Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10–12 September 2005, Aix-en-Provence, France, pp. 13-16.

A comparison of disfluency patterns in normal and stuttered speech

Timothy Arbisi-Kelm & Sun-Ah Jun

Department of Linguistics, UCLA, Los Angeles, CA, USA

Abstract

While speech disfluencies are commonly found in every speaker's speech, stuttering is a language disorder characterized by an abnormally high rate of speech aberrations, including prolongation, cessation, and repetition of speech segments [5]. However, despite the obvious differences between stuttered and normal speech, identifying the crucial qualities that identify stuttered speech remains a significant challenge. A story-telling task was presented to four stutterers and four non-stutterers in order to analyze the prosodic patterns that surfaced from their spontaneous narrations. Preliminary results revealed that the major difference between stutterers' and non-stutterers' disfluencies —aside from the total number—is the type of disfluency and the context affected by the disfluency. Disfluencies in both groups included *prolongation, pause* and *cut*, but stutterers' disfluencies also include *repetition* and combinations of the three (e.g., *cut* followed by *pause*). In addition, stutterers' disfluencies were accompanied by more prosodic irregularities (e.g. pitch accent on function words, creating a prosodic break with degraded phonetic cues) prior to the actual disfluency than non-stutterers' disfluencies, indirectly supporting the overvigilant self-monitoring hypothesis [14, 16, 17].

1. Introduction

Previous stuttering literature has defined disfluencies as the "repetitions or prolongations in the utterance of short speech elements, namely sounds, syllables, and words of one syllable" [18], but the variety of disfluency types have not been translated into a contemporary model of linguistic intonation. Word break patterns are a potentially revealing source of data in stuttered speech, since pauses and prolongations have been claimed to be evidence of processing delays, whether they be early (e.g., conceptual or lexical) or late (e.g., phonological or articulatory). Furthermore, unnatural break patterns may also provoke deviations from normal prosodic patterns, which can in turn result in the loss of information crucial to the intonational meaning of an utterance.

As shown in a number of earlier studies, metrical stress appears to be a trigger of disfluencies in stuttered speech [8, 13]. Arbisi-Kelm [1] manipulated the intonation structure of various sentence types adopting the Autosegmental-metrical model of intonational phonology proposed by Pierrehumbert and her colleagues ([3, 11, 12]) and asked adult stutterers to read the sentences. He transcribed the prosodic patterns of utterances using the English ToBI conventions [2], and found that, with the exception of focused nuclear pitch accent, disfluency patterns follow the hierarchical metrical scale of prominence: that is, stuttered disfluencies are a function of metrical prominence, predictable by properties of intonation [2, 3]. Since pitch accented syllables bear a greater degree of prominence than non-pitch accented syllables, it is predicted that they should attract a higher rate of stuttering disfluency. Consequently, disfluencies should have disruptive effects on the intonation structure in which they occur.

Though some believe that stuttered disfluencies are distinct from that of normal speakers [9, 10], others believe that the difference is simply a matter of degree [4]. Vasic and Wijnen [16] also support the latter view. Adopting Levelt's theory [6] that both overt speech and an internal speech plan are monitored, they claim that the difference between stutterers and non-stutterers are in the degree of self-monitoring. That is, stuttering is a "direct result of an overvigilant monitor". They believe that repairs made often introduce disfluencies rather than prevent them ([17], cited in [14]). Recently, Russell, Corley, & Lickley [14] found that stutterers are more sensitive than non-stutterers in perceiving the disfluencies made by other people, supporting the self-monitoring hypothesis.

In this paper, we will examine the disfluency patterns in normal and stuttered speech and compare the types of disfluencies as well as the effect of each disfluency types on the tonal context and the phrasing. The results will suggest whether the difference in the disfluencies produced by the stutterers and non-stutterers are qualitative or quantitative.

2. Method

2.1. Subjects

Four age-matched adult persons who stutter (stutterers) and four age-matched normal speakers (non-stutterers) were selected to participate in a storytelling task. Participants were recruited from the greater Los Angeles area, through local universities and stuttering support groups. Stuttering level of each participant was moderate-severe, as determined by assessments provided by licensed speech language pathologists.

2.2. Procedure

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 is 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 picture. In order to facilitate the creation of a narrative structure, subjects were instructed to peruse the book before the task and form a general idea of the story.

2.3. Data Analysis

All data samples were audiotaped, sampled at a rate of 11025 Hz, and stored digitally. Using the *PitchWorks* signal analysis software program (SCICON R&D), data files were coded following the ToBI (Tones and Break Indices) transcription conventions for English intonation and juncture (Beckman & Ayers [2]). For tones, five pitch accents (H*, L*, L+H*, L*+H, H+!H*) plus downstepped H tones (e.g., !H*, L+!H*, L*+!H) were used as well as phrase accents (H-, L-, !H-) and boundary tones (H%, L%). For break indices, the existing ToBI labelling for the juncture between words was used: '1' for the default, phrase-medial, word boundary, '3' for the juncture corresponding to an Intermediate phrase boundary, and '4' for the juncture corresponding to an Intonation phrase boundary.

The disfluency labelling in English ToBI is done in two tiers -- break indices (BI) and miscellaneous (misc) tiers. The labelling in the BI tier includes the p diacritic in conjunction with a break index 1, 2, or 3. '1p' (an abrupt cutoff before an actual repair), '2p' (a hesitation pause or prolongation of segmental material without having an intermediate phrase tone), and '3p' (a hesitation pause or a pause-like prolongation where there is an intermediate phrase accent in the tone tier). The misc tier is used when the disfluency spans over a word and is difficult to locate its exact timing. In this case, both the beginning and the end of the event is marked by the diacritics '<' and '>', respectively after the disfluency word such as 'disfl< ... disfl>' or 'repair< ... repair>'.

Since the current disfluency labelling does not cover the variety of disfluent types occurring in stuttered speech, we extended the disfluency diacritics as shown in (1) and labelled in the break index tier after the break index number corresponding to the different degrees of juncture due to this disfluency. For example, when a part of a word is cut, '0c' is used, when a phrase-medial word boundary is cut, '1c' is used, and when a word at the end of an intermediate phrase is cut, '3c' is used. Similarly, '1ps' is used to label a disfluent pause after a word boundary, and '3pr.ps' is used to label a disfluent prolongation followed by a pause after an ip boundary.

(1) cut (c)
  prolongation (pr)
  pause (ps)
  repetition (t)
  combinations of these:
    cut followed by pause (c.ps)
    prolongation followed by pause (pr.ps)
    prolongation followed by cut (pr.c)
    prolongation followed by cut followed by pause (pr.c.ps)

3. Results and Discussion

3.1. Prosodic phrasing and frequency of disfluencies

Labellings have been done for three speakers in each group, but data from two speakers in each group have been analyzed so far and are reported here. Table 1 shows the total number of words, the total number of Intermediate phrases (ip), and the total number of disfluencies in the story narrated by four speakers: sm1 and sm2 are stutterers and cm1 and cm2 are non-stutterers. As expected, stutterers have significantly more number of disfluencies than non-stutterers. All speakers (except for sm1) used a similar number of words in narrating the story. Speaker sm1 used more words to describe the pictures, thus producing a greater number of disfluencies, pitch accents, and ips than the other speakers. However, the average number of words and pitch accents per ip was slightly smaller than those of the other stutterer, sm2, suggesting that sm1's phrase would sound choppier than sm2's. But, since sm1's average number of disfluencies per ip was slightly smaller than that of sm2 (0.89 vs. 1.14 per ip), sm1 may not necessarily be a more severe stutterer than sm2.

On the other hand, the other stutterer, sm2, though he has a large number of disfluencies, seems not much different from cf1, one of the control speakers, in terms of the number of words and pitch accents per ip. This suggests that stutterers may not be qualitatively different from non-stutterers in phrasing words and producing prominent words within the phrase. Rather, it seems that there is a continuum of fluency. Stutterers have fewer words in a phrase than non-stutterers. This may have happened because the disfluency interrupts producing a longer phrase.

Table 1: Total number of words, intermediate phrases (ip), disfluencies, and pitch accents (PA) in the story produced by two stutterers (sm1, sm2) and two non-stutterers (cm1, cf1), and the average and standard deviation of disfluencies and PAs in an ip.

	sm1	Sm2	cm1	cf1
# words	1148	535	492	489
# ip	341	150	109	126
Avg. wd/ip	3.4	3.6	4.5	3.9
#disfl.	314	168	8	9
Avg. disfl./ip	0.92	1.14	0.07	0.07
sd of disfl	0.97	1.01	0.29	0.26
#PA	535	288	272	232
Avg. PA/ip	1.56	1.92	2.39	1.83
sd of PA	0.88	1.07	1.13	0.93

3.2. Types of disfluencies

Though phrasing and pitch accent data seem to suggest a continuum of fluency between stutterers and non-stutterers, the types of disfluency data show a qualitative difference between these two groups. Table 2 lists the types of disfluencies produced by each speaker. It is shown that disfluency types differ between the stutterers and non-stutterers in two ways. First, while *prolongation, pause*, and *cut* types of disfluency occur in both groups, *repetition* occurs only in stutterers' speech. Second, the combination of *prolongation, pause*, and *cut* was observed mainly in the stutterers' speech, though not as frequently as the single type disfluency. In non-stutterers' speech, the combination was found only once by Speaker cm1 (i.e., prolong-pause). However, data show that among the types of disfluency, the proportion of *prolongation, pause* and *cut* is similar between stutterers and non-stutterers. For both groups, *prolongation* or *pause* was the most common and *cut* was the least common.

Table 2: Types of disfluency (single and combinations) for each speaker. In each type, the top row shows a raw number and the bottom row shows the percentage.

Disfluency types	sm1	sm2	cm1	cf1
Prolong (pr)	73	67	2	5
	24%	40%	25%	56%
Pause (ps)	115	39	4	3
	36%	23%	50%	33%
Cut (c)	46	22	1	1
	15%	13%	12.5%	11%
Repetition (t)	11	4	0	0
	4%	2.4%		
Cut-pause (c.p)	7	2	0	0
	2.2%	1.2%		
Prolong-cut	4	2	0	0
(pr.c)	1.3%	1.2%		
Prolong-pause	52	26	1	0
(pr.ps)	16.6%	15%	12.5%	
Prolong-cut-	0	1	0	0
Pause pr.c.ps)	0%	0.6%		
Total	314	168	8	9

3.3. The effect of disfluency on the adjacent contexts

Stutterers' and non-stutterers' speech also differed qualitatively in how the disfluency affected the adjacent tonal contexts and phrasing. Table 3 shows the types of disfluency and their effects (behaviors) and the frequency of each behavior for each speaker. Since the frequencies of each disfluency type produced by non-stutterers are very small, a direct comparison between groups cannot be made. The table is here to provide an idea of how each disfluency type affects the tonal pattern, duration, and phrasing.

In general, the stutterers' disfluencies disturbed the phrasing and tonal context in more various ways than those by the non-stutterers. For example, for both groups, disfluent *prolongation* often created a downstepped pitch accent on the following word and created an intermediate phrase (ip) boundary by extending f0 rise or fall on the prolonged word. However, in stutterers' speech, disfluent prolongations often created pitch accent on the prolonged word and rushed the following words to compensate for the time lost, but this was not the case with non-stutterers' speech. Similarly, for the existence of disfluent *pauses*, both groups often placed an ip boundary at the pause location and produced downstep-like pitch accent on the following word, but stutterers tended to produce pitch reset on the post-pause word and often did not fully produce pre-boundary cues such as phrase-final lengthening or phrase accent. Finally, when disfluent *cuts* were made, the phrasing was rarely affected by speakers of both groups. But one stutterer (sm1) produced pitch reset 28% of the time and the other stutterer (sm2) produced unintended pitch accent on the following word 25% of the time (probably to reinforce failed articulations).

Another noticeable influence of disfluency on prosody was to produce pitch accented function words and rising pitch accent on the disfluent word (e.g. L+H*), creating an incorrect information structure or meaning of contrast. This pattern was found more often in stutterers' speech than non-stutterers' (though we need more non-stutterers' data to confirm this). In sum, the errors in tone and duration and various unorthodox ip patterns found in stutterers' speech suggest that the grouping of pitch accents was jeopardized because of tune-text alignment deficit (i.e., improperly anchored PAs).

Finally, data show that, in stutterers' speech, different types of disfluency were found *before* the target word (e.g. disfluency on the function word before the following content word) as well as on the target word itself. For stutterers, half of the dislfuency data was found before the target word, but for non-stutterers, almost all disfluencies were found during the target word. The target word was always a pitch accented word, and stutterers produced disfluencies before producing the prominent word, i.e., pitch accented target word. This created more disruptions in the prosodic contour and in the listener's understanding of the information structure, resulting in the perception of more disfluent speech. This may be interpreted as evidence that stutterers monitor their speech in advance to produce the target word, and their early repairs result in more disfluencies, thus supporting the self-monitoring hypothesis.

In sum, the data suggest that the disfluencies produced by stutterers and non-stutterers are qualitatively as well as quantitatively different. The data also suggest that analyzing disfluent utterances prosodically (in terms of tonal patterns, breaks, and phrasing) provides valuable data to the studies of disfluent speech.

4. References

[1] Arbisi-Kelm, Timothy. 2004. An Intonational Analysis of Disfluency Patterns in Chronically Stuttered Speech. A poster presented at the ASA.

[2] Beckman, Mary & Gayle Ayers-Elam. 1994. Guidelines for ToBI Labelling.

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

[4] Bloodstein, O. 1970. Stuttering and normal nonfluency: A continuity hypothesis. British J. of Disorders of Communication. 1970, 30-39.

[5] DSM-IV-R. 1994. Diagnostic and statistical manual of mental disorders: IV. Washington, D.C. : American Psychiatric Association.

[6] Levelt, Willem J. M. 1989. *Speaking. From Intention to Articulation*. Cambridge, Massachusetts: MIT Press.

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

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

[9] Perkins, W. H. 1990. What is stuttering? *J. of Speech and Hearing Disorders*, 55:370-382.

[10] Perkins, W. H. 1995. *Stuttering and science*. San Diego, CA: Singular Publishing Group.

[11] Pierrehumbert, J. 1980. The Phonology and Phonetics of English Prosody, Doctoral dissertation, MIT.

[12] Pierrehumbert, J. & M. E. Beckman. 1988. *Japanese Tone Structure*, Cambridge, MA: MIT Press.

[13] Prins, D., Hubbard, C., & Krause, M. 1991. Syllabic Stress and the Occurrence of Stuttering. JSHR 34, 5, 1011-1016.

[14] Russell, M., Corley, M. & Lickley, R. 2005. Magnitude Estimation of disfluency by stutterers and nonstutterers. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F.

Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech*. Hove (East Sussex): Psychology Press.

[15] Shriberg, Elizabeth. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of Berkeley, California.

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

[17] Wijnen, 2000. Stotteren als resultaat van inadequate spraakmonitoring [Stuttering as the result of inadequate speech monitoring]. *Stem-, Spraak- en Taalpathologie*, 9.

[18] Wingate, M. E. 1964. A Standard Definition of Stuttering. Journal of Speech and Hearing Disorders. November 29: 484-9.

Table 3: The effect of disfluency on the adjacent contexts for each speaker.

Disfl patterns	Behavior	sm1	sm2	cm1	cf2
Prolongations	force Pitch Accent (PA) on prolonged word	38	62	0	0
	create timing compensation errors (i.e., rushed after disfluency)	8	7	0	0
	force ip break	64	23	0	1
	force downstep on following word	19	18	1	0
	force pitch reset	3	2	0	0
	force continuation f0 rise/falls	32	48	0	1
	no effect on ip	6	9	3	0
Pauses	force PAs on following word	2	19	0	1
	create timing compensation errors	7	1	0	0
	force ip break	118	27	3	1
	force downstep on following word	8	11	1	0
	force pitch reset	6	3	0	0
	degrade pre-boundary cues (onset closure=block)	46	1	0	0
	no effect on ip	23	10	3	2
Cuts	force PAs on following word	6	7	0	0
	force ip break	9	5	0	0
	force downstep on following word	1	2	0	0
	force pitch reset	18	3	0	0
	no effect on ip	29	9	2	1
Repetitions	force ip break (before or after boundary)	3	1	0	0
	force downstep on same word	1	0	0	0
	force pitch reset	4	0	0	0
	no effect on ip	13	7	0	0
PA on function words		27	62	1	1
Disfluency on target (Nuclear Pitch Accent/Pitch Accent)		**206**	**67**	**8**	**8**
Disfluency NOT on target		**108**	**101**	**0**	**1**