H% (p=0.009) contours in pre-disfluency position compared with their overall distribution, they produced more !H-H% (p=0.011) contours pre-disfluently. Controls, meanwhile, produced more !H-H% (p=0.036) and !H-L% (p=0.023) contours in pre-disfluent position, but fewer L-L% (p=0.004) contours:

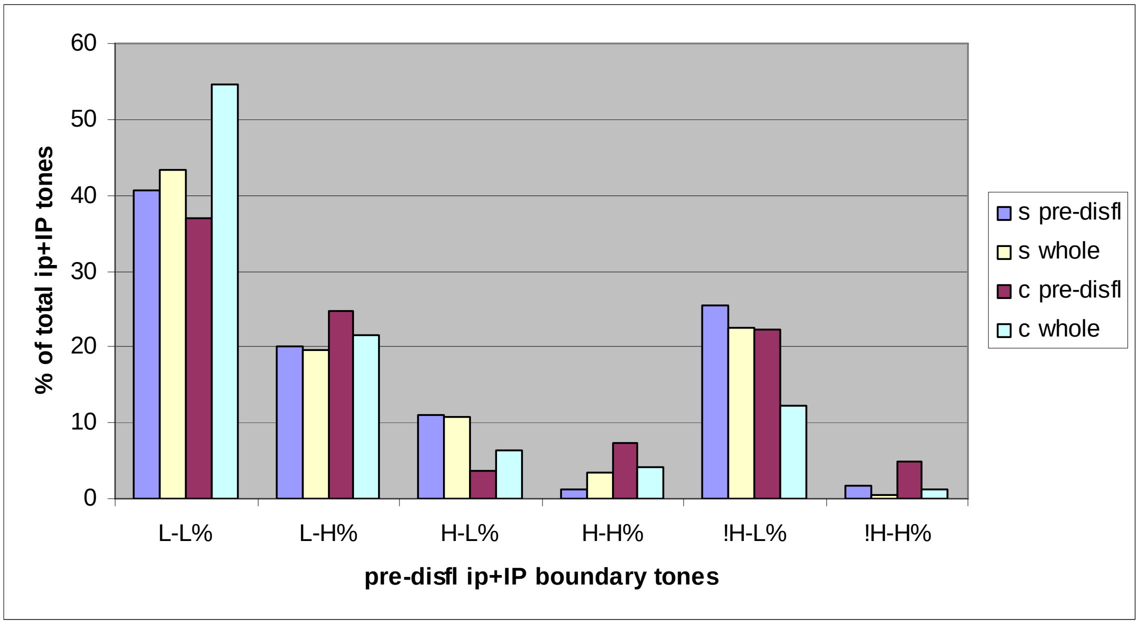


Figure 45: phrase accent and boundary tone contours: pre-disfl vs. whole—both groups

In order to test whether the differences above corresponded with broader tonal contour patterns, group comparisons were made separating boundary contour types into four tonal classes: low (L-L%), mid (!H-L%), high (H-L%), and rising (L-H%, H-H%, !H-H%):

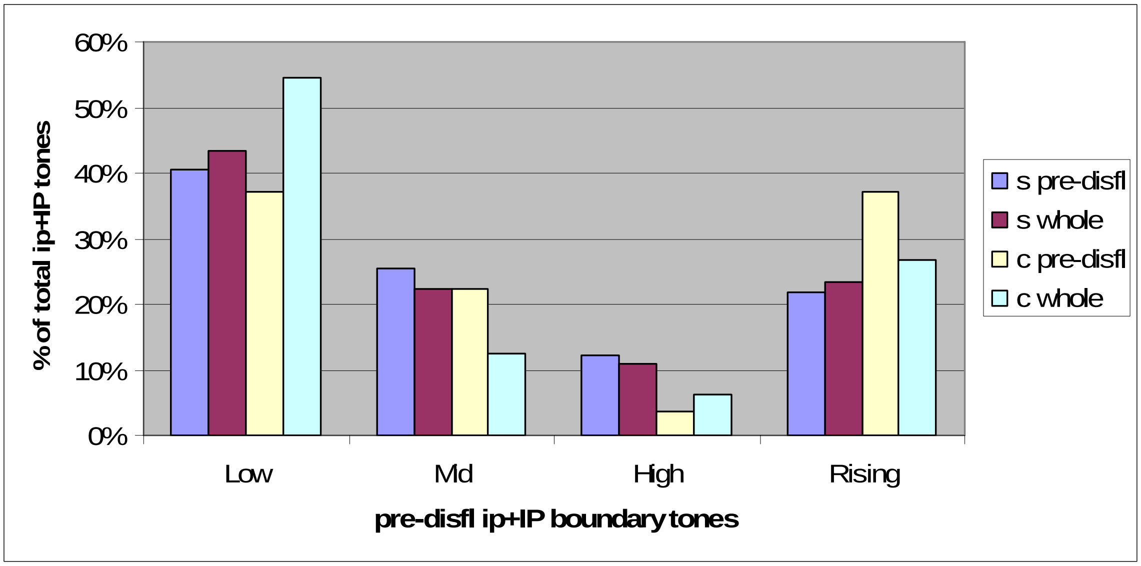


Figure 46: comparing rising and falling contours: stutterers vs. controls

As Figure 46 shows, when tonal contours are combined into these more general categories, they further underscore the strong tendency for control subjects to produce more—though not significantly—rising tonal contours in pre-disfluent position (37.0%) than in their data as a whole (26.7%). Control subjects did generate significantly more ($p=0.023$) mid-tone contours and fewer ($p=0.004$) low contours in pre-disfluent positions than they did overall. In contrast, statistically significant differences which surfaced in the more specific contour comparisons for stutterers (i.e., fewer !H-H% and H-H% contours) disappeared in the more general comparisons. Nevertheless, the data in Figure 46 clearly show stutterers' opposite behavior from that of control subjects with respect to the production of rising contours in pre-disfluent position.

Tones realized on or at disfluencies were similarly recorded for all subjects, following the guidelines outlined at the beginning of this section. Again, overall pitch accent tonal distribution was compared with the distribution of all tones occurring on disfluencies for each group, as shown below in Figure 47:

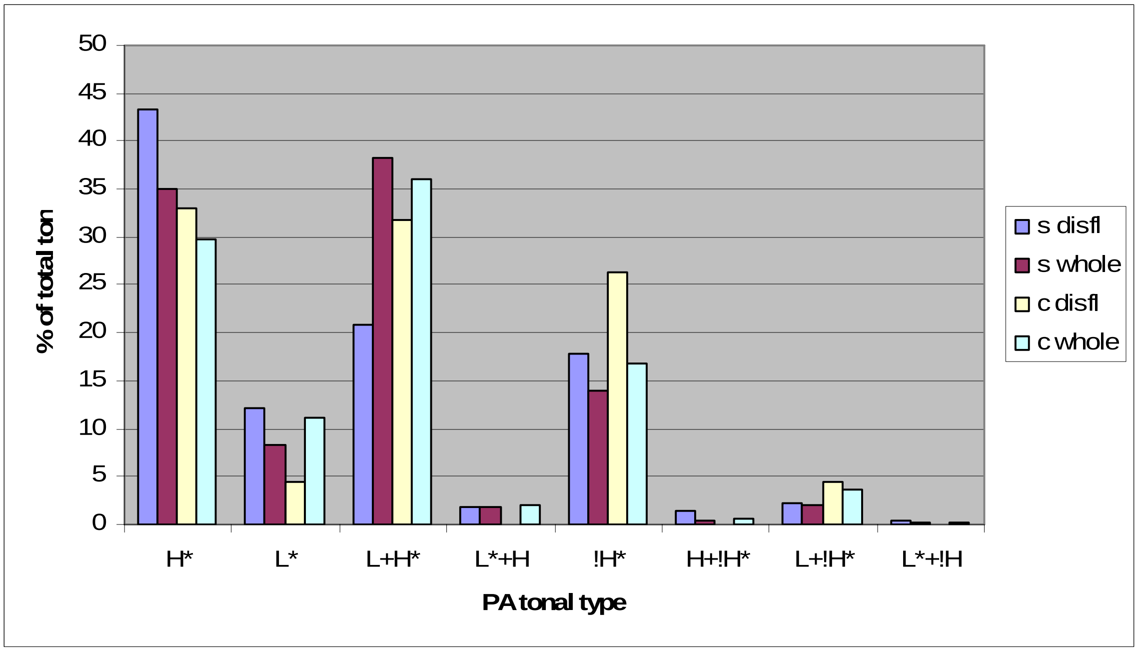


Figure 47: pitch accent tone distribution, all subjects—disfluency and whole data

Pairwise comparisons were conducted for each tonal type, revealing significant differences between total and disfluent pitch accent distribution for stutterers' productions of H*, L*, L+H*, !H*, and H+!H*. H* (p<0.0001), L* (p=0.0005), and H+!H* (p=0.003) all surfaced more frequently on disfluent forms, while L+H* (p<0.0001) occurred less often on disfluencies. Control subjects showed fewer significant differences, although two tonal types—L* and !H*—were produced disfluently at significantly different rates: L* less often (p=0.0504), and !H* more often (p=0.03) on disfluencies.

Stuttered disfluency rates on tonal patterns were then compared with control disfluency rates, and two significantly different tonal types were found: stutterers produced L* significantly more and !H* significantly less frequently on disfluencies than did controls ($p=0.025$ and 0.024, respectively).

As illustrated by Figure 48, both speaker groups revealed trends in the same direction for all comparisons. For phrase accent comparisons, stutterers produced L- phrase accents significantly less often ($p=0.0008$) and !H- accents more often ($p<0.0001$) on disfluencies than in the whole data. Control subjects also produced !H- accents significantly more often ($p<0.0001$) on disfluencies, while generating H- accents significantly less often ($p=0.045$) on disfluencies than in their whole distribution:

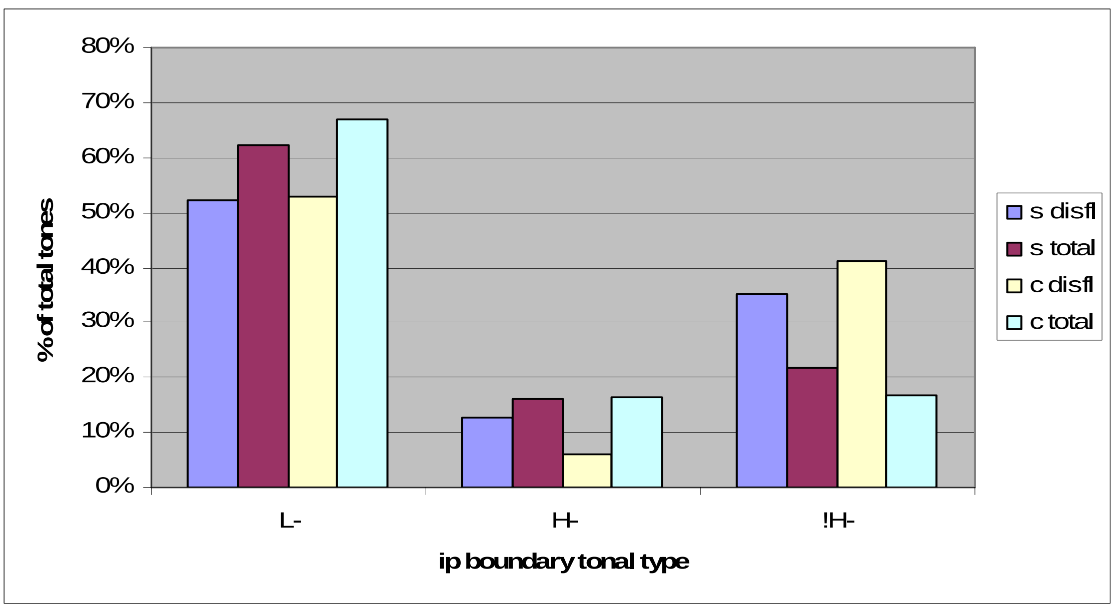


Figure 48: ip boundary tone distribution, all subjects—total and disfluency

Finally, as in the pre-disfluency tonal analysis, comparisons were made for both groups with respect to boundary tone contour types:

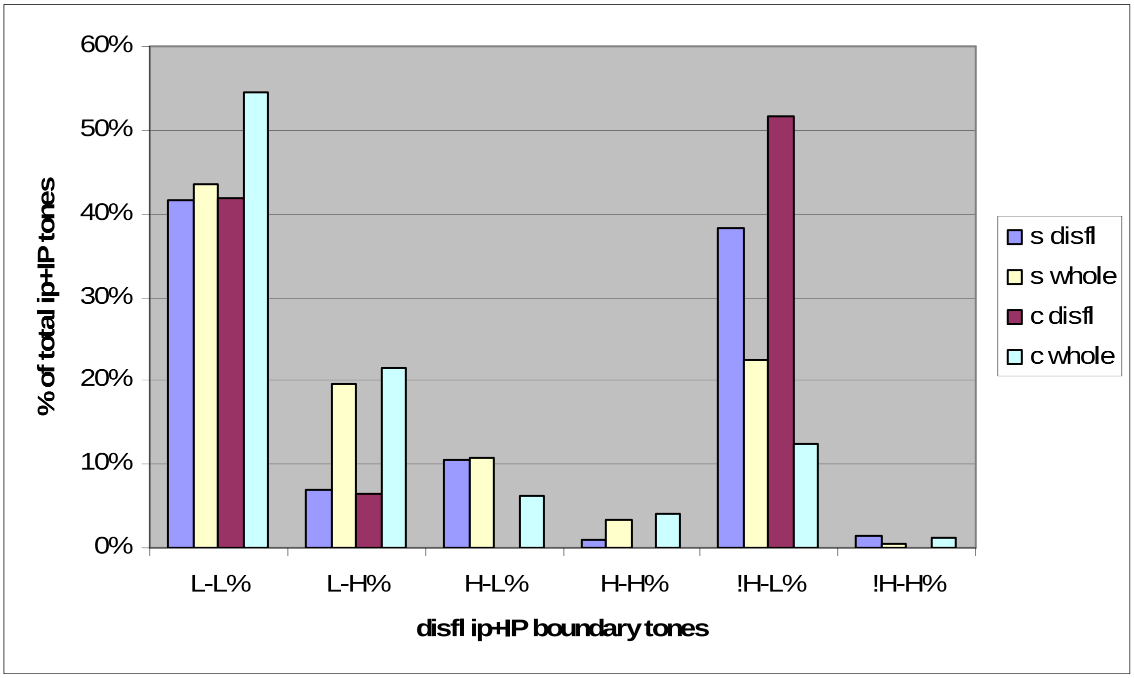


Figure 49: phrase accent and boundary tone contours: stutter disfl vs. control disfl

As seen above in Figure 49, both stutterers and controls produced significantly fewer L-H% contours on disfluencies than compared to their whole distribution (Stutterers: $p<0.0001$; Controls: $p=0.018$). Stutterers generated only 7% of disfluent IP boundaries as L-H%, compared with 19.5% in the data as a whole. Control subjects produced a slightly smaller rate of L-H% contours on disfluencies (6.5%), though their disfluent data again were fewer in number. In addition, both groups produced significantly more !H-L% contours on disfluencies than overall ($p<0.0001$), while for control subjects significantly fewer L-L% contours also occurred with disfluencies than they did in their whole distribution (p=0.037).

Because control subjects overall produced such a low number of disfluencies at IP (Intonation Phrase) boundaries, group comparisons were not made for general tonal contour types (i.e., low, mid, high, and rising contours).

4.2.4 Disfluencies and word class

The final disfluency comparisons made for the narrative data were between disfluency and word class—specifically, whether the disfluency occurred more often on a content or function word. Pairwise comparisons were made for stutterers and their age-matched controls, as seen in the raw-data figure below:

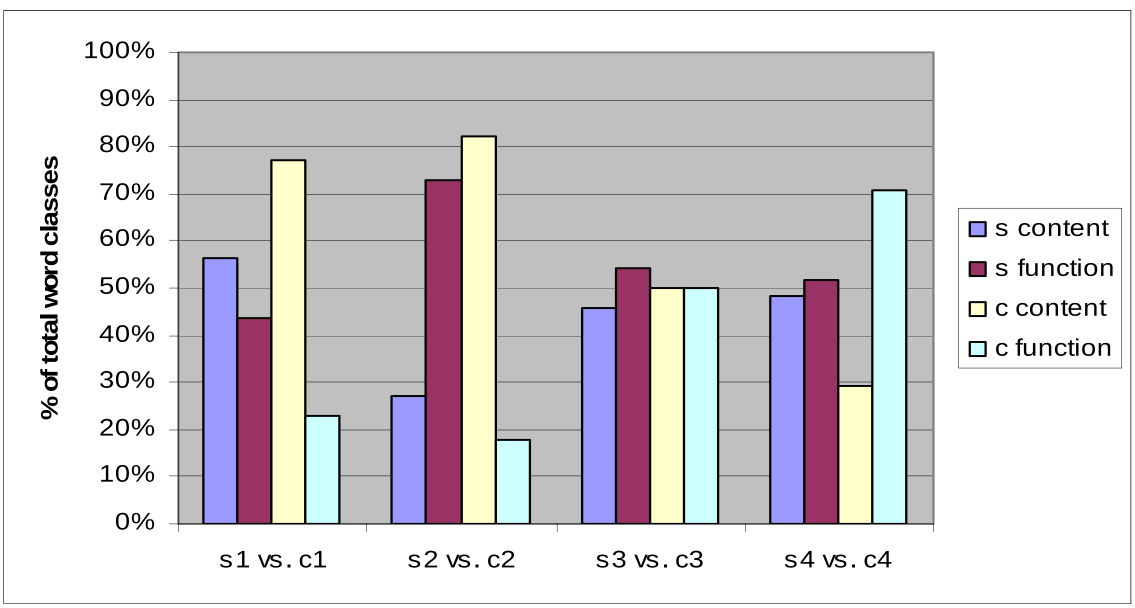


Figure 50: disfluency and word class: stutterers vs. controls (individual)

Direction of difference varied on the whole, as neither group behaved consistently. One subject pair differed significantly: S2 produced 73.1% of his disfluencies on function words, compared with only 17.6% for C2's disfluencies ($p<0.0001$). The anomalous behavior of C4, who produced disfluencies on content words only 29.1% of the time (compared with 77.3%, 82.4%, and 50% for subjects C1-C3, respectively), skewed controls' overall rates of disfluencies with content words.

While correlation variation was high among subjects, a general pattern emerged when subjects were collapsed in order to compare groups as a whole. Overall, both stutterers and controls produced more disfluencies on function words, although stutterers produced a higher percentage of disfluencies on function words than did controls. The difference between groups was nevertheless not significant ($p=0.646$).

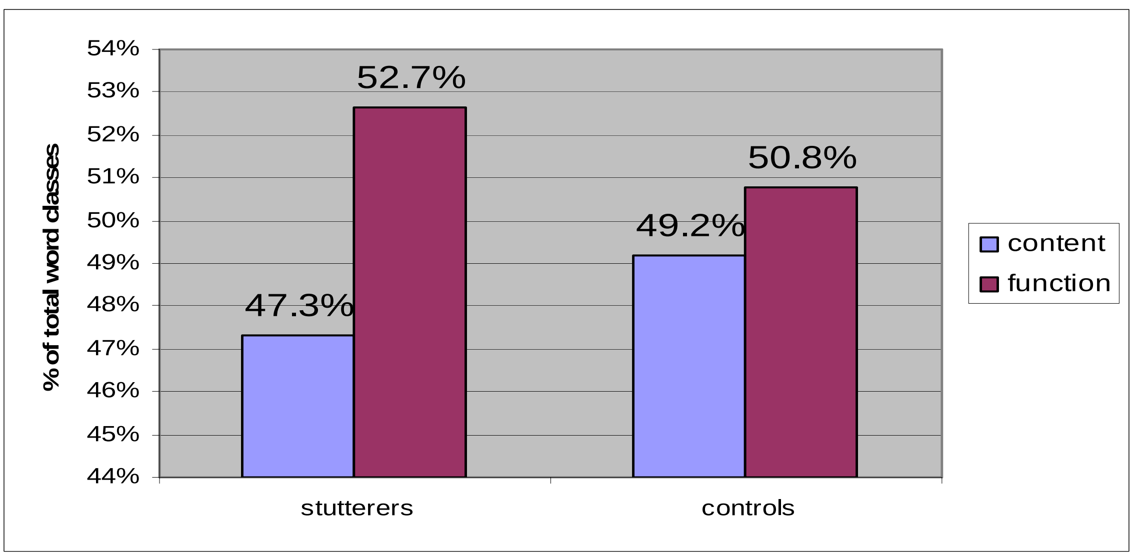


Figure 51: disfluency and word class: stutterers vs. controls (group)

4.3 Summary of spontaneous narration data results

A number of significant patterns appeared in the analysis of the narrative data. Regarding basic descriptive comparisons, disfluency type categories occurred in similar distributions for both stutter and control groups; stutterers, however produced significantly more prolongations and combinations than did control subjects, while significantly fewer pauses. Regarding phrase length, stutterers produced shorter ips than did controls both with respect to average number of words and pitch accents per ip.

Prominence patterns of disfluencies differed between the two groups, as stutterers produced pitch-accented disfluencies (PA + NPA) more often than unaccented forms (64.5% vs. 35.5%), while controls produced fewer PA'd (PA + NPA) disfluencies than unaccented ones (43% vs. 57%). Comparisons of disfluency rates for each prominence level showed no differences for controls as a group, as disfluency rates ranged from 7.1% to 8.9% for the three prominence types, while stutterers generated considerably higher disfluency rates on PA'd words (66.5%) and NPA'd words (31.1%) than on unaccented words (22.9%).

Position of disfluencies within an intermediate phrase also differed, with stutterers generating the majority of their pause disfluencies (61%) at intermediate phrase boundaries and 39% ip-medially, while controls yielded the opposite result of more ip-medial pauses (70%) than ip-boundary pauses (30%). Both stutterers and controls subjects produced prolongations and cuts more often ip-medially, and fillers most frequently in ip-final position. Restarts, on the other hand, were produced least frequently by both groups, and while stutterers generated more restarts ip-medially (71%) than initially (26%), controls reflected the opposite order with slightly more ip-initial (50%) than medial (47%) restarts. Finally, about two-thirds of all combinations were attested in ip-initial position for both groups.

The types of tones surfacing before a disfluency did not differ very reliably among controls, but stutterers produced significantly fewer L+H* and more H* immediately preceding disfluencies than in other locations. The difference between stutterers' and control subjects' pre-disfluency L+H* productions also reached significance. Both groups produced !H- phrase accents significantly more often than other accent types when the following word was a post-boundary disfluency. Meanwhile, with respect to boundary tones, while controls produced IP-final rises 32.1% of the time, stutterers produced final rises for only 21.2% of their boundary tones.

Tone types generated on or at the disfluency[13] again showed a statistically significant difference for the stuttering group, but not the control group: stutterers produced significantly fewer L+H* pitch accents on disfluencies than they did in other locations in general. Once again !H- was the most common phrase accent tone produced by both groups when disfluencies were present, and both groups also generated a higher percentage of mid-tone contours (!H-L%) during a disfluent production.

Looking at the target data from script analyses is crucial for testing the robustness and relevance of these patterns, since essentially all of the conditions—with the exception of word position—rely on determining the intended (targeted) structures if their interactions

[13] Again, for non-pauses this refers to the pitch accent of the word on which the disfluency occurs, while for pauses this refers to the pitch accent of the word immediately following the pause.

with disfluencies are to be fully understood. Evidence of a planned nuclear pitch accent, for example, may become opaque if a preceding disfluency resulted in a derivative ip break—thus rendering the pitch accent phrase-initial instead of phrase-final.

CHAPTER 5

RESULTS 2: SCRIPT DATA

This chapter continues the analysis of the picture book narrations, focusing on the comparison between stutterers' data and verbatim readings of the stutterers' data by their age-matched controls. While the comparisons of the spontaneous narrations of both stuttering and control groups revealed a number of significant patterns and trends, the fact that many factors of spontaneous output cannot be carefully controlled inevitably limits the generalizability of the comparisons. For instance, the tendency for stutterers to produce shorter ips than did controls is confounded with the fact that stutterers also generated higher overall disfluency rates, since, as the analyses of Chapter 4 showed, disfluencies often caused ip breaks. With respect to the prosodic conditions of disfluencies and their environments, also analyzed in Chapter 4, one crucial effect is that of the disfluency on *preceding* material in the phrase—in particular, the possibility of triggering further disfluencies. A clear example of this is illustrated in Figure 52 (re-printed from Chapter 3, Figure 16). A speaker produces a combination disfluency ('4p.ps') in what appears to be a lexical search error, since there is no trace of phonetic content after the '4p.ps', and since the speaker restarts the same syntactic phrase before producing the previously missing word "cliff":

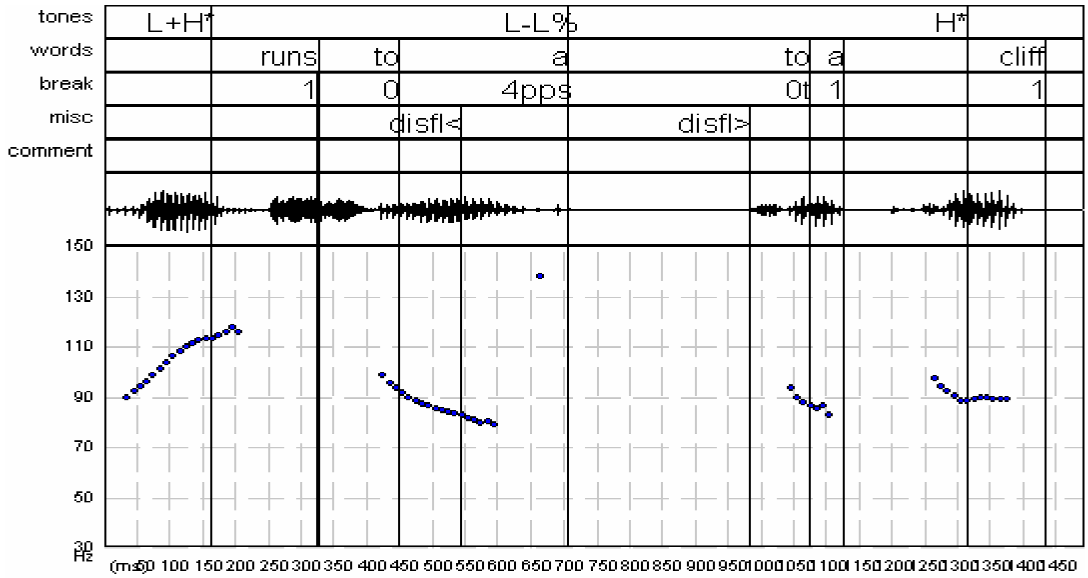


Figure 52: Pitch track illustrating a combination (4pps) and restart (1t): "runs to a—to a cliff"

What is obvious from the diagram is that, crucially, the speaker did not exhibit a production difficulty with the locus of the 'p.ps' disfluency itself. Rather, the underlying problem can be traced to an element later in the phrase—presumably, to the production of "cliff"[14]. By examining the prosodic conditions of the disfluency and its environment, it becomes apparent that this disparity between a disfluency's underlying trigger and its (often) earlier manifestation can result in a misleading interpretation which fails to capture this important distinction. For example, assuming that the intended word was "cliff", it is visible from the second and successful iteration of the prepositional phrase "to a cliff" that there is neither an ip nor IP break between "a" and "cliff". Coding this data literally, however, would indicate that the disfluency occurred at an ip boundary,

[14] Of course, it cannot with certainty be determined what the exact nature of the breakdown was, including whether the word "cliff" was the problematic element, since the speaker may have been, for example, re-constructing syntactic structure later in the utterance. What is uncontroversial, however, is that the word "a" was not the actual cause of the 'p.ps' disfluency, since the word has been successfully produced.

when, in fact, if the origin of the disfluency is in the production of ip-medial “cliff” (since “cliff” is preceded and followed by break index ‘1’), then it is ultimately the case that the underlying disfluency is ip-medial. Simply put, if the speaker’s intended utterance was “runs to a cliff”, and a production breakdown occurred in ip-medial position (i.e., on “cliff”), the realization of the breakdown not only surfaces prior to the breakdown (i.e., on “a”), but the earlier surfacing disfluency results in an unintended ip break at the point of disfluency. Thus, the fact that the breakdown was actually ip-internal is lost in the prosodic effect of the surfacing disfluency.

A similar masking effect can be observed in the analysis of the prominence relationships in Figure 52. Again, if the origin of the disfluency is in production of the word “cliff”, and the intended phrase has no ip boundary between “a” and “cliff”, then there should be no nuclear pitch accent (i.e., final pitch accent before an ip boundary) in this portion of the utterance. According to the actual data, however, a nuclear pitch accent does indeed appear on “runs”, as a result of the IP break after the disfluency realized on “a”. That is, by virtue of the fact that an unplanned IP boundary surfaces after the ‘4p.ps’ disfluency, the pitch accent realized on “runs” is rendered the NPA—the final pitch accent in the IP.

It is at this point that the categorical distinction between ‘target’ and ‘anticipatory’ disfluencies, as outlined in Chapter 2, becomes important to the analysis. In order to examine the underlying effects of a production breakdown on the prosodic conditions analyzed in Chapter 4, it is first necessary to determine which disfluencies are triggered by detection of the problematic material (i.e., ‘anticipatory’), and which disfluencies represent the actual breakdown. As was illustrated by the example in Figure 52, once this distinction is made, it becomes possible to ascertain the relationship between the underlying breakdown and essential components of the prosodic structure such as prominence and phrasal position.

Thus, as briefly outlined in the Method section, a second task was designed in which control subjects read scripts of stutterers' productions. The purpose of this was to provide a reference prosody for each stutterer. In addition to holding constant the lexical and segmental content of the controls' output, the scripts functioned as reference markers of both prominence relations and phrase boundaries. This provided a method for deducing the stutterers' prosodic targets, thereby determining more accurately the possible triggers of the disfluencies.

Only scripts of three of the four stuttering subjects from the narration task—speakers S1, S2, and S4—were analyzed for the target comparisons. Speaker S3 was not selected because he and S2 belonged to the same age group. Thus, three balanced age brackets were maintained for the target comparisons: 30+ years, 50+ years, and 70+ years.

5.1 Coding script data

Script data were coded similarly to natural data, with a few important exceptions. While combination disfluencies in natural data were broken into constituents if the latter formed separate *events*, all script data disfluencies were counted as single disfluencies. For example, a prolongation + pause ('pps') consists of two events over time: a prolonged nucleus of a word, followed by a pause after the prolongation. Since in this case the target is the same for both prolongation and pause, script coding would classify this as a single disfluency, or *attempt*. Similarly, in the natural data, single restarts occurring immediately following other disfluencies were still counted descriptively as separate disfluencies. However, in the script data they were disregarded because they marked the onset of fluent speech; in other words, they constituted the resolution of the disfluency rather than the trigger.

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 crucially separated from the lexical word (or further satellite material) by a '0' break index boundary.

An example demonstrating both definitions is shown in Figure 53 below, where the speaker is attempting to utter the phrase “goes to sleep”:

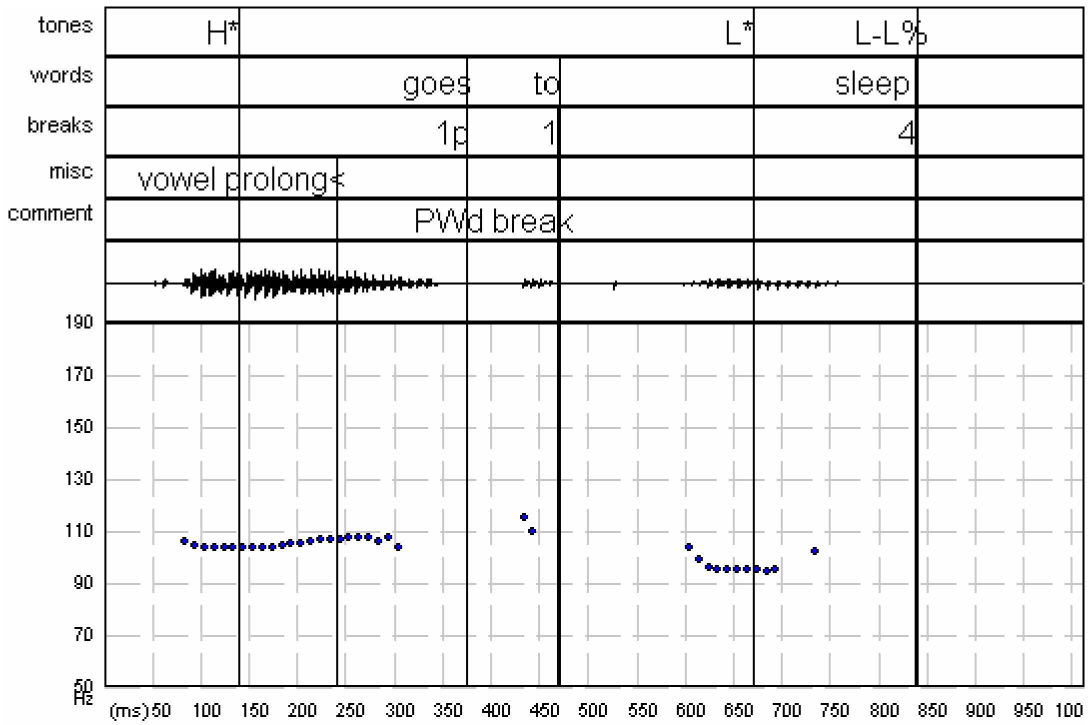


Figure 53: Post-disfluency target

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. Determining the precise size of all prosodic constituents and their breaks is possible by referring to the prosody produced by a control speaker. In this way it becomes possible to hypothesize the word and phrase boundaries intended by the stuttering speaker. In some cases a disfluency will obviously fundamentally alter an intonation structure if a speaker is forced to re-start a word or phrase, as was seen in Figure 52. In Figure 53 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". This argument alone would simply be conjecture, but the reference prosody provided by the control speaker gives evidence for this continued delay:

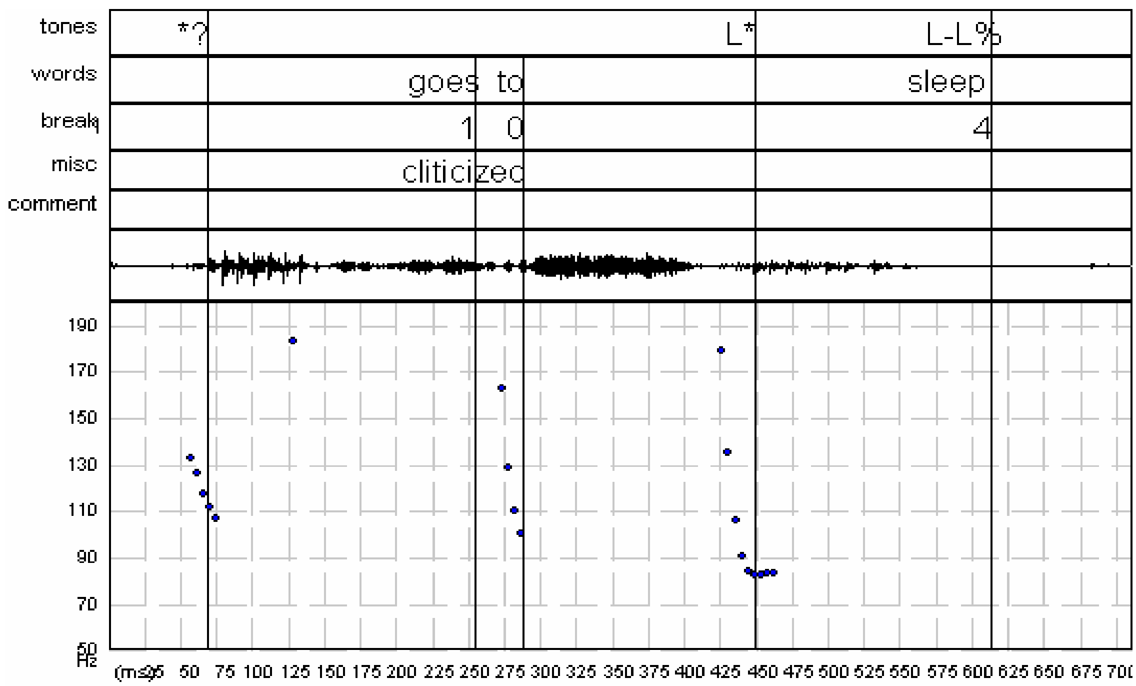


Figure 54: Prosodic Word example

In the example shown in Figure 54, the reference prosody crucially reveals a break index '0' immediately preceding the final word "sleep", which receives 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 crucial 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 literal evidence in the natural data, 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 53, 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 Wingate’s “fault line” analysis discussed in Chapter 2, 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 (Wingate 1976). 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[15]. Applying this metric

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

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.

More examples and further elaboration of this target determination process will be examined later in this chapter.

5.2 Natural vs. target comparisons

This section will compare the disfluent environments as they surfaced naturally in the spontaneously produced data described in Chapter 4, with the target environments as determined by the script data.

Prosodic features of the 'target' word were analyzed for each subject with respect to the tonal type, word class, prominence level, and phrasal position of the target word, and syllable location of the disfluency in the target word. Unlike in Chapter 4, where comparisons were made among the natural data, in this chapter the disfluencies found in the natural data were re-interpreted by referring to the reference prosody of the script. For example, the phrase in Figure 54 represents the reference prosody of the script reading by control speaker C1, corresponding to the spontaneously-produced phrase by stutterer S1, shown in Figure 53. Since Definitions 1 and 2 indicated that the target was the Prosodic Word "to sleep", features of the target word would therefore be the following: tonal type= H*; grammatical class= content word[16]; prominence level= NPA; phrasal position= ip-final.

Looking at the target data from script analyses is crucial for testing the robustness and relevance of these patterns, since essentially all of the conditions—with the exception of word position—rely on determining the intended (targeted) structure, if their interactions with disfluencies are to be fully understood. Evidence of a planned nuclear pitch accent, for example, may become opaque, if a preceding disfluency resulted in a new ip break (henceforth called 'derivative' ip)—thus rendering the pitch accent phrase-initial instead of phrase-final.

5.2.1 Characteristics of reference prosody

The basic descriptive characteristics of the reference prosody for each subject are presented in Table 9, listing the raw numbers of words, intonation phrases (IP), intermediate phrases (ip), disfluencies, and pitch accents (PA).

	S1 Natural	S1 Script	S2 Natural	S2 Script	S4 Natural	S4 Script
# words	1212	1005	571	508	2814	1828
# ip	357	232	153	90	701	329
Avg. wd/ip	3.53	5	3.76	5.91	4.13	5.56

[16] Although a Prosodic Word can begin with a function word (e.g., "to"), it is proposed that the word class of a Prosodic Word target be the head of the Prosodic Word—i.e., a content word (e.g., "sleep")—since the function word in this case is cliticized to the content word. In accord with Definitions 1 and 2, if the function word is not cliticized to another word (i.e., is not separated from an adjacent word by a 0-sized prosodic break), then the target would be the function word itself.

# PA	566	470	310	258	1257	824
Avg. PA/ip	1.65	1.98	2.05	1.63	1.85	2.14

Table 9: Reference prosody characteristics for all subjects: # of words, ips, IPs, and PAs

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. For instance, any time a speaker restarted a whole word, each restart represented a separate attempt of the *same* target. Thus, in the example in Figure 52, the speaker produced the phrase "runs to a—to a cliff" in his narrative. A script indicating the intended utterance, however, would require removal of the extraneous restart to produce the crucially shorter intended phrase, "runs to a cliff". Thus, in this most common case, a script interpretation resulted in less prosodic material (e.g., fewer words) than was in the original narrative.

5.2.2 Tonal type comparisons

The first comparison between target and narrative data was in tonal quality of the pitch accent. Pitch accent tone comparisons did not differ appreciably from the overall tonal distribution of each stutterer's narrative data: L+H* and H*, the most commonly occurring tones for all stutterers, were also the tones produced most frequently for each disfluency type. For the target (i.e., script) analysis, however, the proportion of these tones was reversed: L+H* tones were the most common targets for each speaker, while H* production was considerably lower for script targets than for the natural data:

PA target tone	H*	L*	L+H*	L*+H	!H*	H+!H*	L+!H*	L*+!H
S1	50	9	104	2	17	0	8	1
%	**26.2%**	4.7%	**54.5%**	1.0%	8.9%	0%	4.2%	0.5%
S2	16	7	42	2	16	0	8	0
%	**17.6%**	7.7%	**46.2%**	2.2%	17.6%	0%	8.8%	0%
S4	216	36	255	9	147	3	24	6
%	**31.0%**	5.2%	**36.6%**	1.3%	21.1%	0.4%	3.4%	0.9%

Table 10: Target pitch accent tone of disfluency: all stutterers

As the table above reveals, target L+H* was produced more often than target H* for all speakers. Although the difference between the two tones was smaller for S4 than for the other two speakers, it was nevertheless significant (S4: $p=0.0313$; S1, S2: $p<0.0001$).

When comparisons were made between speakers' natural and target PA tones, differences were significant for L+H* and H* for S1 ($p<0.0001$) and S4 ($p<0.0001$), but only constituted a trend for S2's L+H* ($p=0.075$). Figures 55-57 illustrate clearly the much higher frequency of L+H*, and lower rate of H* pitch accents, for targets in the script data:

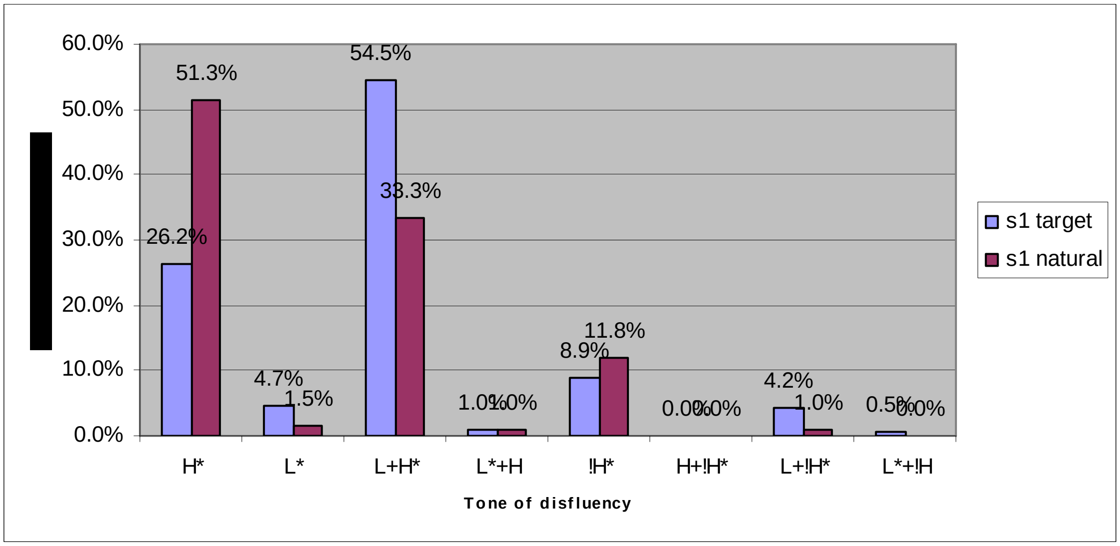


Figure 55: Target pitch accent tone of disfluency: s1

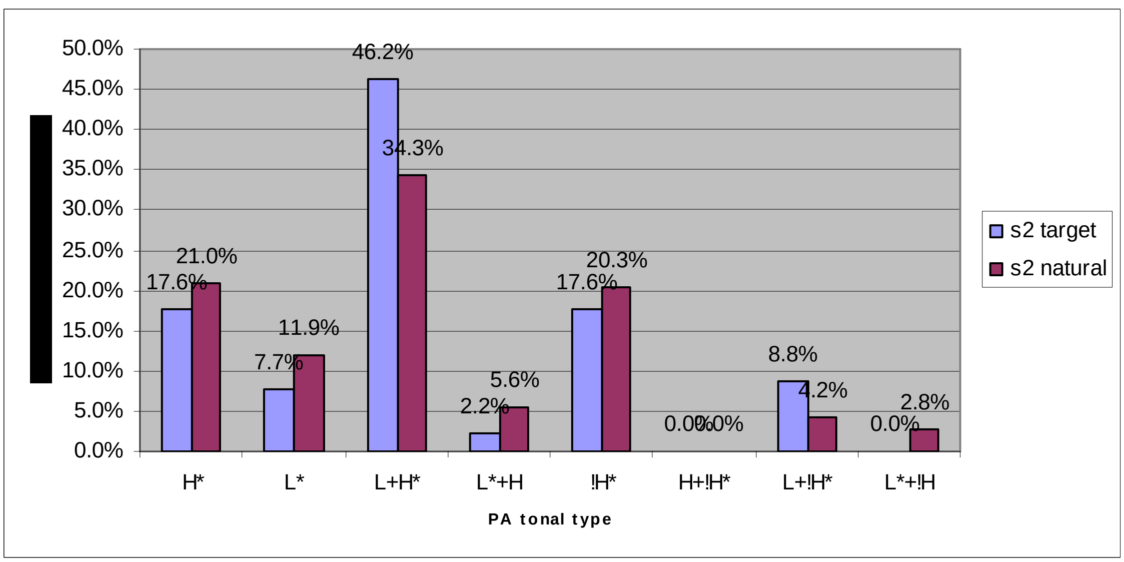


Figure 56: Target pitch accent tone of disfluency: s2

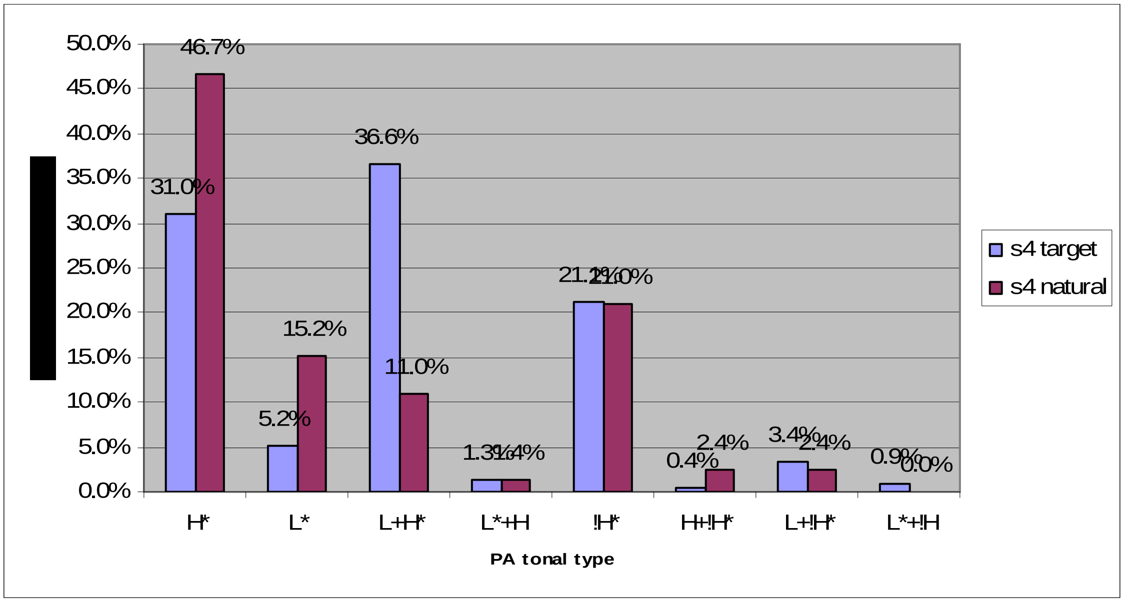


Figure 57: Target pitch accent tone of disfluency: s4

Tonal quality of phrase accents also differed consistently for all stutterers. In natural production, phrase accents realized after disfluencies most frequently carried !H- tones; targets produced in the script readings, however, preceded significantly fewer !H- phrase accents for script targets than for natural disfluencies (S1: $p=0.030$; S2: $p=0.006$; S4: $p=0.0001$). The frequency of L- phrase accents, meanwhile, was significantly greater for targets than for natural disfluencies for speakers S2 ($p=0.006$) and S4 ($p<0.0001$), but not for S1.

In the graphs below, comparisons of phrase accents for each speaker are shown for three environments: in the overall ('whole')[17] distribution, on disfluencies in natural speech ('disfl'), and on the disfluency targets as determined by the reference prosody ('target') produced by the age-matched control subject:

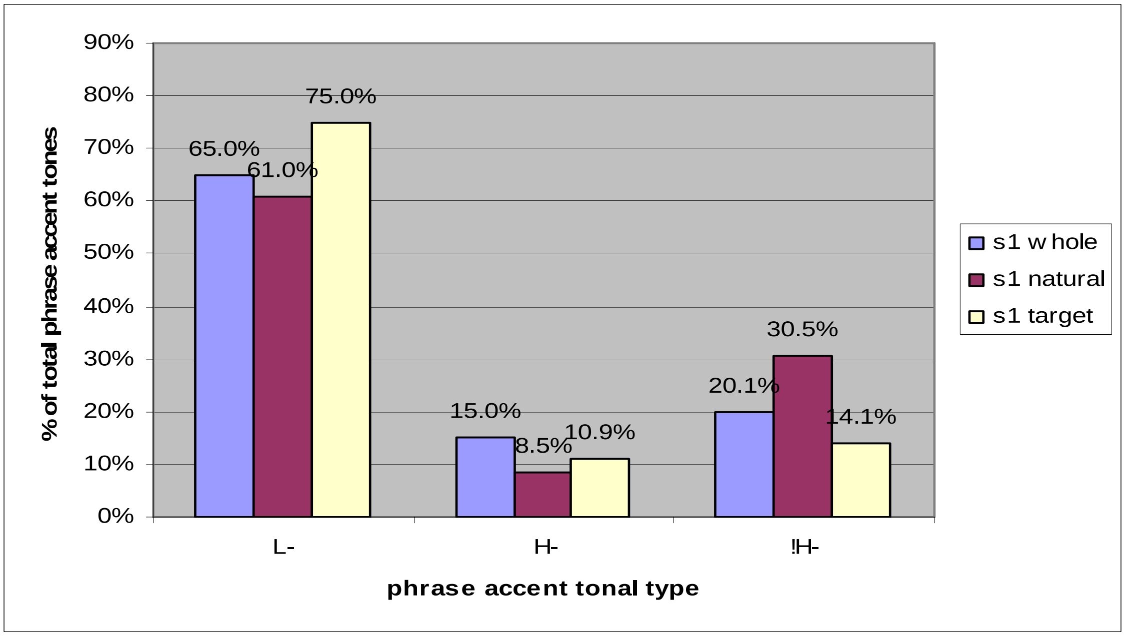


Figure 58: phrase accent tonal type comparisons: s1

[17] As in Chapter 4, 'whole data' refers to the distribution of the prosodic phenomenon in question (in this case, phrase accents) throughout the entire narrative, disfluencies and fluent productions alike.

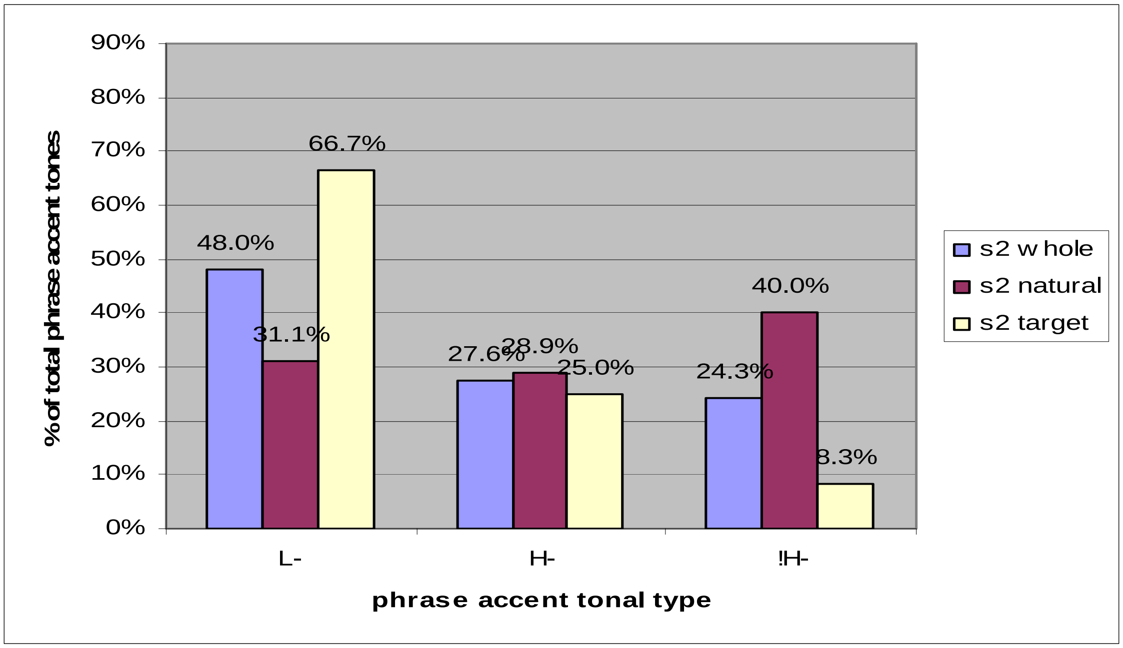


Figure 59: phrase accent tonal type comparisons: s2

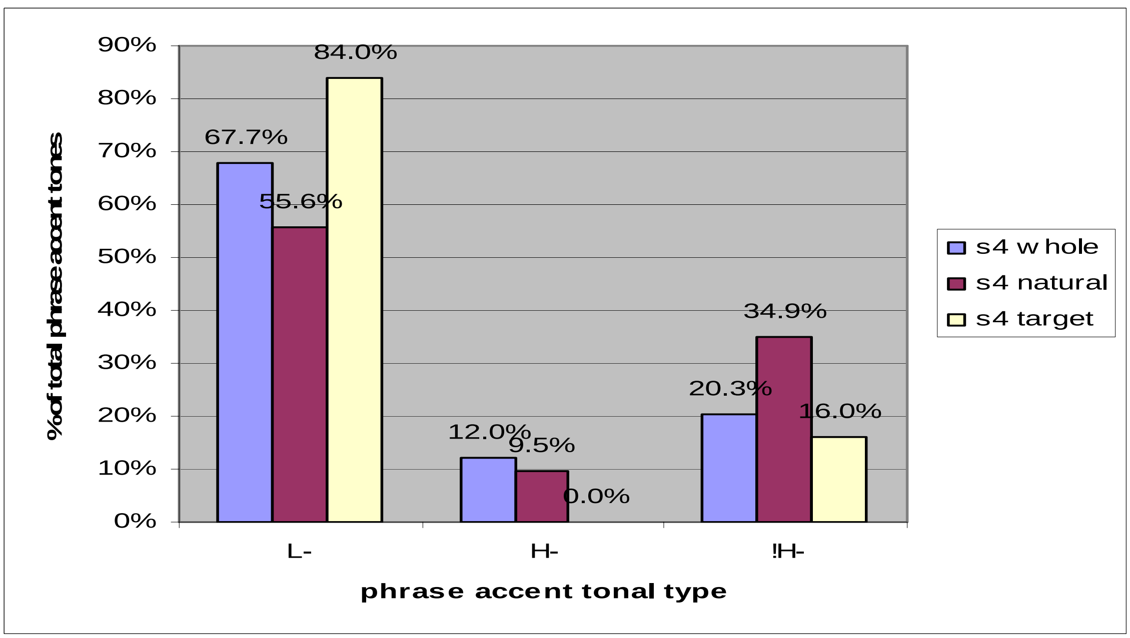


Figure 60: phrase accent tonal type comparisons: s4

The differences analyzed in Chapter 4—between normal environments and disfluencies—reveal the same trend for each subject: a decrease in L- production, and an increase in !H- production in natural disfluencies as compared to that of the whole data. A comparison of the natural disfluencies with their targets, however, follows the pattern of the whole data, with higher production levels of L- and a corresponding lowering of !H- rate.

All combinations of boundary tone contours were again compared for each stuttering subject, and one pattern similar to that in phrase accents emerged: the anomalous behavior of post-disfluency !H-L% contours discussed in Chapter 4 was underscored by the reversion to near-normal tonal levels in the target generation. As observed in Figures 61-63, while all stutterers showed an increase of !H-L% boundaries in disfluent environments relative to the whole data, target !H-L% tone contours were produced at rates closer to—though still below—those found in the data as a whole. For individual speakers, none of the differences between target and whole !H-L% contours were significant. To the contrary, for all speakers, !H-L% boundaries occurred on targets significantly less often than they did on natural disfluencies (S1: $p=0.002$; S2: $p=0.027$; S4: $p=0.010$).

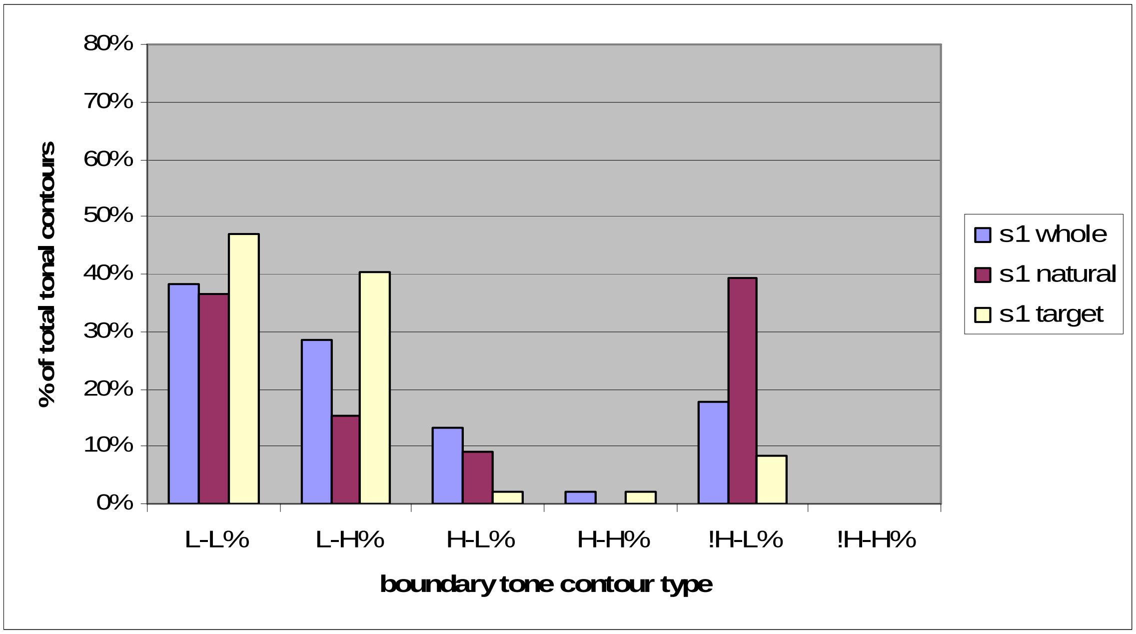


Figure 61: whole vs. natural vs. target boundary contour types: s1

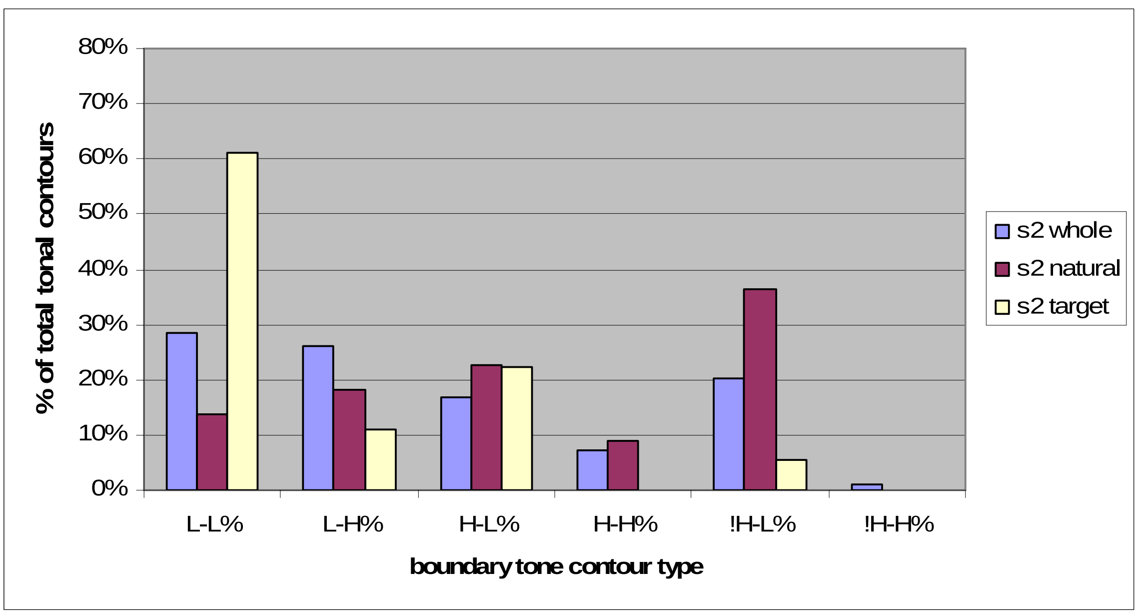


Figure 62: whole vs. natural vs. target boundary contour types: s2

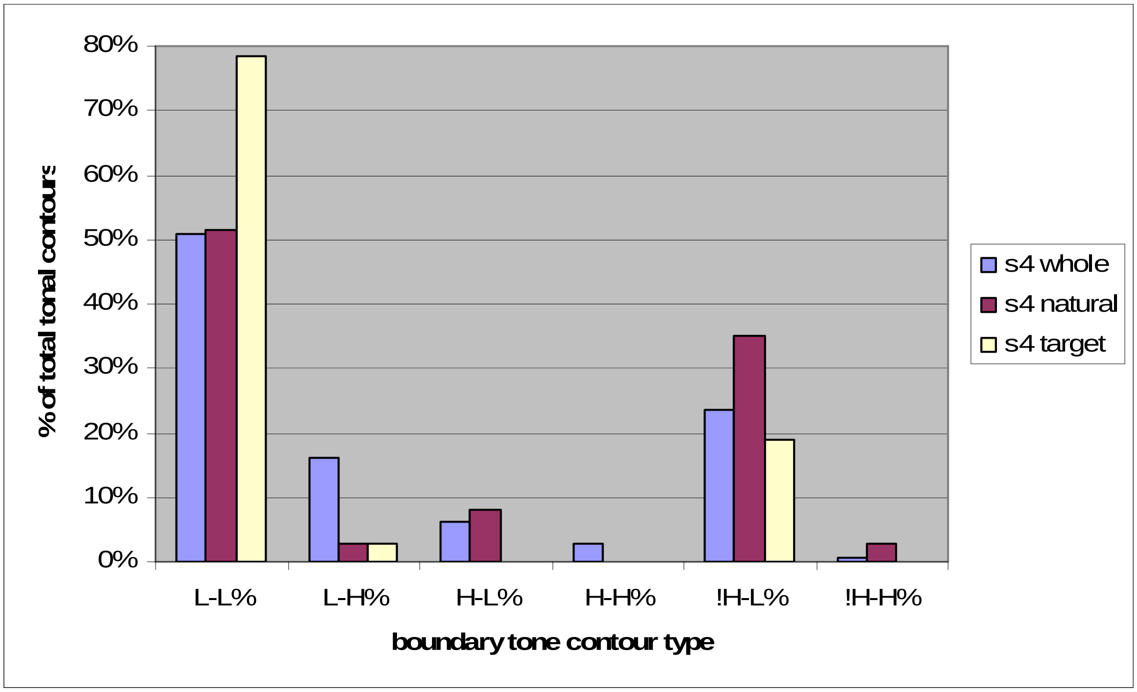


Figure 63: whole vs. natural vs. target boundary contour types: s4

Although the compiled data for the three subjects in Figure 64 reflect a higher rate of L-H% boundary tone contours for targets ('s target') than for disfluencies in the natural data ('s disfl'), this rate was inflated by the script data of S1, as Figure 61 shows. However, L-L% tone contours surfaced for all three subjects at a much higher rate after target words, than for both the disfluent words in the natural data ('s disfl') and in the general distribution of the data as a whole ('s whole'). The difference between L-L% target and whole data distribution was significant for S2 ($p=0.013$) and S4 ($p<0.0001$), while the difference between target and natural L-L% was also significant just for speakers S2 ($p=0.003$) and S4 ($p<0.0001$).

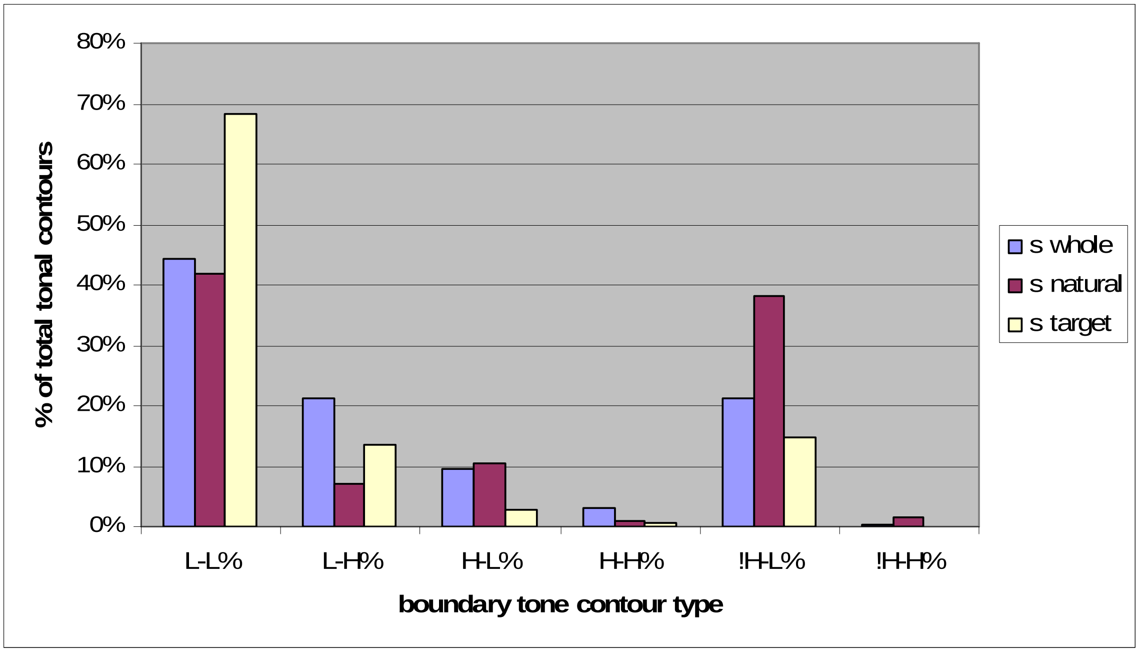


Figure 64: whole vs. natural vs. target boundary contour types: all stutterers (3)

Targets often reflected a speaker's average—rather than disfluent—production patterns, but for Speakers S2 and S4 a low relative percentage of L-H% tones were produced in both natural and target disfluency position, as Figure 62 and Figure 63 show. For S1, on the other hand, L-H% contours in target position represented a higher relative percentage of total boundary tones than did L-H% contours in both natural disfluencies and in the data as a whole. The exact cause notwithstanding, these competing trends among speakers resulted in a less consistent picture of boundary contour comparisons when the latter were categorized as a function of f0, as shown in Figure 65. As in Chapter 4, boundary contour types were again divided into four tonal classes: low (L-L%), mid (!H-L%), high (H-L%), and rising (L-H%, H-H%, !H-H%). Although target boundary contour patterns varied among subjects, one salient pattern found in all subjects was a higher percentage of low target contours relative to other tone types. For all three subjects, the percentage of low target contours was significantly greater than mid (S1: $p<0.0001$; S2: $p=0.0001$; S4: $p<0.0001$) and high contours (S1: $p<0.0001$; S2: $p=0.041$; S4: $p<0.0001$). For low vs. rising comparisons, the difference was significant only for S2 ($p=0.004$) and S4 ($p<0.0001$).

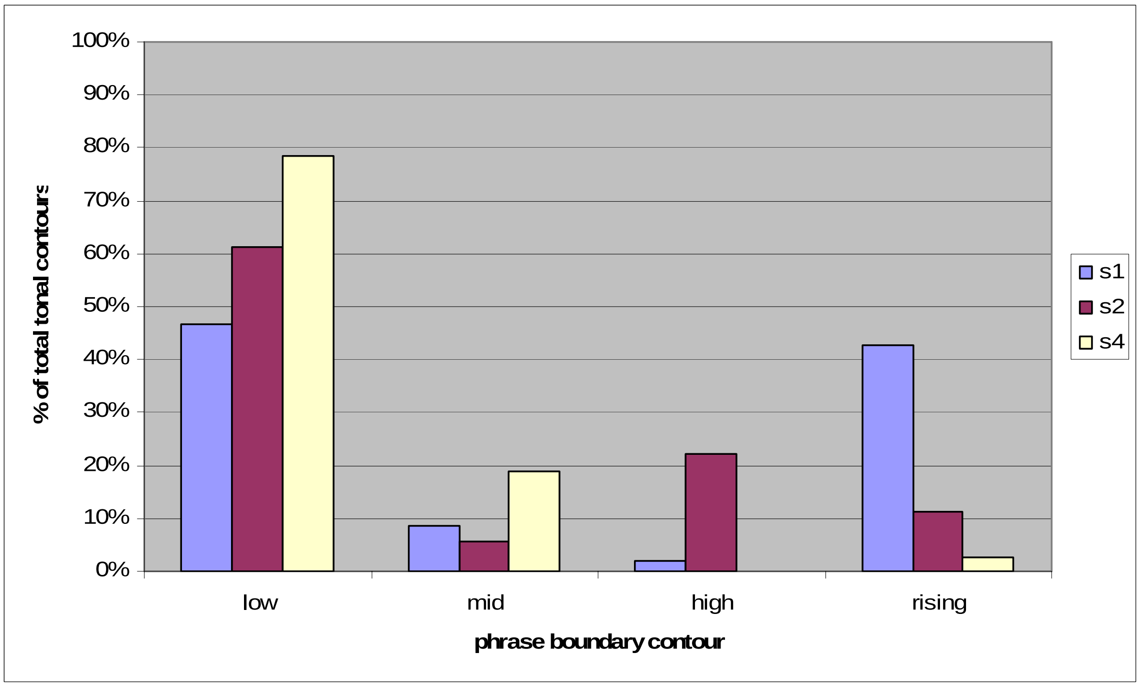


Figure 65: Target boundary contour comparisons: all stutterers

5.2.3 Word class

Natural data analyses in the previous chapter revealed that stutterers as a group produced more disfluencies on function words than on content words, while control subjects produced the opposite trend. When the three stutterers' disfluency data and their scripted targets are broken down, it is apparent that the disproportionately high number of disfluencies produced by Speaker S2 on function words is responsible for much of this bias. Figures 66-68 show target and natural comparisons for each individual stutterer:

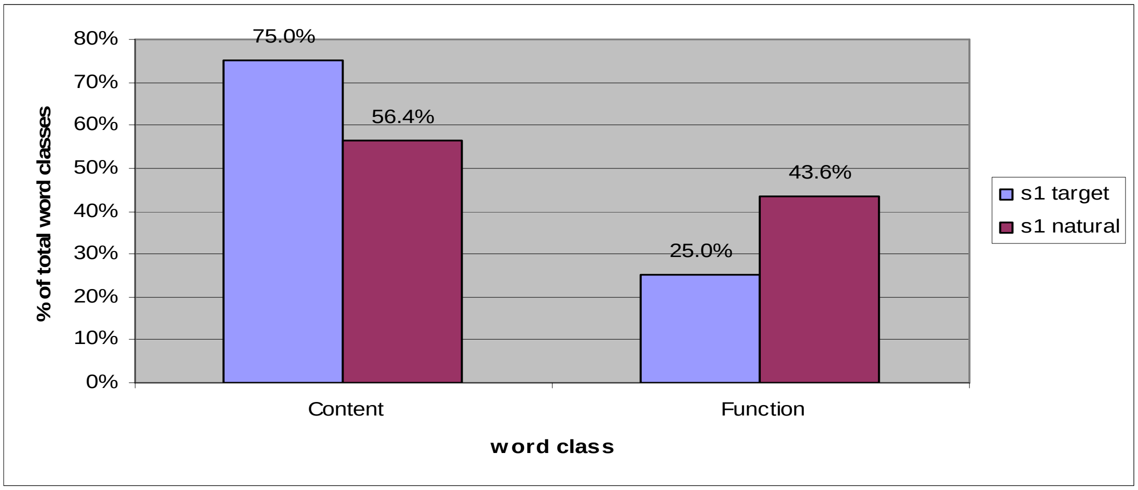


Figure 66: word class target vs. natural: S1

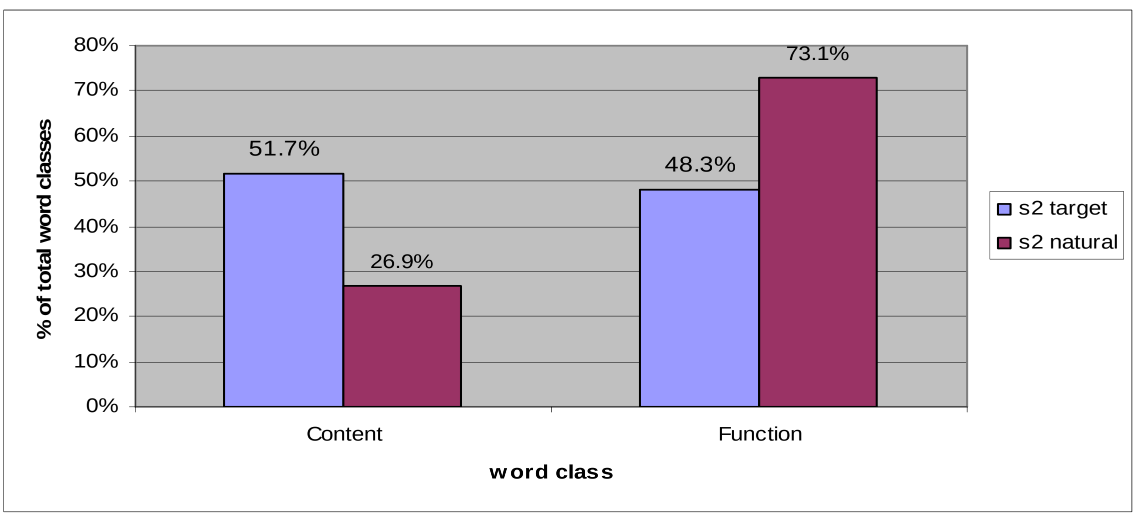


Figure 67: word class target vs. natural: S2

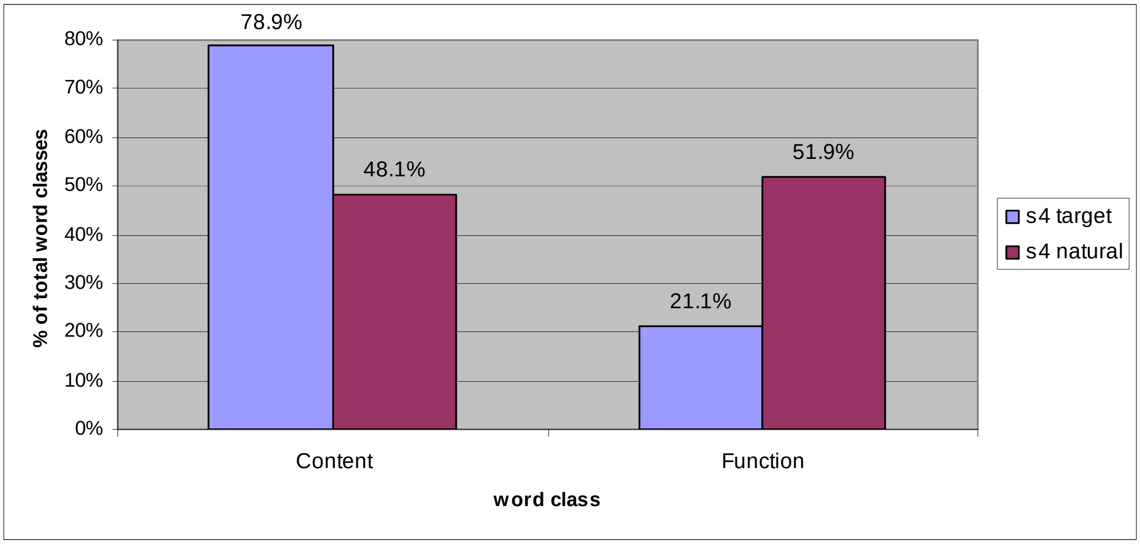


Figure 68: word class target vs. natural: S4

Despite the dissimilar pattern in S2's natural data, the data of all three stutterers interpreted with respect to their corresponding script readings yielded statistically significant ($p<0.0001$) differences in grammatical classes in the same direction—i.e., a larger proportion of content than function words in target position.

5.2.4 Prominence level

It was reported in the previous chapter that stutterers and controls differed considerably in their generation of pitch accents in disfluent environments. Controls produced disfluencies on unaccented words more than half the time (57%) while stutterers' disfluencies occurred significantly more often on pitch-accented and nuclear pitch-accented words (64.5%) than on unaccented forms. Nevertheless, it is feasible that when unaccented words are produced disfluently, they receive pitch accent as a derivative effect. 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.

Such concerns rendered even more crucial the decision to use a reference prosody to which natural prominence patterns could be compared. For the three stutterers, target word prominence levels were higher across the board, as targets were either pitch-accented or nuclear pitch-accented with greater frequency than they were unaccented. 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 evidence. The combined results for all subjects are shown in Figure 69, while Figures 70 and 72 illustrate how two speakers (S1, S4) followed this general pattern: i.e., more targets were pitch-accented than unaccented. Conversely, as Figure 71 shows, Speaker S2's target PA rate was lower than his natural disfluency PA rate:

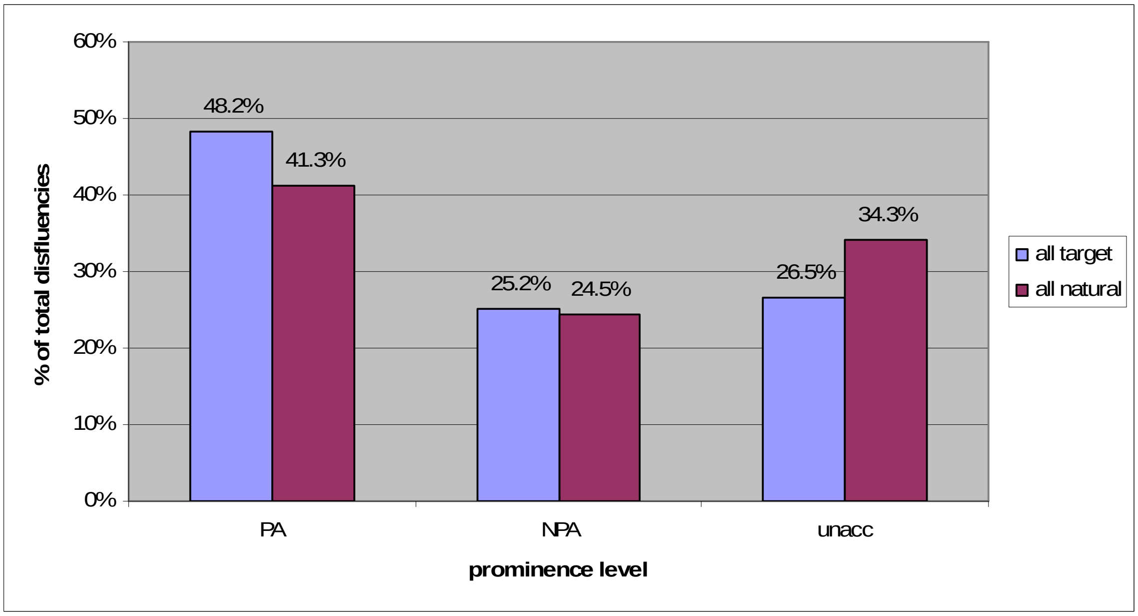


Figure 69: Target vs. natural prominence level—all speakers (3)

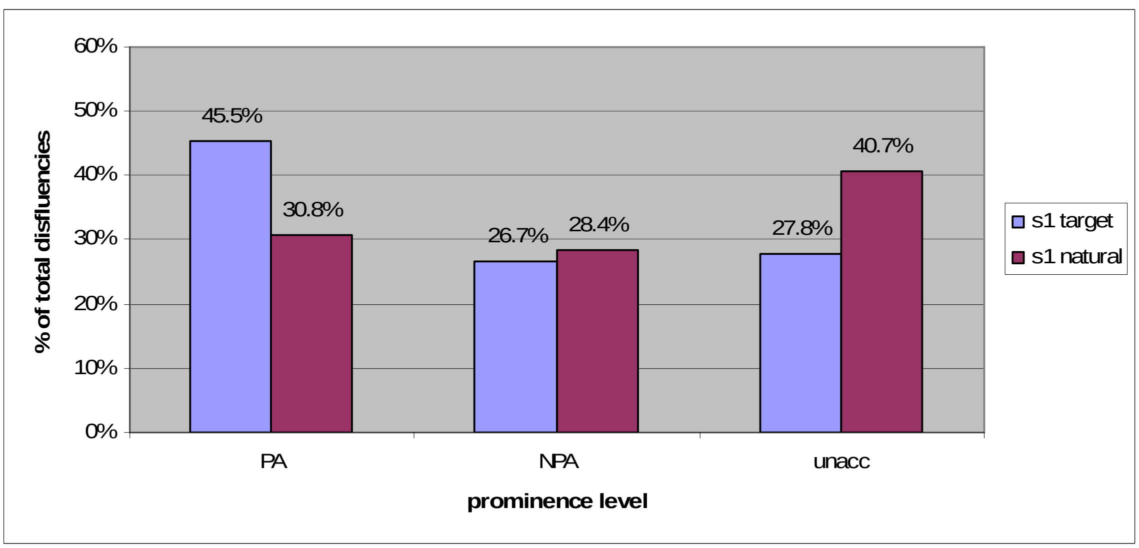


Figure 70: Target vs. natural prominence level—S1

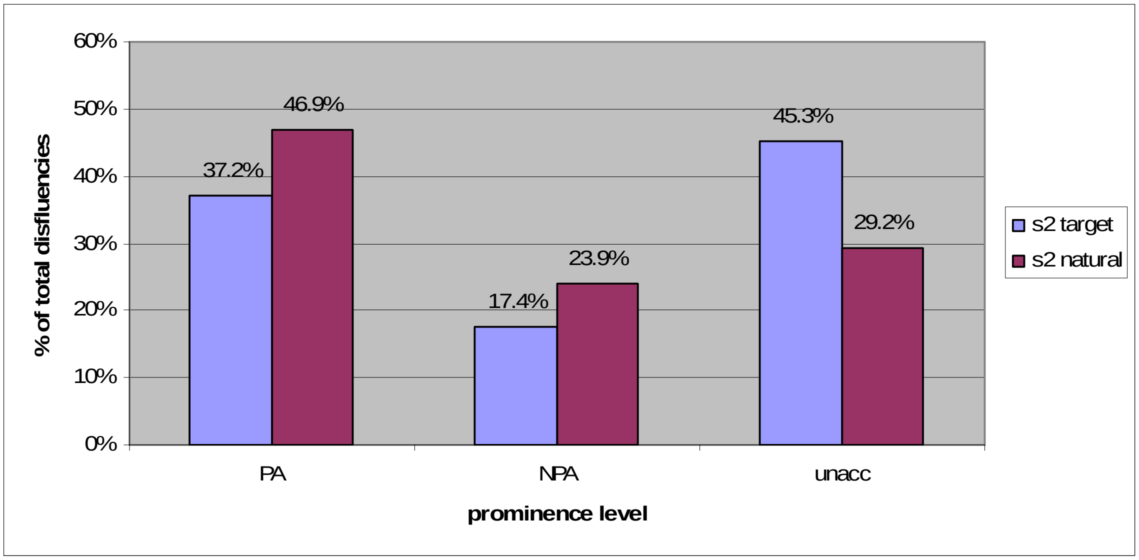


Figure 71: Target vs. natural prominence level—S2

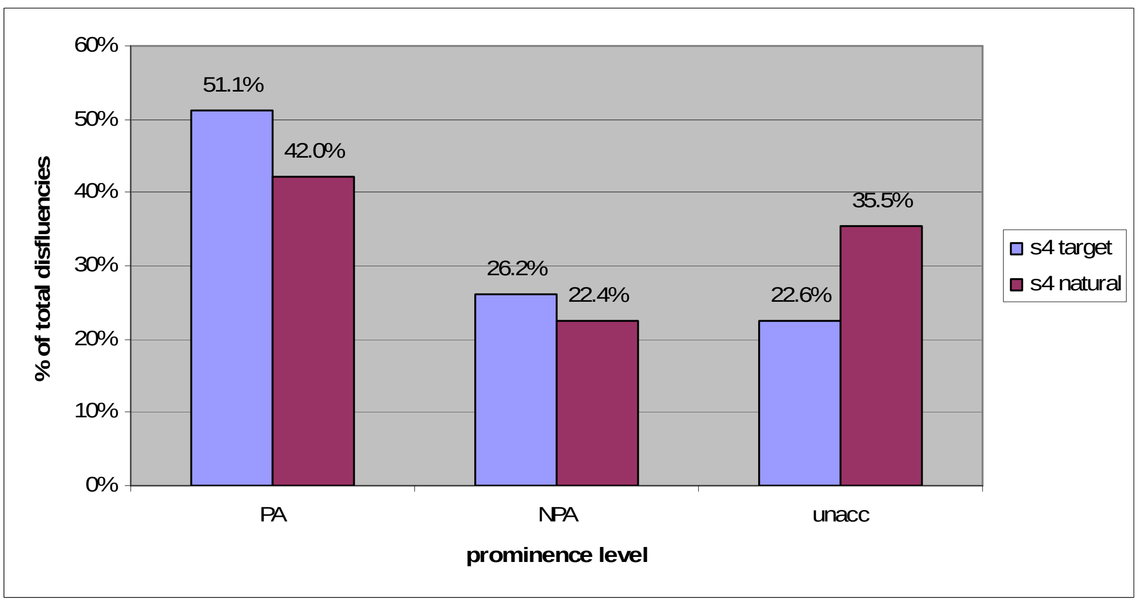


Figure 72: Target vs. natural prominence level—S4

The general trend in the data from Figure 69 shows that pitch-accented targets made up a larger percentage of target accent types than did pitch-accented disfluencies in the natural data (target= 48.2%; natural= 41.3%). Unaccented targets, meanwhile, constituted only 26.5% of target accents, while unaccented disfluencies represented 34.3% of all disfluent accents. Both differences were statistically significant ($p=0.0002$; $p<0.0001$).

For S1, differences between target and natural prominence rates were highly significant for pitch-accented ($p=0.0003$) and unaccented ($p<0.0001$) words. In the natural data, 30.8% of disfluent words were pitch-accented while 40.7% were unaccented; the same environments when examined in the script readings, meanwhile, resulted in target words being pitch-accented 45.5% of the time, and unaccented in just 27.8% of the cases.

As illustrated by Figure 71, S2 showed a similar pattern in NPA production, but pitch-accented target words made up a smaller percentage of target data (37.2%) compared to the proportion of pitch-accented disfluent words within the natural data (46.9%). In addition, unlike for S1, unaccented words made up a higher percentage of S2's targets (45.3%) than of his disfluent natural data (29.2%); raw data comparisons revealed this difference to be statistically significant ($p=0.0013$).

Figure 72 reveals a pattern similar to that of S1, as S4's unaccented target words constituted a smaller proportion of total target accents (22.6%) compared with disfluent words from the natural data (35.5%) ($p<0.0001$). Likewise, pitch-accented target words made up a greater percentage of S4's total script data accents (51.1%) than did pitch-accented disfluent words of the natural data (42%) ($p=0.0007$).

The above comparisons only measure the percentage of each prominence level rate relative to all other prominence levels. However, since the high percentage of any prominence level can be due to the overall high occurrence of the prominence type, disfluency rates for each prominence level were calculated relative to all the occurrences of the same prominence type. Combined speaker data in Figure 73 reveal a highly significant difference ($p<0.0001$) between the disfluency rates of NPAs when determined by the target data compared to naturally occurring data: 53.3% of target-defined NPAs were disfluent overall, while only 31.5% of NPAs in the natural data were disfluent. Target pitch accents also had a higher rate of disfluency (73.6%) than natural pitch accents (70.0%; $p=0.0959$), while unaccented words were disfluent less often in the target data than in the natural data (target= 20.4%; natural= 23.2%; $p=0.3228$), though neither of these comparisons reached statistical significance.

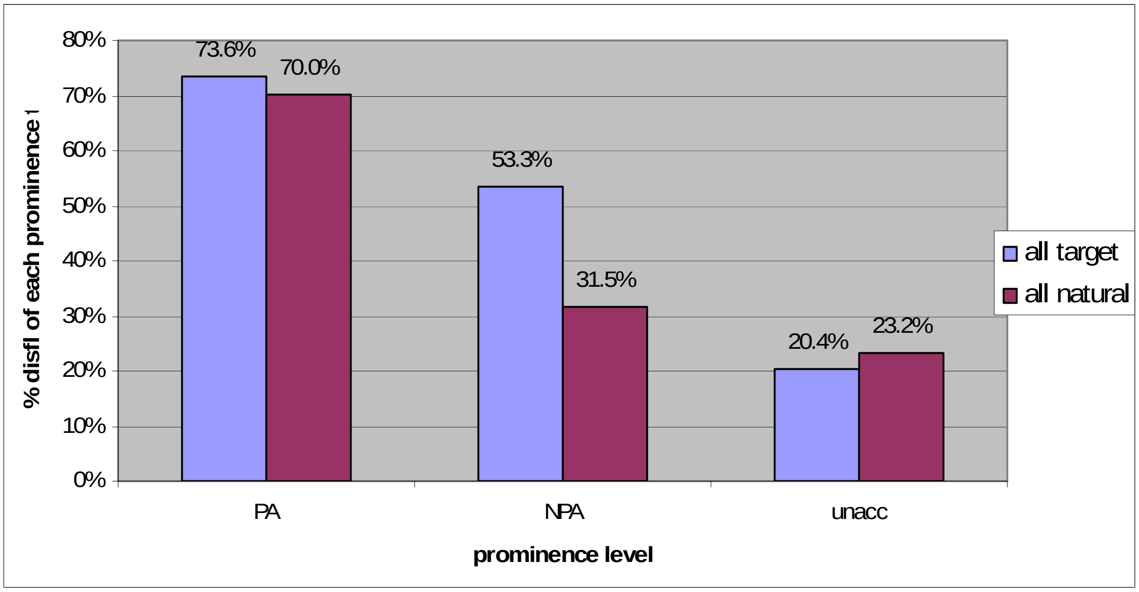


Figure 73: % disfluent of each prominence type relative to the total occurrence of the same prominence type—all speakers (3)

Target and natural disfluency rates for individual speakers, however, were consistent with the group trend only for Speakers S1 (Figure 74) and S4 (Figure 76): i.e., both speakers produced more disfluent target than natural pitch-accented words, more disfluent target than natural nuclear pitch-accented words, and fewer disfluent target than natural unaccented words. Differences between target and natural unaccented words were significant for both S1 ($p=0.008$) and S4 ($p=0.043$), while differences between target and natural PA, and target and natural NPA, were significant only for S4 ($p<0.0001$). As Figure 75 illustrates, Speaker S2 varied from this pattern by producing significantly fewer disfluencies on target pitch accents (38.1%) than on natural pitch accents (62.4%; $p<0.0001$), and more disfluencies on unaccented targets (31.2%) than on unaccented words in the natural data (23.4%; $p=0.059$).

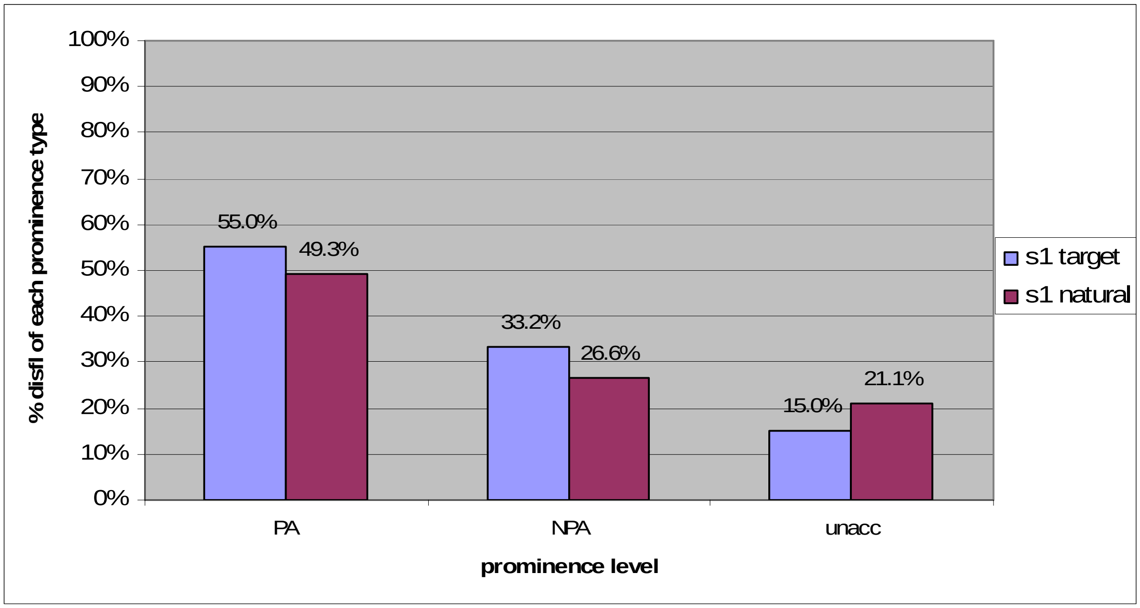


Figure 74: % disfluent of each prominence type relative to the total occurrence of the same prominence type: S1

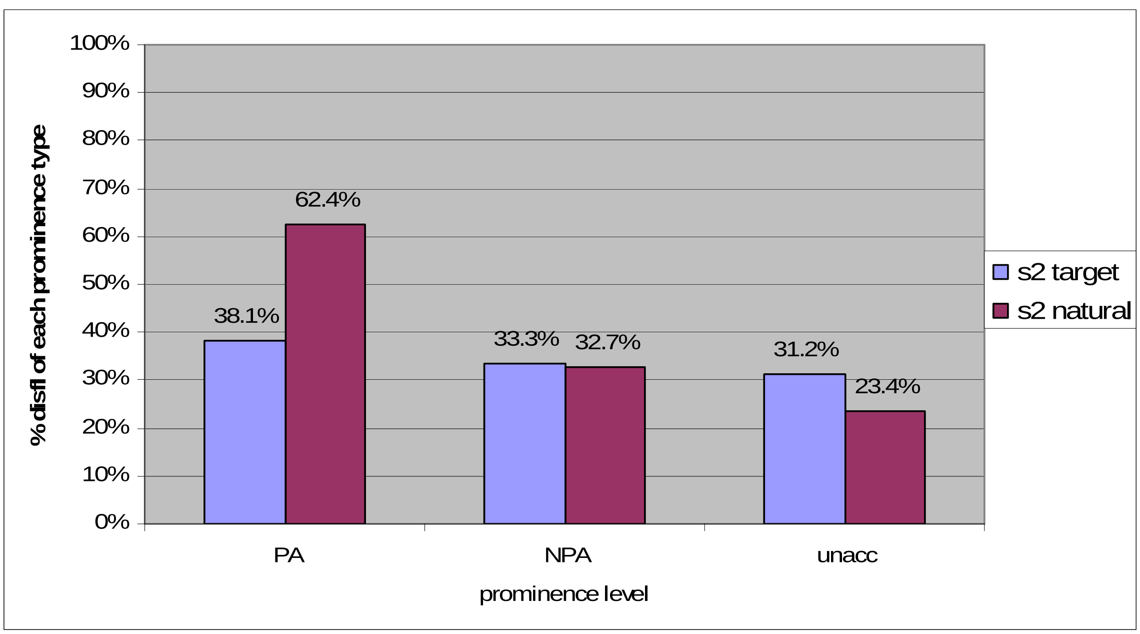


Figure 75: % disfluent of each prominence type relative to the total occurrence of the same prominence type: S2

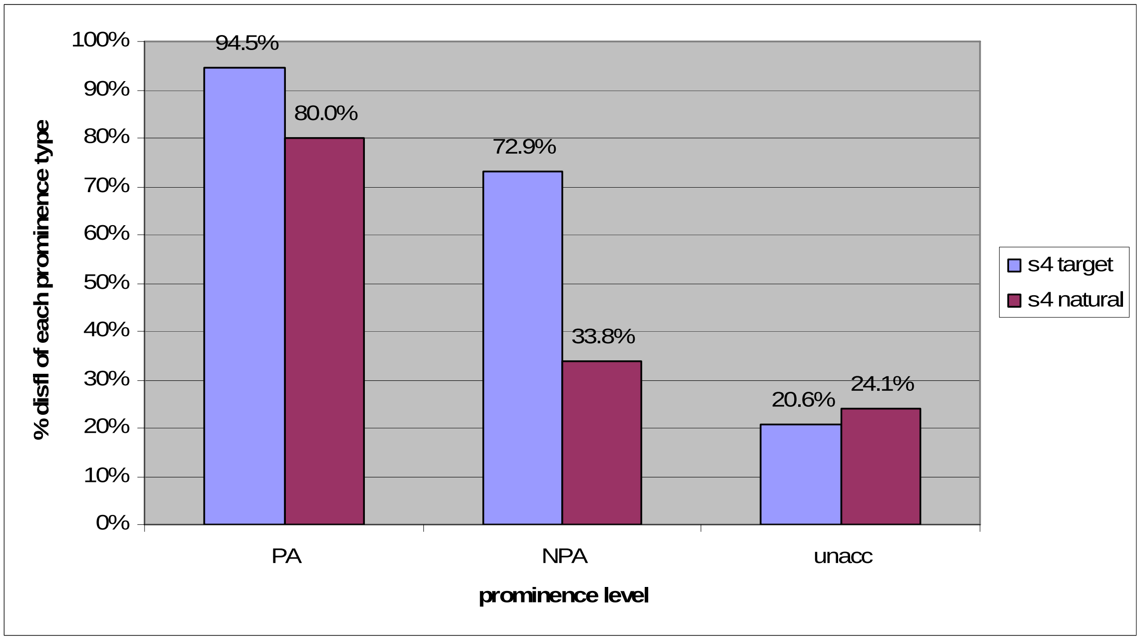


Figure 76: % disfluent of each prominence type relative to the total occurrence of the same prominence type: S4

In order to determine more specifically the effects of metrical prominence, the coding categories for prominence were expanded from the three used in the natural data analysis (pitch-accented, nuclear pitch-accented, and unaccented) to seven, thus adding to the original three the following four prominence categories: pre-nuclear pitch accent (pre-NPA), post-nuclear pitch accent (post-NPA), pre-pitch accent (pre-PA), and two words preceding a PA[18] (2-pre-PA). The motivation behind this more detailed classification was to capture any patterned effects of smaller prosodic constituents—i.e., Prosodic Words as defined at the beginning of this chapter. In addition to Prosodic Words, in the previous literature on prosodic structure a further distinction is often proposed between Prosodic Words and Phonological Phrases, the latter based on syntactic constituency as

[18] This category was collapsed to include all pitch accents, including nuclear pitch accents.

elaborated by Selkirk (1986), Nespor & Vogel (1986), and Hayes (1989). 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. Satellite prosodic material may be either stressed or unstressed, but crucially not pitch-accented, and the entire structure often—though not necessarily—coincides with syntactic phrase structure.

For the purpose of this analysis, it will be maintained that while the constituents of Prosodic Words cannot be separated from one another by a break index larger than '0', the constituents of Phonological Phrases may be separated either by breaks of size '0' or '1'. For instance, in the phrase uttered in Figure 77 below, the prepositional phrase "for the frog" also forms a Phonological Phrase, with the nucleus (pitch accent) occurring on "frog" and its satellite material "for the" directly preceding it:

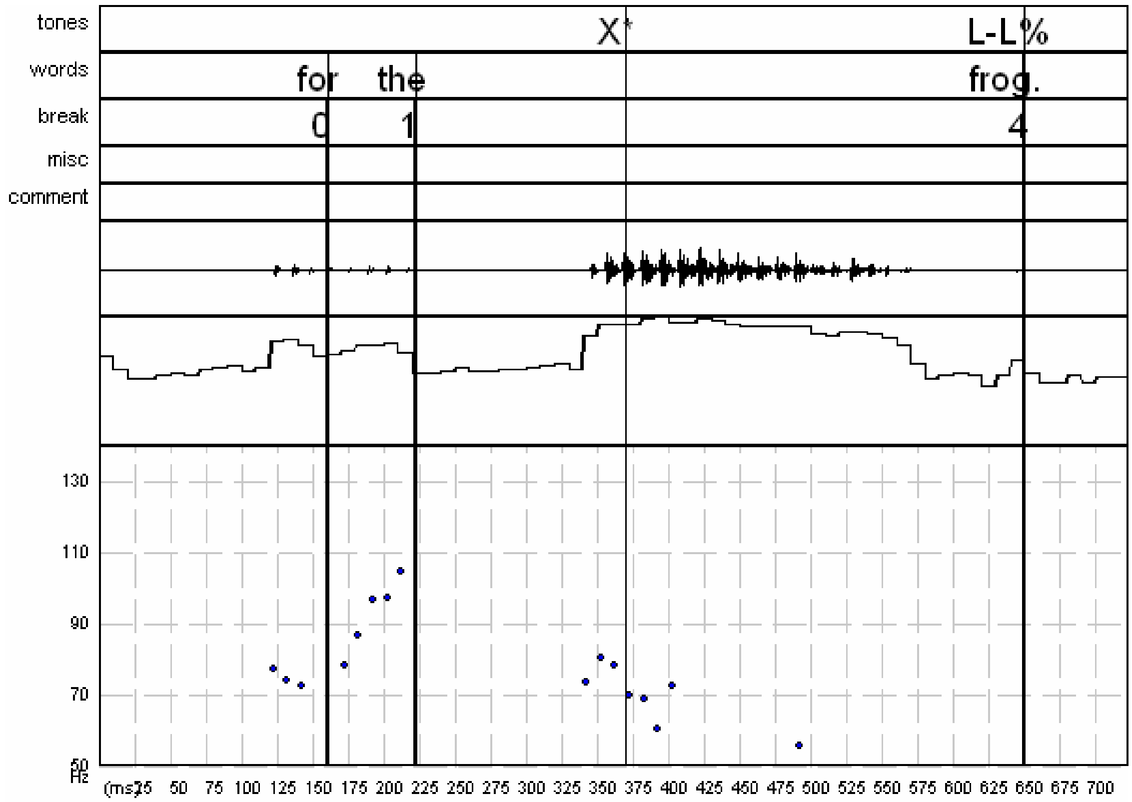


Figure 77: Phonological phrase example

Although “for” and “the” are separated only by a clitic-sized break (BI ‘0’), the entire phrase is a Phonological Phrase and not a Prosodic Word, since the nucleus “frog” is separated from its satellite “for the” by a word-sized break (BI ‘1’).

Putting all of these facts together, and using the more detailed prominence classifications proposed above, a disfluency realized on the word “for” as a vowel prolongation would—despite any further derivative disfluent effects following the prolongation—have as its target, the next word: ‘the’. Unlike the earlier examples in Figure 53 and Figure 54, however, the resulting target in Figure 77 is not the entire phrase “the frog”, but rather just the immediately post-disfluency word “the”, since the 1-sized break following “the” prevents any subsequent material from prosodically cliticizing to it. The prominence classification for the target of a prolongation disfluency occurring on the vowel in “for”, therefore, would in this case be ‘pre-NPA’, since the target of “for” is “the”, and the latter immediately precedes the NPA “frog”.

Applying these expanded prominence classifications to the data, it was found that for all subjects, targets were realized with pitch accents significantly more often than they were with other prominence types, as Figure 78 shows:

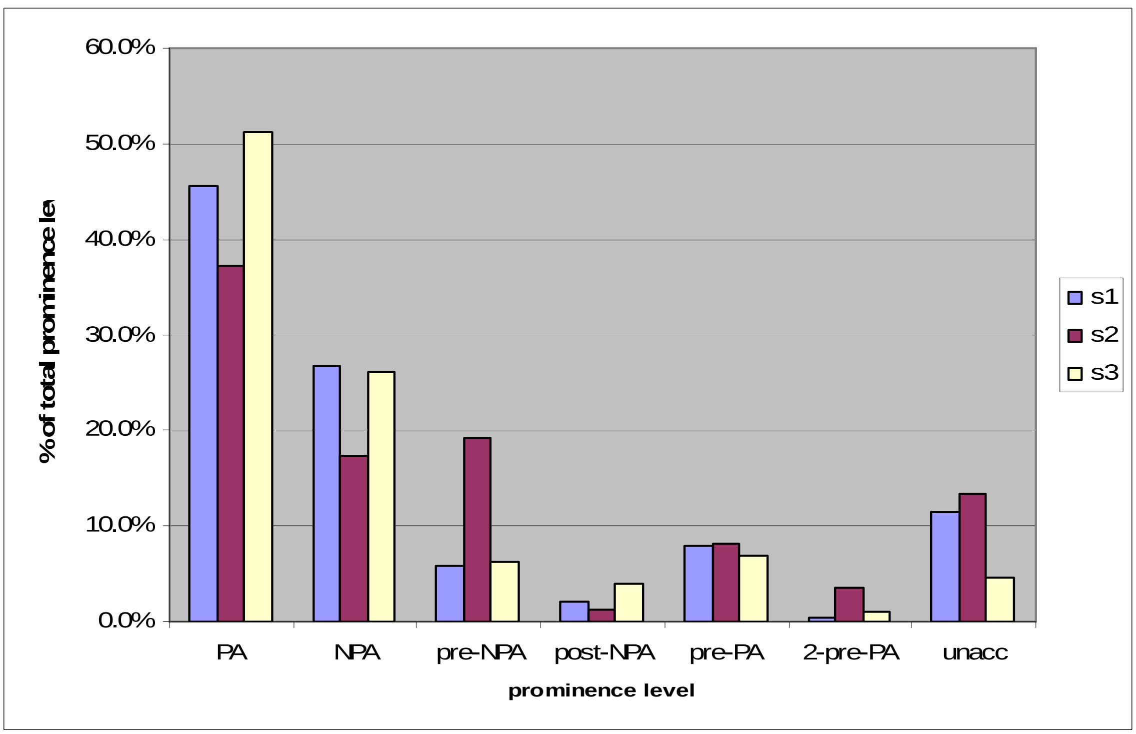


Figure 78: target prominence level: all stutterers

The majority of pairwise comparisons between prominence levels for each subject revealed significant differences, particularly so for speakers S1 and S4. Comparisons which were not statistically significant include the following: for all speakers, pre-PA vs. unaccented was non-significant; for S1 and S4, pre-PA vs. pre-NPA (p=, pre-PA vs. unaccented, and unaccented vs. pre-NPA were non-significant; for S1 and S2, post-NPA vs. 2-pre-PA failed to reach significance; and for individual speakers, NPA vs. pre-NPA (S2), NPA vs. unaccented (S2), pre-PA vs. 2-pre-PA (S2), and unaccented vs. post-NPA (S4) were also not significant. Tables Table 11-Table 13 list for each subject the p-values of all prominence category comparisons.

	PA vs.	**NPA vs.**	**pre-NPA**	**post-NPA**	**pre-PA**	**2-pre-PA**
	NPA					
p-value	0.001					
	pre-NPA	pre-NPA				
p-value	<0.0001	<0.0001				
	post-NPA	post-NPA	post-NPA			
p-value	<0.0001	<0.0001	0.0445			
	pre-PA	pre-PA	pre-PA	pre-PA		
p-value	<0.0001	<0.0001	ns	0.01		
	2-pre-PA	2-pre-PA	2-pre-PA	2-pre-PA	2-pre-PA	
p-value	<0.0001	<0.0001	0.0016	ns	0.0002	
	unacc	unacc	unacc	unacc	unacc	unacc
p-value	<0.0001	<0.0001	ns	0.0004	ns	<0.0001

Table 11: statistical comparisons of prominence types for S1 targets

	PA vs.	NPA vs.	pre-NPA	post-NPA	pre-PA	2-pre-PA
	NPA					
p-value	<0.0001					
	pre-NPA	pre-NPA				
p-value	0.0003	Ns				
	post-NPA	post-NPA	post-NPA			
p-value	<0.0001	<0.0001	<0.0001			
	pre-PA	pre-PA	pre-PA	pre-PA		
p-value	<0.0001	0.0147	0.0043	0.0034		
	2-pre-PA	2-pre-PA	2-pre-PA	2-pre-PA	2-pre-PA	
p-value	<0.0001	<0.0001	<0.0001	ns	ns	
	unacc	unacc	unacc	unacc	unacc	unacc
p-value	<0.0001	Ns	ns	<0.0001	ns	0.0015

Table 12: statistical comparisons of prominence types for S2 targets

	PA vs.	NPA vs.	pre-NPA	post-NPA	pre-PA	2-pre-PA
	NPA					
p-value	<0.0001					
	pre-NPA	pre-NPA				
p-value	<0.0001	<0.0001				
	post-NPA	post-NPA	post-NPA			
p-value	<0.0001	<0.0001	0.0328			
	pre-PA	pre-PA	pre-PA	pre-PA		
p-value	<0.0001	<0.0001	ns	0.0069		
	2-pre-PA	2-pre-PA	2-pre-PA	2-pre-PA	2-pre-PA	
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	
	unacc	unacc	unacc	unacc	unacc	unacc
p-value	<0.0001	<0.0001	ns	ns	0.044	<0.0001

Table 13: statistical comparisons of prominence types for S4 targets

Of the highly significant differences, the key comparisons between pitch-accented (PA and NPA) targets and the lowest prominence targets (post-NPA and unaccented) reveal an overwhelming tendency for targets to bear pitch accents. Post-NPA and unaccented positions are both without metrical prominence—the latter by definition and the former because of de-accenting rules in English intonation. Therefore, a low rate of occurrence in target position suggests that disfluencies are rarely triggered by unaccented environments. Conversely, the high accordance between pitch accents and disfluency targets supports the hypothesis that accent is a potential disfluency trigger.

Sorting through just the comparisons which occur less frequently throughout the entire data, however, shows that post-NPA prominence still rarely surfaces on targets, even relative to less frequent but higher-prominence levels. For instance, pre-NPA and pre-PA positions, in accord with the definition of Prosodic Word proposed in this analysis (i.e., a nucleus—'pitch accent'—and its satellite material), would at least in part inherit metrical prominence from the nucleus of the PWd of which they are constituents. In other words, pre-NPA and pre-PA positions, by virtue of their forming prosodic constituents (i.e., Prosodic Words) with the target word that follows them, would ostensibly share the proneness to disfluency manifested in the target; i.e., in a literal sense, pre-NPA and pre-PA positions are part of the target word itself. Thus, both their higher level of prominence and higher rate of occurrence in target position (compared to post-NPA level) would again suggest a link between prominence and disfluency-triggering.[19]

5.2.5 Phrasal position

In order to compare target and natural results with respect to phrasal position, all data were organized into the same three phrasal classes: ip-initial, ip-medial, and ip-final.

The most salient difference from the phrasal comparisons is that between ip-initial and ip-medial positions, as illustrated by Figures 79-81. For all three subjects, significantly more targets were generated ip-medially than were natural disfluencies; complementarily, targets occurred significantly less often in ip-initial position than did their natural counterparts:

[19] Since the extended prominence coding system was only applied to script data analysis, target-natural comparisons were limited to PA, NPA, and unaccented prominence types.

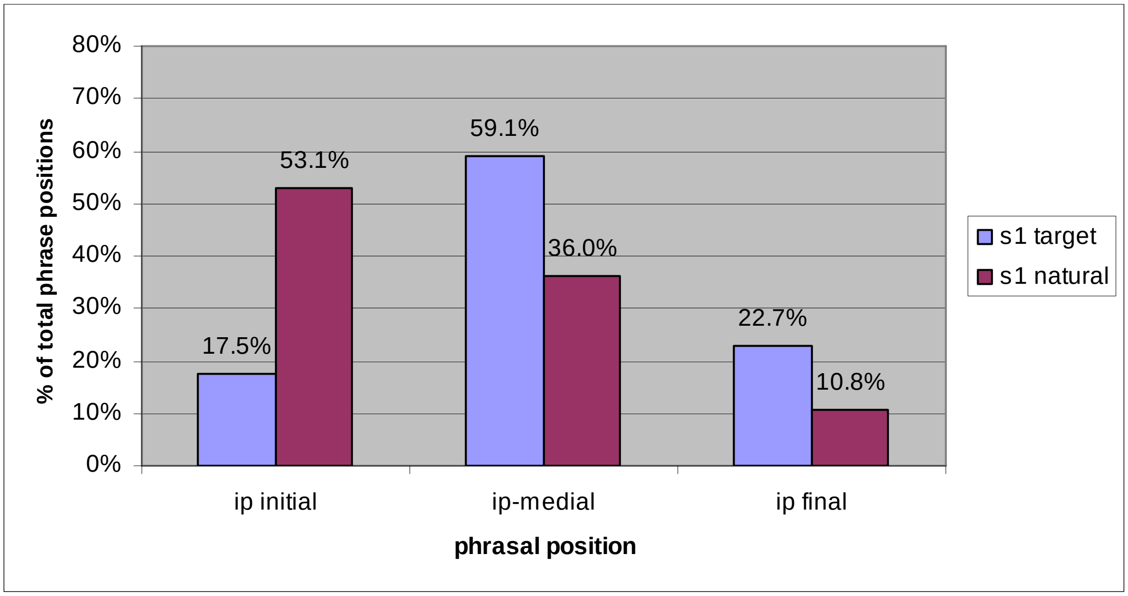


Figure 79: target vs. natural ip-position: s1

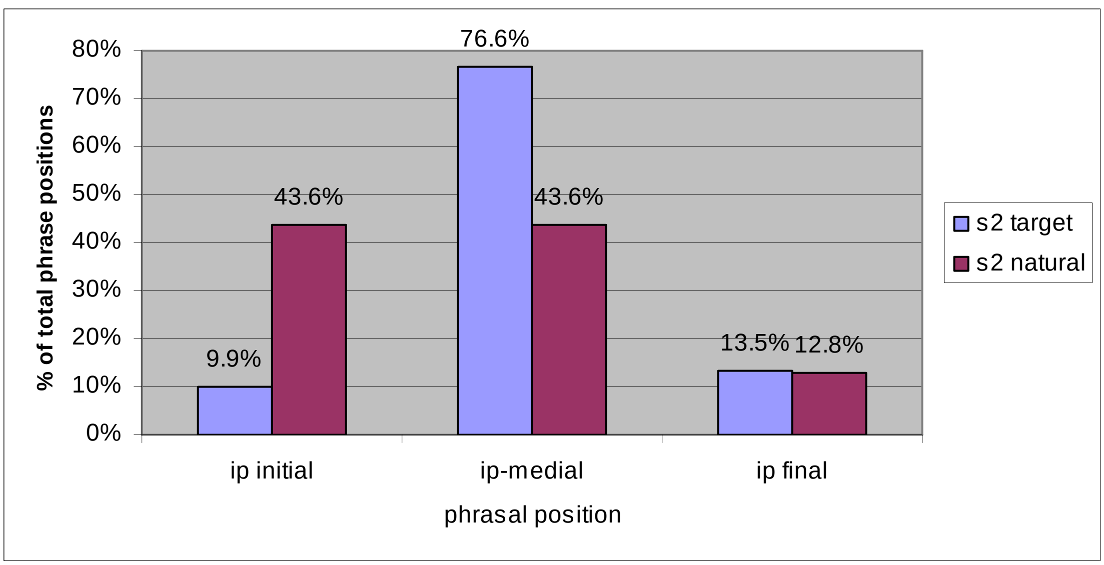


Figure 80: target vs. natural ip-position: s2

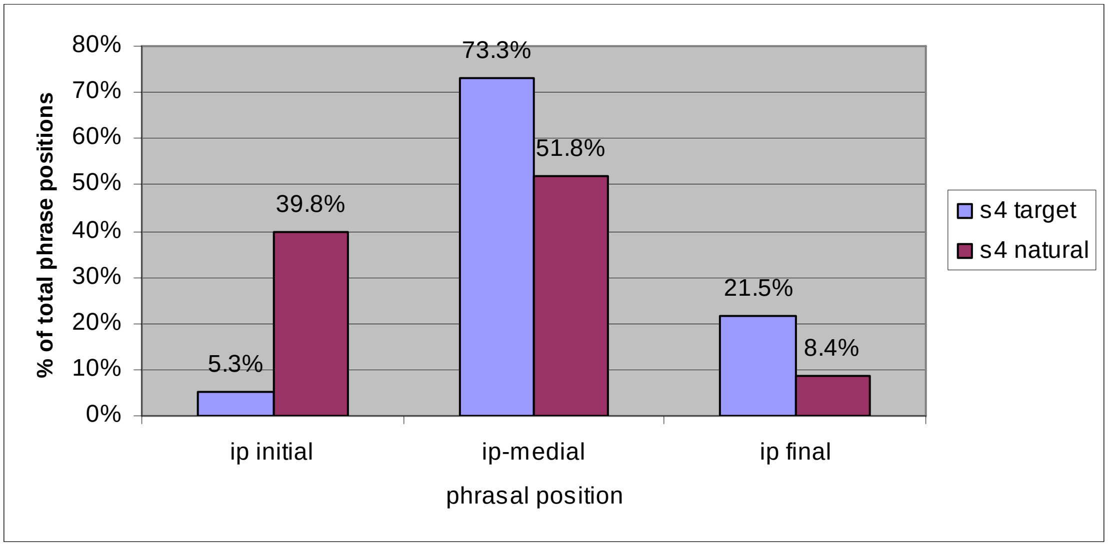


Figure 81: target vs. natural ip-position: s4

This difference is further amplified, since pauses were classified either as ip-medial or ip-initial; in other words, even factoring in the potential artificial boost to ip-initial target numbers when classifying non-medial pauses as ip-initial (rather than a neutral "ip-boundary"), ip-medial position for targets is still highly significant ($p<0.0001$) for each individual subject.

As in the prominence comparisons from the previous section, a more detailed analysis was necessary in order to reveal more accurately any interactions between disfluency patterns and phrasal position. 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, using the definition of Phonological Phrase introduced in the previous section. Phrase length ranged from a single PhP (1 pitch accent) to six PhPs (six pitch accents). Figure 82 shows the percentage of target disfluencies located in each PhP for all three 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.

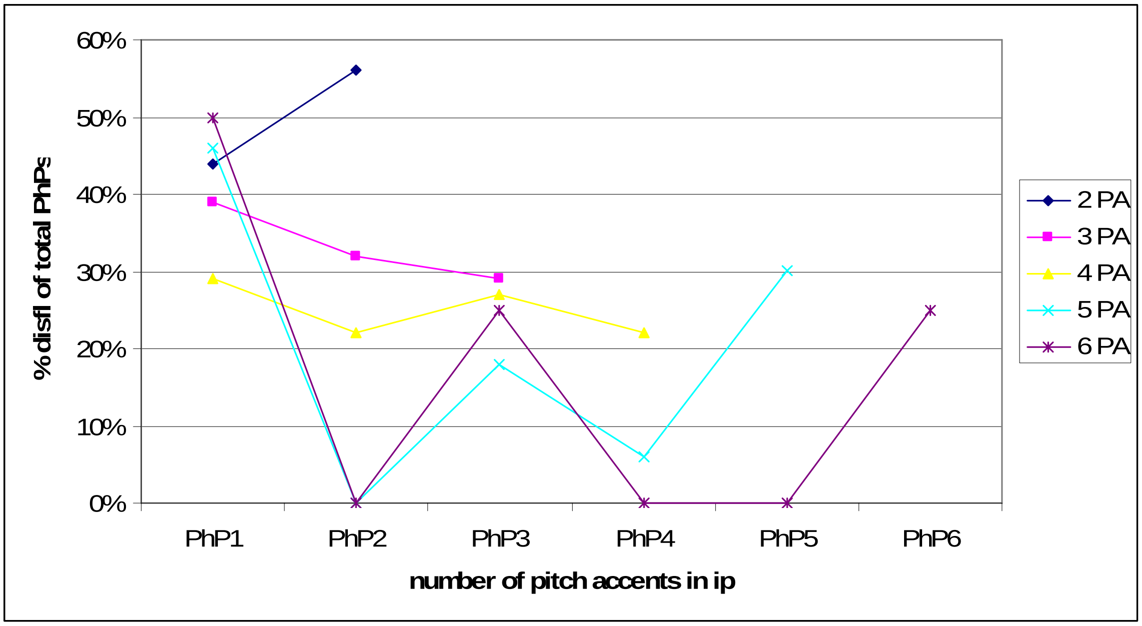


Figure 82: target disfluency rate for Phonological Phrases

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 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 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, 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.

Pairwise comparisons revealed significant differences between the percentages of disfluencies in the first and second PhPs when ips consisted of two PAs ($p=0.004$) and five PAs ($p<0.0001$), and in the first and third PhPs when ips contained three PAs ($p=0.03$) and five PAs ($p=0.001$). Significant differences also were found for 5-PA ips between disfluency percentages of the first and fourth PhPs ($p<0.0001$), second and third PhPs ($p=0.006$), second and fifth PhPs ($p<0.0001$), and fourth and fifth PhPs ($p=0.0009$).[20]

In short, an analysis of target disfluency distribution within ips divided into Phonological Phrases reveals a more nuanced picture. While the coarser analysis illustrated by Figures 79-81 showed that the ip's actual initial position—i.e., the onset of the ip—was the location of few disfluencies overall, 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.

[20] Because tokens for 6-PA ips were very few in number, statistical tests were not performed on these comparisons.

5.2.6 Syllable location

The final condition analyzed was the location of the disfluency within the syllable. Unlike the other four conditions, syllable location refers not to the disfluency target but rather to the location of the disfluency itself. In later comparisons, however, this locational information will be used in conjunction with other conditions to draw more detailed conclusions regarding the behavior of the target.

Analyzing syllable location of disfluencies does not directly lend itself to a target vs. natural comparison, since this is the only condition that requires that a disfluency be present. In other words, all words—natural disfluencies and fluent targets alike—embody some tonal type, prominence level, word class, and phrasal position, but syllable location refers solely to the locus of a disfluency with respect to a syllable's constituents.

Group comparisons revealed a statistically significant difference ($p=0.0003$) between stutterers' and controls' disfluency production with respect to location within the syllable. As Figure 83 shows, 76.3% of control group disfluencies were realized on syllable onsets; for the stuttering group, meanwhile, only 61.7% of disfluencies occurred on syllable onsets. Group behavior was consistent among subjects S2 and S4, who both produced fewer than 60% of their disfluencies on syllable onsets. S1, however, behaved more similarly to the three control subjects, generating 77.7% of his disfluencies in onset position. Figures 84 and 85 display the individual comparisons for stutterers and controls, respectively:

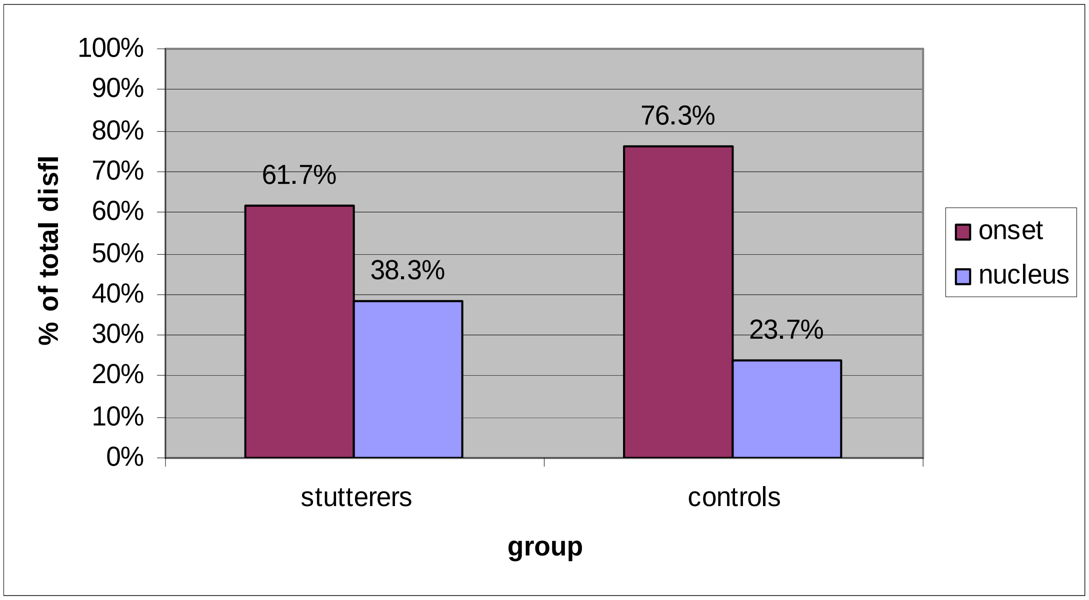


Figure 83: disfluency location within the syllable: stutterers vs. controls

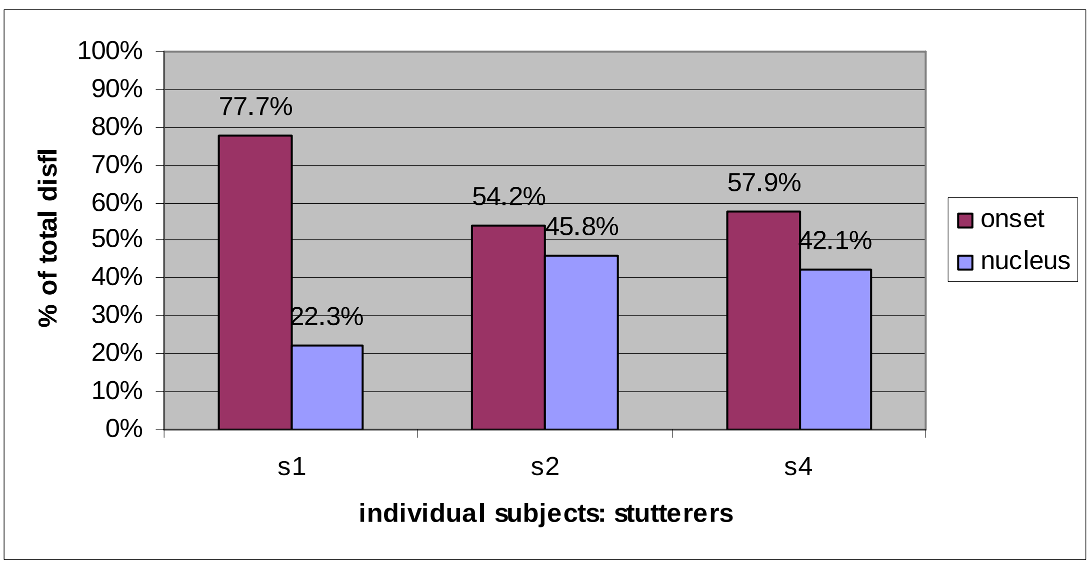


Figure 84: disfluency location within the syllable: individual stutterers

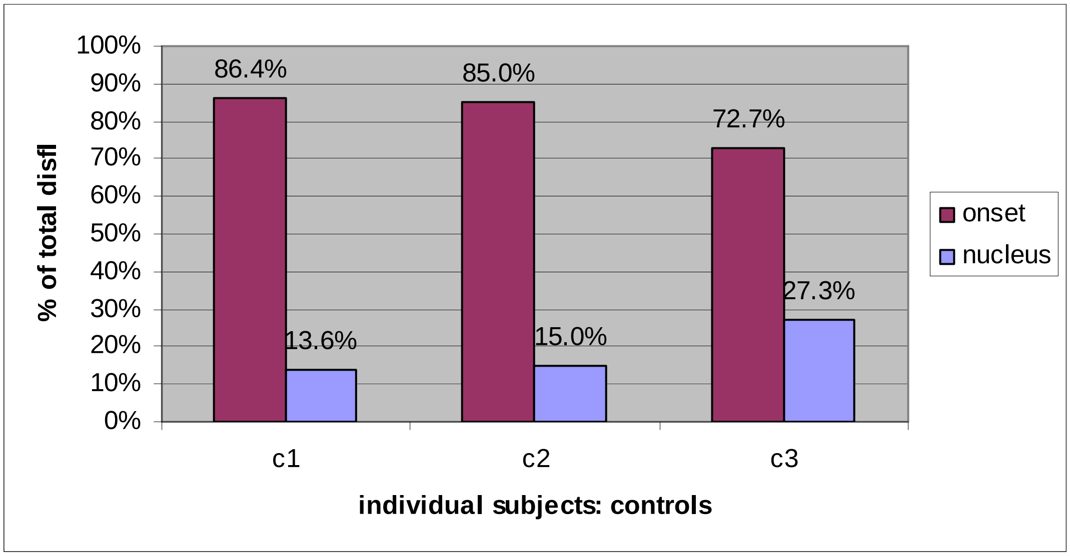


Figure 85: disfluency location within the syllable: individual controls

Adding a second condition—prominence level—to the comparisons of disfluency rate and syllable location, uncovers a subtle trend distinguishing stuttered from control disfluency patterns. Looking first at all subjects combined for each group, the disfluency rates at onset position for both PA (stutterers: 23.2%; controls: 23.4%) and NPA forms (stutterers: 14.8%; controls: 15.2%) were extremely similar for both groups. The largest difference ($p<0.0001$) 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%:

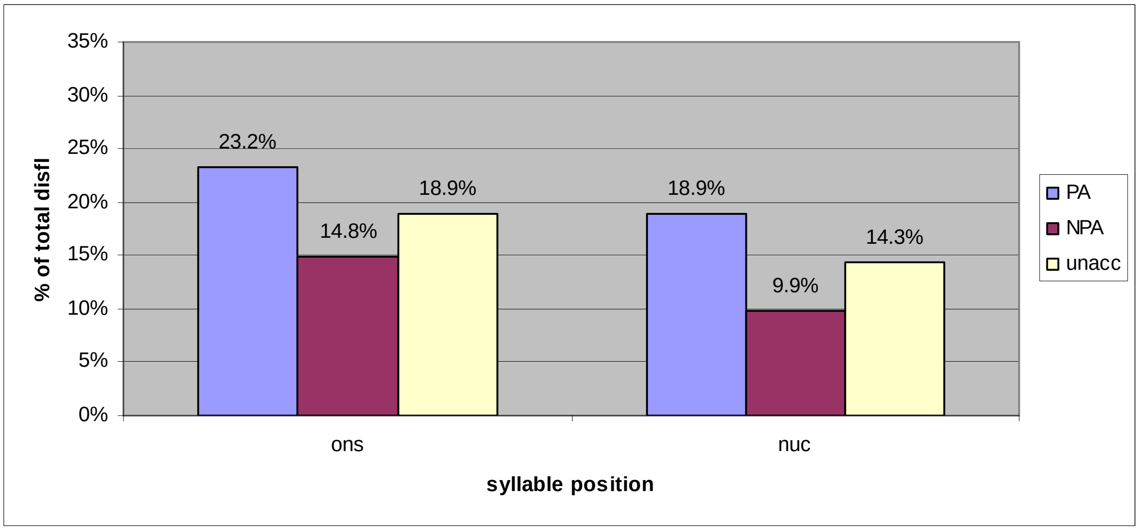


Figure 86: prominence vs. syllable position: all stutterers (3)

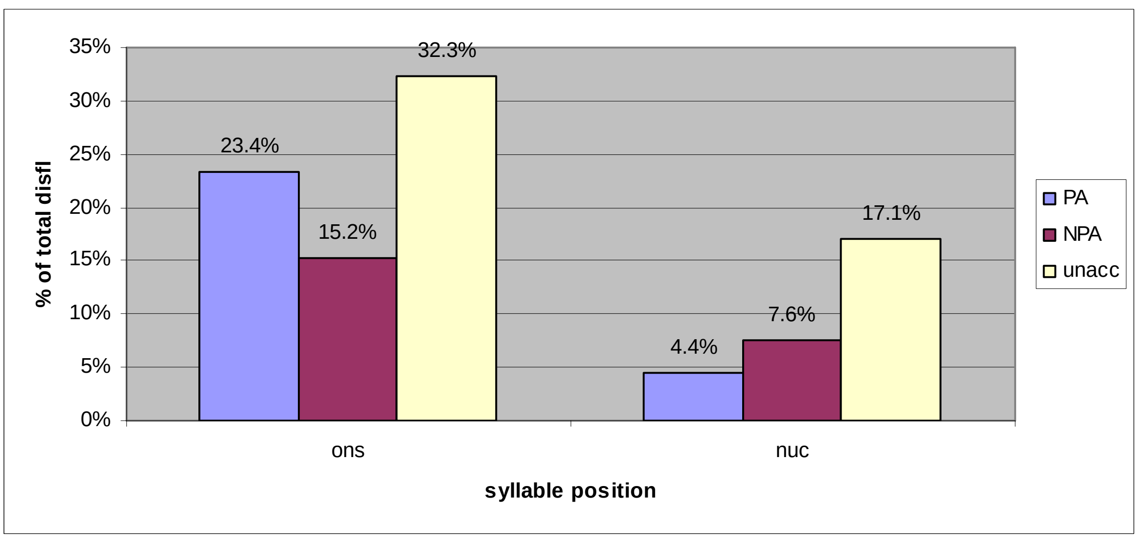


Figure 87: prominence vs. syllable position: all control subjects (3)

The disfluency rate difference between NPA'd nuclei (9.9% vs. 7.6%) was also significant (p=0.00015).

Individual differences were slightly more variable, though this was mostly a result of the low relative rates of disfluencies produced by Speaker S1 for PAs and NPAs in nuclear position, and higher rate of disfluencies in S1's unaccented onsets. While combined PA + NPA nucleus disfluency rates for S2 (37.2%) and S4 (32%) each reached much higher levels than for C1 (11.5%), C2 (15.0%), and C4 (11.6%), S1's combined rate (13.5%) was much closer to that of the controls. Among the control subjects, C1 and C2 produced relatively low disfluency rates on all unaccented syllable positions, while C4 demonstrated the opposite tendency in producing the majority of his disfluencies on both onsets (42%) and nuclei (21.4%) of unaccented syllables.

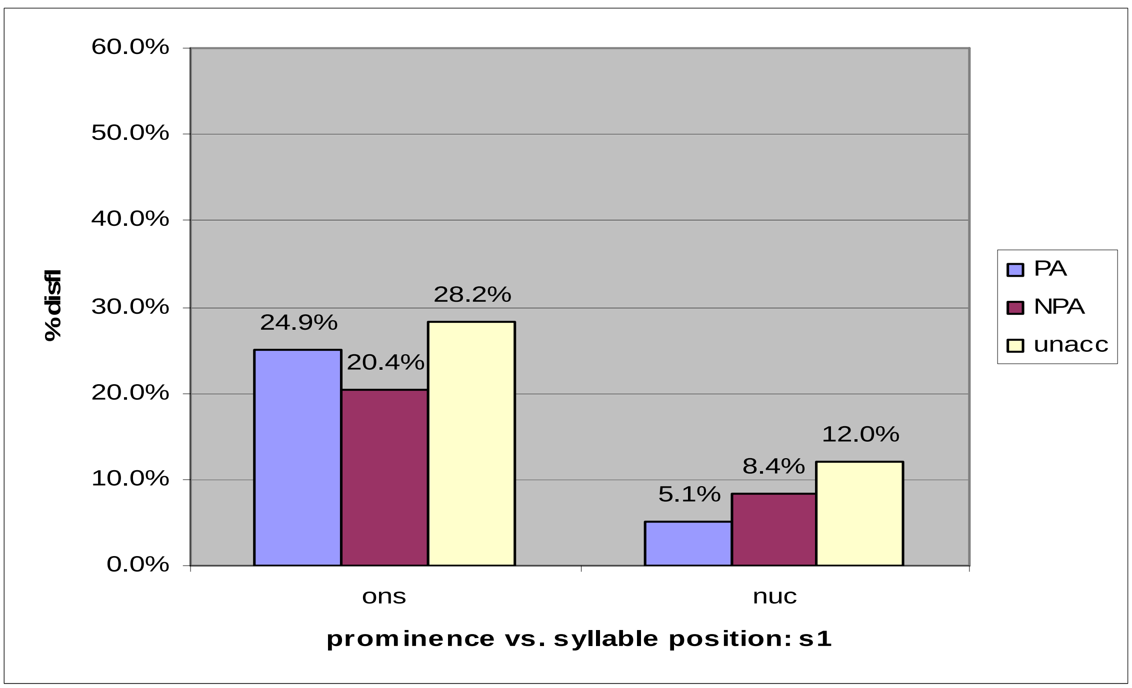


Figure 88: prominence vs. syllable position (S1)

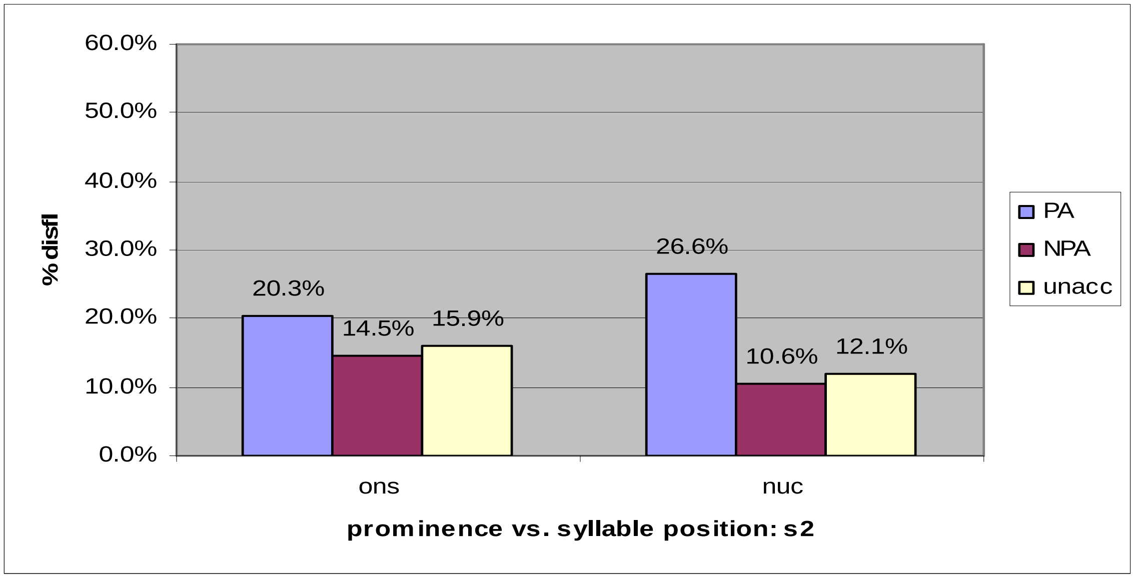


Figure 89: prominence vs. syllable position (S2)

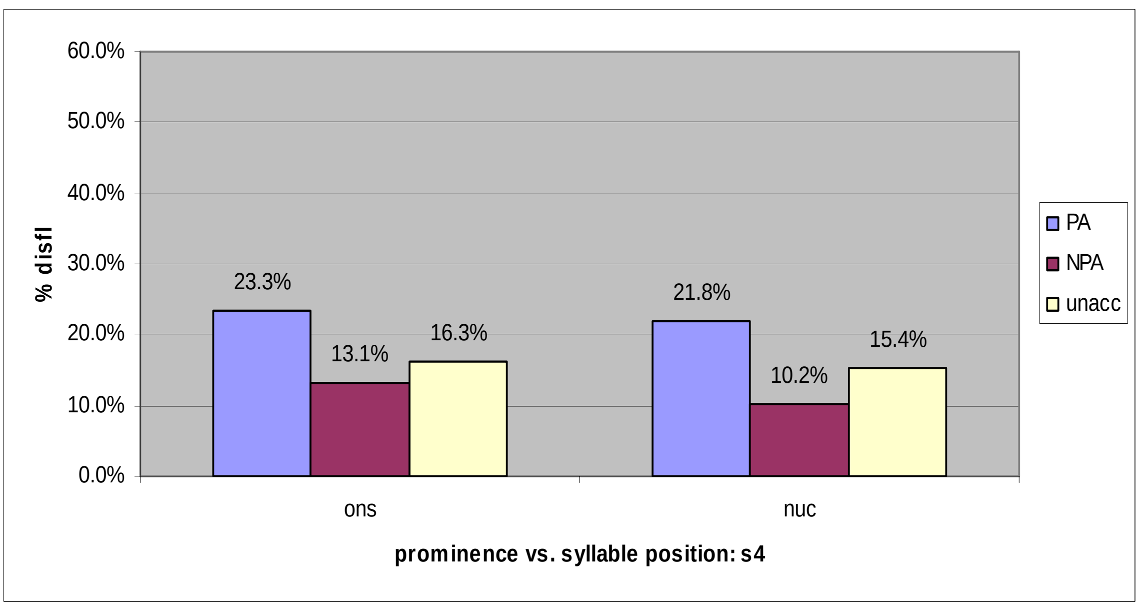


Figure 90: prominence vs. syllable position (S4)

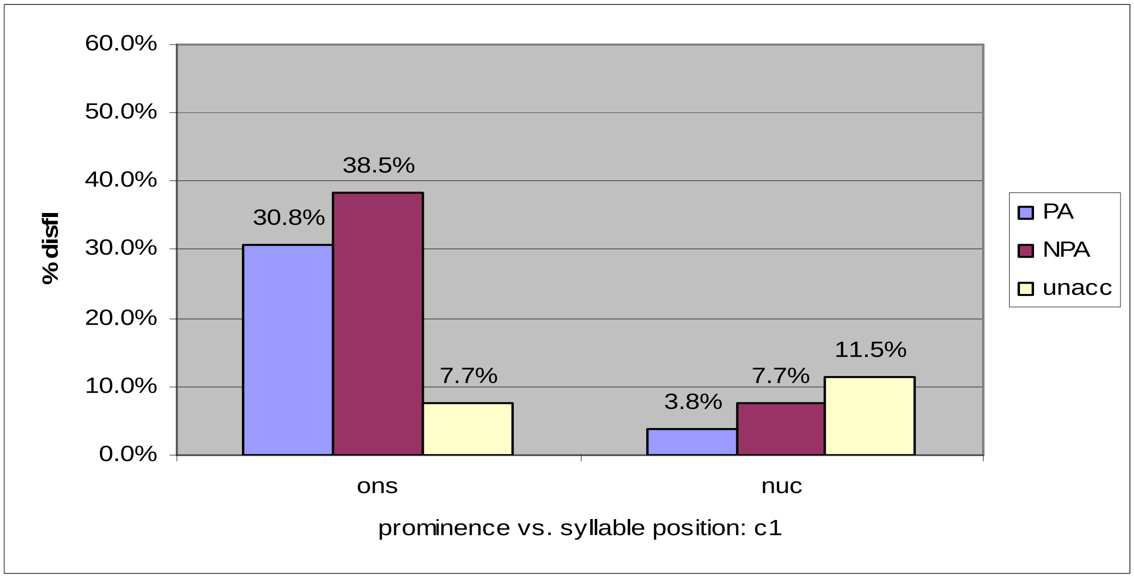


Figure 91: prominence vs. syllable position (C1)

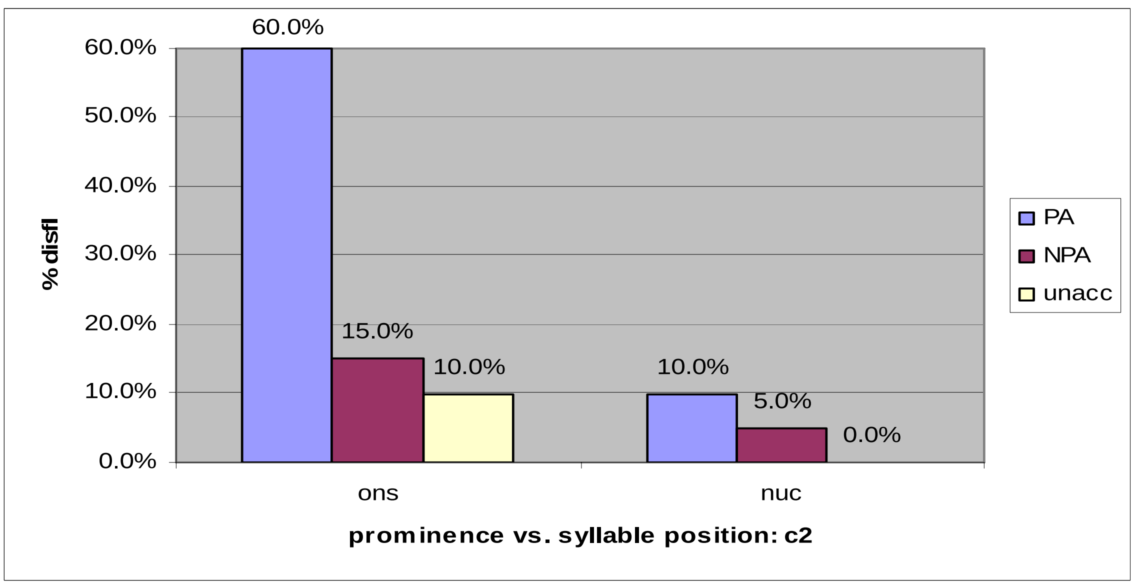


Figure 92: prominence vs. syllable position (C2)

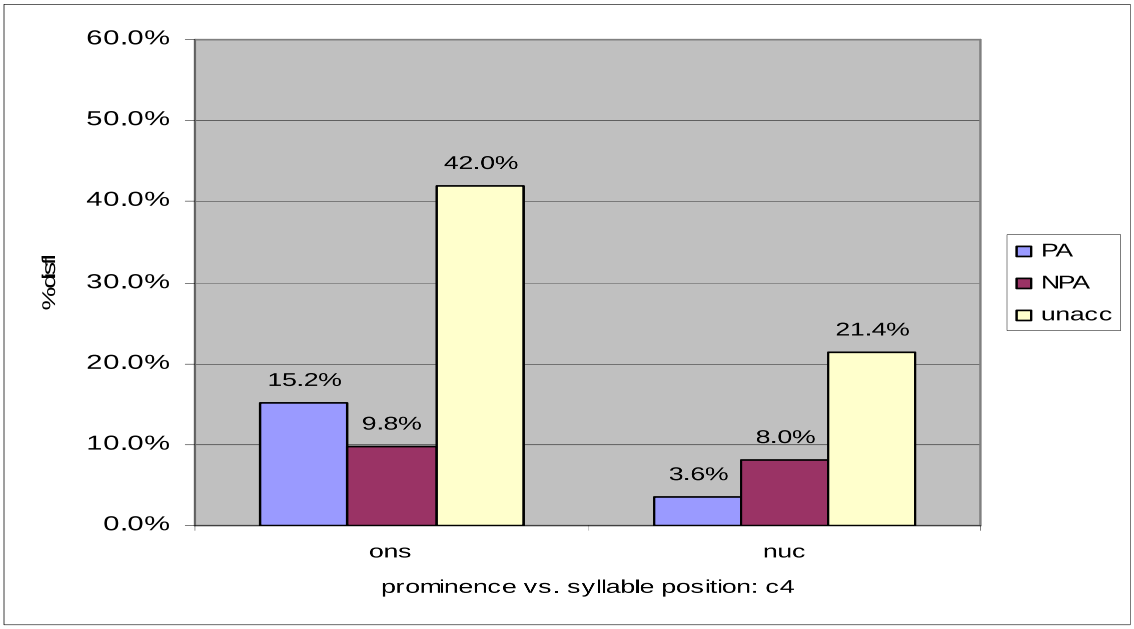


Figure 93: prominence vs. syllable position (C4)

5.3 Prosodic conditions and disfluency types of target

In this section, like in section 5.2, comparisons will again be made between the prosodic conditions in which disfluencies from the narrative data surfaced, and the prosodic conditions established by the reference data readings. The purpose here, however, is to examine whether any specific disfluency types patterned with any of the prosodic conditions of the target word.

5.3.1 Tonal type

Speaker 1 (S1) displayed a relatively wide distribution of disfluency types with respect to the full range of target pitch accent tones. As noted in the general tonal analysis in section 5.2.2, L+H* and H* were the two most common pitch accent tone types for all three stutterers. Table 8 below shows the complete pattern of disfluency type and target tone distribution for S1:

S1 Disfl \ PA tone	H*	L*	L+H*	L*+H	!H*	H+!H*	L+!H*	L*+!H
Disfluency	50	9	104	2	17	0	8	1
%	**26.2%**	4.7%	**54.5%**	1.0%	8.9%	0%	4.2%	0.5%
restart	0	0	2	0	0	0	0	0
%	0%	0%	**100%**	0%	0%	0%	0%	0%
prolong	13	3	28	1	2	0	2	0
%	**26.5%**	6.1%	**57.1%**	2.0%	4.1%	0%	4.1%	0%
cut	6	0	10	0	0	0	2	0
%	**33.3%**	0%	**55.6%**	0%	0%	0%	11.1%	0%
pause	15	4	35	0	8	0	2	1
%	**23.1%**	6.2%	**53.8%**	0%	12.3%	0%	3.1%	1.5%
filler	1	0	2	1	0	0	0	0
%	**25.0%**	0%	**50%**	25.0%	0%	0%	0%	0%
comb	15	2	27	0	7	0	2	0
%	**28.3%**	3.8%	**50.9%**	0%	13.2%	0%	3.8%	0%

Table 14: Distribution of disfluency types for each targeted pitch accent type: s1

While some disfluency types such as prolongations and pauses occurred with greater frequency overall than others, the distribution of all types nevertheless clustered predictably around L+H* and H*—the two most common tones both for disfluencies and the natural data as a whole, as was observed in Chapter 4. Figure 94 shows S1's distribution patterns by percentage for all target disfluency types:

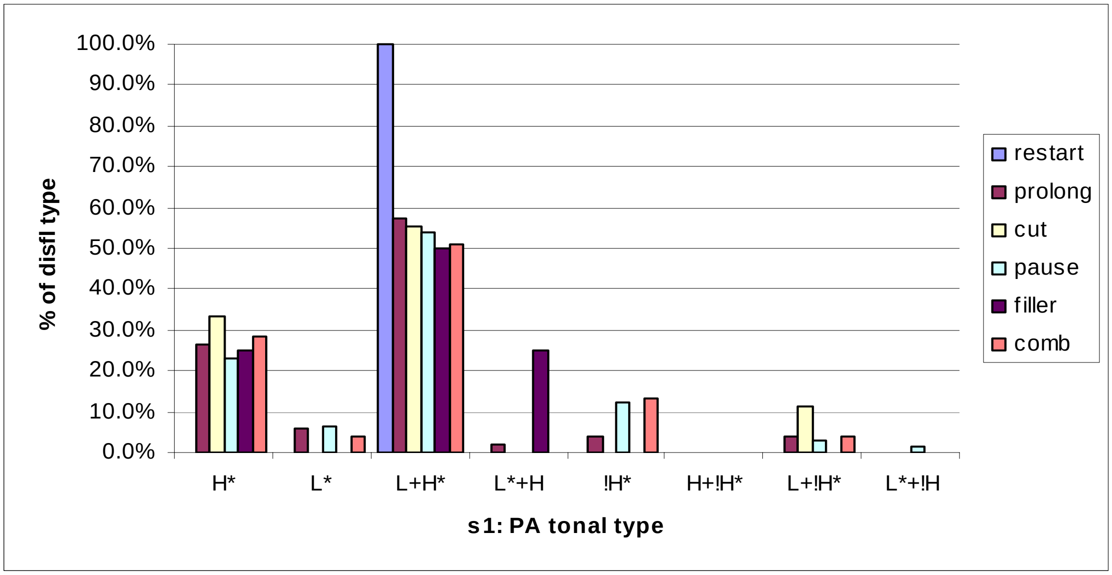


Figure 94: Distribution of disfluency types in percentage for each targeted pitch accent type: s1

Minor patterns surfaced for !H*, as 15 of its 17 disfluent occurrences were only in pauses and combinations. High rates of L+H* tones on targets of restarts, as well as L*+H on targets of fillers, stand out from the other disfluency types for these respective tones. However, the very low number of fillers and restarts produced overall by S1 inflate these proportions: e.g., S1 produced only two restarts, both of which were realized in the target data as L+H*.

Quite similar distribution patterns were found for Speaker 2 (S2), where all disfluency types clustered around target pitch accents H*, L+H*, and, to a lesser degree, !H*. Distribution patterns for S2's target pitch accent types are displayed in Table 15:

S2 Disfl \ PA tone	H*	L*	L+H*	L*+H	!H*	H+!H*	L+!H*	L*+!H
Disfluency	16	7	42	2	16	0	8	0
%	**17.6%**	7.7%	**46.2%**	2.2%	**17.6%**	0%	8.8%	0%
restart	0	0	0	1	0	0	0	0
%	0%	0%	0%	100%	0%	0%	0%	0%
prolong	5	3	9	1	1	0	2	0
%	**23.8%**	14.3%	**42.9%**	4.8%	4.8%	0%	9.5%	0%
cut	1	2	4	0	5	0	2	0
%	7.1%	14.3%	**28.6%**	0%	**35.7%**	0%	14.3%	0%
pause	6	2	8	0	4	0	2	0
%	**27.3%**	9.1%	**36.4%**	0%	**18.2%**	0%	9.1%	0%
filler	0	0	1	0	0	0	0	0
%	0%	0%	**100%**	0%	0%	0%	0%	0%
comb	4	0	20	0	6	0	2	0
%	**12.5%**	0%	**62.5%**	0%	**18.8%**	0%	6.3%	0%

Table 15: Distribution of disfluency types for each targeted pitch accent type: s2

As reported in the previous chapter, S2 produced fewer disfluencies overall than did S1. The disfluency type distribution pattern among target pitch accent tones, however, was quite similar. Slightly less uniformity was found in the L+H* and H* clusters: for example, a smaller proportion of S2's restarts (S2=0%; S1=100%) and cuts (S2=28.6%; S1=55.6%) co-occurred with L+H*, while fewer of S2's cuts (S2=7.1%; S1=33.3%) and fillers (S2=0%; S1=25%) were produced with H* pitch accents. Figure 95 shows the disfluency type distribution for S2's target pitch accent types in percentage:

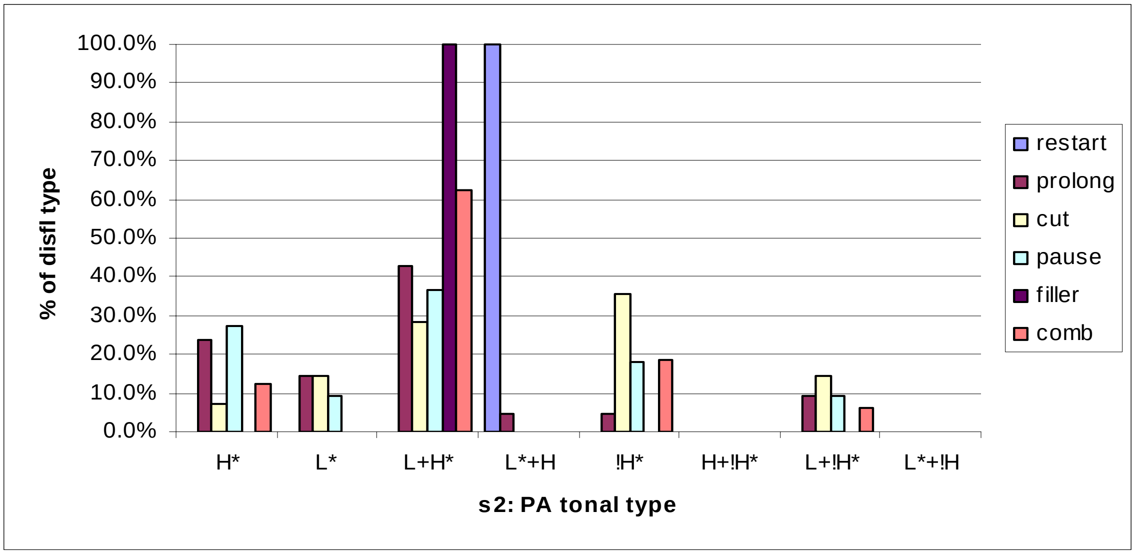


Figure 95: Distribution of disfluency types in percentage for each targeted pitch accent type: s2

Finally, Speaker 4 (S4) followed a highly similar pattern as the other speakers, with the same concentration of all disfluency types on H*, L+H*, and !H*. Table 15 and Figure 96 display the distribution patterns for S4's target pitch accent types.

S4 Disfl \ PA tone	H*	L*	L+H*	L*+H	!H*	H+!H*	L+!H*	L*+!H
Disfluency	72	12	85	3	49	1	8	2
%	**31.0%**	5.2%	**36.6%**	1.3%	**21.1%**	0.4%	3.4%	0.9%
restart	8	0	3	0	1	0	0	0
%	**66.7%**	0%	**25.0%**	0%	8.3%	0%	0%	0%
prolong	12	3	14	0	5	0	1	0
%	**34.3%**	8.6%	**40%**	0%	**14.3%**	0%	2.9%	0%
cut	7	1	11	0	2	0	1	0
%	**31.8%**	4.5%	**50%**	0%	9.1%	0%	4.5%	0%
pause	6	1	8	1	8	0	2	0
%	**23.1%**	3.8%	**30.8%**	3.8%	**30.8%**	0%	7.7%	0%
filler	5	1	11	1	2	0	1	0

%	**23.8%**	4.8%	**52.4%**	4.8%	9.5%	0%	4.8%	0%
comb	34	6	38	1	31	1	3	2
%	**29.3%**	5.2%	**32.8%**	0.9%	**26.7%**	0.9%	2.6%	1.7%

Table 16: Distribution of disfluency types for each targeted pitch accent type: s4

Disfluencies were distributed more evenly among the three most frequent tonal types—H*, L+H*, and !H*—particularly for pauses and combinations. Restarts were also represented more frequently among H* (66.7%) than for the other two speakers:

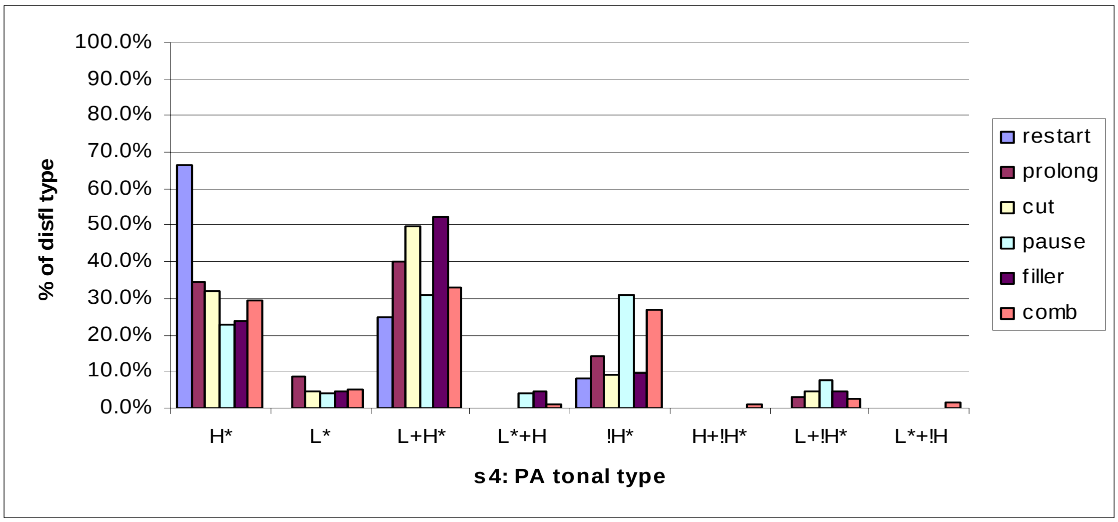


Figure 96: Distribution of disfluency types in percentage for each targeted pitch accent type: s4

An analysis of target phrasal tones reveals yet another disfluency pattern representative of the normal tonal distribution. In Chapter 4 it was found that L- tones were the most frequently occurring phrase accents in pre-disfluent, disfluent, and non-disfluent (i.e., normal) speech. This pattern was upheld for disfluency type distribution, as well, as seen in Table 17 beginning with S1:

S1 Disfl \ phrasal tone	L-	H-	!H-	L%	H%
Disfluency	48	7	9	56.3%	43.8%
%	**75.0%**	10.9%	14.1%	1	0
restart	1	0	0	100%	0%
%	**100%**	0%	0%	3	8
prolong	10	2	1	27.3%	**72.7%**
%	**76.9%**	15.4%	7.7%	1	3
cut	5	1	0	25.0%	**75.0%**
%	**83.3%**	16.7%	0%	13	4
pause	19	2	4	**76.5%**	23.5%
%	**76.0%**	8.0%	16.0%	0	0
filler	0	0	0	0%	0%
%	0%	0%	0%	9	6
comb	13	2	4	**60%**	40%
%	**68.4%**	10.5%	21.1%	56.3%	43.8%

Table 17: Distribution of disfluency types for each targeted phrasal tone type: s1

Five of the six disfluency types (fillers were not produced at any phrase boundaries for this speaker) were produced most frequently before target L- phrase accents, while no higher than 16% of any disfluency type was found before either targeted H- or !H-. Figure 97 shows the data in percentage.

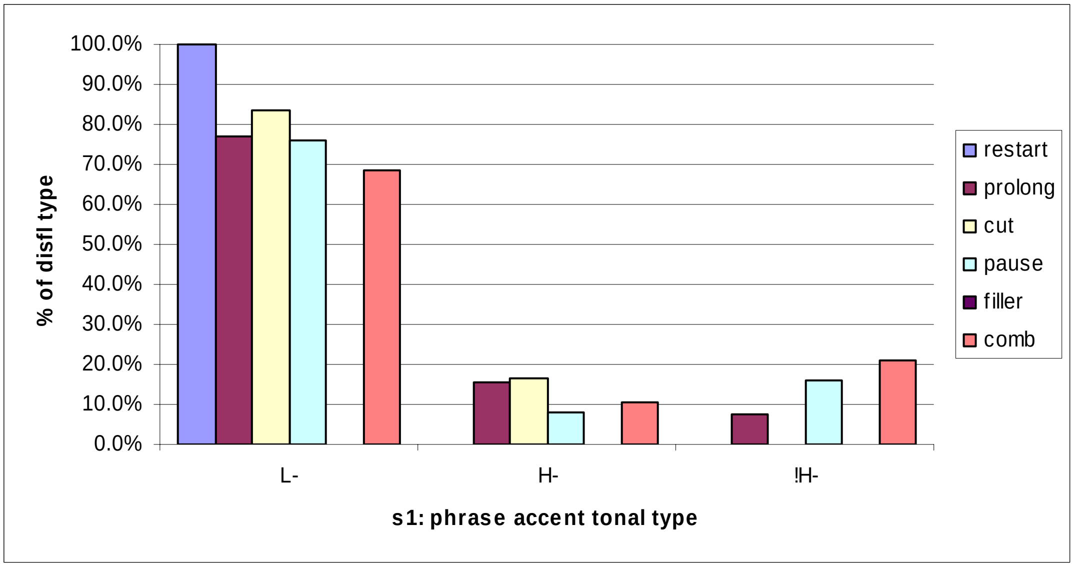


Figure 97: Distribution of disfluency types in percentage for each targeted phrasal tone type: s1

As clearly seen in Table 17, target boundary tones did not pattern with all disfluency types evenly. While prolongations (72.7%) and cuts (75%) occurred more often with target H% boundary tones, pauses and combinations were produced with L% tones more frequently (76.5% and 60%, respectively).

Disfluency types were somewhat less predictably distributed for Speaker 2, as shown in Table 18 and Figure 98. S2 produced markedly fewer prolongations (S2=25%; S1=76.9%) and restarts (S2=0%; S1=100%) before target L- phrase accents compared with Speaker 1:

S2 Disfl \ phrasal tone	L-	H-	!H-	L%	H%
Disfluency	16	6	2	14	2
%	66.7%	25.0%	8.3%	**87.5%**	12.5%

restart	0	0	0	0	0
%	**0.0%**	0.0%	0.0%	0.0%	0.0%
prolong	2	4	2	5	0
%	**25.0%**	50.0%	25.0%	**100.0%**	0.0%
cut	3	1	0	2	0
%	75.0%	25.0%	0.0%	**100.0%**	0.0%
pause	2	0	0	2	0
%	100.0%	0.0%	0.0%	**100.0%**	0.0%
filler	0	0	0	0	0
%	0.0%	0.0%	0.0%	0.0%	0.0%
comb	9	1	0	5	2
%	90.0%	10.0%	0.0%	**71.4%**	28.6%

Table 18: Distribution of disfluency types for each targeted phrasal tone type: s2

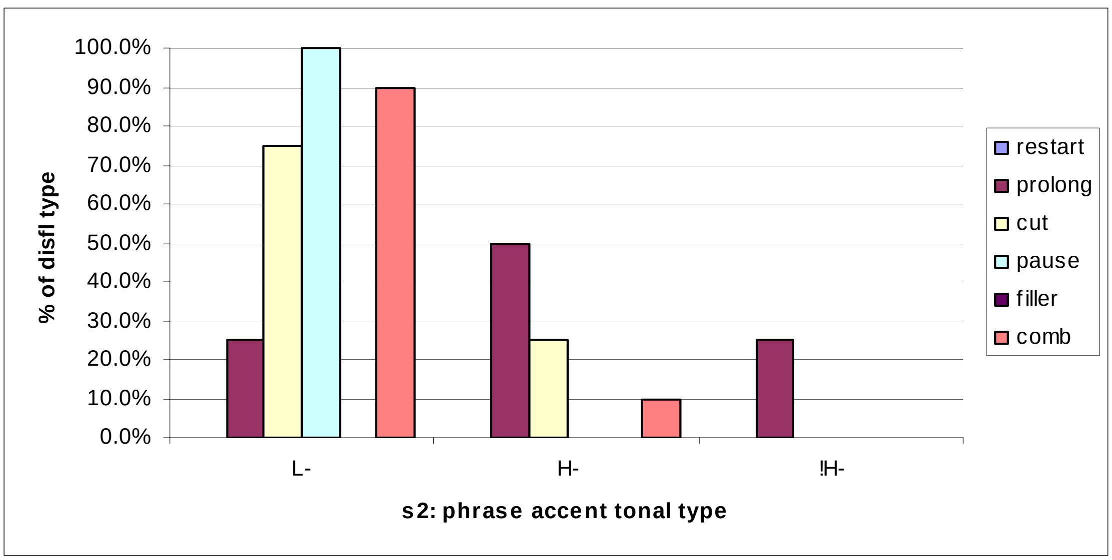


Figure 98: Distribution of disfluency types in percentage for each targeted phrasal tone type: s2

Target boundary tones, on the other hand, differed from those produced by Speaker 1 in that almost all disfluency types patterned exclusively with L%—only combination disfluencies were attested before H%.

Speaker 4 shared with Speaker 1 a more restricted patterning for phrase accents, and, with Speaker 2, a similar clustering before L%. Unlike S1, however, S4 produced no disfluencies before targeted H- phrase accents, as shown in Table 19 and Figure 99:

S4 Disfl \ phrasal tone	L-	H-	!H-	L%	H%
Disfluency	126	0	24	108	3
%	84.0%	**0%**	16.0%	**97.3%**	2.7%
restart	9	0	0	9	0
%	100%	**0%**	0%	**100%**	0%
prolong	21	0	6	12	0
%	77.8%	**0%**	22.2%	**100%**	0%
cut	18	0	3	9	0
%	85.7%	**0%**	14.3%	**100%**	0%
pause	12	0	6	12	0
%	66.7%	**0%**	33.3%	**100%**	0%
filler	6	0	6	12	0
%	50%	**0%**	50%	**100%**	0%
comb	60	0	3	54	3
%	95.2%	**0%**	4.8%	**94.7%**	5.3%

Table 19: Distribution of disfluency types for each targeted phrasal tone type: s4

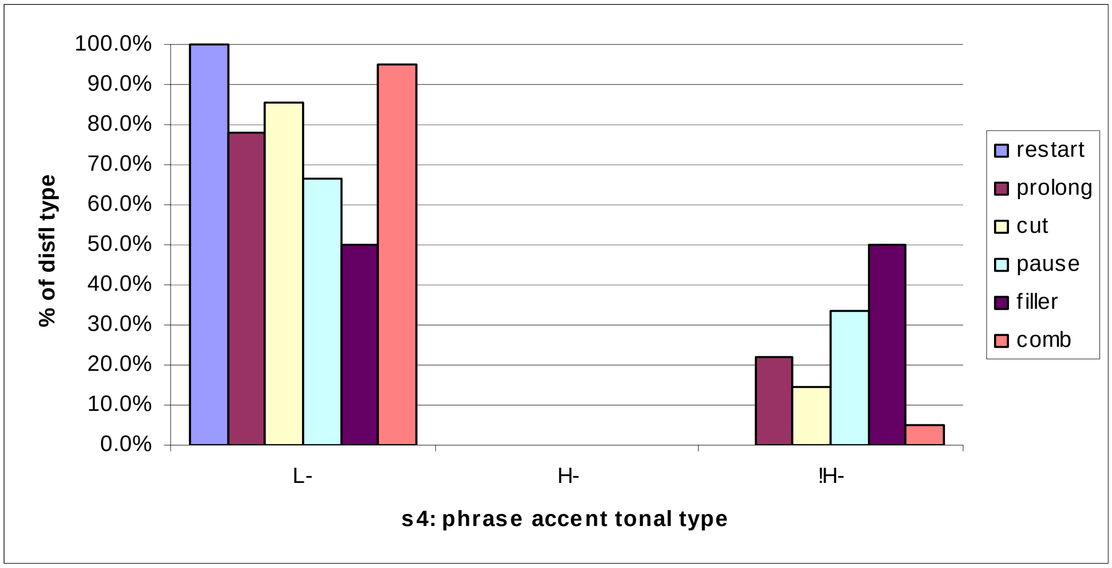


Figure 99: Distribution of disfluency types in percentage for each targeted phrasal tone type: s4

The final condition analyzed was the patterning of disfluency types with targeted phrasal tone contours—i.e., the final contours produced by phrase accents followed by boundary tones. Consistent with the distribution frequencies of contours found in speakers' overall data, L-L% was the most common target boundary type for all speakers. Speaker 1 also produced a high percentage of disfluency types before target L-H% (40.4%), with cuts (75%), prolongations (60%), and combinations (40%) occurring most often before this target contour. Table 20 and Figure 100 show the data.

S1 Disfl \ boundary contour	L-L%	L-H%	H-L%	H-H%	!H-L%	!H-H%
Disfluency	22	19	1	1	4	0
%	**46.8%**	**40.4%**	2.1%	2.1%	8.5%	0%
restart	1	0	0	0	0	0
%	**100%**	0%	0%	0%	0%	0%
prolong	3	6	0	1	0	0

%	**30%**	**60%**	0%	10%	0%	0%
cut	1	3	0	0	0	0
%	**25.0%**	**75.0%**	0%	0%	0%	0%
pause	12	4	0	0	1	0
%	**70.6%**	23.5%	0%	0%	5.9%	0%
filler	0	0	0	0	0	0
%	0%	0%	0%	0%	0%	0%
comb	5	6	1	0	3	0
%	**33.3%**	**40%**	6.7%	0%	20%	0%

Table 20: Distribution of disfluency types for each targeted boundary tone contour type: s1

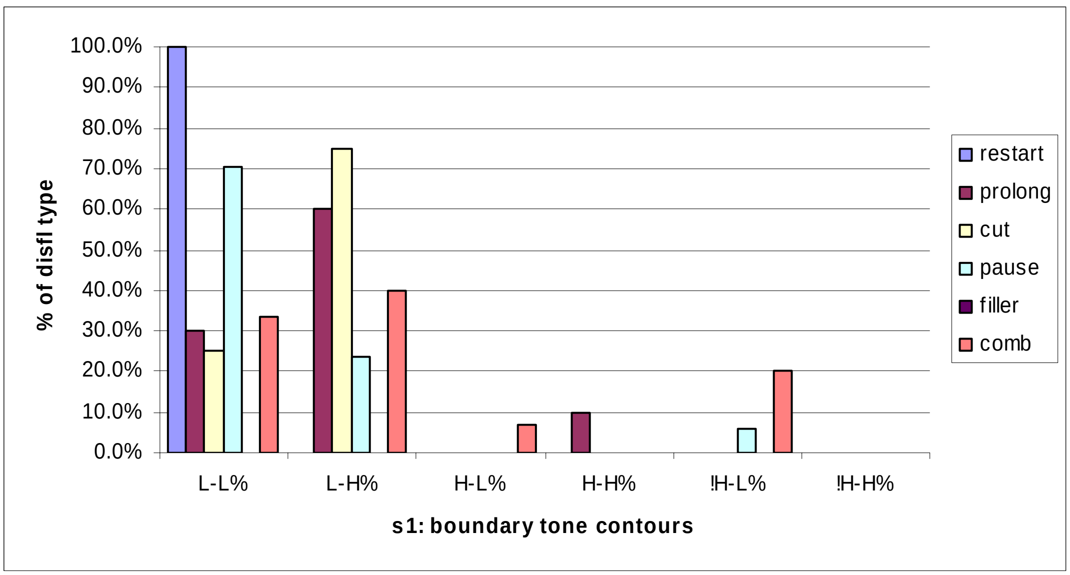


Figure 100: Distribution of disfluency types in percentage for each targeted boundary tone contour type: s1

Speaker 2 shared with Speaker 1 both the fact that L-L% was the most common tonal contour target of the majority of disfluency types, and that pauses were more frequent than prolongations when L-L% was the targeted tone contour. Unlike S1, however, S2's target disfluencies occurred before L-L% (61.1%) contours much more often than before L-H% (11.1%), as shown in Table 21 and Figure 101.

S2 Disfl \ boundary contour	L-L%	L-H%	H-L%	H-H%	!H-L%	!H-H%
Disfluency	11	2	4	0	1	0
%	**61.1%**	**11.1%**	22.2%	0%	5.6%	0%
restart	0	0	0	0	0	0
%	0%	0%	0%	0%	0%	0%
prolong	2	0	2	0	1	0
%	**40%**	0%	40%	0%	20%	0%
cut	2	0	1	0	0	0
%	**66.7%**	0%	33.3%	0%	0%	0%
pause	2	0	0	0	0	0
%	**100%**	0%	0%	0%	0%	0%
filler	0	0	0	0	0	0
%	0%	0%	0%	0%	0%	0%
comb	5	2	1	0	0	0
%	**62.5%**	25.0%	12.5%	0%	0%	0%

Table 21: Distribution of disfluency types for each targeted boundary tone contour type: s2

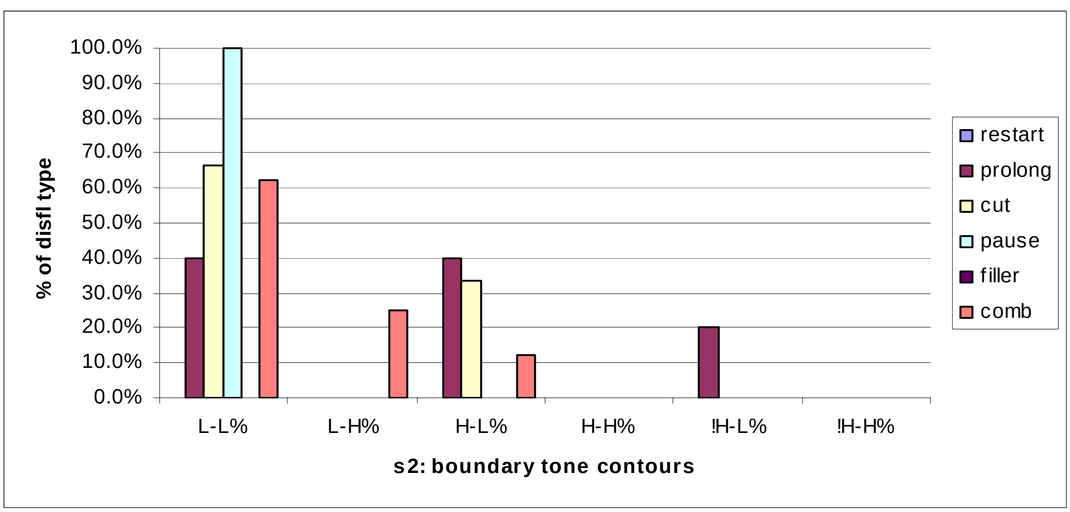


Figure 101: Distribution of disfluency types in percentage for each targeted boundary tone contour type: s2

Finally, as seen in Table 22 and Figure 102, Speaker 4 supported the majority pattern once again by producing all disfluency types most often before targeted L-L% contours (78.4%). In addition, like for S2, considerably fewer L-H% than L-L% contours occurred in target position for S4 (2.7%). S4 also produced a moderate number of prolongations, cuts, pauses, and fillers before !H-L% contours:

S4 Disfl \ boundary contour	L-L%	L-H%	H-L%	H-H%	!H-L%	!H-H%
Disfluency	87	3	0	0	21	0
%	**78.4%**	**2.7%**	0%	0%	18.9%	0%
restart	9	0	0	0	0	0
%	**100%**	0%	0%	0%	0%	0%
prolong	6	0	0	0	6	0
%	**50%**	0%	0%	0%	**50%**	0%
cut	6	0	0	0	3	0
%	**66.7%**	0%	0%	0%	33.3%	0%
pause	6	0	0	0	6	0
%	**50%**	0%	0%	0%	**50%**	0%
filler	6	0	0	0	6	0
%	**50%**	0%	0%	0%	**50%**	0%
comb	54	3	0	0	0	0
%	**94.7%**	5.3%	0%	0%	0%	0%

Table 22: Distribution of disfluency types for each targeted boundary tone contour type: s4

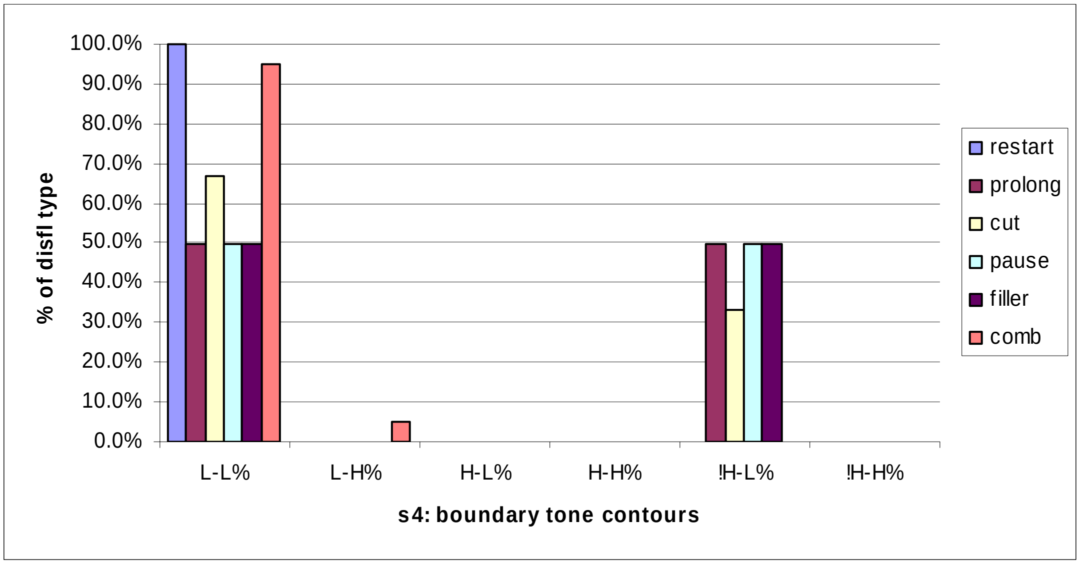


Figure 102: Distribution of disfluency types in percentage for each targeted boundary tone contour type: s4

In summary, no consistent disfluency-type patterns surfaced before tonal targets for any of the three stutterers. To the contrary, all disfluency types were almost unfailingly produced in greatest proportion before the most frequent tonal targets, thereby simply reflecting regular tonal distribution patterns. Thus, it appears that tone differences did not independently affect the type of disfluency that surfaced in that environment.

5.3.2 Word class

Targets for word class followed an interesting overall trend for the stutterers. While it was shown in Chapter 4 that the realization of disfluencies for the stuttering group occurred slightly more often on function words (52.7%) than content words (47.3%), words produced in target position were more often content words than function words:

Disfl \ word class	S1		S2		S4	
	Content	function	content	function	content	function
Disfluency	213	71	89	83	720	192
%	**75.0%**	25.0%	**51.7%**	48.3%	**78.9%**	21.1%
restart	2	0	1	0	36	6
%	**100.0%**	0.0%	**100.0%**	0.0%	**85.7%**	14.3%
prolong	53	14	21	36	102	60
%	**79.1%**	20.9%	36.8%	63.2%	**63.0%**	37.0%
cut	23	5	11	9	69	24
%	**82.1%**	17.9%	**55.0%**	45.0%	**74.2%**	25.8%
pause	73	27	24	13	84	42
%	**73.0%**	27.0%	**64.9%**	35.1%	**66.7%**	33.3%
filler	4	3	1	3	63	15
%	**57.1%**	42.9%	25.0%	75.0%	**80.8%**	19.2%
comb	58	22	31	22	366	45
%	**72.5%**	27.5%	**58.5%**	41.5%	**89.1%**	10.9%

Table 23: Distribution of disfluency types for each target word class: all subjects' targets

As Table 23 shows, for speakers S1 and S4, content words were three times more frequent than function words in disfluency target position, while S2 produced a nearly equal number of target content and function words. Similarly, as illustrated by Figures 103-105, all disfluency types for S1 and S4 occurred before more target content than function words, while greater variation in S2's production resulted in more prolongations (63.2%) and fillers (75%) before target function words:

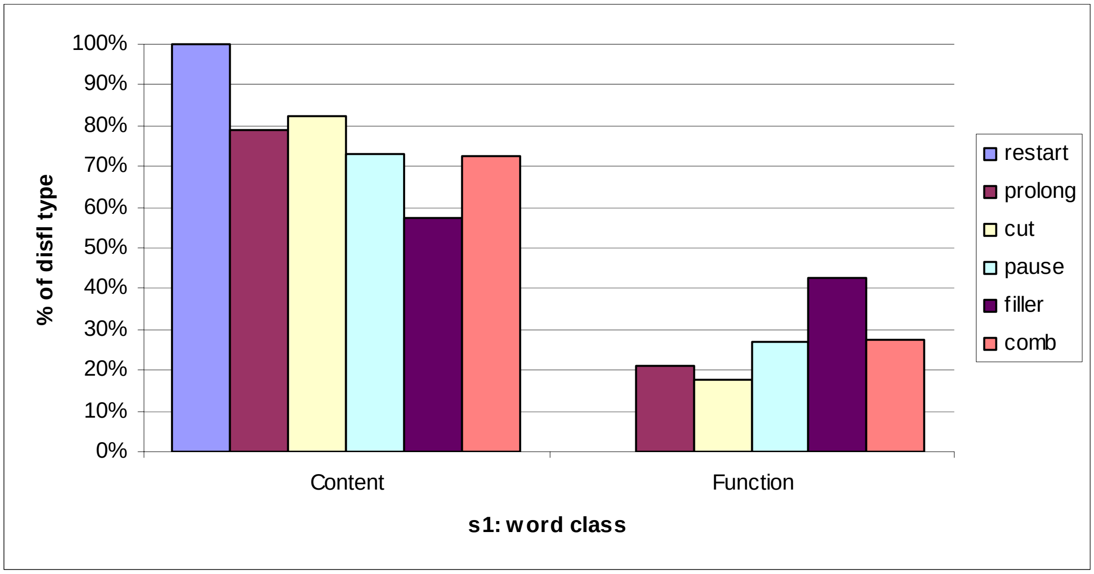


Figure 103: Distribution of disfluency types in percentage for each target word class: s1

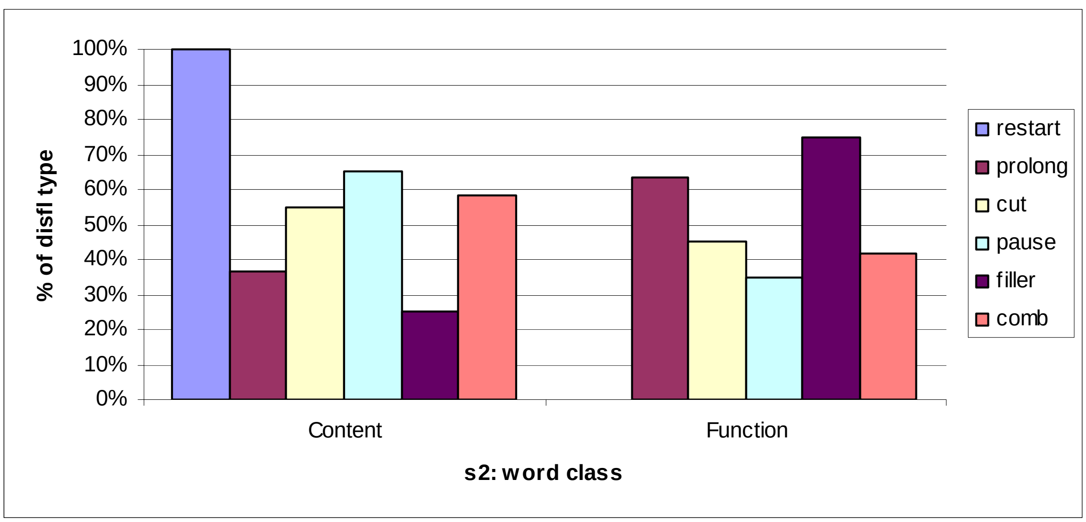


Figure 104: Distribution of disfluency types in percentage for each target word class: s2

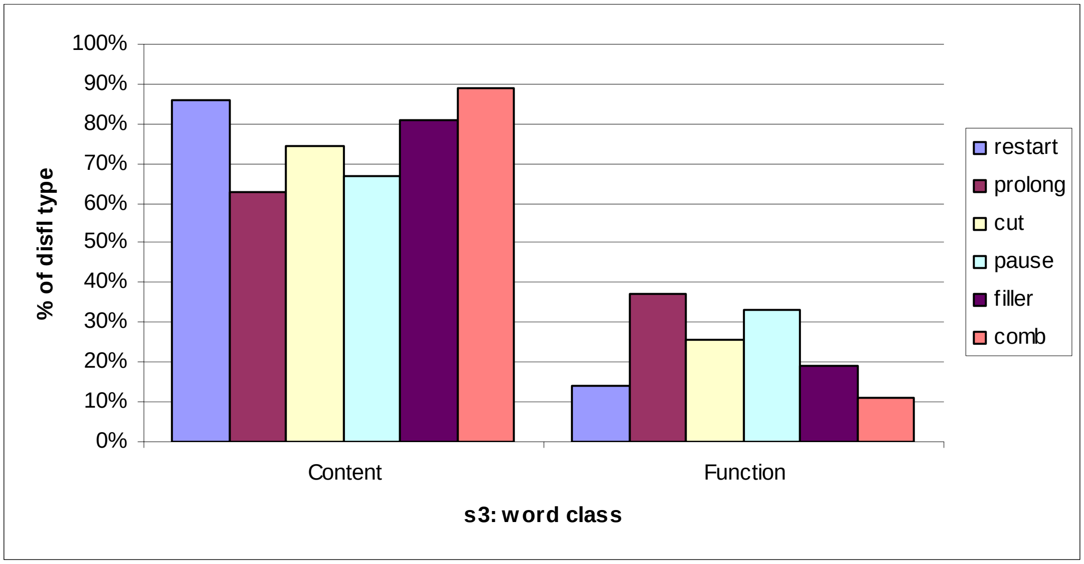


Figure 105: Distribution of disfluency types in percentage for each target word class: s4

The distribution differences for grammatical categories in script (target) and natural data will be looked at in greater detail in section 5.2.

5.3.3 Prominence level

In order to unveil more subtle prominence differences among target productions themselves, comparisons were made using the more detailed prominence classifications proposed earlier in the chapter.

With respect to prominence level of their targets, the distribution of disfluency types again clustered around the most attested target type—in this case, pitch-accented and nuclear pitch-accented targets:

S1 Disfl \ prominence level	PA	NPA	pre-NPA	post-NPA	pre-PA	2-pre-PA	unacc
Disfluency	131	77	17	6	23	1	33
%	**45.5%**	**26.7%**	5.9%	2.1%	8.0%	0.3%	11.5%
restart	1	1	0	0	0	0	0
%	**50%**	**50%**	0%	0%	0%	0%	0%
prolong	35	19	2	1	6	0	5
%	**51.5%**	**27.9%**	2.9%	1.5%	8.8%	0%	7.4%
cut	14	7	1	1	2	0	4
%	**48.3%**	**24.1%**	3.4%	3.4%	6.9%	0%	13.8%
pause	39	31	8	2	7	1	13
%	**38.6%**	**30.7%**	7.9%	2.0%	6.9%	1.0%	12.9%
filler	3	1	2	0	1	0	0
%	**42.9%**	14.3%	**28.6%**	0%	14.3%	0%	0%
comb	39	18	4	2	7	0	11
%	**48.1%**	**22.2%**	4.9%	2.5%	8.6%	0%	13.6%

Table 24: Distribution of disfluency types for each target prominence type: s1

One slight trend appeared in the behavior of fillers, which, despite their low overall frequency, occurred in pre-NPA position in two instances. This is consistent with the fact that fillers are nonlexical prosodic forms, and therefore have no semantic or syntactic restrictions on their distribution. No other robust patterns surfaced for Speaker S1, as illustrated in Figure 106:

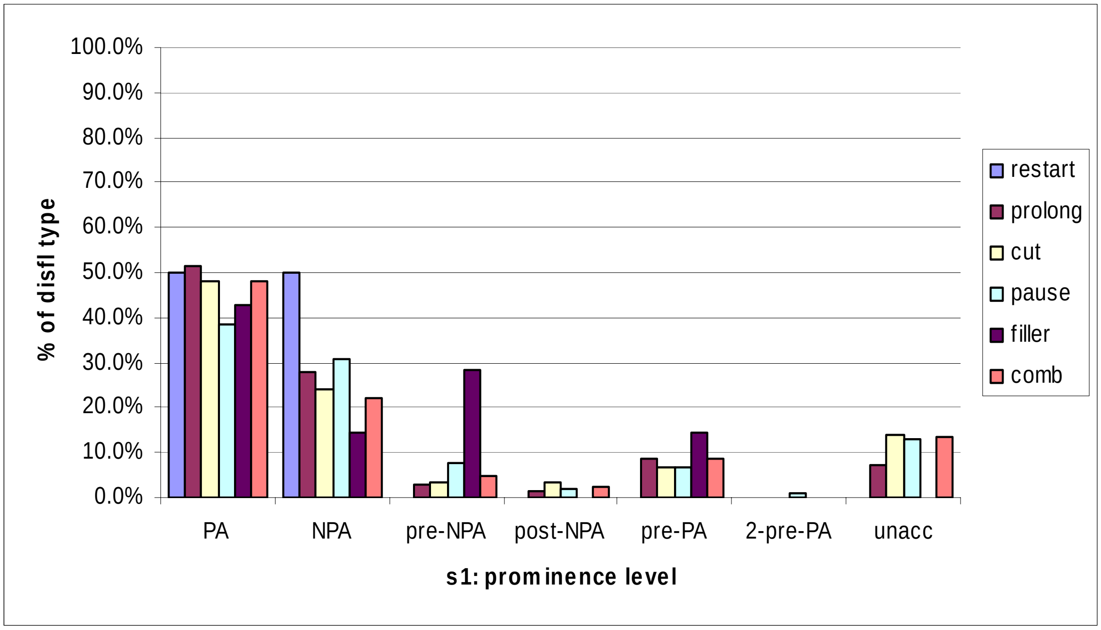


Figure 106: Distribution of disfluency types in percentage for each target prominence type: s1

Speaker S2 differed from Speaker S1 in producing slightly more variability in its disfluency type patterns, with prolongations occurring more often before pre-NPA targets (36.8%) than before PA (22.8%) or NPA (12.3%) targets. 2-pre-PA target position—nearly absent from S1's data—was represented by a small number of pauses in S2's production:

S2 Disfl \ prominence level	PA	NPA	pre-NPA	post-NPA	pre-PA	2-pre-PA	unacc
Disfluency	64	30	33	2	14	6	23
%	37.2%	17.4%	19.2%	1.2%	8.1%	3.5%	13.4%
Restart	1	0	0	0	0	0	0
%	100%	0%	0%	0%	0%	0%	0%
Prolong	13	7	21	0	6	1	9
%	**22.8%**	**12.3%**	**36.8%**	0%	10.5%	1.8%	15.8%

Cut	10	4	1	0	1	0	4
%	50%	20%	5.0%	0%	5.0%	0%	20%
Pause	16	8	2	2	2	4	3
%	43.2%	21.6%	5.4%	5.4%	5.4%	**10.8%**	8.1%
Filler	2	0	0	0	0	0	2
%	50%	0%	0%	0%	0%	0%	50%
Comb	22	11	9	0	5	1	5
%	41.5%	20.8%	17.0%	0%	9.4%	1.9%	9.4%

Table 25: Distribution of disfluency types for each target prominence type: s2

Once again fillers were a minor but noticeable disfluency type in that they occurred preceding unaccented targets half the time:

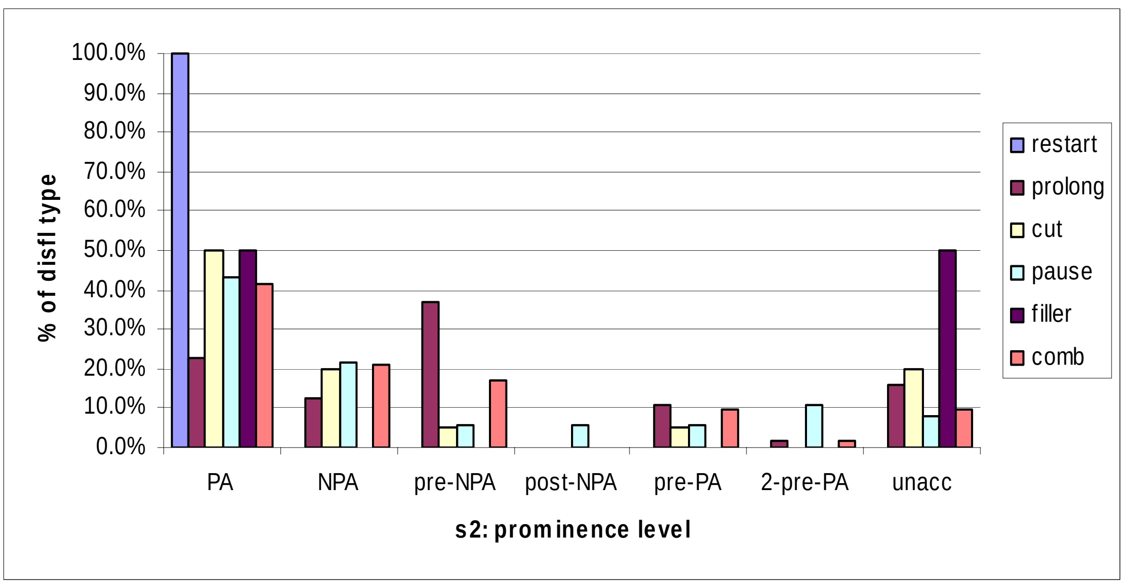


Figure 107: Distribution of disfluency types in percentage for each target prominence type: s2

Finally, Speaker 4 followed the same general patterns as the other two speakers—i.e., producing all disfluency types most often before the most frequent target tones PA and

NPA. Nevertheless, similar to Speaker 2, S4's prolongations showed less restricted behavior than other types, as a substantial percentage of them (11.1%) again occurred before pre-NPA targets:

S4 Disfl \ prominence level	PA	NPA	pre-NPA	post-NPA	pre-PA	2-pre-PA	unacc
Disfluency	468	240	57	36	63	9	42
%	51.1%	26.2%	6.2%	3.9%	6.9%	1.0%	4.6%
Restart	27	9	3	0	3	0	0
%	64.3%	21.4%	7.1%	0%	7.1%	0%	0%
Prolong	57	48	18	3	18	3	15
%	35.2%	29.6%	**11.1%**	1.9%	11.1%	1.9%	9.3%
Cut	45	24	6	6	3	0	9
%	48.4%	25.8%	6.5%	6.5%	3.2%	0%	9.7%
Pause	60	27	9	12	6	6	9
%	46.5%	20.9%	7.0%	9.3%	4.7%	4.7%	7.0%
Filler	**42**	**21**	**9**	0	3	0	6
%	**51.9%**	**25.9%**	**11.1%**	0%	3.7%	0%	7.4%
Comb	237	111	12	15	30	0	3
%	58.1%	27.2%	2.9%	3.7%	7.4%	0%	0.7%

Table 26: Distribution of disfluency types for each target prominence type: s4

S4 also produced considerably more fillers than either of the other two speakers, although its distribution patterned more similarly to other well-represented disfluency types:

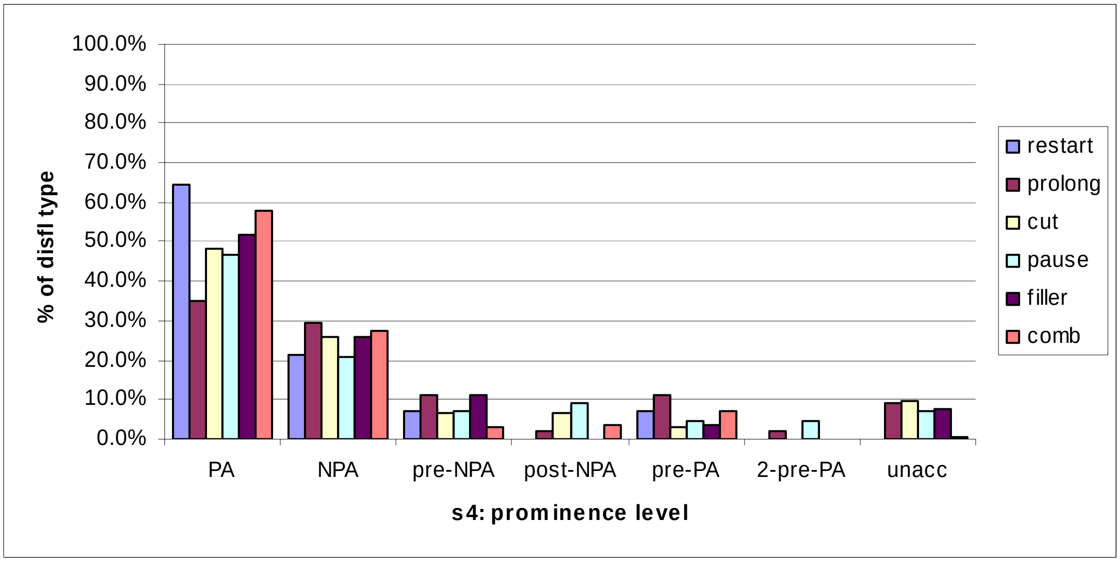


Figure 108: Distribution of disfluency types in percentage for each target prominence type: s4

The overall behavior of disfluencies with respect to the prominence level of their targets again reflects a general clustering around the distribution norm. In other words, disfluency types do not appear to pattern differentially before targets of different prominence level.

5.3.4 Phrasal position

The effects of target words from different phrasal environments on preceding disfluencies were measured by using the same classification categories used in the natural data analysis: ip-initial, ip-medial, and ip-final positions. As discussed in Chapter 4, pauses necessitate different treatment because unlike other disfluencies they are not realized directly on words or segments. However, in order to analyze them together with other disfluencies, pauses were adapted to the three-part classification above and categorized as either ip-medial or ip-initial.

Speaker 1 produced disfluency types with distributional regularity, revealing no robust co-occurrence patterns between individual disfluency types and target position within the phrase, as shown in Table 27 and Figure 109.

S1 Disfl \ ip phrasal position	ip-initial	ip-medial	ip-final
Disfluency	50	169	65
%	17.5%	**59.1%**	22.7%
Restart	0	1	1
%	0%	**50%**	50%
Prolong	13	45	12
%	18.6%	**64.3%**	17.1%
Cut	7	15	6
%	25.0%	**53.6%**	21.4%
Pause	15	59	26
%	15.0%	**59.0%**	26.0%
Filler	4	2	1
%	**57.1%**	28.6%	14.3%
Comb	13	47	19
%	16.5%	**59.5%**	24.1%

Table 27: Distribution of disfluency types for each target phrasal position: s1

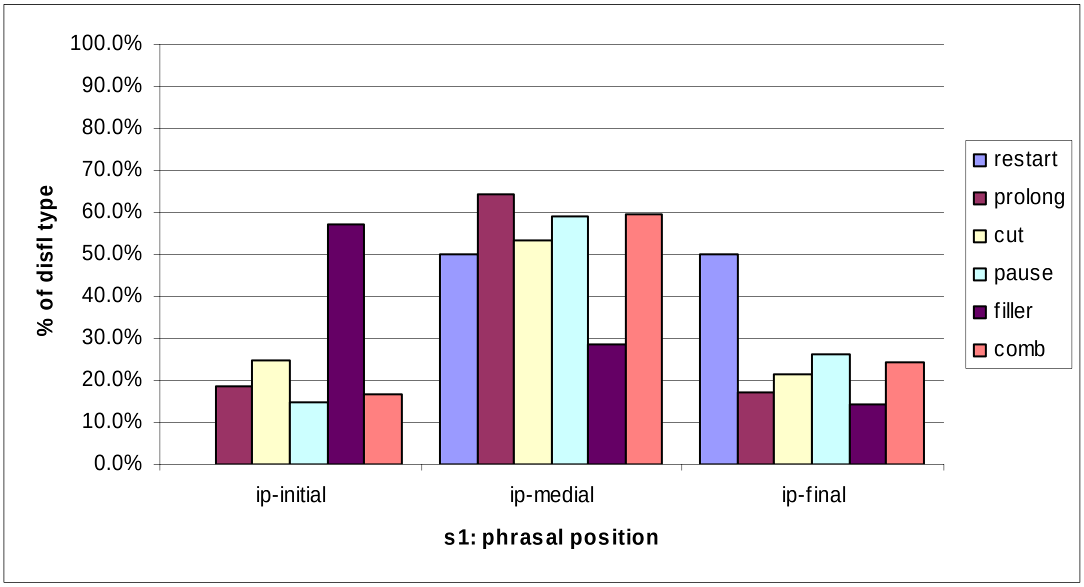


Figure 109: Distribution of disfluency types in percentage for each target phrasal position: s1

One minor exception was filler disfluencies, which, similar to their behavior in relation to target prominence, deviated notably from the overall behavior of disfluencies in this environment, as well. While 59.1% of S1's disfluencies were followed by ip-medial targets, more than half (57.1%) of fillers preceded ip-initial targets. Since fillers are literally considered filled pauses, their high rate of occurrence before ip-initial targets suggests that the filler disfluency may have been employed as a natural delay strategy while one or another element of the upcoming production material was being processed and/or repaired.

Both Speakers S2 and S4, meanwhile produced even more uniform distributions, as significantly more disfluency types preceded ip-medial targets than they did targets in either initial or final position:

S2 Disfl \ ip phrasal position	ip-initial	ip-medial	ip-final
Disfluency	17	131	23
%	9.9%	**76.6%**	13.5%
Restart	0	1	0
%	0%	**100%**	0%
Prolong	6	42	8
%	10.7%	**75.0%**	14.3%
Cut	4	13	3
%	20%	**65.0%**	15.0%
Pause	1	33	3
%	2.7%	**89.2%**	8.1%
Filler	0	4	0
%	0%	**100%**	0%
Comb	6	38	9
%	11.3%	**71.7%**	17.0%

Table 28: Distribution of disfluency types for each target phrasal position: s2

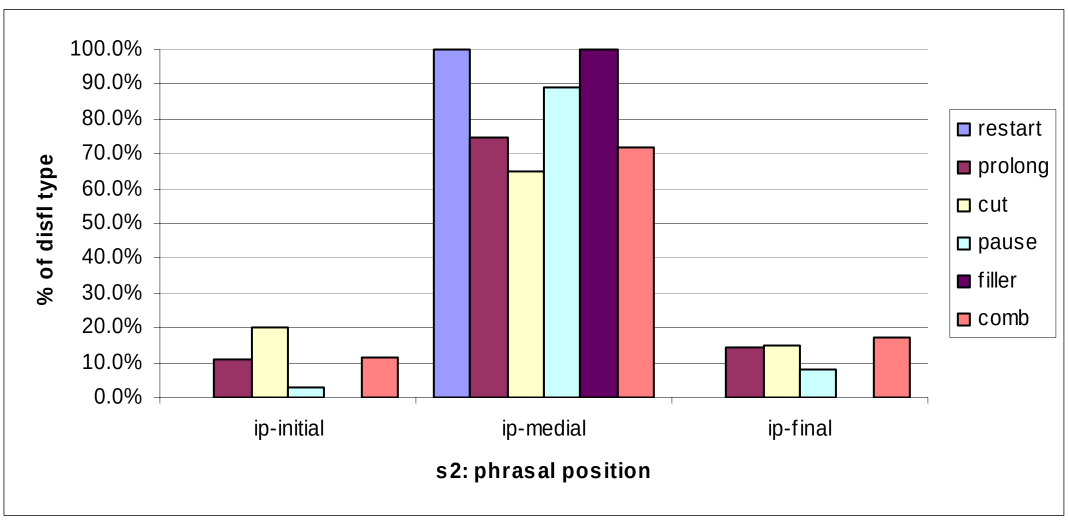


Figure 110: Distribution of disfluency types in percentage for each target phrasal position: s2

S4 Disfl \ ip phrasal position	ip-initial	ip-medial	ip-final
Disfluency	48	666	195
%	5.3%	**73.3%**	21.5%
Restart	3	30	9
%	7.1%	**71.4%**	21.4%
Prolong	9	114	39
%	5.6%	**70.4%**	24.1%
Cut	9	60	21
%	10%	**66.7%**	23.3%
Pause	6	99	21
%	4.8%	**78.6%**	16.7%
Filler	15	48	18
%	18.5%	**59.3%**	22.2%
Comb	6	315	87
%	1.5%	**77.2%**	21.3%

Table 29: Distribution of disfluency types for each target phrasal position: s4

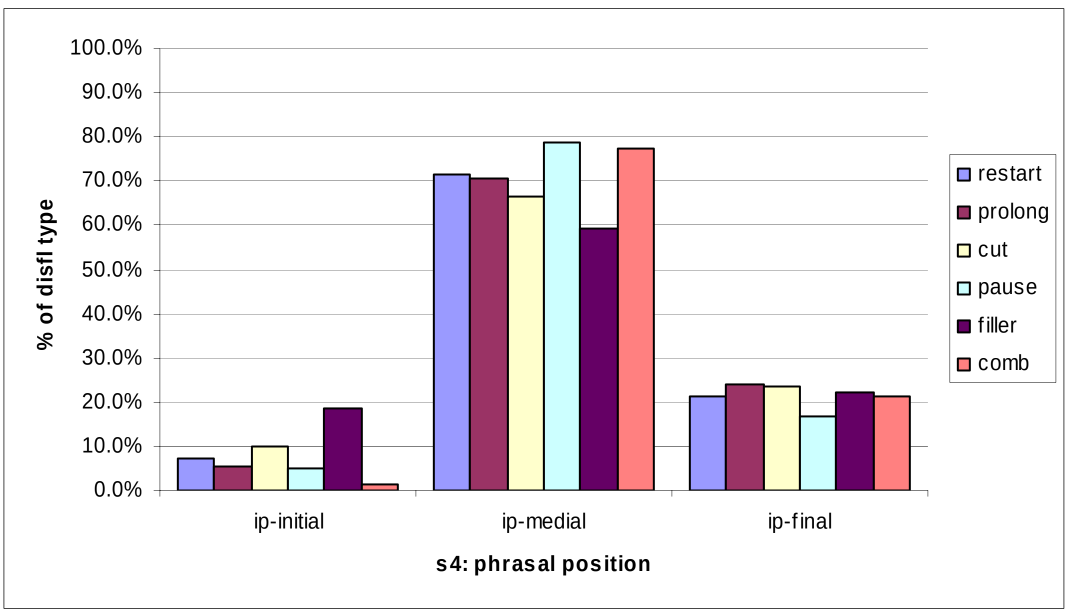


Figure 111: Distribution of disfluency types in percentage for each target phrasal position: s4

Aside from a couple of exceptions, therefore, no individual disfluency types patterned uniquely with specific target phrasal positions.

5.3.5 Syllable location

The earlier analysis of syllable location from section 5.2.6 revealed that more stutterers than controls (38.3% vs. 23.7%) produced disfluencies on syllable nuclei than on syllable onsets. Although this overall effect was somewhat diluted by inter-subject differences, another noteworthy trend with respect to syllable location occurred for a number of specific disfluency types. While most disfluency types were realized on the onsets of syllables, the majority of prolongations for speakers S2 and S4 surfaced on a syllable nucleus. Table 30 below shows the total number and percentage of each disfluency type for the three speakers:

Disfl \ location in the syllable	S1		S2		S4	
	onset	nucleus	onset	nucleus	onset	nucleus
Disfluency	285	82	129	109	615	447
%	**77.7%**	22.3%	**54.2%**	45.8%	**57.9%**	42.1%
restart	49	5	12	5	141	21
%	**90.7%**	9.3%	**70.6%**	29.4%	**87.0%**	13.0%
prolong	45	22	22	35	12	147
%	**67.2%**	32.8%	38.6%	**61.4%**	7.5%	**92.5%**
cut	25	5	15	4	51	48
%	**83.3%**	16.7%	**78.9%**	21.1%	**51.5%**	48.5%
pause	97	9	24	14	120	6
%	**91.5%**	8.5%	**63.2%**	36.8%	**95.2%**	4.8%

filler	0	0	1	2	0	3
%	0%	0%	33.3%	**66.7%**	0%	**100%**
comb	69	41	55	49	291	222
%	**62.7%**	37.3%	**52.9%**	47.1%	**56.7%**	43.3%

Table 30: Distribution of disfluency types for each syllable location (natural data): all subjects

Figures 112-114 show syllable location data (in percentage) for each speaker. Again, while all disfluency types surface more frequently on S1's syllable onsets than nuclei, for S2 and S4 this pattern is reversed for prolongations and fillers, which occur more often on their syllable nuclei.

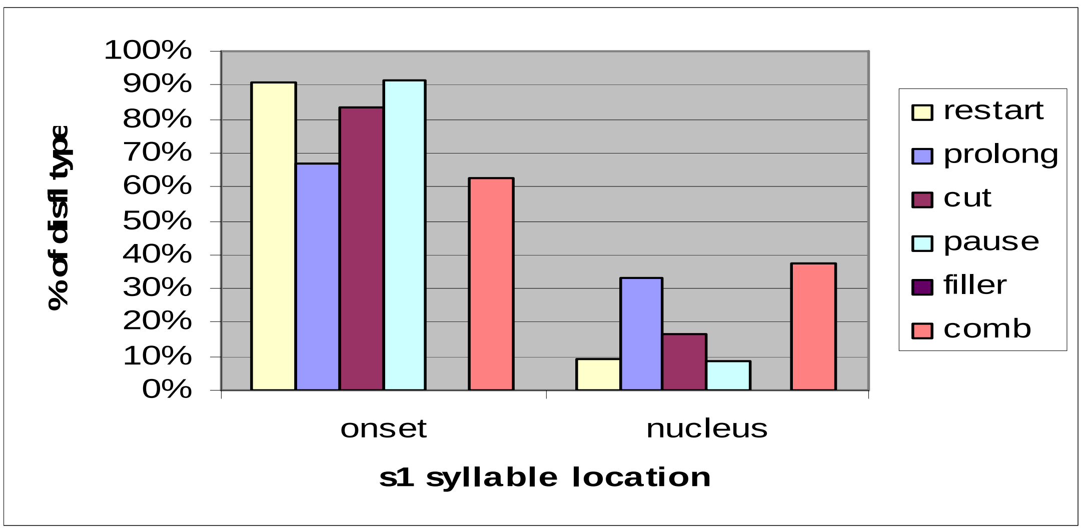


Figure 112: Distribution of disfluency types in percentage for each syllable location (natural data): s1

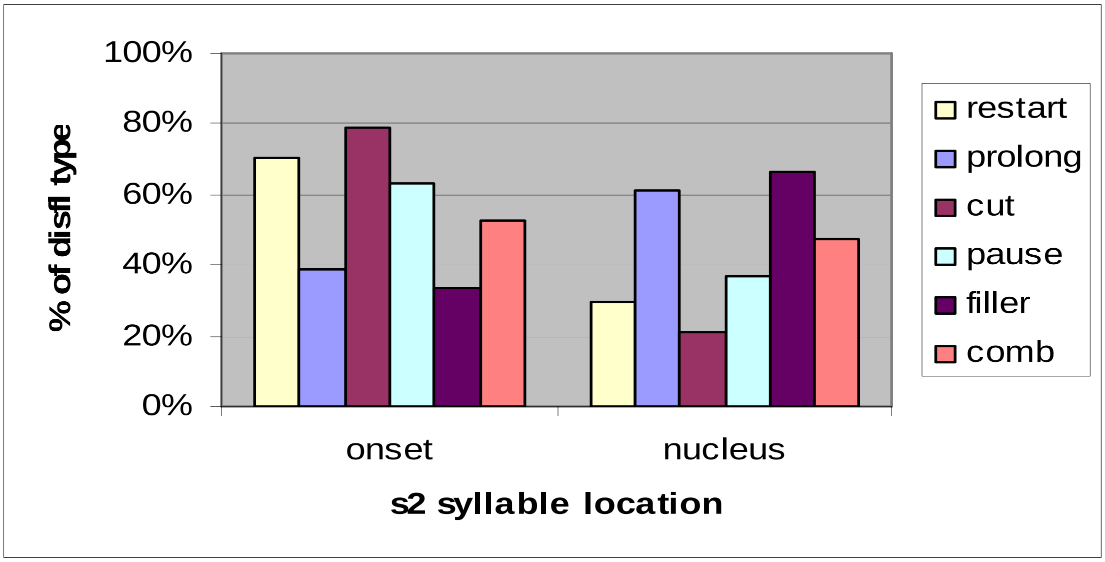


Figure 113: Distribution of disfluency types in percentage for each syllable location (natural data): s2

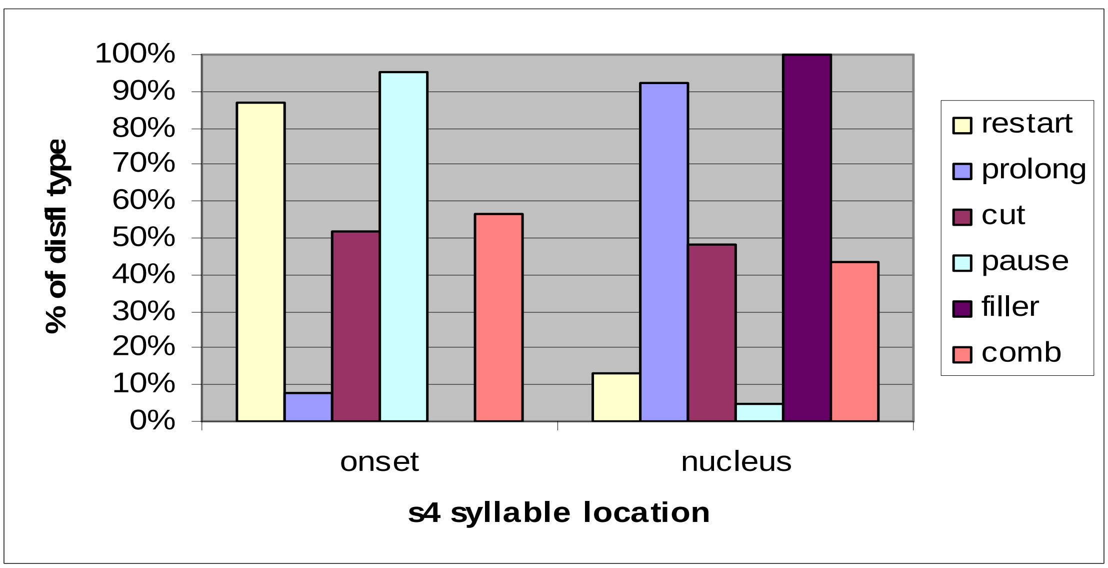


Figure 114: Distribution of disfluency types in percentage for each syllable location (natural data): s4

The reason for this disparity is apparent when it is taken into account that disfluencies such as pauses and cuts involve an arrest of production at or immediately preceding a syllable onset; prolongations, on the other hand, while also commonly occurring in onset position (e.g., "fffrog"), were more often attested for S2 and S4 in nuclear position. The curious aspect of this pattern is that it clearly contradicts previous analyses that have found evidence supporting the hypothesis that stuttered disfluencies are most often triggered by syllable onsets—or by the difficulty in transitioning from the onset to the nucleus (Wingate 1976).

In summary, when analyzed by specific disfluency type, all three stutterers generated more pauses preceding onsets than preceding nuclei, while two stutterers (S2 and S4) produced more prolongations of nuclei than onsets. Thus an interaction of disfluency type and syllable location is evident from the data.

5.4 Summary of script data results

A number of both robust and subtler patterns were revealed by the comparisons of natural disfluency data with the targets of scripted reference prosodies for each natural corpus. Comparisons were made with respect to five conditions necessary for determining the potential influence of linguistic—and specifically prosodic—factors on disfluency rates in both stutterers and controls. These comparisons revealed a number of differences both confirming key trends found in the Chapter 4 analyses, as well as uncovering new ones.

Tonal analyses showed that the significantly lower rate of L+H* produced on disfluencies in the natural data was not an artifact, since target L+H* pitch accents were produced significantly more often in the target data—hence behaving more like L+H* pitch accents in the data as a whole. This supports the hypothesis that the pitch-accented words which stutterers attempted during a disfluency were more likely planned to receive an L+H* tone than any other tone. L+H* pitch accents are extreme tonal targets, in that a local f0 rise must be achieved within a very short time span; furthermore, L+H* is often produced as the final pitch accent (Dainora 2001)—i.e., the nuclear pitch accent—and thus the added prominence it receives could arguably result in a more challenging prosodic target for stutterers.

A higher rate of !H- phrase accents at or immediately after disfluencies was another significant effect for both stuttering and control subjects, and the significantly lower rate of !H- tones in the matched target data motivates the interpretation of !H- as a disfluency-induced phrasal tone. The accompanying higher rate of L- tones in the target data can be said to lend support to this interpretation, especially in light of the L+H* results, since the successful production of the latter allows all normal boundary cues (i.e., pre-boundary lengthening, phrase-final tone) to be realized. When a disfluency is produced, however, these cues are often aborted or distorted. For example, if a tonal target is an L+H* followed by a phrase-final L-, a disfluency that partially cuts the final word would prevent the L+H* from reaching its low target, thus rendering it a H*. Furthermore, the failed realization of the final fall of the L- would instead result in a tone more similar to a !H-. Hence, the increased proportions of H* and !H-, together with their concomitant reductions of L+H* and L-, can be seen as results commonly predicted by the presence of a disfluent event.

The behavior of target boundary contours was similar to that of phrase accent tones, insofar as boundary tones which were either reduced (e.g., L-H%) or increased (e.g., !H-L%) in disfluent production of the natural data returned to normal distribution levels in target tones. Again, one possible interpretation is that contours such as the rising tone of L-H%, or its co-production with a preceding pitch accent target, represent a challenging target for stutterers, thus triggering a disfluency. As a result, the failed attempt at producing the target surfaces as a partially produced contour, again lacking the crucial duration and f0 cues that a boundary tone contour requires.

A second and very different possibility, however, is that the increase of mid or near-mid tones such as !H- and !H-L% might instead represent a *delay,* through which speakers postpone overt production, ostensibly in order to complete or repair anticipated disfluencies. In other words, whether the challenging target was the pitch accented form, the boundary contour, or both, what appear to surface as degraded cues of a disfluent attempt might instead constitute an anticipatory delay upon detection of *later* problematic material. Such an interpretation is consistent with various versions of speech monitoring hypotheses (e.g., Postma & Kolk 1993), in which it is proposed that speakers' disfluencies are actually attempted repairs of aberrant or incomplete encoding processes which, crucially, are detected in advance of articulation.

Results of comparisons with respect to word class may in fact support this latter monitoring hypothesis. Target productions from all subjects differed from their counterparts in the natural data by occurring at content words significantly more often than function words. Disfluently produced words in the natural data, on the other hand, while exhibiting more between-subject variability, as a whole were realized as function words more than half of the time. The noteworthy aspect to this difference is the fact that while the question of whether content or function words represent a greater production challenge for stutterers has been controversial and hitherto unresolved, a number of recent studies have proposed more detailed parsing models for handling the nuances of function/content word relationships in prosodic structure. For example, Au-Yeung, et al., (1998) propose a categorical distinction between function word and content word disfluencies, whereby function word disfluencies represent a stalling technique upon breakdown detection, while content word disfluencies constitute the locus of the disfluency itself. Equally crucial is the fact that disfluent function words precede the content words with which they form a syntactic and prosodic constituent. This is because the authors found that while function words preceding content words are produced disfluently by stutterers with regularity, very few function words that follow content words are disfluent. Thus, the notion of syntactic—and arguably more importantly, prosodic—constituency would appear to be a fundamental starting point for ultimately determining how conditions such as word class affect disfluency distribution.

The prominence comparisons provided further evidence of the influence of prosodic structure on disfluency patterns. Regarding prominence level proper, high-prominence words (i.e., pitch-accented and nuclear pitch-accented words) occurred significantly more often in target position than unaccented words. Once again, this result suggests some related possible interpretations. On the one hand, the high percentage of pitch-accented targets is strong evidence that these high-prominence words were both originally planned and insufficiently processed. The actual effect of this processing problem could have then elicited a number of repair strategies: as postulated above, if the disfluency-triggering problematic target directly preceded a planned intonation phrase boundary, a speaker's production of this boundary was often prosodically degraded (e.g., shortened, missed f0 target). Alternatively, if the detected breakdown was phrase-internal, a speaker may have nevertheless utilized a stalling technique in order to gain processing time for the upcoming target. Consequently, either unstressed material was added or the upcoming target was de-stressed (or both), in essence effecting an unplanned phrase boundary, while also replacing the planned pitch-accented target with an unaccented form.

Thus, in short, prominence clearly appears to play a role in triggering disfluencies, but information regarding the structures in which these prominent targets are generated must be utilized in order to determine how the triggering process actually operates. This necessity motivated the more detailed classification of prominence levels, since not only is it possible that prominence effects are gradient, thus triggering disfluencies to varying degrees, but prominence must also interact in some testable way with the prosodic structure of the utterance, if it is indeed the case that a monitoring mechanism is responsible for detecting upcoming production problems before they are articulated.

For instance, the observation by Au-Yeung, et al., (1998) that function words attract disfluencies only in certain positions (viz., before the content word with which they form a syntactic and prosodic constituent) implies that *how* and *when* disfluencies surface is inevitably determined by rules of prosodic structure. In a similar way, the present study proposed a more detailed series of prominence categories which could be used to examine the environments both preceding and following high-prominence (pitch-accented) positions. As discussed earlier, this fine-grained approach indeed revealed higher frequencies of targets occurring on words immediately preceding pitch accents (pre-PA) or nuclear pitch accents (pre-NPA) than immediately following them (post-NPA). This method also had some challenges, however, in that some positions such as post-NPA are simply less represented in the overall corpus, thereby limiting how frequently they would even be available as targets.

In a domain larger than the Prosodic Words which were indirectly analyzed in the prominence condition, phrasal position comparisons tested the relative influence of various locations of the entire intermediate phrase—specifically, initial, medial, and final positions. The results of the natural data analysis revealed ip-initial position to be the most frequent locus of disfluency realization. In addition, a number of studies in the previous literature have found similar support for this result. However, before it can be concluded that disfluencies reliably surface phrase-initially, the effects of the disfluency itself on the intonational structure must be accounted for. For example, a word generated and designated for ip-medial position could ostensibly trigger a disfluency on an earlier word due to detection by the monitor, thus causing a disfluency immediately before the target word's articulation. Depending on a number of factors, including the speaker's awareness and access to resolution strategies, the result could be a prematurely terminated intermediate phrase. Crucially, then, the word targeted for ip-medial position ultimately is realized ip-initially. Thus, not taking into account the prosodic conditions analyzed here—which effectively serve as clues to what might be the disfluency triggers themselves—may easily lead to an incorrect analysis of which ip-position attracts the highest rate of disfluencies.

As the target vs. natural comparisons here have indicated, ip-initial position ended up showing no evidence of triggering a high rate of disfluencies. To the contrary, medial and final positions—the latter which, incidentally, often function as the site of nuclear pitch accents—were instead the target positions which attracted more disfluencies. However, analyzing disfluency distribution patterns as a function of Phonological Phrase structure revealed that while the onset of an ip (i.e., the onset of the ip's first word) was rarely disfluent, disfluencies surfaced most often on the pitch-accented words of the first and final PhPs, regardless of the number of pitch accents within the ip. In addition, a subtler pattern emerged in which disfluencies occurred in second and fourth PhPs less frequently as total PA number increased. The percentage of disfluencies surfacing in the third PhP, however, remained relatively stable. This behavior suggests that prominence level and ip structural properties interacted to constrain the distribution of disfluencies in a patterned way. Specifically, the fact that target disfluencies regularly occurred in the final PhP, which is the location of default nuclear pitch accents, partially supports the prediction of Hypothesis 1 that NPAs will trigger the highest percentage of disfluencies in an intermediate phrase. Furthermore, the consistent generation of target disfluencies in the initial, third, and final Phonological Phrases of an ip suggests that a prosodic constituent larger than a PhP, but smaller than an ip, functioned to constrain the distribution of stutterers' disfluencies. This possibility will be explored in greater detail in the next chapter.

Analyzing disfluencies in terms of their location in the syllable revealed patterns consistent with the hypothesis that disfluency triggers originate in points of high prominence, as well as the related hypothesis that these triggers are part of larger prosodic phrases whose internal structure is crucial to determining when a speaker—perhaps via monitoring—detects an impending breakdown. Overall, stutterers produced more disfluencies on nuclei (39.9% of their total) than did controls (26.7% of their total). A closer analysis of the disfluency data, with respect to their range of disfluency types, revealed that stuttered prolongations surfaced most often on syllable nuclei, while most stuttered cuts and pauses were realized before or on syllable onsets. That disfluencies realized on the syllable nucleus were most frequently prolongations of the vowel itself represented solid evidence that the point of disfluency was not necessarily the deficient point—i.e., trigger—of the disfluency. Additionally, the results of the word class comparisons presented further evidence that the realization of stutterers' disfluencies on function words is more likely an anticipatory effect of a later breakdown, since targets were more often content words (which are pitch-accented more frequently than function words). One possible interpretation that follows from this finding is that stutterers detect disfluency triggers more often than do controls. It is still a point of debate whether what is detected is an actual disfluency resulting from deficient phonological encoding (Postma & Kolk 1993), or the result of a "hypervigilant speech monitor" (Vasic &

Wijnen 2005) which, essentially, detects false positives and treats them as disfluencies by interrupting the speech flow in an attempt to repair them.

A final interaction was discovered when prominence and syllable position were analyzed together for stutterers and controls. In addition to the results described above where stutterers detect more disfluencies than do normal speakers, and thus end up producing derivative disfluencies such as prolongations at the point of detection (e.g., function words), the interaction between prominence and syllable position revealed that stutterers also end up pitch-accenting these anticipatory disfluencies significantly more often than controls. Ultimately, then, not only can an anticipatory disfluency produce unplanned phrase boundaries in certain environments, but it can also attract pitch accent if the cues proper to pitch accents are salient enough during the anticipatory disfluency.

CHAPTER 6

GENERAL DISCUSSION

6.1 Summary of narrative and script results

It was hypothesized in this thesis that stuttered disfluencies are predictable by properties of intonation. On the one hand, it was predicted that these properties constitute a prosodic deficit which is ultimately responsible for stuttered disfluencies. Based on the theory of intonational phonology, which assumes a hierarchical structure of metrical prominence relations among items of a prosodic structure, the first hypothesis predicted that words bearing a higher level of prominence would be more prone to disfluent production in stutterers' speech than words bearing lower metrical prominence. At the same time, it was also predicted that stutterers' disfluencies would be accompanied by more prosodic irregularities prior to the underlying disfluency than would non-stutterers' disfluencies. The motivation for this claim was that if stuttering is ultimately rooted not in lexical or phonological, but rather sentence-level stress effects—specifically, the planning and production of metrical prominence—then the distribution of derivative disfluencies which surface when stutterers detect these forthcoming breakdowns should be consistent with the rules of prosodic phrase structure.

Results from the natural data supported the first hypothesis that metrically prominent, or pitch-accented, words would attract a higher rate of stuttered disfluencies than unaccented words. This was not the case for control subjects, whose disfluencies occurred with greater frequency on unaccented words; thus, the first hypothesis is stuttering-specific. The second important result from the natural data supported the second hypothesis that stutterers would generate categorically distinct "anticipatory" disfluencies triggered by impaired structure later in the phrase—i.e., implicit "target" disfluencies. Stutterers produced a significantly higher percentage of their disfluencies on the nuclei of words than did controls. Unlike onset disfluencies, which prevented successful production of a syllable, disfluencies in syllable nuclei allowed both major constituents of the syllable (i.e., onset and rime) to surface. The successful production of syllable components in these nuclei disfluencies was interpreted as evidence that production difficulties did not originate in the syllable in which they surfaced, but rather resulted from a speaker's detection of problematic material later in the speech stream. Irregularly prolonged nuclei, which often served to attract pitch accents significantly more often than did those of controls, provided further evidence of a qualitatively different disfluency type produced solely by stutterers.

When the natural data were compared with the script data, the results revealed again that high-prominence words occurred significantly more often in the script-interpreted target position (i.e., where script-reading controls produced pitch accents) than did unaccented words. Furthermore, of all possible ip positions, targets of disfluencies were produced most often ip-medially, on the pitch accent of the first Phonological Phrase, and ip-finally, on the nuclear pitch accent. The location of remaining disfluencies suggested the constraining influence of smaller prosodic structures on the disfluencies' distribution patterns.

A detailed analysis of prominence categories revealed that words immediately preceding pitch accents (pre-PA) and nuclear pitch accents (pre-NPA) were targets more frequently than were words immediately following nuclear pitch accents (post-NPA). This supports the second hypothesis that in addition to prominence level of a target word, the internal structure of a prosodic phrase will play a role in determining patterns of disfluency.

A summary of the results of both experiments is provided in Table 31. Each condition is listed by row, sub-divided into the two tasks (i.e., 'narrative' and 'target'). The first three columns summarize the main effects found for each condition as a function of total disfluencies. The first column lists the results for stutterers, the second column for controls, and the third summarizes the comparison between groups. In the narrative task, the group comparisons are between stuttered ('S') and control subject ('C') disfluencies, while in the script task the comparisons are between the target ('T') and narrative ('N') disfluencies. Finally, the last two columns indicate whether the results of a given comparison supported ('+'), contradicted ('-'), partially supported/contradicted ('+/-'), neither supported nor contradicted ('neith'), or were not evaluated with respect to either of the two main hypotheses (N/A).

	Stutterers	Controls	S vs. C / T vs. N	Hyp 1	Hyp 2
Prominence (% disfluent)					
Narrative	PA>NPA>unacc	unacc>PA>NPA		+/-	N/A
Target	PA>NPA>unacc pre-NPA > post-NPA	N/A	T>N: PA N>T: unacc	+/- N/A	N/A +
Phrasal position (% disfluent)					
Narrative	pause: boundary>medial; non-pause: medial>initial>final	pause: medial>boundary; non-pause: medial>initial>final		N/A -	N/A N/A
Target	medial>final>initial PhP: initial>final>medial	N/A N/A	T>N: medial, final N>T: initial N/A	+/- +/-	N/A N/A
Pre-disfluency tone					
Pitch accent					
Narrative	more H*, !H* fewer L+H*, L*	fewer L*+!H	S>C: H*, !H* C>S: L+H*, L*	N/A	+
Target	N/A	N/A			
Phrase accent					
Narrative	more !H-	more !H- fewer L-	S=C	N/A	-
Target	N/A	N/A			
Boundary tones					
Narrative	more !H-H% fewer H-H%	more !H-H%, !H- L% fewer L-L%		N/A	-
Target	N/A	N/A			
Boundary contours					
Narrative	no change	more mid fewer low	S<C: rising	N/A	+
Target	N/A	N/A			
Disfluency tone					
Pitch accent					

Narrative	more H*; L*, !H*, H+!H* fewer L+H*	more !H* fewer L*	S>C: L* C>S: !H*	neith	N/A
Target	more L+H* fewer H*	N/A	T>N: L+H* N>T: H*	-	N/A
Phrase accent					
Narrative	more !H-; fewer L-	more !H-; fewer H-	S=C	+	N/A
Target	more L-, fewer !H-	N/A	T>N: L- N>T: !H-	+	N/A
Boundary tones					
Narrative	more !H-L% fewer L-H%	more !H-L% fewer L-H%	Not measured	+	N/A
Target	more L-L% fewer !H-L%	N/A	T>N: L-L% N>T: !H-L%	+	N/A
Boundary contours					
Narrative		Too few data			
Target		N/A			
Content vs. function word (% disfluent)					
Narrative	more function disfl	more function disfl	S>C: function	N/A	+
Target	more content disfl	N/A	T>N: content	+	N/A
Syllable location of disfluency					
Narrative	onset > nucleus	onset > nucleus	S>C: nucleus	N/A	+
Target	N/A	N/A	N/A	N/A	N/A

Table 31: Summary of results

Taking together the results from both the natural and script analyses of this thesis, a number of related conclusions can be drawn regarding how disfluencies are both triggered and detected in stutterers' language production. These generalizations emerge by reviewing several key patterns uncovered in both the narrative and script analyses, and by attempting to answer the questions these patterns raise in greater detail, as well as by exploring the broader implications they have for language processing. The questions they raise are as follows: 1) why are rising IP boundary tones in pre-disfluency position produced less frequently by stutterers than by control subjects? 2) why do stutterers produce disfluencies more often on function words than do nonstuttering speakers in natural speech? 3) why are disfluencies which consist of prolonged nuclei more common in stutterers' speech than in control subjects' speech? 4) why are disfluencies produced in natural speech less frequently in ip-final position, and more frequently in ip-initial position relative to other locations? 5) why do more stuttered target disfluencies surface on pitch-accented words of initial and final Phonological Phrases than on PAs of other PhPs?

6.2 Disfluency effects in prosodic context

As was discussed in Chapter 2, in the theory of intonational phonology, intermediate and Intonation Phrase boundaries are indicated by crucial disjuncture and tonal cues. Intonation Phrase boundaries are preceded by lengthening of the final word of the phrase, as well as a tonal contour, which is realized generally on this final lengthening. The final word of the Intonation Phrase is then followed by a pause (or greater lengthening), after which a new Intonation Phrase is signaled by a resetting of F0 at the next pitch-accented word. Intermediate phrases share most of these cues, but as they are themselves constituents of Intonation Phrases, they contain only a pre-boundary phrase accent tone (rather than both a phrase accent and boundary tone) and are not followed by a pause.

The major prosodic phrase boundary cues can be summarized as follows:

1) disjuncture cue 1: pre-boundary lengthening
2) disjuncture cue 2: post-boundary pause
3) tonal cue 1: pre-boundary f0 excursion
4) tonal cue 2: post-boundary f0 reset

Intonation Phrases are signaled by the presence of all four cues, while all cues but the second (post-boundary pause) function to mark an intermediate phrase boundary.

Disfluencies often result in producing false, degraded, or missing boundary cues. For instance, prolongations, by virtue of extending the duration of a word, may become confusable with pre-boundary lengthening (disjuncture cue 1). In addition, since, intonationally, whatever tone realized on a word will necessarily continue its pitch excursion with a segment's added duration (provided the segment is tone-bearing), a phrase-internal word will attract a tone confusable with a pre- (or post-) boundary tonal cue. Thus, a prolongation disfluency may result in creating an unintended ip boundary cue. In data from Speaker S2 in Figure 115, a function word ("to") following an ip boundary ('H-') is disfluently prolonged:

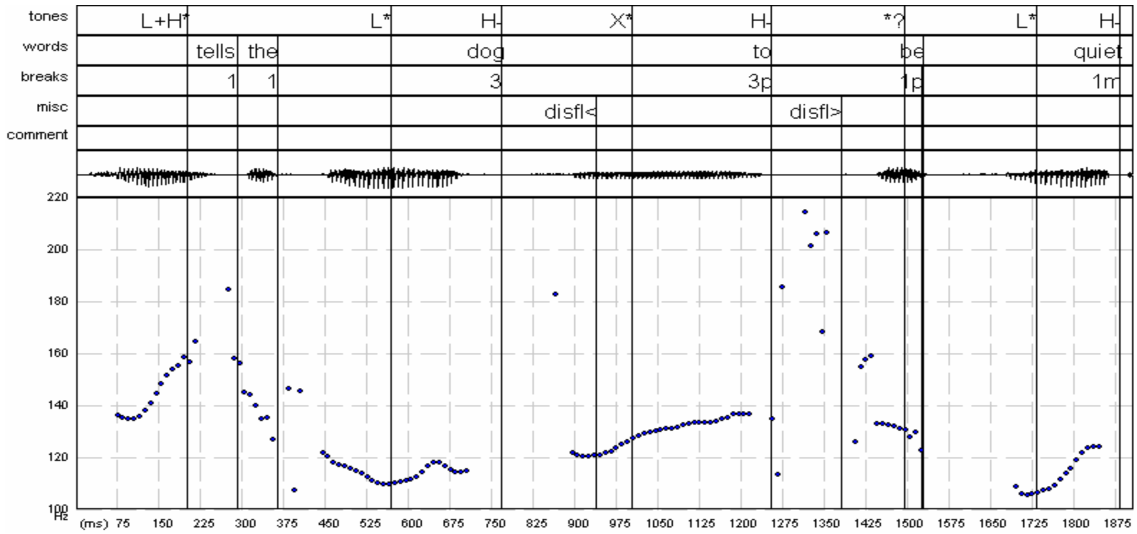


Figure 115: Prolongation example (3p): adding a false boundary cue (S2)

As exemplified by the slowly rising f0 on the prolonged “to”, the arguably unintended prosodic effect is to manifest an intermediate phrase boundary cue (‘H-’). Furthermore, the combination of increased duration from the prolongation and the upward f0 excursion effectively produce a pitch accent on the word “to”. Thus, a single prolongation disfluency results in transforming an ip-internal unstressed function word into a nuclear pitch-accented word in ip-final position.

The cue distortions are readily apparent when S2’s production shown in Figure 115 is compared with that of his age-matched control subject C2 in the latter’s reading of S2’s script. Figure 116 below highlights the crucial differences between the two productions, as, in the absence of prolongation, no tonal or disjuncture cues are realized on the word “to”:

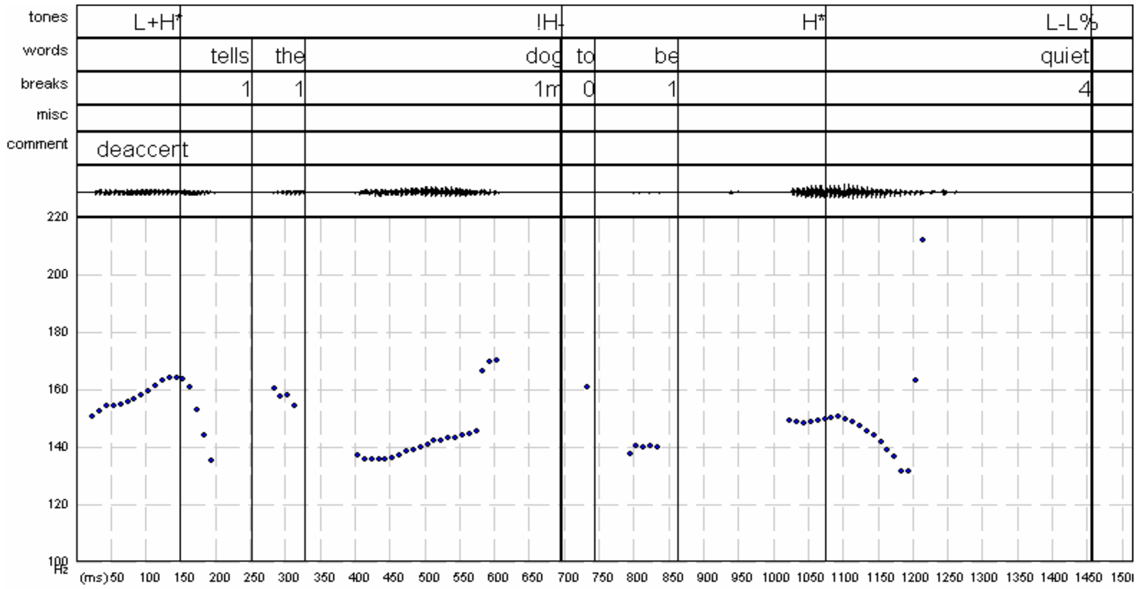


Figure 116: Control (C2) Script of S2, example 1

Not only are no confusable boundary cues present in the vicinity of "to", but the normal cliticized production of "to" does not interfere with the gradually rising interpolation from "be" to the f0 peak of NPA target word "quiet".

A further interesting effect which sometimes results from a stuttered prolongation disfluency is a speaker's subsequent *subtraction* of a later targeted boundary cue. Figure 117 expands the phrasal environment of Figure 115, showing what happens after the '3p' prolongation surfacing on "to":

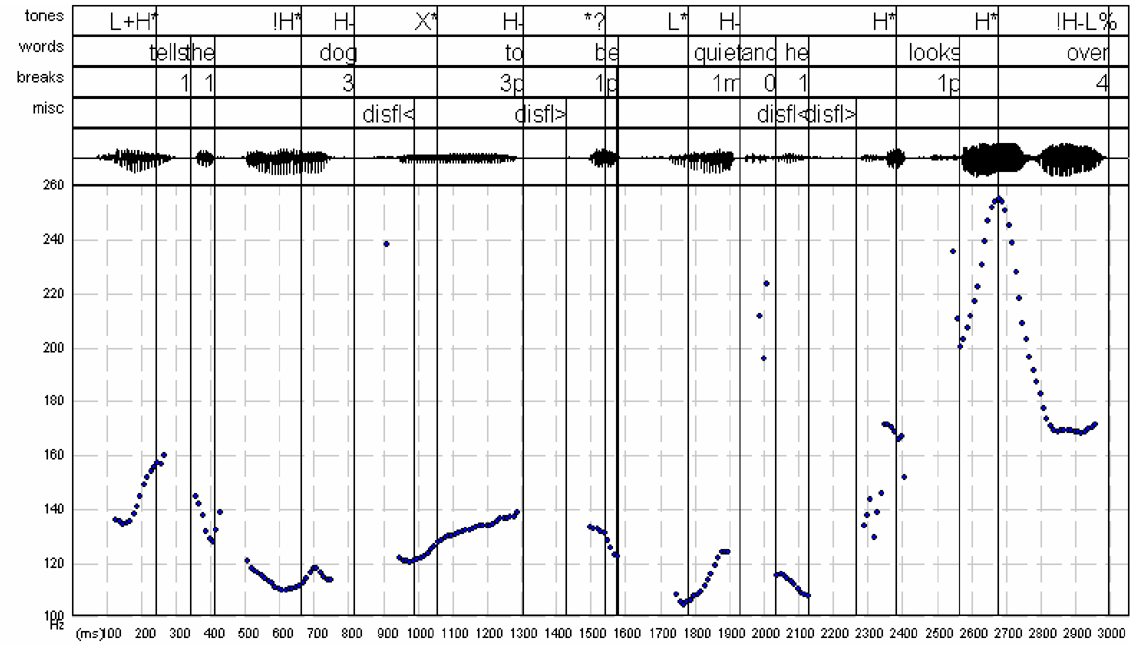


Figure 117: Example of post-prolongation timing compensation (1m) for speaker S2: "tells the dog to—be quiet and he looks over"

As Figure 117 illustrates, two words after the disfluent prolongation on "to", a '1m' mismatched boundary after "quiet" signifies that the speaker produced mismatched cues at an intermediate phrase boundary: in this case, while both pre-boundary tonal cue 1 ('H-') and post-boundary tonal cue 2 (reset f0 at the following H* on "looks") are present, the pre-boundary lengthening of disjuncture cue 1 is missing. Factoring in the effect of the preceding '3p' prolongation on the speaker's temporal structure, it can be inferred that the '1m' constitutes a *timing compensation*—i.e., the speaker compensated for the time lost during the prolongation on "to" by not providing the lengthening cue expected at the planned ip break after "quiet".

Like prolongations, pauses are also responsible for producing contradictory prosodic phrase boundary cues, depending on the type of pause that surfaces. Pauses which result from a failure to produce a syllable's onset—known as 'blocks' in the literature on clinical treatment (Bloodstein 1995)—almost inevitably will produce conflicting cues, given that the pause resulted not from a planned phrase boundary break, but rather from a disfluency. In the utterance in Figure 118, three pause disfluencies surface when the speaker attempts the onsets of the words, "dog", "looking", and "in". The first two pauses appear to result in ip breaks ('3ps?'), while the third pause occurs ip-medially ('1ps'). The post-pausal F0 resets on "dog" and "looking" provide evidence that the pauses demarcate either ip or IP boundaries. However, despite the presence of both tonal and disjuncture *post*-boundary cues, the lack of clear *pre*-boundary tonal and lengthening cues before either of these two pauses militates against the interpretation that these pauses represent either ip or IP boundaries:

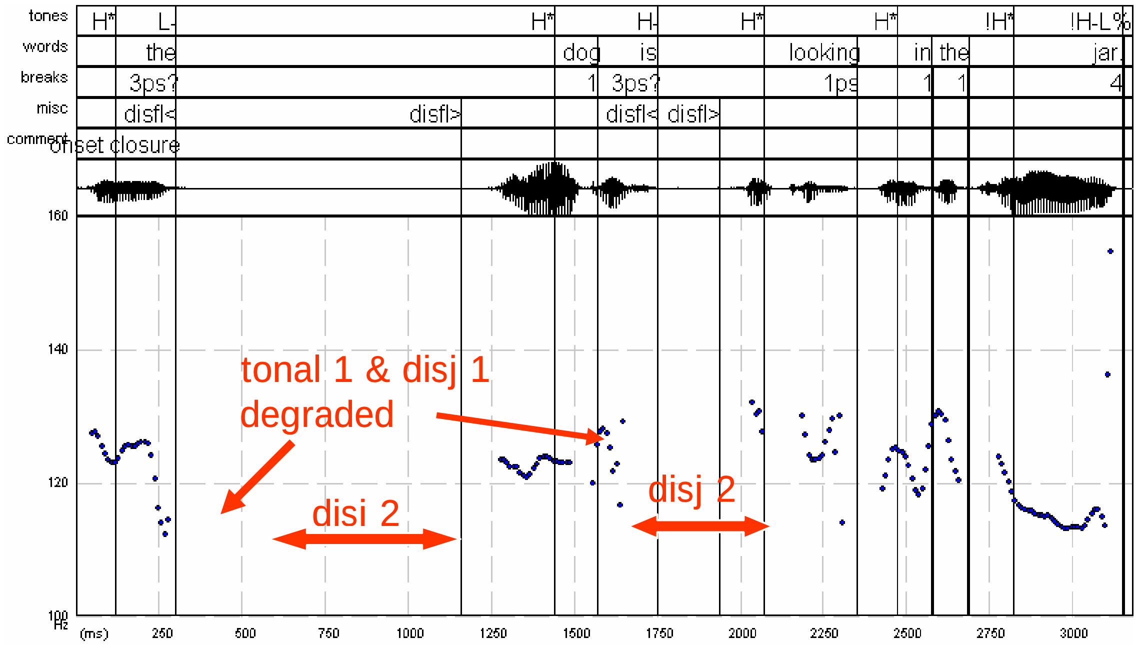


Figure 118: Pause example (3p): adding unintended boundary cues after "the" and "is" (S2): "the dog is looking in the jar"

Thus, the prosodic evidence in Figure 118 presents a scenario in which contradictory cues prevent a straightforward analysis of the between-word boundaries. Evaluating the utterance strictly on syntactic grounds, the pauses are clearly unintended in this environment, since they both separate a determiner from a noun. The degraded pre-boundary cues, however, also serve as indicators of a disfluency, since it is necessary that both pre- and post-boundary cues be realized in order to signal an ip or IP boundary. The classification '3ps?' (rather than '4ps') was motivated by the fact that the evidence of a pre-boundary tone is slight at best; that is, the unlengthened words "the" and "is" are hardly sufficient for full ip boundary tone realization—much less for the longer tonal contours necessary to produce an IP boundary.

Speaker C2's script reading of the same utterance provides a reference prosody against which S2's original utterance may be evaluated. As Figure 119 shows, fluent production of the phrase "the dog is looking in the jar" occurs as a single ip (and IP), without either the tonal or disjuncture post-boundary cues found in S2's natural production.

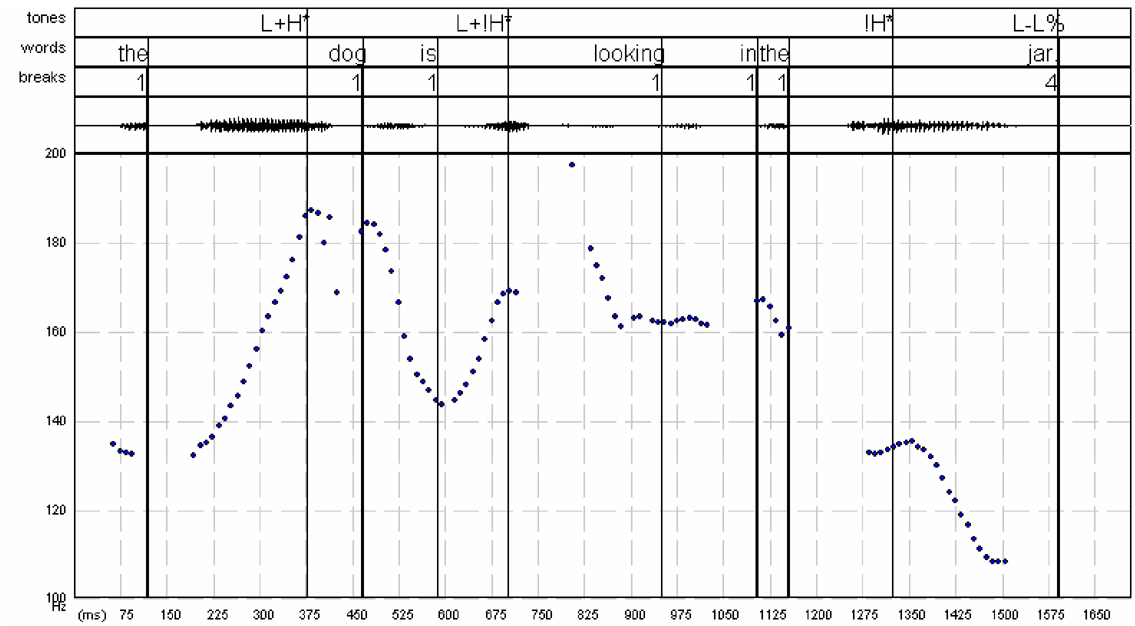


Figure 119: Control (C2) Script of S2, example 2

Models of disfluency production in normal speech, as discussed in Chapter 2, can account for these phonetic aberrations by decomposing a disfluency's structure into various key elements. Assuming a structure consisting of the three components originally proposed in Levelt (1983)—i.e., reparandum, editing phase, and repair—Shriberg (1999) noted that during both reparandum and editing stages, speakers may preserve the intonation patterns and f0 relationships that were established in the prior context. For instance, speakers often produced filled pauses which, through interpolated pitch excursion, preserved ip/ IP boundaries despite the presence of the disfluency. The increased duration of these filled pauses resulted in flattened f0 patterns which mirror the results found in the present study. As the boundary tone contour data from Chapter 5 showed, stutterers and control subjects behaved similarly in producing !H-L% contours during disfluencies significantly more often than other contour types. The !H-L% tones are consistent with the flattened f0 of filled pauses produced during the editing phase by speakers in Shriberg (1999).

The pre-disfluency data in Chapter 4, however, which presumably correspond to the reparandum in Levelt's model, reveal different behavior in stuttering and control groups. While stutterers did not differ significantly in their pre-disfluency boundary contour productions compared to their boundary contours overall, they nevertheless produced considerably fewer rising F0 contours than did control subjects. One interpretation of this contrastive behavior is that stutterers maintained a lower f0 in anticipation of an approaching disfluency, while control subjects demonstrated no such early awareness of imminent production difficulties. Furthermore, while it was found in Shriberg (1999) that the vast majority of single-word repetitions were followed by pauses without any phonetic cues of the following word, stutterers' transitions in the present study were often detectable even after filled and unfilled pauses. Figure 120 shows a pitch track from a stuttering speaker (S2) who, after producing a series of both prolongation and pause disfluencies—and therefore presumably producing a lexical search error—nevertheless shows evidence of a phonetic error after the restarted word "the":

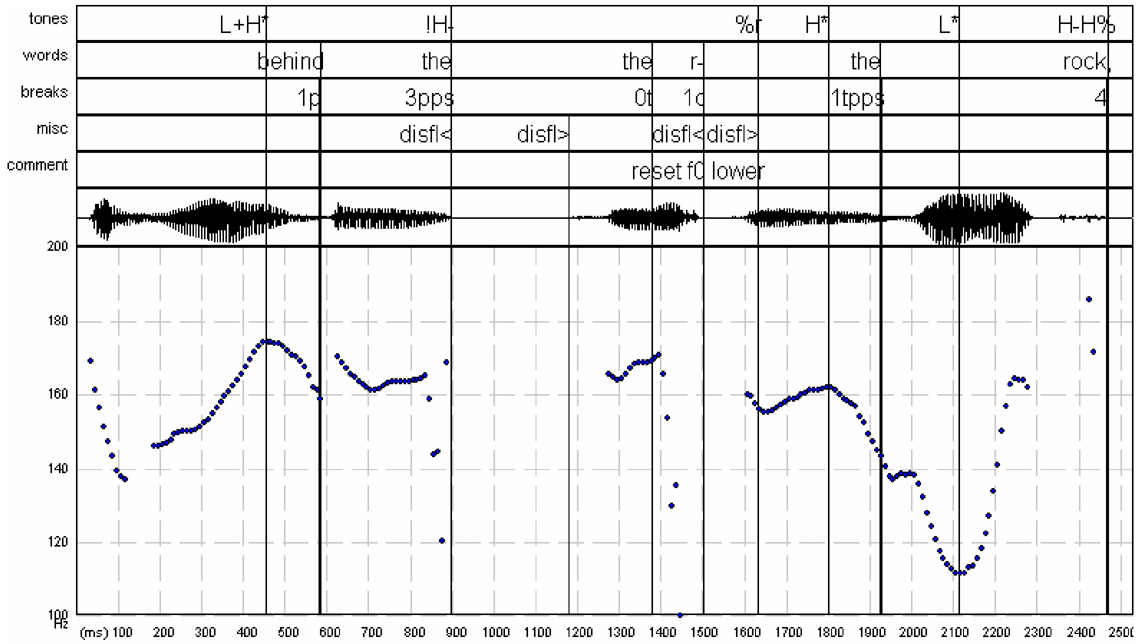


Figure 120: Phonetic cues of disfluency (S2)

That is, after disfluent prolongation of the vowels in "behind" and "the", followed by an unfilled pause, an eventual restart results in a cut word beginning with the segment [r]. The crucial feature of this string of disfluencies is the fact that phonetic evidence of the following word (i.e., the [r] of "rock") surfaced even after pauses and prolongations, which would imply that not only was the problematic element already detected two words previously ("behind"), but the problematic word's phonological information was at least partially encoded.

A comparable control subject example in Figure 121 shows a crucial distinction after the disfluency beginning with "a": while C4 produces a function word prolongation +

pause combination disfluency (‘4p.ps’) similar to S2’s ‘3p.ps’ in Figure 120, C4 shows no subsequent evidence of phonetic breakdown during or after the restart at ‘1t’:

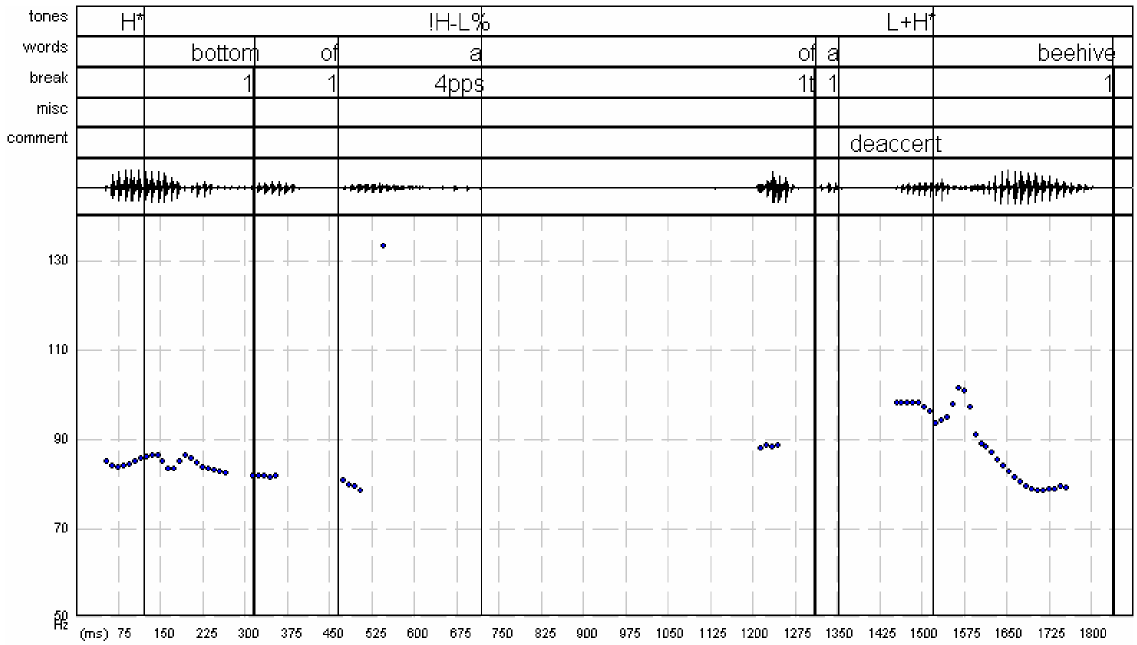


Figure 121: No phonetic cues of disfluency (C4)

In summary, prosodic effects in a disfluency’s immediate environment reveal similar behavior between stutterers and control subjects: in both cases, F0 tends to flatten out to a mid-tonal range, resulting in a larger percentage of !H-L% boundary tones produced during the disfluency (i.e., editing phase). At the same time, the lower frequency of rising boundary contours produced by stutterers *before* a disfluency suggest that stutterers are responding to a detection of an upcoming disfluency by maintaining a lower F0. Furthermore, as shown in such examples as Figure 120, it appears that stuttering speakers crucially have access to some level of a word’s phonological form upon detection of the disfluency (during the reparandum). In contrast, disfluencies produced by normal subjects in Shriberg (1999) crucially lacked transitional evidence of upcoming phonetic material. This suggests that either the disfluency detection process, or the nature of the underlying disfluency itself, is qualitatively different for stutterers as compared with that of normal speakers. The next section will further address this question by looking at the evidence of other prosodic comparisons.

6.3 Anticipatory vs. target disfluencies

Several previous studies in the stuttering literature indicated that, compared with normal speakers, stutterers show linguistic evidence of early disfluency detection (e.g., Viswananth 1989, Au-Yeung, et al., 1998). The finding by Au-Yeung, et al. (1998) that stutterers frequently produce disfluencies on function words immediately preceding content words was consistent with the results of the present thesis. As the group results from the word class comparisons of Chapter 4 showed, stutterers produced a slightly greater percentage of disfluencies on function words than did control subjects.

Stutterers' higher disfluency rate on function words does not, of course, constitute sufficient evidence that function word disfluencies are necessarily anticipatory. However, the observation by Au-Yeung, et al. (1998) that stutterers' function word disfluencies occurred almost exclusively before the content words with which they formed prosodic units (what the authors call *phonological words*) clearly reveals that the relative location of the function word is critical to its disfluent production. For example, in a phrase such as "the boy took it", the first function word, "the", prosodically cliticizes to the following content word, “boy”, while the second function word, “it”, cliticizes to the *previous* content word, “took”. The results from Au-Yeung, et al. (1998) showed that only the function words cliticized to a following content word (e.g., “the boy”) were likely to attract disfluencies. As shown in the prominence analysis of Chapter 4, the results of the present thesis uncovered a similar asymmetrical bias toward the position crucially preceding pitch-accented and nuclear pitch-accented target words, as compared with the position following nuclear pitch-accented words. Thus, it was not the word class of the function word per se that was associated with high disfluency rates, but rather the prosodic position in which the function word surfaced. Moreover, the fact that this position was predictably before the pitch-accented word lends support to the anticipatory interpretation of this disfluency type.

The results of the syllable position analysis in Chapter 5 revealed further 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 reason for this latter claim will be discussed below in greater detail.

An example of a function word both surfacing disfluently and attracting pitch accent is illustrated below in Figure 122. First, the function word “and” is prolonged and pitch-accented, followed by an L- phrase accent. Next, the function word “in” is both prolonged and followed by a post-IP pause, thus receiving the classification ‘4p.ps’:

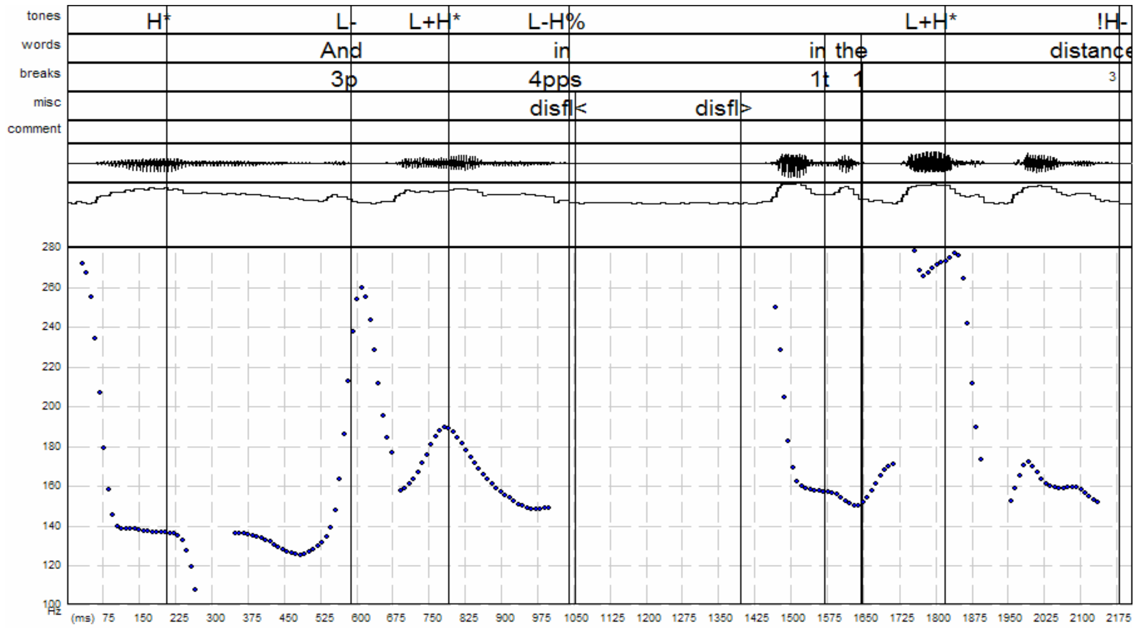


Figure 122: Disfluency on nucleus of pitch-accented function word (S2)

Along with increased duration, the second prolongation causes a pitch accent (L+H*) to be realized on the disfluent function word "in". The fact that neither the pitch accent on "and" nor "in" was intended, but rather was a result of the prolongation, is strongly suggested by the next fluent iteration of "in" in the restart (1t), which precedes the nuclear pitch-accented word "distance." What is particularly striking about this example is the realization not only of pitch accent on the prolonged function word "in", but of the entire final Intonation Phrase boundary contour L-H%. Rather than producing a normal pitch excursion interpolating to the target pitch-accented word "distance", as occurs in the second fluent attempt, the speaker realizes the entire phrase's contour solely on the function word "in". Thus, not only does this prolongation example appear consistent with an anticipatory analysis, since nucleus prolongations imply a postponement of later material, but the realization of upcoming tonal information—i.e., the L+H* pitch accent and subsequent phrase accent—on the prolonged function word gives further evidence that, despite halted segmental production, the speaker had crucially planned a prosodic contour for the rest of the utterance.

6.4 Monitoring and lookahead

In Chapter 2, a general model of language production (Levelt 1989; Levelt, et al., 1999) was reviewed—specifically with respect to the relationship between prosodic and phonological structure generation. A competing model of prosody generation (Keating & Shattuck-Hufnagel 2002) was also outlined, which motivated two key questions for this thesis: 1) How much prosodic structure is available before a word is phonologically encoded? 2) At what level(s) in the language production process are stutterers' and nonstutterers' disfluencies produced? It was hypothesized that an analysis of the patterns of disfluency and their relationship to specific aspects of prosodic phrase structure would provide crucial information for answering these questions. As examples like Figure 120 and Figure 122 show, the presence of segmental and tonal information amid multiple disfluencies can provide clues to determining the locus of a production breakdown, as well as how far in advance a speaker may have detected the breakdown.

In this section, the salient evidence of anticipatory disfluencies found in the stutterers' data of this study will be looked at with respect to contemporary models of speech monitoring.

As discussed briefly in Chapter 2, an influential monitoring model called the "Covert Repair Hypothesis" asserts that stuttering and normal speakers can covertly detect and repair a production error (Kolk 1991; Postma, et al., 1990; Postma & Kolk 1993). The authors' further claim was that the errors the monitor detects are the result of a phonological encoding deficit; that is, while stutterers and non-stutterers both share the same monitoring architecture, and produce disfluencies that are qualitatively identical, stutterers simply produce more disfluencies because a phonological encoding impairment results in markedly more segmental errors in their production.

As has been shown here, the results of the present thesis strongly support the hypothesis that a monitoring system plays some role in locating disfluencies, as evidenced in particular by the distinct type of anticipatory disfluency surfacing in highly restricted environments. However, since the error detection is ostensibly covert, it is trickier to assess to what degree (if any) the breakdown is rooted in dysfunctional phonological encoding.

An example continuing the utterance of Speaker S2 from Figure 122 is illustrated below in Figure 123, again revealing a string of anticipatory disfluencies before a clear target disfluency. The two disfluency types are distinguished by the type and location of disfluency: the first three instances—the prolongation + filler combination ("a"), the post-filler pause (after "uh"), and the restart ("a")—all contain fully produced vowels, and thus are classified as anticipatory disfluencies. The cut onset [b], on the other hand, reveals a failed attempt to produce the syllable's nucleus, thus rendering it a target disfluency.

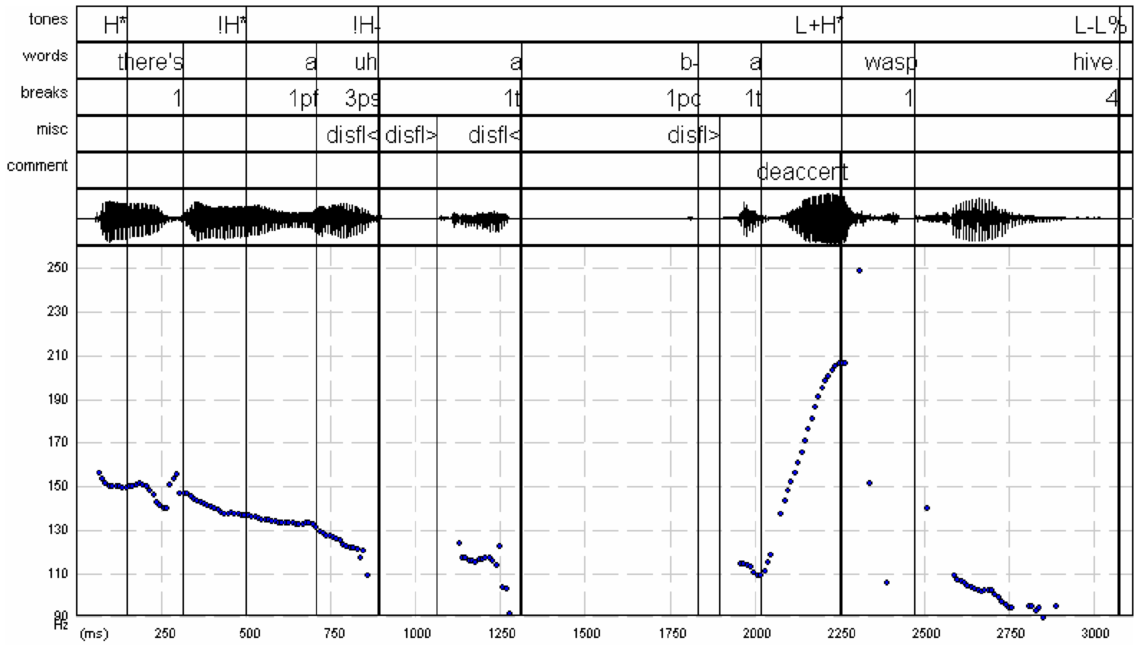


Figure 123: lookahead examples (S2)

Once again, like in Figure 120 and Figure 122, the anticipatory disfluencies signal a production problem not with the words on which they occur, but rather with material coming later in the prosodic structure. As the target analyses revealed, the material detected after the anticipatory disfluency in the speech stream (i.e., the target disfluency) was most frequently metrically prominent, and thus predictable by the intonational structure of the utterance. Such is the case in Figure 123, as the postponed nuclear pitch-accented word "b__" (presumably "beehive") was ultimately produced with an overt disfluency on the prolonged and cut onset [b]. A further interesting detail of this example is the speaker's near-immediate substitution of a semantically related word ("wasphive") for the problematic target word "beehive".

Figure 123 essentially reveals its own target word through both its phonetic clues and its preserved Intonational Phrase structure (i.e., neither the ‘1pc’ prolongation + cut nor the ‘1t’ restart force a new ip/IP). Figure 124, meanwhile, shows the reference prosodic structure as produced by control subject C2:

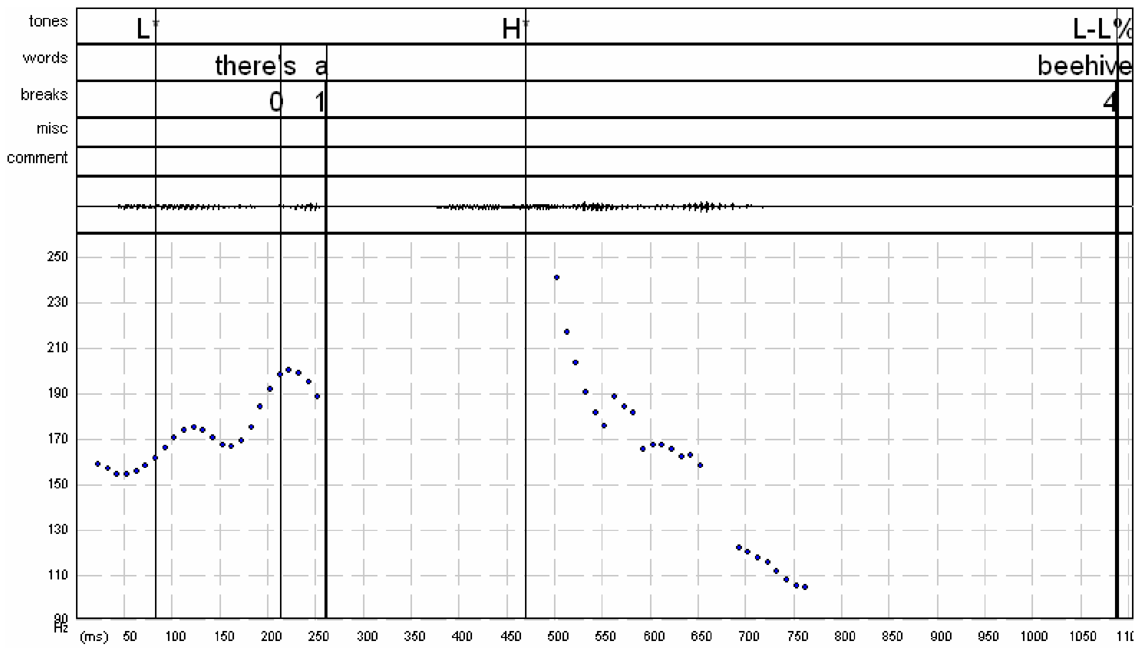


Figure 124: Control (C2) Script of S2, lookahead example

The results of this thesis are 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 will be explored here. On the one hand, the evidence of the current thesis supports 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. Regardless of whether the impairment lies in production or monitoring of prosodic structure, the evidence provided in this study indicates that stutterers are sensitive to the metrical prominence of words located at least one Prosodic Word in advance. The behavior of anticipatory disfluencies suggests that by the time such disfluencies occur, speakers have already accessed key syntactic and semantic information of a given word or phrase (i.e., a word's lemma). But unlike lexical search and retrieval errors, in which a speaker demonstrates no evidence of completed production beyond the lemma, stuttering anticipatory disfluencies indicated prosodic evidence such as realization of upcoming intonational contours (e.g., Figure 122) and continued phonetic disfluency on target words (e.g., Figure 120 and Figure 123). The implication of this evidence for normal production is that before such lexical and post-lexical processes as word-form encoding, speakers have access to a sizable amount of prosodic structure. In the same way, the fact that speakers are aware of production problems in advance of their articulation argues for the influence of prosodic structure in locating these triggering events, presumably via a speech monitoring system.

As was revealed in the distributional analysis of Phonological Phrases in Chapter 5, the predictable distribution of target disfluencies also provided 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.

A number of tentative conclusions may be drawn with respect to these distributional patterns. First of all, the fact that disfluencies were not produced at equal rates for all PhPs implies that certain phrasal positions are more predisposed to triggering disfluencies than others. Moreover, the consistent patterns of disfluency distribution suggest a constraining influence of prosodic phrasal structure. Table 32 lists a series of example ips divided into an increasing number of Phonological Phrases, which are shaded to represent the average number of disfluencies produced in each PhP (reproduced from Figure 31 of Chapter 5). The darkest shade indicates the PhPs with highest number of target disfluencies, while the unshaded boxes indicate PhPs with the lowest number. Pitch-accented syllables are underlined, and nuclear pitch accents are capitalized.

Total PhPs	*PhP1*	*PhP2*	*PhP3*	*PhP4*	*PhP5*	*PhP6*
1	He SAW him					
2	He saw him	in the PARK				
3	He saw him	in the park	with the FROG			
4	He saw	the boy	in the park	with the FROG		
5	He saw	the boy	walking	in the park	with the FROG	
6	He saw	the little	boy	walking	in the park	with the FROG

Table 32: Target disfluency rate for Phonological Phrases

As Table 32 shows, except for in ips consisting of two PhPs, the highest disfluency rates were found in the initial PhP. As PhPs are added, relative disfluency rates decrease for PhP2, PhP4, and PhP5. PhP3, however, interestingly maintains a fairly steady disfluency rate in ips of all lengths. Interpreting these results in terms of a monitoring hypothesis, the consistently high percentage of disfluencies in the first PhP suggests that stutterers detect a problem very early in the utterance, regardless of the number of PhPs within a given ip. One possible reason would be that a speaker may have access to prominence patterns late in the utterance in order to plan a prominence pattern for earlier PhPs. This is true of the predictions of the KSH (2002) model, exemplified in particular by the Beat Movement examples discussed in Chapter 2. Since the nuclear pitch accent is the most prominent element within an ip, a speaker would therefore need to minimally plan an entire ip—i.e., the NPA and all preceding material—before beginning an utterance. As the monitor scans the entire ip to compare input (i.e., originally intended

NPA) with the output (i.e., the result of attempted NPA assignment), a detected lack of agreement between the two forms results in a disfluency at the point of detection—that is, during the attempted articulation of PhP1. Thus, assuming a lookahead window equal to the size of an ip, a problem in nuclear pitch accent detection could explain the high frequency of target disfluencies in the initial PhP.

However, taking into account the moderately high disfluency rates found in PhP3, a second possibility is that a prosodic structure smaller than an ip functions as a secondary planning unit. Assuming still that nuclear pitch accent must first be assigned in order to determine the boundaries of an ip structure, high disfluency rates in stutterers' final PhPs are evidence of a difficulty in assigning this anchor of the ip's right boundary. At the same time, high disfluencies at the left edge of the ip suggest that what is ultimately deficient is the building of the initial ip-sized planning unit. The high percentage of disfluencies generated in PhP3, however, indicate that setting the initial boundary of smaller constituents is also impaired. In particular, a planning unit composed of two PhPs appears to mark the secondary domain of planning difficulty in stuttering. Thus, combining this evidence, it would appear that determining the boundaries necessary for building prosodic structure is problematic in stuttering.

Following the example utterance in Table 32, the first complex ip, "He <u>saw</u> him in the <u>PARK</u>", would allow a lookahead window spanning the entire ip, since it consists of only two PhPs. Nuclear pitch accent would be assigned to the lemma "park" in the second PhP, designating the ip's right edge, and a second pitch accent would be accent to the first PhP's lemma "saw", to mark the left edge. If stutterers are unable to properly mark these boundaries, then the ip structure remains unformed, resulting in a disfluency upon attempting production of *both* PhPs—thus explaining the high disfluency rates for PhP1 and PhP2 in an ip composed of two PhPs. For an ip consisting of 3 PhPs, once again the problem target would lie in forming the two edge PhPs.

Once an ip reaches four PhPs in length, the effects of the secondary planning unit, consisting of two PhPs, become explicit. Continuing with the proposed model, initiating an ip composed of four PhPs again requires assignment of the nuclear pitch accent and first pitch accent in order to set the boundaries of an ip. As in the shorter ips, a disfluency would be expected during attempted production of PhP1, as a result of failed pitch accent assignment on PhP1 and PhP4 during the planning of these boundary PhPs. Since, in shorter ips, PhP1 represents the left edge of both an ip and a secondary unit composed of two PhPs, the effects of improperly setting the left boundary of the secondary unit would be masked by the boundary breakdown of the larger ip. In an ip containing four PhPs, however, these secondary boundary effects would be observable in a high disfluency rate on PhP3, which is what the data in Table 32 show. Compared to PhP2, in which a relatively low percentage of disfluencies were produced, PhP3 represents the beginning of a new secondary planning unit, and thus reflects impaired building of prosodic structure through a higher disfluency rate.

The final two examples—for ips containing five and six PhPs, respectively—again reflect disfluency distribution patterns consistent with the proposed planning units. The relationship between PhP3 and PhP4 again mirrors that of PhP1 and PhP2, as the first PhP in each unit realizes the disfluency from a failure to assign the boundaries of the smaller planning unit. Once again, since the first PhP represents the point at which both an ip boundary and a smaller, dual-PhP boundary, must be determined, its disfluency rate is markedly higher than that of PhP3, whose production requires only planning of a dual-PhP boundary. The only aberration is the absence of disfluency in PhP5, which, since it presumably initiates a new planning unit, should behave similarly to PhP1 and PhP3. However, in accord with previous analyses of English stress assignment (e.g., Selkirk 1984), speakers often avoid stress clashes by shifting stress, or avoiding the assignment of stress to two heads of a prosodic phrase. For instance, although PhP5 would represent the head of a new planning unit, the fact that the very next PhP is already assigned nuclear pitch accent ostensibly blocks PhP5 from initiating a new unit. This would therefore account for the low percentage of disfluencies occurring in PhP5 in this context.

In summary, the model outlined here presupposes a basic planning unit of a single ip, and an additional secondary planning unit consisting of two PhPs. In order to build prosodic structure, a speaker must first set the boundaries of the larger unit (ip), and then incrementally set the left boundary of the smaller unit (dual-PhP). Therefore, if stutterers have a deficit in assigning boundary pitch accents to build prosodic structure, the disfluency effects resulting from this impairment will be most marked in the first and final PhPs of an ip, since an ip represents the primary processing unit. Since further boundaries must be assigned to smaller units—presumably consisting of two PhPs—disfluencies will also surface upon initiating these smaller units, albeit at a lower rate than that resulting from ip boundary assignment. This disfluency pattern then repeats itself with attempts to plan ips composed of increasing numbers of PhPs, resulting in disfluencies manifested at the onset of each of these dual-PhP prosodic units.

With respect to the question of determining the exact relationship between stuttering and pitch accent production, the distribution patterns from Table 32 suggest that stutterers have an impairment not only in planning the boundaries of prosodic structures, but also in realizing NPAs in actual production. Evidence for the latter is demonstrated by the presence of some level of disfluency in the final PhP of all ips, regardless of total PhP number. In particular, cases where disfluencies would not be expected—such as in PhP4 and PhP6 when both occur ip-finally—nevertheless reflect some degree of disfluency. This could be accounted for if what is impaired is not the monitoring mechanism per se, but rather the actual planning of metrically prominent material. In other words, while the primary disfluency would surface when the monitor *covertly* detects the incorrect output (i.e., a lemma without an assigned NPA), a second disfluency would nevertheless surface in the attempted production of the NPA since what the monitor detected—a lemma without an assigned NPA—was never corrected.

The evidence shown here, which reveals that stutterers appear to be sensitive to prosodic breakdowns well before articulation of the problematic material has ensued, also has 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, et al., 1999), the prosody generation model presented here is more similar to those developed by Ferreira (1993) and Keating & Shattuck-Hufnagel (2002). Non-incrementalist models presuppose the accessibility of larger prosodic constituent structures relatively early in the production process—crucially, before phonological encoding. As discussed in the previous section on monitoring, the implication this has for stuttering models is that if higher levels of prosodic constituent structure are available to speakers in pre-lexical production planning, then the ability to employ "lookahead" into later constituents—and, therefore, monitor the planning of these constituents—should also be accessible to a stutterer. Such a model would therefore predict that if stutterers' production breakdowns are due to a component critical to construction of an intonation structure—such as impaired nuclear pitch accent generation and construction of the internal prosodic structure—then the effects of such a breakdown should manifest themselves when they are detected in the speech planning process. Crucially, by the predictions of these models, this would occur pre-lexically, and thus effectively account for the anticipatory disfluencies commonly observed in this thesis.

In addition, the fact that control subjects demonstrated significantly fewer anticipatory disfluencies than stutterers—as well as no consistent disfluency distribution patterns with respect to PhP location—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). In section 6.2 of this chapter, the analysis of disfluencies within their prosodic contexts provides further evidence of this difference in disfluency triggering. For example, control subjects produced significantly 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, errors in which speakers must delay production without having access to any phonological content of the intended word, are most likely either again producing a disfluency resulting from a phonological error (i.e., during attempted lexical retrieval of an entire word form), or potentially from a semantic error produced in early conceptual planning stages. Crucially, however, in errors such as these there is no consistent evidence of prosodically constrained distributional patterns, as the evidence in both anticipatory and target disfluencies in stutterers has shown.

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 ('dual-PhPs').

6.5 Timing and rhythmicity

The final section of this chapter will investigate a phenomenon of metric speech facilitation observed in the speech of stutterers both in this and in former studies.

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 (Fransella & Beech

1965; Fransella 1967; Brady 1969; Brayton & Conture 1978; Howell & El-Yaniv 1987; 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 thesis. On several occasions, all four 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.

In Figure 125, speaker S1 not only successfully produces the phonetic structure of each word in his utterance, but most crucial aspects of intonational structure—in particular, pitch accent realization and prosodic constituent grouping—are fully intact. The only irregularity is in the prominence relations established. In other words, by strictly parsing the utterance with an isochronous foot structure (i.e., trochaic/iambic meter), the speaker adds an atypical pitch accent to the function word "to":

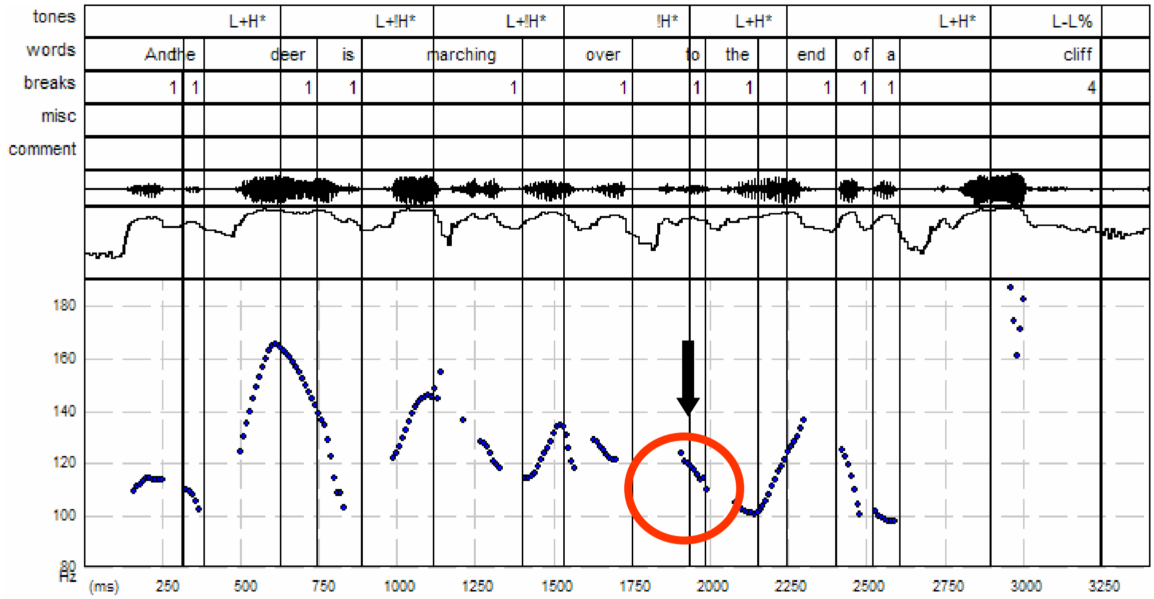


Figure 125: Example 1 (S1) of a rhythmic constraint: isochronous foot structure

Conversely, the control script in Figure 126 reveals both the asynchronous and naturally unpredictable relationship among pitch-accented words that would be expected for this utterance. The function word "to", which was inappropriately pitch-accented in the stuttering data, is produced in the script as both unaccented and with a reduced vowel.

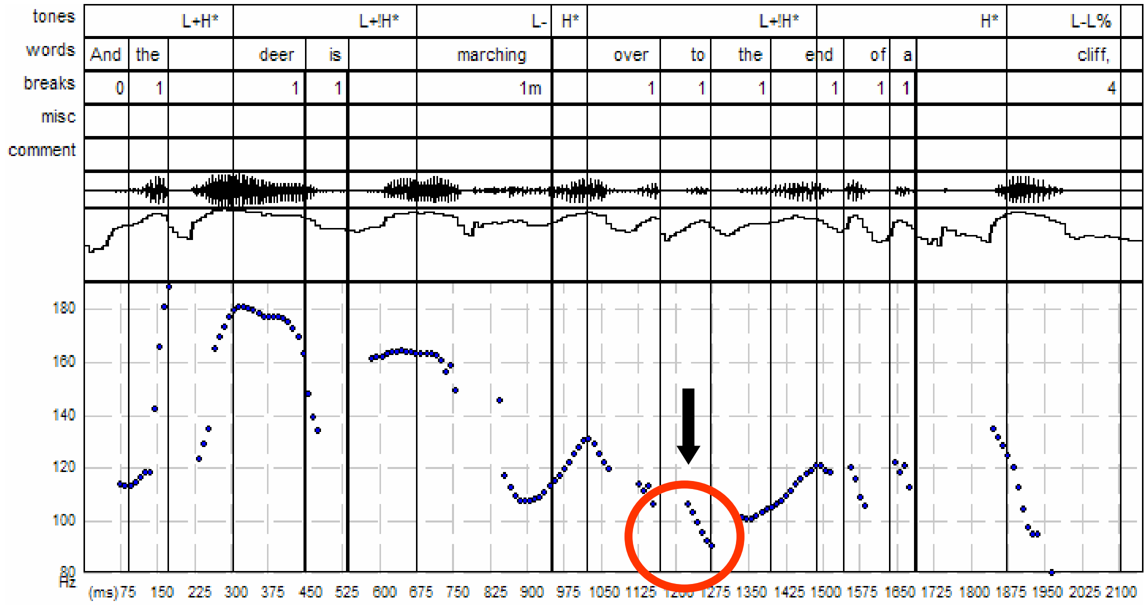


Figure 126: Control (C1) Script of S1, lookahead example

A similar example can be seen in speaker S2's production of a fairly long Intonational Phrase in Figure 127. After a series of early disfluencies, the speaker begins a fluent string of isochronously distributed pitch-accented words. Again, although phonetic production is fully fluent for all words in the remainder of the IP, the application of a strict rhythmic parsing results in an unexpected pitch accent on function words "in" and "of":

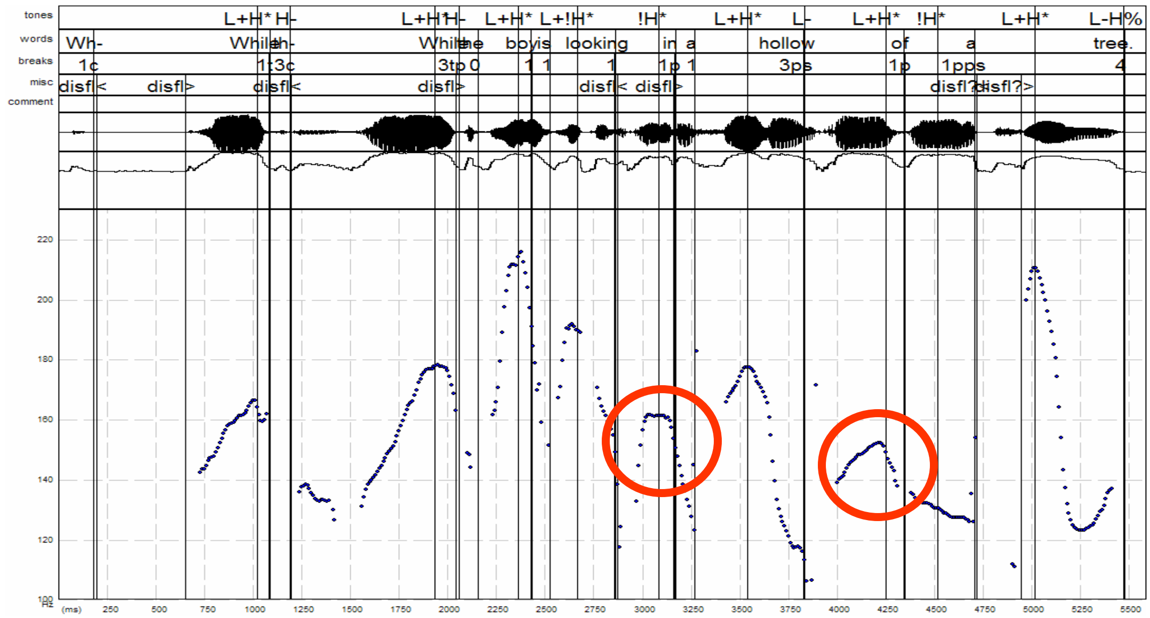


Figure 127: Example 2 (S2) of a rhythmic constraint: isochronous foot structure

While the hypothesis pursued in Vasic & Wijnen (2005)—that the speech monitor in stutterers is improperly responsive to temporal variation and discontinuity in the speech plan—accurately captures data such as the isochronous examples shown above, it is nevertheless still possible that the monitor is detecting a genuine breakdown in prosodic structure, as was initially hypothesized in this thesis and discussed in detail in the last section. 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[21]) 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 crucially 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.

Furthermore, proposing a timing component is also consistent with the stuttering disfluency model proposed in the previous section. In essence, a timing component would obviate the need to assign nuclear pitch accents two Phonological Phrases into the

[21] In the Intonational Phonology model assumed in this thesis, the utterance-final PWd is equivalent to the NPA (if no other PWds are contrastively focused).

planning structure, since a strict, foot-based isochrony requires only that a speaker assign a single pitch accent to a single PhP. That is, each iambic (or trochaic) foot would represent the entire prosodic structure, as strictly-applied isochrony simply repeats individual disyllabic feet until the end of the utterance. It is also possible that a speaker would not need to assign any pitch accents at all, since this pre-determined rhythmic unit would be simply accessed as a stored, pre-planned structure.

Alternatively, if processing breakdowns in stuttering originate not in the prosodic structure but rather in the speech monitoring architecture, as proposed in the over-vigilant speech monitor hypothesis (Vasic & Wijnen 2005), then the fluent examples of isochrony may be the result not so much of a distracted monitor, but of a monitor which compares speech output to a stricter series of prosodic input constraints. Thus, rather than scanning the planning unit for lack of agreement between lemmas with and without assigned NPA, as was proposed for the model in Section 6.4, a rhythmically-constrained monitor would compare two-syllable units of the output with disyllabic trochaic (or iambic) feet from the input. If the output does not conform to this rigid trochaic structure, then the monitor detects this as an error, resulting in an overt disfluency.

6.6 Implications and future work

The results presented in this thesis have implications for theoretical models of stuttering, as well as for models of language production in general. While it is feasible that stuttering is rooted in a complex interaction of factors not limited to linguistic processing, it is nevertheless clear that the linguistic patterns which emerge can help us understand more accurately the origins of a stutterer's production breakdown.

The disfluency patterns found in both the stuttering and control subjects alike can be used to more accurately determine the internal structure of a disfluency. Moreover, these patterns serve as partial windows to the processing mechanisms underlying fluent as well as non-fluent speech. It is hoped that the results found in the experiments here, as well as the rudimentary prosody generation model outlined in this thesis, can contribute to the continued development of current language production models, particularly in traditionally less-understood areas such as the generation of prosodic structure.

The wealth of disfluency data obtained from the experiment tasks made it necessary to expand considerably the ToBI transcription conventions used to code these disfluencies within the model of Intonational Phonology. This was crucial for capturing the many important distinctions between disfluency types in both stutterers' and controls' utterances. In addition, this more elaborate categorization system made it possible to account for and measure the interactions these disfluencies had with various prosodic elements of entire Intonational Phrase.

Finally, the results of this thesis have clinical implications for the field of speech pathology. Linguistic analysis is necessary to help uncover the factors responsible for language breakdown in stuttering, and, ultimately, to inform the direction of therapeutic methodologies used in clinical practice. It is hoped that the work here can contribute to accomplishing this greater goal, and thus in some way make a difference in the lives of communicatively impaired individuals.

REFERENCES

Adams, M., and Hayden, P. (1976). The ability of stutterers and nonstutterers to initiate and terminate phonation during production of an isolated vowel. *Journal of Speech and Hearing Research*, 19, 290-296.

Adams, M. (1987). Voice onsets and segment durations of normal speakers and beginning stutterers. *Journal of Fluency Disorders*, 12, 133-139.

Agnello, J. (1975). Voice onset and voice termination features of stutterers. In L.M. Webster & L.C. Furst (Eds.), *Vocal tract dynamics and dysfluency*. New York: Speech and Hearing Institute.

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

Au-Yeung, J., Howell, P., and Pilgrim, L. (1998). Phonological words and stuttering on function words. *Journal of Speech, Language, and Hearing Research*, 41, 1019-1030.

Beckman, M., and Pierrehumbert, J. (1986). Intonational Structure in Japanese and English. *Phonology Yearbook* 3, 255-309.

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

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

Bergmann, G. (1986). Studies in stuttering as a prosodic disturbance. *Journal of Speech and Hearing Research*, 29, 290-300.

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

Blomgren, M., Nagarajan, S.S., Lee, J.N., Li, T., and Alvord, L. (2003). Preliminary results of a functional MRI study of brain activation patterns in stuttering and nonstuttering speakers during a lexical access task. *Journal of Fluency Disorders*, 28, 337-356.

Blomgren, M., Robb, M., and Chen, Y. (1998). A note on vowel centralization in stuttering and nonstuttering individuals. *Journal of Speech, Language, and Hearing Research*, 41, 1042-1051.

Bloodstein, O. (1995). *A Handbook on stuttering* (5th ed.). Chicago: The National Easter Seal Society.

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

Boutsen, F. (1995). A comparative study of stress timing of stutterers and nonstutterers. *Journal of Fluency Disorders*, 28, 145-155.

Brady, J.P. (1969). Studies on the metronome effect on stuttering. *Behavior Research and Therapy*, 7, 197-204.

Braun, A.R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J.M., Carson, R.E., and Ludlow, C.L.(1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering. *Brain*, 120, 761-784.

Brayton, E.R., and Conture, E.G. (1978). Effects of noise and rhythmic stimulation on the speech of stutterers. *Journal of Speech and Hearing Research*, 21, 276-284.

Brown, S.F. (1938). Stuttering with relation to word accent and word position. *Journal of Abnormal & Social Psychology*. Volume 33, 112-120.

Brown, S.F. (1945). The loci of stutterings in the speech sequence. *Journal of Speech Disorders* 10, pp. 181–192.

Burger, R., and Wijnen, F. (1998). *The effects of accent, linear word position and consonant-vowel transition on stuttering*. Poster presented at the ESCoP Conference, Jerusalem.

Burger, R., and Wijnen, F. (1999). Phonological Encoding and Word Stress in Stuttering and Nonstuttering Subjects. *Journal of Fluency Disorders*, 24, 91-106.

Büring, D. (2003). *Semantics, Intonation and Information Structure.* Handbook article, submitted.

Chang, E., Ohde, R.N., and Conture, E.G (2002). Coarticulation and formant transition rate in young children who stutter. *Journal of Speech, Language, and Hearing Research*, 45, 676-688.

Dayalu, V.N., Kalinowski, J., Stuart, A., Holbert, D., and 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.

Dainora, A. (2001). *An empirically based probabilistic model of intonation in American English*. Ph.D. dissertation, University of Chicago.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 293-321.

Dell, G., Chang, F., and Griffin, Z. (1997). Connectionist models of language production: lexical access and grammatical encoding. *Cognitive Science*, 23, 4, 517-542.

DSM-IV-R (1994). *Diagnostic and statistical manual of mental disorders: IV. Washington*, D.C.: American Psychiatric Association.

Dworzynski, K., Howell, P., and Natke, U. (2003). Predicting stuttering from linguistic factors for German speakers in two age groups. *Journal of Fluency Disorders*, 28, 2, 95-113.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 2, 233-253.

Fox, P.T., Ingham, R.J., Ingham, J.C., Zamarripa, F., Xiong, J., and Lancaster, J.L. (2000). Brain correlates of stuttering and syllable production: A PET performance-correlation analysis. *Brain* 123, 1985-2004.

Fransella, F., and Beech. H.R. (1965). An experimental analysis of the effect of rhythm on the speech of stutterers. *Behaviour Research and Therapy*, 3, 195-201.

Fransella, F. (1967). Rhythm as a distractor in the modification of stuttering. *Behaviour Research and Therapy*, 5, 253-255.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27-52.

Fougeron, C., and Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101, 3728-3740.

Hahn, E.F. (1942). A study of the relationship of stuttering occurrence and phonetic factors in oral reading. *Journal of Speech Disorders*. 7:143-151.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20, 253-306.

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

Hillman, R.E., and Gilbert, H.R. (1977). Voice onset time for voiceless consonants in the fluent reading of stutterers and nonstutterers. *Journal of the acoustical society of America*, 61, 610-611.

Howell, P., Au-Yeung, J., Sackin, S. (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language, and Hearing Research*, 42, 2, 345.

Howell, P., and El-Yaniv, N. (1987). The effects of presenting a click in syllable-initial position on the speech of stutterers: comparison with a metronome click. *Journal of Fluency Disorders*, 12, 249-256.

Howell, P., and Sackin, S. (2000). Speech rate manipulation and its effects on fluency reversal in children who stutter. *Journal of Developmental and Physical Disabilities*, 12, 291-315.

Hubbard, C. (1998). Stuttering, stressed syllables, and onsets. *Journal of Speech, Language, and Hearing Research*, 41, 4, 802-807.

Ingham, R.J. (2001). Brain imaging studies of developmental stuttering. *Journal of Communication Disorders*. Volume 34, Issue 6, 493-516.

Ingham, R.J., Fox, P.T., Ingham, J.C., and Zamarripa, F. (2000). Is Overt Stuttered Speech a Prerequisite for the Neural Activations Associated with Chronic Developmental Stuttering? *Brain and Language* 75, 163-194.

Ingham, R.J., Fox, P.T., Ingham, J.C., Zamarripa, F., Martin, C., Jerabek, P., and Cotton, J. (1996). Functional-lesion investigation of developmental stuttering with PET. *Journal of Speech, Language, and Hearing Research*, Vol. 40, 1334-40.

Jayaram, M. (1983). Phonetic influences on stuttering in monolingual and bilingual stutterers. *Journal of Communication Disorders*, 16, 4, 287-297.

Kaasin, K., and Bjerkan, B. (1982). Critical Words and the Locus of Stuttering in Speech. *Journal of Fluency Disorders*, 7, 433-446.

Kadi-Hanifi, K., and Howell, P. (1992). Syntactic analysis of the spontaneous speech of normally fluent and stuttering children. *Journal of Fluency Disorders*, Volume 17, Issue 3, 151-170.

Keating, P., and Shattuck-Hufnagel, S. (2002). A Prosodic View of Word Form Encoding for Speech Production. *UCLA Working Papers in Phonetics*, 101, 112-156.

Kleinow, J., and Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research*, 43, 548–559.

Klouda, G., and Cooper. W. (1988). Contrastive stress, intonation, and stuttering frequency. *Language and Speech*, 31, 1, 3-20.

Kolk, H. (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.

Koopmans, M., Slis, I.H., and Rietveld, T.C.M. (1992). Stotteren als uiting van spraakplanning. Een vergelijking tussen voorgelezen en spontane spraak. *Stem-, Spraak- en Taalpathologie* 1, 87-101.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, Mass: MIT Press.

Levelt, W.J.M., Roelofs, A., and Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.

Liberman, Mark. (1975). The intonational structure of English. Doctoral dissertation, MIT. [Distributed by Indiana University Linguistics Club.]

Maner, K.J., Smith, A., and Grayson, L. (2000). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research*, 43, 2, 560-573.

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

Melnick, K, and Conture, E. (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, U., Grosser, J., Sandrieser, P., and Kalveram, K.T. (2002). The duration component of the stress effect in stuttering. *Journal of Fluency Disorders*, Volume 27, Issue 4, 305-318.

Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Packman, P., Onslow, M., Coombes, T., and Goodwin, A. (2001). Stuttering and lexical retrieval. *Clinical Linguistics and Phonetics*, Volume 15, no.6, 487-498.

Pitchworks Software, version 6.4.n. (1999-2003). Sciconrd.com.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT, published 1988 by IULC.

Pierrehumbert, J., and Hirschberg, J.(1990). The meaning of intonational contours in the interpretation of discourse. In P.Cohen, J. Morgan, and M. Pollack, (Eds.), *Intentions in Communication*. MIT Press: Cambridge, MA.

Pierrehumbert, J., and Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90-117). Cambridge, UK: Cambridge University Press.

Postma, A., Kolk, H., and Povel, D.J. (1990). Speech planning and execution in stutterers. *Journal of Fluency Disorders*, 15, 49–59.

Postma, A., and Kolk, H. (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, A., and Kolk, H. (1992b). Error monitoring in people who stutter. Evidence against auditory feedback defect theories. *Journal of Speech and Hearing Research*, 35, 1024-1032.

Postma, A., and Kolk, H. (1993). The Covert Repair Hypothesis: Prearticulatory repair Processes in Normal and Stuttered Disfluencies. *Journal of Speech and Hearing Research*, 36, 472-487.

Prins, D., Main, V., and Wampler, S. (1997). Lexicalization in Adults Who Stutter. *Journal of Speech, Language, and Hearing Research*, 40, 373-384.

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

Quarrington, B., Conway, J., and Siegel, N. (1962). An experimental study of some properties of stuttered words. *Journal of Speech and Hearing Research*, 5, 387-394.

Robb, M., Blomgren, M., and Chen, Y. (1998). Formant frequency fluctuation in stuttering and nonstuttering adults. *Journal of Fluency Disorders*, 23, 73-84.

Schiffman, Harold F. (1983). *A Reference Grammar of Spoken Kannada*. Seattle: University of Washington Press.

Selkirk, E.O. (1984). *Phonology and Syntax: The Relation between Sound and Structure*, The MIT Press, Cambridge, MA.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology*, 3, 371-405.

Shriberg, E. (1999). Phonetic consequences of speech disfluency. In *Proceedings of the International Congress of Phonetic Sciences* (ICPhS-99), San Francisco, Vol. I, 619-622.

Silverman, K., and J. Pierrehumbert (1990). The timing of prenuclear high accents in English. In J. Kingston and M. Beckman (Eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge University Press, Cambridge. 72-106.

Soderberg, G.A. (1962). Phonetic influences upon stuttering. *Journal of Speech and Hearing Research*, 5, 315-320.

Stager, S.V., Denman,D.W., and 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.

Subramanian, A., Yairi E., and Amir, O. (2001). Second formant transitions in fluent speech of persistent and recovered preschool children who stutter. *Journal of Communication Disorders*. Volume 36, Issue 1, 59-75.

Taylor, I.K. (1966). The properties of stuttered words. *Journal of Verbal Learning and Verbal Behavior* 5, 112-118.

Tornick, G., and Bloodstein, O. (1976). Stuttering and Sentence Length. *Journal of Speech and Hearing Research*, 19, 4, 651-654.

Vasic, N., and Wijnen, F. (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, N. (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.

Watson, B.C., and Alfonso, P.J. (1982). A comparison of LRT and VOT values between stutterers and nonstutterers. *Journal of Fluency Disorders*, 7, 219-241.

Weiner, A.E. (1984). Stuttering and syllable stress. *Journal of Fluency Disorders*, 9, 4, 301-305.

Wijnen, F., and Boers, I. (1994). Phonological Priming Effects in stutterers. *Journal of Fluency Disorders*, 19, 1-20.

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

Wingate, M.E. (1988). *The structure of stuttering: A psycholinguistic approach*. New York: Springer-Verlag.

Yaruss, J.S., and Conture, E.G. (1993). F2 transitions during sound/syllable repetitions of children who stutter and predictions of stuttering chronicity. *Journal of Speech and Hearing Research*, 36, 883-896.

Yaruss, J.S. (1999). Utterance length, syntactic complexity, and childhood stuttering. *Journal of Speech, Language, and Hearing Research*, 42, 2, 329-344.