Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10–12 September 2005, Aix-en-Provence, France, pp. 53-58.

Prosodic parallelism as a cue to repetition and error correction disfluency

Jennifer Cole, Mark Hasegawa-Johnson, Chilin Shih, Heejin Kim, Eun-Kyung Lee, Hsin-yi Lu, Yoonsook Mo, Tae-Jin Yoon

University of Illinois at Urbana-Champaign, USA

Abstract

Complex disfluencies that involve the repetition or correction of words are frequent in conversational speech, with repetition disfluencies alone accounting for over 20% of disfluencies. These disfluencies generally do not lead to comprehension errors for human listeners. We propose that the frequent occurrence of parallel prosodic features in the reparandum (REP) and alteration (ALT) intervals of complex disfluencies may serve as strong perceptual cues that signal the disfluency to the listener. We report results from a transcription analysis of complex disfluencies that classifies disfluent regions on the basis of prosodic factors, and preliminary evidence from F0 analysis to support our finding of prosodic parallelism.

1. Acoustic-prosodic correlates of disfluency

Disfluency occurs in spontaneous speech at a rate of about one every 10-20 words, or 6% per word count [17], yet this interruption of fluent speech does not generally lead to comprehension errors for human listeners. Recent research has shown that important cues to disfluency can be found in the syntactic and semantic structures conveyed by the word sequence, and in the phonological and phonetic structures signaled by acoustic features local to the disfluency interval. These cues identify the components of the disfluent region --- the reparandum (REP), edit phrase (EDT), and alteration (ALT) --- and their junctures. Work on automatic disfluency detection has shown that the most successful approach combines both lexical and acoustic features, with explicit models of the lexical-syntactic and prosodic features that pattern systematically with disfluent intervals [1,6].

Of the acoustic-prosodic correlates of disfluency, the post-reparandum pause (filled or unfilled) has been studied the most extensively. Nakatani & Hirschberg's [12] detailed acoustic and classification studies examine duration, F0 and energy, and also report unusual patterns of lengthening, coarticulation, and glottalization near the interruption point of a disfluency. In this paper we examine the nature of prosodic correlates of disfluency in the characteristic patterns of F0, duration and energy that identify and distinguish among various types of disfluency involving word repetition and error correction.

There are distinct types of disfluency that can be characterized in terms of their form and function. Shriberg [16, 17] classifies the disfluencies of the Switchboard corpus into six categories: filled pause ("uh" and "um"), repetition (of one or more words, without correction), substitution (repetition of zero or more words, followed by the correction of the last word in the disfluent interval), insertion, deletion, and speech error. Other work identifies abandonment (fresh start) disfluencies, in addition [6,11,18]. These distinct types of disfluency may be caused by different psychological processes. Levelt [9] suggests that corrections of a single word may result from monitoring of the phonetic plan, while corrections that involve repair or abandonment of an entire phrase may result from monitoring of the pre-syntactic message. Clark & Fox Tree [3] and Clark & Wasow [4] propose a different psychological account for filled pause and repetition disfluencies. In these accounts filled pauses like "uh" and "um" are phonological words that are used by the speaker to signal a delay in the preparation of the upcoming speech. Repetition disfluencies occur when the speaker makes a premature commitment to the production of a constituent, perhaps as a strategy for holding the floor, and then hesitates while the appropriate phonetic plan is formed. The continuation of speech is marked by "backing up" and repeating one or more words that precede the hesitation, as a way of restoring fluent delivery. Henry & Pallaud [7] support the findings of Clark & Wasow [4] by demonstrating that morphological, syntactic, and structural features strongly differentiate repetition disfluencies from word fragment disfluencies. Clark & Wasow [4] note that repetition disfluencies are four times as common as repair disfluencies; they suggest that a small number of repetition disfluencies may be "covert repairs" [9], but that most repetitions are more closely related to filled pause disfluencies than to speech repairs.

The acoustic-prosodic features that serve to cue disfluency vary according to the type of disfluency. Levelt & Cutler [10] observe a contrastive emphasis on the repair segment of an error-correcting disfluency, manifest in increased F0, duration and amplitude. Shriberg [15] and Plauché & Shriberg [13] find that F0 contours, word durations, and the distribution of pauses serve to differentiate among three types of repetition disfluencies. Shriberg [15] describes repetition disfluencies that signal covert repair as having a characteristic reset of the F0 contour to a high, phrase-initial value at onset of the alteration. Similarly, Savova & Bachenko [14] propose an "expanded reset rule," according to which "alteration onsets are dependent on both reparandum onsets and reparandum offsets," echoing the observation of Shriberg [15] that when speakers modify the duration of a repeated word in a repetition disfluency, "they tend to do so in a way that preserves intonation patterns and local pitch range relationships."

In our study of prosody and disfluency in the Switchboard corpus of conversational telephone speech, we observe parallelism in the prosodic features of the REP and ALT phases as characteristic of repetition and error correction disfluencies. Highly similar F0 patterns express a parallel intonation structure that cues the relationship between the REP and ALT for the majority of repetition and error correction disfluencies we have observed. We propose an extended typology of repetition disfluencies in this paper, based on prosodic comparison of REP and ALT. Section 2 describes the methods of our transcription study of disfluency in Switchboard, and section 3 presents frequency data on five sub-categories of repetition and error correction disfluency that are prosodically distinguished based on a comparison of the prosodic features of the REP and ALT intervals. Section 4 reports on preliminary quantitative evidence from F0 data that support our analysis based on perceptual transcription.

2. Method

2.1. Corpus

Switchboard is a corpus which consists of 2500 spontaneous informal telephone conversations [5]. We selected 70 sound files from those conversations, representing 58 different speakers. Within each file we used a random process to excerpt a two minute sound segment. These short files were transcribed for disfluency intervals by the authors, all of whom are trained in acoustic phonetics with prior experience in prosodic transcription using ToBI annotation conventions. 3 transcribers labeled disfluencies for the entire two-minute duration of 10 files each (for a total of 60 minutes of speech) and 5 transcribers labeled only for the first talker turn of duration ranging from 3 to 60 seconds in each of 10 files (for approximately 25 minutes of speech). All eight labelers participated in a series of three group training sessions to assure consistency of labeling criteria, and two group sessions were held for the resolution of problem cases raised by individual labelers.

2.2. Labeling Criteria

Disfluencies are classified by their function into two types, hesitation and repair. These functional categories divide into several subtypes based on lexical and prosodic form. Hesitation disfluencies are classified as repetition, lengthening, silent pause and filled pause. Repair disfluencies are classified as error correction and abandonment. Classification was based on lexical, syntactic, and prosodic factors. Lexical factors are the presence of a repeated word, an error-correcting word substitution, or a filled-pause phrase like "um" or "ah". Syntactic criteria were used to identify instances of phrase abandonment followed by fresh restart and to identify the REP-ALT correspondence in error-correction. Prosodic factors were used to identify lengthening, and provided additional evidence for some cases of error correction (with prosodic emphasis on ALT) and abandonment (with truncation of an intonational tune at the abandoned edge). Labeling was done on the basis of listening and visual inspection of the waveform, spectrogram, F0 and intensity contours, using Praat [2]. The disfluency labels were entered on two tiers in the TextGrid associated with each wave file, and disfluency intervals (REP, EDT, ALT) were aligned with the beginnings and endings of the associated word intervals. Table 1 shows the typology of disfluencies by function and form and the labeling conventions used.

Table 1. Typology of Disfluencies and Labeling Convention

Type of Disfluency		Labeling: 1st Tier	Labeling: 2nd Tier		
Hesitation	Repetition	hesi-r	REP	EDT	ALT
			REP	ALT	
	Lengthening	hesi-l			
	Silent Pause	hesi-s			
	Filled Pause	hesi-f			
Repair	Error Correction	repair-e	REP	EDT	ALT
			REP	ALT	
	Abandonment	repair-a	REP	EDT	
			REP		

Labels on the first disfluency tier identify the type of disfluency (e.g., hesi-r for Hesitation Repetition), while the components of complex disfluencies were individually segmented on the second disfluency tier. A complex disfluency always includes a reparandum (REP) and an alteration (ALT), and may also include an edit phrase (EDT). Hesitation Repetition disfluency labeling is illustrated in Figure 1. The hesi-l label marks hesitation lengthening that can not be attributed to prosodic phrase-final lengthening based on tonal evidence and perceived disjuncture.

In addition, hesi-s denotes a sentence internal silence that interrupts an otherwise fluent phrase, and hesi-f marks an independent occurrence of filled pause expressions such as "um", "uh". For the repair category, repair-e marks an error followed by a self-correction (e.g. "he can stri- he can swing") and repair-a denotes a semantic and syntactic abandonment of the phrase (e.g. "they you know you can't live in Dallas").

…the kids	instead of	[sil]	instead of	teaching…
	hesi-r			
	REP	EDT	ALT	

Figure 1: TextGrid tiers for a Hesitation Repetition disfluency [Switchboard file: SW03719A.wav]

The EDT label marks the occurrence of filled pauses, silent pauses and editing expressions (e.g., "I mean", "you know") between REP and ALT.

Hesitation Repetition and Repair Error Correction disfluencies, the two disfluency types that have both REP and ALT intervals, were further broken down into five sub-classes based on comparison of prosodic features between REP and ALT. These five sub-classes, listed in Table 2, were proposed on the basis of our earlier exploratory analyses with Switchboard samples; the present study was designed to test the adequacy and acoustic correlates of the proposed classification scheme. Data used in the exploratory analysis were not included in the present study. Prosodic features were assessed on the basis of listening in conjunction with visual inspection of the F0 and intensity contours, spectrogram and waveform. Repetitions in which the ALT and REP were judged to have highly similar prosodic patterns, with identical intonation features in a ToBI transcription, were assigned the label suffix '-same'. An example pitch track from a Repetition-Same disfluency is shown in Figure 2. The '-fp' label was used to label examples where the ALT interval had prosody characteristic of a filled pause: low intensity, and low, flat F0, with reduced consonant or vowel articulations. (This pattern was not observed with Repair Error Correction disfluencies.) The '-ip' label was used to label cases where the REP was perceived as the final word in a well-formed intermediate phrase, based on the F0 contour and perceived disjuncture between REP and the onset of ALT. The '-exaggerated' label was applied to examples in which the ALT displayed a similar but exaggerated version of the prosodic pattern of the REP, typically with increased duration, intensity and higher F0 values. In many cases these examples would receive the same ToBI transcription for REP and ALT, with differences in F0 scaling. Finally, the label '-change' was used for examples where the ALT differed prosodically from the REP in its accentuation (different type or location of accent, or presence vs. absence of accent). Hesitation Repetitions labeled in the change subcategory sounded much like error corrections, where the correction was at the level of pragmatic meaning expressed through accent, rather than at the level of word or syntactic meaning. For all disfluency types, the REP interval was further identified as ending in a word fragment (-frag), or a complete word (-nonfrag), but comparison of fragment and non-fragment tokens is not presented here.

Table 2. Types of Repetition: Prosodic Classification

Hesitation-Repetition	**Repair-Error Correction**
hesi-r-same	repair-e-same
hesi-r-fp	repair-e-fp
hesi-r-ip	repair-e-ip
hesi-r-exaggerated	repair-e-exaggerated
hesi-r-change	repair-e-change

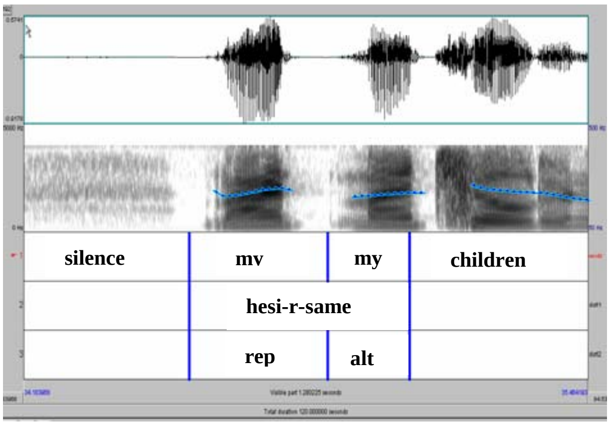

Figure 2. Example of highly similar F0 tracks on REP and ALT (my…my) in Hesitation-repetition-same disfluency: "[sil] my my children…" [Switchboard file: SB03633b]

3. Results

Table 3 provides the number of tokens of each type of disfluency labeled in the corpus, pooling data from all labelers. The most frequent type of disfluency in this corpus is Hesitation, with Silence the most frequent sub-type. Repetitions and Filled Pauses are also frequently occurring Hesitation types. Among Repair disfluencies, Abandonment is fairly common, while error correction and lengthening are infrequent.

Table 3. Distribution of the types of disfluency

		Frequency	Percentage
Hesitation	Silence	299	31.54 %
	Repetition	205	21.62 %
	Filled pause	208	21.94 %
	Lengthening	52	5.48 %
Repair	Abandonment	123	12.97 %
	Error correction	50	5.27 %
Total		948	

Table 4 presents the total number of REP, EDT and ALT intervals, automatically extracted from Hesitation-Repetition and Repair-Error Correction disfluencies in the transcription files. The number of REP and ALT intervals are not equal, due to the occurrence of multiple repetition tokens that contain more than two instances of the repeated word (e.g., "I I uh I tried to…"). For multiple repetitions all but the non-final repetition are coded as independent REP intervals, with the final repetition coded as ALT.

Table 4. Distribution of REP, EDT, and ALT for Hesitation Repetition and Repair Error Correction

	REP	EDT	ALT
Repetition	216	94	205
Error Correcction	51	18	50
Total	267	112	255

The distribution of disfluencies in our corpus over the 10 prosodically-defined sub-classes is shown in Table 5. Repetitions in which REP and ALT have the same prosody (*same*) are the most numerous, and are as frequent as the total of repetitions that mimic filled pauses, cross intermediate phrase boundaries, display exaggerated prosody, or display changed prosody on ALT. These results indicate that while a variety of prosodic patterns are observed over REP and ALT in complex disfluencies, the most common pattern perceived by labelers is that of prosodic parallelism, corresponding to the *same* label, which occurs most frequently with hesitation-repetition disfluencies, but which also occurs as the most frequent pattern with repair-error correction disfluencies.

Table 5. Number of Hesitation Repetition and Repair Error Correction examples by prosodic sub-class.

	same	**fp**	**ip**	**exag**	**change**
Hesi- r	102	22	21	32	28
Repair-e	12	0	4	5	9

The reliability of the labeling scheme was tested by assessing the agreement between pairs of labelers who labeled the same files. Agreement was assessed on a subset of the files labeled for disfluency. Specifically, one or two files from each labeler's bunch were randomly assigned to each of the other labelers for an independent labeling. This second-pass labeling resulted in a set of 59 files labeled independently by two labelers, utilizing all possible labeler pairs. The files labeled in the second-pass labeling were limited to a short interval of between 1.94 – 58.82 seconds, representing the first talker turn of the file, for a total of 1,298 seconds of speech. These files were labeled by second labelers using the same labeling scheme as the first pass labeling. Based on the second-pass labeling of this subset of files, the agreement rate for disfluency type (e.g., hesi-r or repair-a) was 86.82%. When taking into account the subclasses of disfluency in Table 2, the agreement rate among labelers was 85.07 %.

4. F0 Analysis

F0 values were compared between REP and ALT as an empirical measure of intonational similarity. This section describes the method for extracting smoothed F0 contours, time normalization, and a measure of F0 contour difference.

F0 is calculated from short-term autocorrelation and smoothing with Praat [2]. Null values of F0 at the start of a REP or ALT interval, which reflect silence or voiceless segments, are eliminated before aligning initial non-null zero values of REP and ALT. Non-initial frames with null F0 values are discarded in the comparison of REP and ALT F0 contours. Also discarded are any frames in which delta-F0 after smoothing is unexpectedly high or low (change of more than 100 Hz in 10 ms). Four methods of pitch comparison are used in this study: trimmed F0 difference, time-normalized F0 difference, trimmed F0 distance, and time-normalized F0 distance. Trimming and normalization are two methods we use to guarantee that the F0 contours of REP and ALT that we are comparing have the same length. For trimmed F0 analysis, the F0 trajectories of the REP and ALT are compared, where the longer F0 trajectory is trimmed to match the length of the shorter F0. For time-normalized F0, the shorter F0 trajectory (REP or ALT) is time-normalized to match the length of the longer one by using the linear interval interpolation. Because the trimming and time normalization methods do not result in any significant difference in our analyses, only the trimming method is reported here.

The mean F0 difference of REP and ALT is obtained by:

$$\Delta F0 = \frac{\sum_{i,j=1}^{n} (F_0^{(i)} - F_0^{(j)})}{n}$$

Here, *i* is the *i*th sample of REP and *j* is the corresponding sample of ALT, and *n* is the number of samples in the F0 contours (equal for REP and ALT). The F0 difference value is not squared in the equation, because we want to preserve the sign to distinguish cases where REP F0 is scaled higher than ALT from cases which have the opposite scaling relation. We have visually inspected the F0 contour of the REP and ALT sections to be certain that we do not encounter cases where the F0 contours have opposite slopes. In the F0 difference calculation the first F0 values correspond to those of the reparandum (REP) and the second F0 values correspond to those of the alteration (ALT). Thus, when $\Delta F0 > 0$, REP is higher in F0 than ALT and when $\Delta F0 < 0$, ALT is higher in F0 than REP. Figure 3 shows overlaid time-normalized F0 contours for one REP-ALT pair.

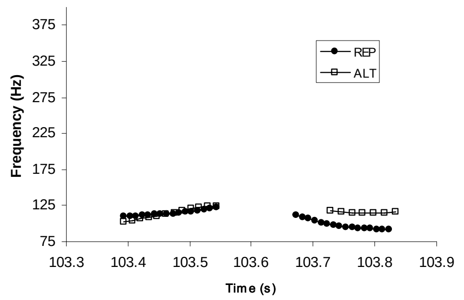

Figure 3. F0 trajectories during the REP (circles) and ALT (squares) segments of a repetition disfluency. Segments are aligned using the time-normalized F0 difference measurement.

Figures 4 and 5 show results of F0 difference comparison that provide evidence for the distinct F0 patterns in our labeling scheme for hesitation repetition and repair error correction, respectively. Both trimmed F0 difference and time-normalized F0 difference show the same trends, thus only box plots of trimmed F0 difference are presented below for the sake of space. Hesi-r-same in Figure 4 and repair-e-same in Figure 5 have mean values close to 0, indicating prosodic parallelism of REP and ALT. Hesi-r-exaggerated and repair-e-exaggerated in Figures 4 and 5 exhibit negative mean values, consistent with our perception that the F0 trajectory of ALT is scaled higher than that of REP, as an exaggeration of the REP F0 contour. The same trend in F0 mean values is shown in hesi-r-ip and repair-e-ip, but in this case the underlying F0 patterns are different than in the exaggerated pattern. An intonational phrase boundary (ip) is perceived at the end of REP, with a pitch reset at the onset of ALT that is responsible for the higher scaling of ALT, especially at the beginning of the ALT interval. The two patterns are further differentiated by the presence or absence of final lengthening and filled or unfilled pauses. The opposite sign (i.e., positive value), for hesi-r-fp is also consistent with our expectation that the ALT in hesi-r-fp functions like a filled pause (e.g., *uh* or *um*), with the characteristically low F0 of a filled pause, scaled lower than the F0 of REP. The effect of hesi-r-change is not strong enough to support our expectation, based on our impression of this sub-class as a kind of prosody repair disfluency, that ALT is scaled higher than REP.

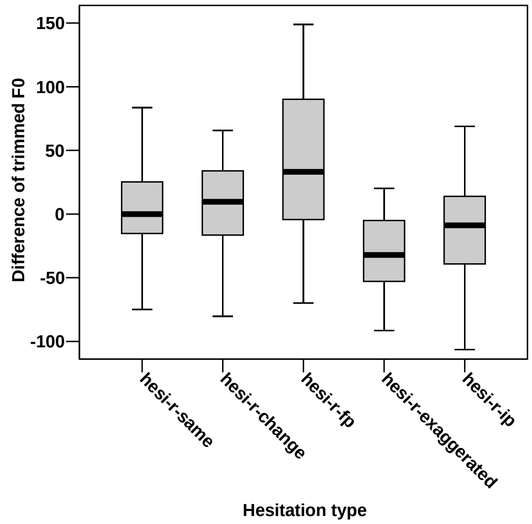

Figure 4. Box plot of trimmed F0 differences between REP and ALT for Hesitation repetition.

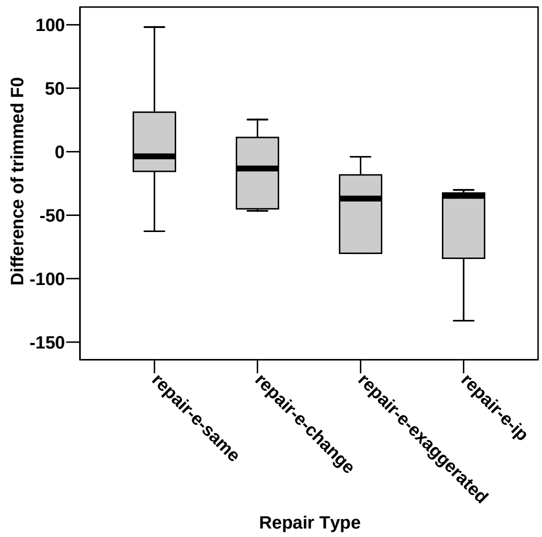

Figure 5: Box plot of trimmed F0 difference between REP and ALT for Repair error correction

The other comparison metric used in this paper is the mean F0 distance metric, which calculates Euclidean distance divided by the number of samples. The F0 difference and F0 distance have the following advantages and disadvantages. F0 difference measures only the difference between average F0 of the REP and average F0 of the ALT; F0 distance, on the other hand, is a measure of the dissimilarity in shape of the two F0 trajectories, when plotted as functions of time. The disadvantage of F0 distance is that it is an unsigned measure: F0 differences may be positive or negative, but all F0 distances are positive. The mean F0 distance of REP and ALT is obtained by:

$$D(F_0) = \sqrt{\frac{\sum_{i,j=1}^{n} (F_0^{(i)} - F_0^{(j)})^2}{n}}$$

Here, as in the mean F0 difference metric, *i* is the *i*th sample of REP and *j* is the corresponding sample of ALT, and *n* is the length of the F0 contours. Division of the Euclidean distance by the square root of the number of samples is done to normalize the effect of sampling length of the tokens in our corpus. (The formula is also known as Root Mean Square (RMS)). As in the difference metric, trimming and time normalization are used for F0 comparison. When D(F0) is close to 0, REP and ALT have similar F0 trajectories, and when the value of D(F0) is large, the F0 trajectories of REP and ALT are different. Thus, when REP and ALT are almost parallel, then the value of D(F0) is expected to be close to 0, and when REP is higher than ALT or ALT is higher than REP, we expect the value of D(F0) to be greater than 0.

Figures 6 and 7 show results of mean F0 distance comparison of hesitation repetition and repair error correction. As noted above, because both trimming and time normalization resulted in similar trends, we present distance of trimmed F0 for hesitation repetition (Figure 6) and repair error correction (Figure 7). Like the results of F0 difference comparison, the F0 distance results confirm our perception of the F0 patterns in the labeling task. In general, hesi-r-same and repair-e-same reveal a small mean distance, near zero in both Figure 6 and Figure 7, while hesi-r-exaggerated and repair-e-exaggerated have the largest mean distance in both Figure 6 and Figure 7. We note that the mean F0 distance of hesi-r-change is also short, indicating that the *change* sub-class exhibits the same prosodic parallelism we predicted and observed for the *same* sub-class.

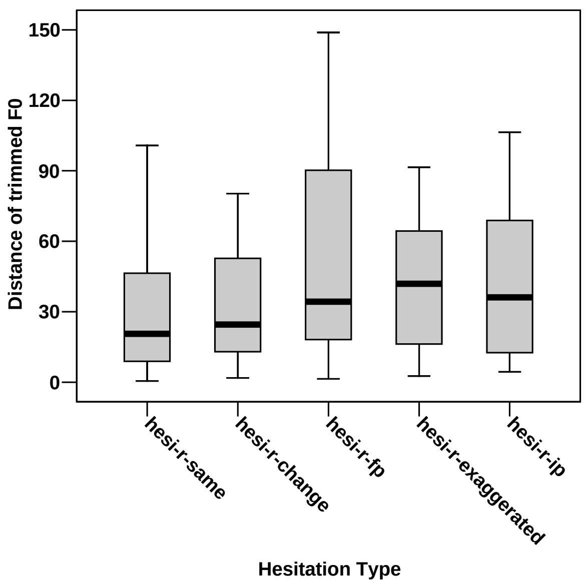

Figure 6: Box plot of trimmed F0 distance between REP and ALT for Hesitation repetition.

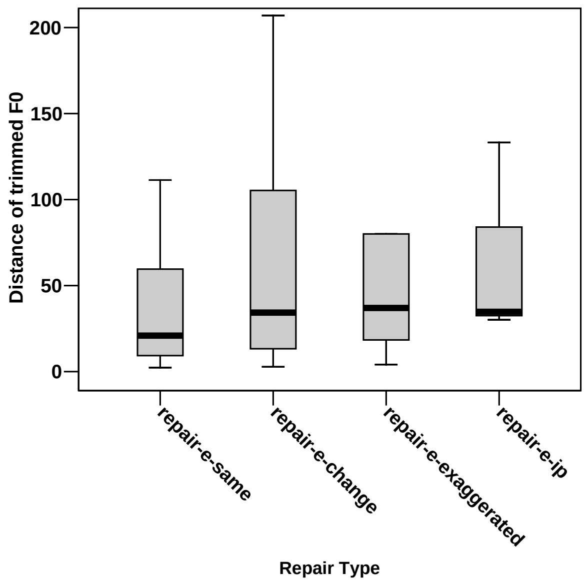

Figure 7: Box plot of trimmed F0 distance between REP and ALT for Repair error correction.

In this section we presented quantitative analysis of F0 comparison using difference and distance metrics on the trimmed and time-normalized intervals of REP and ALT. The overall results of the F0 comparison provide quantitative evidence from F0 measurements for the labelers' perception of the prosodic relationship between REP and ALT in complex disfluencies. Prosodic parallelism is evident in small F0 difference and distance measures for the most frequent sub-classes of hesitation and error correction disfluencies. F0 measures also confirm distinct patterns of F0 relationship between REP and ALT for disfluencies in the *exaggerated* and *ip* sub-classes.

5. Discussion and Conclusion

Our labeling of repetition and error correction disfluencies (Table 5) demonstrated the frequency of five distinct intonational patterns that characterize the prosodic relationship between REP and ALT intervals of complex disfluencies. The most frequent pattern (*same*, with 114 tokens) involved the perceived repetition in the ALT segment of the F0 pattern of the REP segment. The *same* pattern represents almost half of the total number of repetition and error correction disfluencies in this study. The remaining four categories each contain between 1.7 - 12.5% of the total number of complex disfluencies, and represent patterns in which ALT is produced with a low flat F0 (*filled pause*), or patterns where ALT is produced with higher F0 due to pitch reset (*ip*) or higher overall F0 scaling on ALT (*exaggerated*).

Our quantitative measures of F0 provide supporting evidence for the F0 patterns described in our perceptual labeling scheme. Prosodic parallelism of REP and ALT is confirmed by highly similar F0 contours for the largest prosodic sub-class of hesitation and error correction disfluencies. The frequent occurrence of this pattern may provide an important perceptual cue to the listener for the occurrence of disfluency, and may help in the online editing of the disfluency.

Acknowledgment

This research is supported by NSF award number IIS-0414117. Statements in this paper reflect the opinions and conclusions of the authors, and are not endorsed by the NSF.

References

Baron, Don, Elizabeth Shriberg, & Andreas Stolcke. 2002. Automatic punctuation and disfluency detection in multi-party meetings using prosodic and lexical cues. *Proc. ICSLP'02*, Denver, CO, vol. 2, pp. 949-952.

Boersma, Paul & David Weenink. 2005. *Praat: doing phonetics by computer* (version 4.3.04) [Computer Program]. Retrieved March 8, 2005 http://www.praat.org.

Clark, Herbert H. & Jean E. Fox Tree. 2002. Using uh and um in spontaneous speaking. *Cognition*, vol. 84, pp. 73-111.

Clark, Herbert H. & Thomas Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology*, vol. 37, pp.201-242.

Godfrey, John J., Edward C. Holliman, & Jane McDaniel. 1992. Telephone speech corpus for research and development. *Proc. the International Conference on Acoustics, Speech, and Signal Processing,* March 1992, San Francisco, CA, pp. 517-520.

Heeman, Peter A. & James F. Allen. 1999. Speech repairs, intonational phrases and discourse markers: Modeling speakers' utterances in spoken dialogue. *Computational Linguistics,* vol. 25(4), pp. 527-571.

Henry, Sandrine & Berthille Pallaud. 2003. Word fragments and repeats in spontaneous spoken French. *Proc. DiSS'03*, 5-8 September 2003, Goeteborg University, Sweden, pp. 77-80.

Lendvai, Piroska, Antal van den Bosch, & Emile Krahmer. 2003. Memory-based disfluency chunking. *Proc. DiSS'03*, 5-8 September 2003, Goeteborg University, Sweden, pp. 63-66.

Levelt, Willem J. M. 1989. *Speaking. From Intention to Articulation*. Cambridge, MA: MIT Press.

Levelt, William J. M. & Anne Cutler. 1983. Prosodic marking in speech repair. *Journal of Semantics*, vol. 2, pp. 205-217.

Liu, Yang, Elizabeth Shriberg, & Andreas Stolcke. 2003. Automatic disfluency identification in conversational speech using multiple knowledge sources. *Proc. Eurospeech*, Geneva, Switzerland, pp. 957-960.

Nakatani, Christine H. & Julia Hirschberg, 1994. A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, vol. 95(3), pp. 1603-1616.

Plauché, Madelaine C. & Elizabeth Shriberg. 1999. Data-driven subclassification of disfluent repetitions based on prosodic features. *Proc. International Congress of Phonetic Sciences*, San Francisco, CA, vol. 2, pp. 1513-1516.

Savova, Guergana & Joan Bachenko. 2003. Prosodic features of four types of disfluencies. *Proc. DiSS'03*, Goeteberg University, Sweden, pp. 91-94.

Shriberg, Elizabeth. 1995. Acoustic properties of disfluent repetitions. *Proc. International Congress of Phonetic Sciences*, Stockholm, Sweden, vol. 4, pp. 384-387.

Shriberg, Elizabeth. 1996. Disfluencies in Switchboard. *Proc. ICSLP'96*, 3-6 October 1996, Philadelphia, PA, vol. Addendum, pp. 11-14.

Shriberg, Elizabeth. 2001. To `errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, vol. 31(1), pp.153-164.