

A SPEECH-FIRST MODEL FOR REPAIR DETECTION AND CORRECTION

Christine Nakatani
Division of Applied Sciences
Harvard University
Cambridge MA 02138

Julia Hirschberg
2D-450, AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill NJ 07974-0636

ABSTRACT

Interpreting fully natural speech is an important goal for spoken language understanding systems. However, while corpus studies have shown that about 10% of spontaneous utterances contain self-corrections, or REPAIRS, little is known about the extent to which cues in the speech signal may facilitate repair processing. We identify several cues based on acoustic and prosodic analysis of repairs in the DARPA Air Travel Information System database, and propose methods for exploiting these cues to detect and correct repairs.

1. INTRODUCTION

Disfluencies in spontaneous speech pose serious problems for spoken language systems. First, a speaker may produce a partial word or FRAGMENT, a string of phonemes that does not form the complete word intended by the speaker. Some fragments may coincidentally match words actually in the lexicon, as in (1); others will be identified with the acoustically closest lexicon item(s), as in (2).¹

- (1) What is the earliest **fl**i- flight from Washington to Atlanta leaving on Wednesday September fourth?
- (2) *Actual string:* What is the fare **fr**o- on American Airlines fourteen forty three
Recognized string: With fare **four** American Airlines fourteen forty three

Even if all words in a disfluent segment are correctly recognized, failure to detect the location of a disfluency may lead to interpretation errors during subsequent processing, as in (3):

- (3) ...Delta leaving Boston seventeen twenty one arriving Fort Worth **twenty two** twenty one forty and flight number ...

Here, '*twenty two twenty one forty*' must somehow be interpreted as a flight arrival time; the system must choose on some basis among '21:40', '22:21', and '22:40'.

¹We indicate the presence of a word fragment in examples by the diacritic '-'. Self-corrected portions of the utterance, or REPARANDA, appear in boldface. Unless otherwise noted, all repair examples in this paper are drawn from the corpus described in Section 4. Recognizer output shown is from the recognition system described in [1] on the ATIS June 1990 test.

Although studies of large speech corpora have found that approximately 10% of spontaneous utterances contain disfluencies involving self-correction, or REPAIRS [2, 3], little is known about how to integrate repair processing with real-time speech recognition and with incremental syntactic and semantic analysis of partial utterances in spoken language systems. In particular, the speech signal itself has been relatively unexplored as a source of processing cues that may facilitate the detection and correction of repairs. In this paper, we present results from a pilot study examining the acoustic and prosodic characteristics of all repairs (146) occurring in 1,453 utterances from the DARPA Air Travel Information System (ATIS) database. Our results are interpreted within a new "speech-first" framework for investigating repairs, the REPAIR INTERVAL MODEL, which builds upon Labov 1966 [4] and Hindle 1983 [2].

2. PREVIOUS COMPUTATIONAL APPROACHES

While self-correction has long been a topic of psycholinguistic study, computational work in this area has been sparse. Early work in computational linguistics included repairs as one type of ill-formed input and proposed solutions based upon extensions to existing text parsing techniques such as augmented transition networks (ATNs), network-based semantic grammars, case frame grammars, pattern matching and deterministic parsing [5, 6, 2, 7, 8]. Recently, Shriberg et al. 1992 and Bear et al. 1992 [3, 9] have proposed a two-stage method for processing repairs that integrates lexical, syntactic, semantic, and acoustic information. In the first stage, lexical pattern matching rules are used to retrieve candidate repair utterances. In the second stage, syntactic, semantic, and acoustic information is used to filter the true repairs from the false positives. By these methods, [9] report identifying 309 repairs in the 406 utterances in their 10,718 utterance corpus which contained 'nontrivial' repairs and incorrectly hypothesizing repairs in 191 fluent utterances, which represents recall of 76% with precision of 62%. Of the 62% containing self-repairs, [9] report finding the appropriate correction for 57%.

While Shriberg et al. promote the important idea that automatic repair handling requires integration of knowledge from multiple sources, we argue that such "text-first" pattern-

matching approaches suffer from several limitations. First, the assumption that correct text transcriptions will be available from existing speech recognizers is problematic, since current systems rely primarily upon language models and lexicons derived from fluent speech to decide among competing acoustic hypotheses. These systems usually treat disfluencies in training and recognition as noise; moreover, they have no way of modeling word fragments, even though these occur in the majority of repairs. Second, detection and correction strategies are defined in terms of ad hoc patterns; it is not clear how one repair type is related to another or how the set of existing patterns should be augmented to improve performance. Third, from a computational point of view, it seems preferable that spoken language systems detect a repair as early as possible, to permit early pruning of the hypothesis space, rather than carrying along competing hypotheses, as in “text-first” approaches. Fourth, utterances containing overlapping repairs such as (4) (noted in [2, p. 123]) cannot be handled by simple surface structure manipulations.

(4) I think that **it you get**— it’s more strict in Catholic schools.

Finally, on a cognitive level, there is recent psycholinguistic evidence that humans detect repairs in the vicinity of the interruption point, well before the end of the repair utterance [10, 11, 12].

An exception to “text-first” approaches is Hindle 1983 [2]. Hindle decouples repair detection from repair correction. His correction strategies rely upon an inventory of three repair types that are defined in relation to independently formulated linguistic principles. Importantly, Hindle allows non-surface-based transformations as correction strategies. A related property is that the correction of a single repair may be achieved by sequential application of several correction rules.

Hindle classifies repairs as 1) full sentence restarts, in which an entire utterance is re-initiated; 2) constituent repairs, in which one syntactic constituent is replaced by another;² and 3) surface level repairs, in which identical strings appear adjacent to each other. Correction strategies for each repair type are defined in terms of extensions to a deterministic parser. The application of a correction routine is triggered by an hypothesized acoustic/phonetic EDIT SIGNAL, “a markedly abrupt cut-off of the speech signal” (Hindle 1983 [2, p. 123], cf. Labov 1966 [4]), which is assumed to mark the interruption of fluent speech.

Hindle’s methods achieved a success rate of 97% on a transcribed corpus of 1,500 sentences in which the edit signal was

²This is consistent with Levelt 1983’s [13] observation that the material to be replaced and the correcting material in a repair often share structural properties akin to those shared by coordinated constituents.

orthographically represented. This rate of success suggests that identification of the edit signal site is crucial for repair correction.

3. THE REPAIR INTERVAL MODEL

In contrast to “text-first” approaches, we introduce an alternative, “speech-first” model for repair detection/correction, the REPAIR INTERVAL MODEL (RIM). RIM provides a framework for testing the extent to which cues from the speech signal itself can contribute to the identification and correction of repair utterances. RIM incorporates two main assumptions of Hindle 1983 [2]: 1) correction strategies are linguistically rule-governed, and 2) linguistic cues must be available to signal when a disfluency has occurred and to ‘trigger’ correction strategies. As Hindle [2] noted, if the processing of disfluencies were not rule-governed, it would be difficult to reconcile the infrequent intrusion of disfluencies on human speech comprehension, especially for language learners, with their frequent rate of occurrence in spontaneous speech. We view Hindle’s results as evidence supporting the first assumption. Our study tests the second assumption by exploring the acoustic and prosodic features of repairs that might serve as some kind of edit signal for rule-governed correction strategies. While text-first strategies rely upon ‘triggers’ of a lexical nature, we will argue that our speech-first model is consistent with psycholinguistic evidence concerning the human detection of repairs, and is therefore cognitively plausible as well as linguistically principled.

RIM divides the repair event into three consecutive temporal intervals and identifies time points within those intervals which are computationally critical. A full repair comprises three intervals, the REPARANDUM INTERVAL, the DISFLUENCY INTERVAL, and the REPAIR INTERVAL. Following Levelt [13], we identify the REPARANDUM as the lexical material which is to be repaired. The end of the reparandum coincides with the termination of the fluent portion of the utterance and corresponds to the locus of the edit signal. We term this point the INTERRUPTION SITE (IS). The DISFLUENCY INTERVAL extends from the IS to the resumption of fluent speech, and may contain any combination of silence, pause fillers (‘uh’), or CUE PHRASES (‘Oops’ or ‘I mean’), which indicate the speaker’s recognition of his/her performance error. RIM extends the edit signal hypothesis that repairs are phonetically signaled at the point of interruption to include acoustic-prosodic phenomena across the disfluency interval. The REPAIR INTERVAL corresponds to the uttering of the correcting material, which is intended to ‘replace’ the reparandum. It extends from the offset of the disfluency interval to the resumption of non-repair speech. In (5), for example, the reparandum occurs from 1 to 2, the disfluency interval from 2 to 3, and the repair interval from 3 to 4.

(5) Give me airlines **1** [flying to Sa-] **2** [SILENCE uh

SILENCE] 3 [flying to Boston] 4 from San Francisco next summer that have business class.

4. ACOUSTIC-PROSODIC CHARACTERISTICS OF REPAIRS

We report results from a pilot study on the acoustic and prosodic correlates of repair events as defined in the RIM framework. Our corpus consisted of 1,453 utterances by 64 speakers from the DARPA Airline Travel and Information System (ATIS) database [14, 15]. The utterances were collected at Texas Instruments and at SRI and will be referred to as the “TI set” and “SRI set,” respectively. 132 (9.1%) of these utterances contained at least one repair, and 48 (75%) of the 64 speakers produced at least one repair. We defined repairs for our study as the self-correction of one or more phonemes (up to and including sequences of words) in an utterance.

Orthographic transcriptions of the utterances were prepared by DARPA contractors according to standardized conventions. The utterances were labeled at Bell Laboratories for word boundaries and intonational prominences and phrasing following Pierrehumbert’s description of English intonation [16, 17]. Disfluencies were categorized as REPAIR (self-correction of lexical material), HESITATION (“unnatural” interruption of speech flow without any following correction of lexical material), or OTHER DISFLUENCY. For RIM analysis, each of the three repair intervals was labeled. All speech analysis was carried out using Entropics WAVES software [18].

4.1. Identifying the Reparandum Interval

From the point of view of repair detection and correction, acoustic-prosodic cues to the onset of the reparandum would clearly be useful in the choice of appropriate correction strategy. However, perceptual experiments by Lickley and several co-authors [10, 11, 12] show that humans do *not* detect an oncoming disfluency as early as the onset of the reparandum. Subjects *were* able to detect disfluencies in the vicinity of the disfluency interval — and sometimes before the last word of the reparandum. Reparanda ending in word fragments were among those few repairs subjects detected at the interruption site (i.e. the RIM IS), but only a small number of the test stimuli contained such fragments [11]. In our corpus, about two-thirds of reparanda end in word fragments.³

Based on these experimental results, the reparandum offset is the earliest time point where we would expect to find evidence of Labov’s and Hindle’s hypothesized edit signal. In RIM, the notion of the edit signal is extended conceptually to include any phenomenon which may contribute to the perception of an “abrupt cut-off” of the speech signal — including phonetic cues such as coarticulation phenomena, word fragments, inter-

³Shriberg et al. found that 60.2% of repairs in their corpus contained fragments.

| Syllables | Tokens (N=117) | % |
|-----------|----------------|-------|
| 0 | 44 | 37.6% |
| 1 | 60 | 51.3% |
| 2 | 11 | 9.4% |
| 3 | 1 | 0.9% |
| 4 | 1 | 0.8% |

Table 1: Length of Reparandum Offset Word Fragments

ruption glottalization, pause, and prosodic cues which occur from the reparandum offset through the disfluency interval. Our acoustic and prosodic analysis of the reparandum interval focuses on identifying acoustic-phonetic properties of word fragments, as well as additional phonetic cues marking the reparandum offset.

To build a model of word fragmentation for eventual use in fragment identification, we first analyzed the length and initial phoneme classes of fragment repairs. Almost 90% of fragments in our corpus are one syllable or less in length (Table 1). Table 2 shows the distribution of initial phonemes for all fragments, for single syllable fragments, and for single consonant fragments. From Table 2 we see that single consonant fragments occur six times more often as fricatives than as the next most common phoneme class, stop consonants. However, fricatives and stops occur almost equally as the initial consonant in single syllable fragments. So (regardless of the underlying distribution of lexical items in the corpus), we find a difference in the distribution of phonemic characteristics of fragments based on fragment length, which can be modeled in fragment identification.

We also analyzed the broad word class of the speaker’s intended word for each fragment, where the intended word was recoverable. Table 3 shows that there is a clear tendency for fragmentation at the reparandum offset to occur on content words rather than function words. Therefore, systems that rely primarily on lexical, semantic or pragmatic processing to detect and correct repairs will be faced with the problem of reconstructing content words from very short fragments, a

| Phoneme Class | % of All Fragments (N=117) | % of Single Syllable Fragments (N=60) | % of Single Consonant Fragments (N=44) |
|--------------------|----------------------------|---------------------------------------|--|
| stop | 21% | 28% | 11% |
| vowel | 15% | 18% | 7% |
| fricative | 44% | 25% | 73% |
| nasal/glide/liquid | 15% | 22% | 9% |
| h | 3% | 7% | 0% |

Table 2: Feature Class of Initial Phoneme in Fragments by Fragment Length

| Lexical Class | Tokens | % |
|---------------|--------|-------|
| Content | 61 | 52.1% |
| Function | 13 | 11.1% |
| Unknown | 43 | 36.8% |

Table 3: Lexical Class of Word Fragments at Reparandum Offset (N=117)

task that even human transcribers find difficult.⁴

One acoustic cue marking the IS which Bear et al. [9] noted is the presence of INTERRUPTION GLOTTALIZATION, irregular glottal pulses, at the reparandum offset. This form of glottalization is acoustically distinct from laryngealization (creaky voice), which often occurs at the end of prosodic phrases; glottal stops, which often precede vowel-initial words; and epenthetic glottalization. In our corpus, 29.5% of reparanda offsets are marked by interruption glottalization.⁵ Although interruption glottalization is usually associated with fragments, it is not the case that fragments are usually glottalized. In our database, 61.7% of fragments are not glottalized and 16.3% of glottalized reparanda offsets are not fragments.

Finally, sonorant endings of fragments in our corpus sometimes exhibited coarticulatory effects of an unrealized subsequent phoneme. When these effects occur with a following pause (see Section 4.2), they could be used to distinguish fragments from full phrase-final words — such as ‘*fli-*’ from ‘*fly*’ in Example (1).

To summarize, our corpus shows that most reparanda offsets end in word fragments. These fragments are usually intended (where that intention is recoverable) to be content words, are almost always short (one syllable or less) and show different distributions of initial phoneme class depending on their length. Also, fragments are sometimes glottalized and sometimes exhibit coarticulatory effects of missing subsequent phonemes. These properties of the reparandum offset might be used in direct modeling of word fragmentation in speech recognition systems, enabling repair detection for a majority of repairs using primarily acoustic-phonetic cues. Besides noting the potential of utilizing distributional regularities and other acoustic-phonetic cues in a speech-first approach to repair processing, we conclude that the difficulty of recovering intended words from generally short fragments makes a text-first approach inapplicable for the majority class of fragment repairs.

4.2. Identifying the Disfluency Interval

In the RIM model, the disfluency interval (DI) includes all cue phrases, filled pauses, and silence from the offset of the

⁴Transcribers were unable to identify intended words for over one-third of the fragments in our corpus.

⁵Shriberg et al. report glottalization on 24 out of 25 vowel-final fragments.

reparandum to the onset of the repair. While the literature contains a number of hypotheses about this interval (cf. [19, 3]), our pilot study supports a new hypothesis associating fragment repairs and the duration of pauses following the IS.

Table 4 shows the average duration of DIs in repair utterances compared to the average length of utterance-internal silent pauses for all fluent utterances in the ATIS TI set. Although, over all, DIs in repair utterances are shorter than utterance-internal pauses in fluent utterances, the difference is only weakly significant ($p < .05$, $t_{stat} = 1.98$, $df = 1325$). If we break down the repair utterances based on fragmentation, we find that the DI duration for fragments is significantly shorter than for nonfragments ($p < .01$, $t_{stat} = 2.81$, $df = 139$). The fragment DI duration is also significantly shorter than fluent pause intervals ($p < .001$, $t_{stat} = 3.39$, $df = 1268$), while there is no significant difference for nonfragment DIs and fluent utterances. So, while DIs in general appear to be distinct from fluent pauses, our data indicate that the duration of DIs in fragment repairs could be exploited to identify these cases as repairs as well as to distinguish them from nonfragment repairs. While Shriberg et al. claim that pauses can be used to distinguish false positives from true repairs for two of their patterns, they do not investigate the use of pausal duration as a primary cue for repair detection.

4.3. Identifying the Repair

Several influential studies of acoustic-prosodic repair cues have relied upon lexical, semantic, and pragmatic definitions of repair types [20, 13]. Levelt & Cutler 1983 [20] claim that repairs of erroneous information (ERROR REPAIRS) are marked by increased intonational prominence on the correcting information, while other kinds of repairs such as additions to descriptions (APPROPRIATENESS REPAIRS) generally are not. We investigated whether the repair interval is marked by special intonational prominence relative to the reparandum for repairs in our corpus.

To obtain objective measures of relative prominence, we compared absolute f_0 and energy in the sonorant center of the last accented lexical item in the reparandum with that of the first accented item in the repair interval.⁶ We found a small but reliable increase in f_0 from the end of the reparandum to the beginning of the repair (mean = 5.2 Hz, $p < .001$, $t_{stat} = 3.16$, $df = 131$). There was also a small but reliable increase in amplitude across the DI (mean = +2 db, $p < .001$, $t_{stat} = 4.83$, $df = 131$). We analyzed the same phenomena across utterance-internal fluent pauses for the ATIS TI set and found no similarly reliable changes in either f_0 or intensity — perhaps because the variation in the fluent population was much greater than the observed changes for the repair population. And when

⁶We performed the same analysis for the last and first syllables in the reparandum and repair respectively; results did not substantially differ from those reported here for accented values.

| Utterance Type | Mean | Std Dev | N |
|------------------------|----------|----------|------|
| Fluent pauses | 513 msec | 15 msec | 1186 |
| All repairs | 389 msec | 57 msec | 146 |
| a) Fragment repairs | 252 msec | 32 msec | 94 |
| b) Nonfragment repairs | 637 msec | 143 msec | 52 |

Table 4: Duration of Disfluency Intervals vs. Utterance-Internal Fluent Pauses

we compared the f_0 and amplitude changes from reparandum to repair with those observed for fluent pauses, we found no significant differences between the two populations.

So, while small but reliable differences in f_0 and amplitude exist between the reparandum offset and the repair onset, we conclude that these differences do not help to distinguish repairs from fluent speech. Although it is not entirely straightforward to compare our objective measures of intonational prominence with Levelt and Cutler's perceptual findings, our results provide only weak support for theirs. While we find small but significant changes in two correlates of intonational prominence from the reparandum to the repair, the distributions of change in f_0 and energy for our data are unimodal; when we separate repairs in our corpus into Levelt and Cutler's error repairs and appropriateness repairs, statistical analysis does *not* support Levelt and Cutler's claim that only the former group is intonationally 'marked'.

Previous studies of disfluency have paid considerable attention to the vicinity of the IS but little to the repair offset. Yet, locating the repair offset (the end of the correcting material) is crucial for the delimitation of segments over which correction strategies operate. One simple hypothesis we tested is that repair interval offsets are intonationally marked by minor or major prosodic phrase boundaries. We found that the repair offset co-occurs with minor phrase boundaries for 49% of TI set repairs. To see whether these boundaries were distinct from those in fluent speech, we compared the phrasing of repair utterances with phrasing predicted for the corresponding 'correct' version of the utterance. To predict phrasing, we used a procedure reported by Wang & Hirschberg 1992 [21] that uses statistical modeling techniques to predict phrasing from a large corpus of labeled ATIS speech; we used a prediction tree that achieves 88.4% accuracy on the ATIS TI corpus. For the TI set, we found that, for 40% of all repairs, an actual boundary occurs at the repair offset where one is predicted; and for 33% of all repairs, no actual boundary occurs where none is predicted. For the remaining 27% of repairs for which predicted phrasing diverged from actual phrasing, for 10% a boundary occurred where none was predicted; for 17%, no boundary occurred when one was predicted.

In addition to these difference observed at the repair offset, we also found more general differences from predicted phrasing over the entire repair interval, which we hypothesize

may be partly understood as follows: Two strong predictors of prosodic phrasing in fluent speech are syntactic constituency [22, 23, 24], especially the relative inviolability of noun phrases [21], and the length of prosodic phrases [23, 25]. On the one hand, we found occurrences of phrase boundaries at repair offsets which occurred within larger NPs, as in (6), where it is precisely the noun modifier — not the entire noun phrase — which is corrected.⁷

- (6) Show me all n- round-trip | flights | from Pittsburgh | to Atlanta.

We speculate that, by marking off the modifier intonationally, a speaker may signal that operations relating just this phrase to earlier portions of the utterance can achieve the proper correction of the disfluency. We also found cases of 'lengthened' intonational phrases in repair intervals, as illustrated in the single-phrase reparandum in (7), where the corresponding fluent version of the reparandum is predicted to contain four phrases.

- (7) What airport is it | is located | what is the name of the airport located in San Francisco

Again, we hypothesize that the role played by this unusually long phrase is the same as that of early phrase boundaries in NPs discussed above. In both cases, the phrase boundary delimits a meaningful unit for subsequent correction strategies. For example, we might understand the multiple repairs in (7) as follows: First the speaker attempts a VP repair, with the repair phrase delimited by a single prosodic phrase '*is located*'. Then the initially repaired utterance '*What airport is located*' is itself repaired, with the reparandum again delimited by a single prosodic phrase, '*What is the name of the airport located in San Francisco*'.

While a larger corpus must be examined in order to fully characterize the relationship between prosodic boundaries at repair offsets and those in fluent speech, we believe that the differences we have observed are promising. A general speech-first cue such as intonational phrasing could prove useful both for lexical pattern matching strategies as well as syntactic

⁷Prosodic boundaries are indicated by '|'.

constituent-based strategies, by delimiting the region in which these correction strategies must seek the repairing material.

5. DISCUSSION

In this paper, we propose a "speech-first" model, the Repair Interval Model, for studying repairs in spontaneous speech. This model divides the repair event into a reparandum interval, a disfluency interval, and a repair interval. We present empirical results from acoustic-phonetic and prosodic analysis of a corpus of spontaneous speech. In this study, we found that most reparanda offsets ended in word fragments, usually of (intended) content words, and that these fragments tended to be quite short and to exhibit particular acoustic-phonetic characteristics. We found that the disfluency interval could be distinguished from intonational phrase boundaries in fluent speech in terms of duration of pause, and that fragment and nonfragment repairs could also be distinguished from one another in terms of the duration of the disfluency interval. For our corpus, repair onsets could be distinguished from reparandum offsets by small but reliable differences in f_0 and amplitude, and repair intervals differed from fluent speech in their characteristic prosodic phrasing. We are currently analyzing a larger sample of the ATIS corpus to test our initial results and to evaluate other possible predictors of repair phenomena.

REFERENCES

1. Lee, C.-H., Rabiner, L. R., Pieraccini, R., and Wilpon, J. Acoustic modeling for large vocabulary speech recognition. *Computer Speech and Language*, 4:127-165, April 1990.
2. Hindle, D. Deterministic parsing of syntactic non-fluencies. In *Proceedings of the 21st Annual Meeting*, pages 123-128, Cambridge MA, 1983. Association for Computational Linguistics.
3. Shriberg, E., Bear, J., and Dowding, J. Automatic detection and correction of repairs in human-computer dialog. In *Proceedings of the Speech and Natural Language Workshop*, pages 419-424, Harriman NY, 1992. DARPA, Morgan Kaufmann.
4. Labov, W. On the grammaticality of everyday speech. Paper Presented at the Linguistic Society of America Annual Meeting, 1966.
5. Weischedel, R. M. and Black, J. Responding to potentially unparseable sentences. *American Journal of Computational Linguistics*, 6:97-109, 1980.
6. Carbonell, J. and Hayes, P. Recovery strategies of parsing extragrammatical language. *American Journal of Computational Linguistics*, 9(3-4):123-146, 1983.
7. Weischedel, R. M. and Sondheimer, N. K. Meta-rules as a basis for processing ill-formed input. *American Journal of Computational Linguistics*, 9(3-4):161-177, 1983.
8. Fink, P. E. and Biermann, A. W. The correction of ill-formed input using history-based expectation with applications to speech understanding. *Computational Linguistics*, 12(1):13-36, 1986.
9. Bear, J., Dowding, J., and Shriberg, E. Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog. In *Proceedings of the 30th Annual Meeting*, pages 56-63, Newark DE, 1992. Association for Computational Linguistics.
10. Lickley, R. J., Bard, E. G., and Shillcock, R. C. Understanding disfluent speech: Is there an editing signal? In *Proceedings of the International Congress of Phonetic Sciences*, pages 98-101, Aix-en-Provence, 1991. ICPhS.
11. Lickley, R. J., Shillcock, R. C., and Bard, E. G. Processing disfluent speech: How and when are disfluencies found? In *Proceedings of the Second European Conference on Speech Communication and Technology*, Vol. III, pages 1499-1502, Genova, September 1991. Eurospeech-91.
12. Lickley, R. J. and Bard, E. G. Processing disfluent speech: Recognising disfluency before lexical access. In *Proceedings of the International Conference on Spoken Language Processing*, pages 935-938, Banff, October 1992. ICSLP.
13. Levelt, W. Monitoring and self-repair in speech. *Cognition*, 14:41-104, 1983.
14. Hemphill, C. T., Godfrey, J. J., and Doddington, G. R. The atis spoken language systems pilot corpus. In *Proceedings of the Speech and Natural Language Workshop*, pages 96-101, Hidden Valley PA, June 1990. DARPA.
15. MADCOW. Multi-site data collection for a spoken language corpus. In *Proceedings of the Speech and Natural Language Workshop*, pages 7-14, Harriman NY, February 1992. DARPA, Morgan Kaufmann.
16. Pierrehumbert, J. B. *The Phonology and Phonetics of English Intonation*. PhD thesis, Massachusetts Institute of Technology, September 1980. Distributed by the Indiana University Linguistics Club.
17. Pierrehumbert, J. B. and Beckman, M. E. *Japanese Tone Structure*. MIT Press, Cambridge MA, 1988.
18. Talkin, D. Looking at speech. *Speech Technology*, 4:74-77, April-May 1989.
19. Blackmer, E. R. and Mitton, J. L. Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39:173-194, 1991.
20. Levelt, W. and Cutler, A. Prosodic marking in speech repair. *Journal of Semantics*, 2:205-217, 1983.
21. Wang, M. Q. and Hirschberg, J. Automatic classification of intonational phrase boundaries. *Computer Speech and Language*, 6:175-196, 1992.
22. Cooper, W. E. and Sorenson, J. M. Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America*, 62(3):683-692, September 1977.
23. Gee, J. P. and Grosjean, F. Performance structure: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15:411-458, 1983.
24. Selkirk, E. O. Phonology and syntax: The relation between sound and structure. In Freyjeim, T., editor, *Nordic Prosody II: Proceedings of the Second Symposium on Prosody in the Nordic language*, pages 111-140, Trondheim, 1984. TAPIR.
25. Bachenko, J. and Fitzpatrick, E. A computational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics*, 16(3):155-170, 1990.