PROSODY/PARSE SCORING AND ITS APPLICATION IN ATIS

N. M. Veilleux *M. Ostendorf*

Electrical, Computer and Systems Engineering
Boston University, Boston, MA 02215

ABSTRACT

Prosodic patterns provide important cues for resolving syntactic ambiguity, and might be used to improve the accuracy of automatic speech understanding. With this goal, we propose a method of scoring syntactic parses in terms of observed prosodic cues, which can be used in ranking sentence hypotheses and associated parses. Specifically, the score is the probability of acoustic features of a hypothesized word sequence given an associated syntactic parse, based on acoustic and "language" (prosody/syntax) models that represent probabilities in terms of abstract prosodic labels. This work reports initial efforts aimed at extending the algorithm to spontaneous speech, specifically the ATIS task, where the prosody/parse score is shown to improve the average rank of the correct sentence hypothesis.

1. INTRODUCTION

Human listeners bring several sources of information to bear in interpreting an utterance, including syntax, semantics, discourse, pragmatics and prosodic cues. Prosody, in particular, provides information about syntactic structure (via prosodic constituent structure) and information focus (via phrasal prominence), and is encoded in the acoustic signal in terms of timing, energy and intonation patterns. Since computer knowledge representations are not as sophisticated as human knowledge, utterances that are straightforward for a human to interpret may be "ambiguous" to an automatic speech understanding system. For this reason, it is useful to include as many knowledge sources as possible in automatic speech understanding, and prosody is currently an untapped resource. In fact, some syntactic ambiguities can be resolved by listeners from prosody alone [1].

One way to incorporate prosody in speech understanding is to score the expected prosodic structure for each candidate sentence hypothesis and syntactic parse in relation to the observed prosodic structure. In a speech understanding system where multiple sentence hypotheses are passed from recognition to natural language processing, the prosody/parse score could be used to rank hypotheses and associated parses, directly or in combination with other scores. The parse scoring approach was proposed in previous work [2], where automatically detected prosodic phrase breaks were scored either in terms of their correlation with prosodic structure predicted from parse information or in terms of their likelihood according to a probabilistic prosody/syntax model. Recently, the parse scoring approach was reformulated [3] to avoid explicit recognition of prosodic patterns, which is a sub-optimal intermediate decision. Specifically, the new score is the probability of a hypothesized word sequence and associated syntactic parse given acoustic features, where both an acoustic model and a "language" (prosody/syntax) model are used to represent the probability of utterance, analogous to speech recognition techniques. The parse scoring formalism was also extended to incorporate phrasal prominence information, in addition to phrase breaks. In previous work, we demonstrated the feasibility of using parse scoring to find the correct interpretation in a corpus of professionally read ambiguous sentences. In this work, we use the parse scoring approach to rerank a speech understanding system's N-best output, specifically in the ATIS task domain, in order to improve sentence understanding accuracy.

In the following section, we describe the parse scoring system and the probabilistic acoustic and prosody/syntax models. Next, we discuss issues that arose in extending the parse scoring algorithm to the ATIS task, including several modifications needed to handle new problems associated with spontaneous speech and the new parser and recognizer. We then present experimental results for the task of reranking the top N recognizer hypotheses and associated parses using prosody/parse scores. Finally, we discuss the implications of the results for future work.

2. PARSE SCORING

2.1. General Formalism

The goal of this work is to reorder the set of N-best recognizer hypotheses by ranking each hypothesis and associated parse in terms of a prosody score. More specifically, the prosody-parse score is the probability of a sequence of acoustic observations $\mathbf{x} = \{x_1, \ldots, x_n\}$ given the hypothesized parse, p(**x**|parse), where **x** is a sequence of

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1993		**00-00-1993 to 00-00-1993**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Prosody/Parse Scoring and Its Application in Atis	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Electrical, Computer and Systems Engineering,Boston University,Boston,MA,02215**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**6**	

duration and f0 measurements associated with the recognizer output. We compute this probability using an intermediate phonological representation of a sequence of abstract prosodic labels $\mathbf{a} = \{a_1, \ldots, a_n\}$:

$$\mathrm{p}(\mathbf{x}|\mathrm{parse}) = \sum_{\mathbf{a}} \mathrm{p}(\mathbf{x}|\mathbf{a})\mathrm{p}(\mathbf{a}|\mathrm{parse}). \quad (1)$$

This representation implies the development of two probabilistic models: an acoustic model of prosodic patterns, $\mathrm{p}(\mathbf{x}|\mathbf{a})$, and a model of the relationship between prosody and syntax $\mathrm{p}(\mathbf{a}|\mathrm{parse})$, analogous to a language model in speech recognition.

The general formalism can accommodate many types of abstract labels in the prosodic pattern sequence **a**. Here, the prosodic labeling scheme is an extension of that proposed in [1] and includes integer break indices, one for each word to indicate prosodic constituent structure, and a binary indicator of presence vs. absence of prominence on every syllable. Thus, the prosodic label sequence is given by $\mathbf{a} = (\mathbf{b}, \mathbf{p})$, where **b** represents the break sequence and **p** represents the prominence sequence. To simplify the current implementation, we assume **b** and **p** are independent. This assumption implies the use of two acoustic models, $\mathrm{p}(\mathbf{x}|\mathbf{b})$ and $\mathrm{p}(\mathbf{x}|\mathbf{p})$, and two prosody/syntax models, $\mathrm{p}(\mathbf{b}|\mathrm{parse})$ and $\mathrm{p}(\mathbf{p}|\mathrm{parse})$. (Relaxation of the independence assumption is discussed in Section 5.)

Both the acoustic and prosody/syntax models make use of (different) binary decision trees. A binary decision tree [4] is an ordered sequence of binary questions that successively split the data, ultimately into sets associated with the tree's terminal nodes or leaves. Decision trees are particularly useful for prosody applications because they can easily model feature sets with both categorical and continuous variables without requiring independence assumptions. During training, the sequence of questions is selected from a specified set to minimize some impurity criterion on the sample distribution of classes in the training data. For typical classification problems, a leaf would then be associated with a class label. In this work, however, leaves are associated with the posterior distribution of the classes given the leaf node, and the tree can be thought of as "quantizing" the feature vectors. Here, the classes are either the different levels of breaks, one after each word, or the binary prominence labels, one for each syllable.

2.2. Acoustic Model

The acoustic models, one for breaks and one for prominences, are based on decision trees originally developed for automatic prosodic labeling [5, 6]. The form of the two models is essentially the same. The break model, for example, represents the probability distribution of the different breaks at a word boundary $\mathrm{p}(b|T_{Ab}(x))$, where $T_{Ab}(x)$ is the terminal node of the acoustic break tree corresponding to observation x. Assuming the observations are conditionally independent given the breaks, the probability of the observation sequence is given by

$$\mathrm{p}(\mathbf{x}|\mathbf{b}) = \prod_{i=1}^{n} \mathrm{p}(x_i|b_i) = \prod_{i=1}^{n} \frac{\mathrm{p}(b_i|T_{Ab}(x_i))\mathrm{p}(x_i)}{\mathrm{p}(b_i)}$$

using the decision tree acoustic model. The probability $\mathrm{p}(\mathbf{x}|\mathbf{p})$ is computed using a similar formula with a separate acoustic tree $T_{Ap}(x)$ trained to model prominence.

The key differences between the two acoustic models are in the labels represented and the acoustic features used. The break model represents several different levels of breaks, while the prominence model represents ± prominence. Breaks are associated with words and prominence markers are associated with syllables, so the observation sequences for the two models are at the word level and syllable level, respectively. Both models rely on features computed from speech annotated with phone and word boundary markers found during speech recognition. Phonetic segmentations facilitate the use of timing cues, that in this work are based on segment duration normalized according to phone-dependent means and variances adapted for estimated speaking rate. The observation vectors used in the break model T_{Ab} [5] include features associated with normalized phone duration and pause duration. The observation vectors used to model prominence T_{Ap} [6] include similar features, as well as F0 and energy measurements.

2.3. Prosody/Syntax Model

The break and prominence prosody/syntax models are also based on decision trees, in this case originally designed for synthesis applications. Hirschberg and colleagues have proposed the use of decision trees to predict presence vs. absence of prosodic breaks [7] and of pitch accents [8], with very good results. Our use of trees for prosody/syntax models differs from this work, in the number of prosodic labels represented, in the use of trees to provide probability distributions rather than classification labels, and in the use of trees for parse scoring rather than prediction. Again, the break and prominence models share the same basic form. The leaves of the prosody/syntax break tree T_{Sb}, for example, are associated with a probability distribution of the breaks given the syntactic feature vector z_i, $\mathrm{p}(b|T_{Sb}(z_i))$. These probabilities are used directly in computing $\mathrm{p}(\mathbf{b}|\mathrm{parse})$, assuming the breaks are conditionally independent given

the quantized features $T_{Sb}(z_i)$:

$$\mathrm{p}(\mathbf{b}|\mathrm{parse}) = \prod_{i=1}^{n} \mathrm{p}(b_i|T_{Sb}(z_i)).$$

Again, the probability p(**p**|parse) can be computed using the same approach but with a separate prosody/syntax prominence tree T_{Sp}.

For all prosody/syntax models, the feature vectors used in the tree are based on part-of-speech tags and syntactic bracketing associated with the hypothesized word sequence. For the break model T_{Sb}, the feature vectors (one for each word) include content/function word labels, syntactic constituent labels at different levels of bracketing, measures of distance in branches from the top and the bottom of the syntactic tree, and location in the sentence in terms of numbers of words. For the prominence model T_{Sp} [9], the feature vectors (one for each syllable) include part-of-speech labels, lexical stress assignment and syllable position within the word.

2.4. Joint Probability Score

Using the acoustic and prosody/syntax models and the independence assumptions described above, the probability of the acoustic observations $\mathbf{x} = (\mathbf{x}^{(b)}, \mathbf{x}^{(p)})$ given an hypothesized parse is:

$$\mathrm{p}(\mathbf{x}|\mathrm{parse}) = \mathrm{p}(\mathbf{x}^{(b)}|\mathrm{parse})\mathrm{p}(\mathbf{x}^{(p)}|\mathrm{parse})$$

where the break models contribute to the term

$$\mathrm{p}(\mathbf{x}^{(b)}|\mathrm{parse}) = \prod_{i=1}^{n_w} \mathrm{p}(x_i) \sum_{b} \frac{\mathrm{p}(b|T_{Ab}(x_i))\mathrm{p}(b|T_{Sb}(z_i))}{\mathrm{p}(b)}$$

and the prominence models contribute a similar term. If the problem is to rank different hypothesized parses for the same word sequence, i.e., the same observation sequence $\mathbf{x}$, then the term $\prod_i \mathrm{p}(x_i)$ can be neglected. However, if different observation sequences are being compared, as is the case for different recognition hypotheses, then an explicit model of the observations is needed. Since the acoustic model readily available to this effort does not provide the $\mathrm{p}(x_i)$ information, we simply normalize for differences in the length of the word sequence (n_w) and of the syllable sequence (n_s):

$$\begin{aligned} S_J &= \frac{1}{n_w} \sum_{i=1}^{n_w} \log \sum_{b} \frac{\mathrm{p}(b|T_{Ab}(x_i))\mathrm{p}(b|T_{Sb}(z_i))}{\mathrm{p}(b)} \\ &+ \frac{1}{n_s} \sum_{i=1}^{n_s} \log \sum_{p} \frac{\mathrm{p}(p|T_{Ap}(x_i))\mathrm{p}(p|T_{Sp}(z_i))}{\mathrm{p}(p)}. \quad (2) \end{aligned}$$

The score given by Equation 2 differs from the probabilistic score reported in previous work [2] primarily in that it uses the probability of breaks at each word boundary rather than a single detected break, but also in that it incorporates information about phrasal prominence.

3. APPLICATION TO ATIS

The speech corpus is spontaneous speech from the ATIS (Air Travel Information Service) domain, collected by several different sites whose efforts were coordinated by the MADCOW group [10]. The ATIS corpus includes speech from human subjects who were given a set of air travel planning "scenarios" to solve via spoken language communication with a computer. Queries made by the subjects are classified differently according to whether they are evaluable in isolation (class A), require contextual information (class D) or having no canonical database answer (class X), but these distinctions are ignored in our work. In the ATIS task domain, speech understanding performance is measured in terms of response accuracy with a penalty for incorrect responses, as described in [11]. Our experiments will not assess understanding accuracy, which is a function of the complete speech understanding system, but rather the rank of the correct answer after prosody/parse scoring.

A subset of the ATIS corpus was hand-labeled with prosodic breaks and prominences for training the acoustic and prosody/syntax models. Since the spoken language systems at the various data collection sites differ in their degree of automation, mode of communication, and display, the training subset was selected to represent a balanced sample from each of four sites (BBN, CMU, MIT and SRI) and from males and females. The October 1991 test set is used in the experiments reported in Section 4.

The prosody/parse scoring mechanism was evaluated in the context of the MIT ATIS system [12], which communicates the top N recognition hypotheses to the natural language component for further processing. The speech recognition component, the SUMMIT system, was used to provide phone alignments for the acoustic model. The SUMMIT system uses segment-based acoustic phone models, a bigram stochastic language model and a probabilistic left-right parser to provide further linguistic constraints [12]. TINA, MIT's natural language component [13], interleaves syntactic and task-specific semantic constraints to parse an utterance. As a result, the parse structure captures both syntactic and semantic constituents. For example, parse tree nodes may be labeled as *CITY-NAME* or *FLIGHT-EVENT* rather than with general syntactic labels. In addition, TINA falls back on a robust parsing mechanism when a complete parse is not found, using a combination of the basic parser and discourse processing mechanism ap-

plied within the utterance [14]. The robust parser enables TINA to handle many more queries, which may be difficult to parse because they contain complex and/or incomplete syntactic structures, disfluencies, or simply recognition errors. The robust parser assigns constituent structure to as much of the utterance as possible and leaves the unassigned terminals in the word string, and therefore generates bracketings with a flatter syntactic structure than that for a complete parse.

In order to port our models and scoring algorithm to the ATIS task, the first change needed was a revision to the prosodic labeling system to handle spontaneous speech phenomena. The changes included the addition of two markers introduced in the TOBI prosodic labeling system [15]. First, the diacritic "p" was added to break indices where needed to indicate that an exceptionally long pause or lengthening occurs due to hesitation [15]. As in our previous work, we used a seven level break index system to represent levels in a constituent hierarchy, a superset of the TOBI breaks. (The binary accent labels represent a simplification or core subset of the TOBI system.) The "p" diacritic is used fairly often: on 5% of the total breaks, on 14% of the breaks at levels 2 and 3, and somewhat more often in utterances that required a robust parse. In addition, a new intonational marker, %r, was added to indicate the beginning of an intonational phrase when the previous phrase did not have a well-formed terminus, e.g. in the case of repairs and restarts. The %r marker was rarely used and therefore not incorporated in the models. Two other prosodic "break" labels were added to handle problems that arose in the ATIS corpus: "L" for linking was added for marking the boundaries within a lexical item (e.g. *San L Francisco*) and "X" for cases where the labelers did not want to mark a word boundary between items (e.g. after an interrupted word such as *fli-*). The different break markers were grouped in the following classes for robust probability estimates in acoustic modeling: (0,1,L), 2, 3, 4-5, 6, (2p,3p), and (4p,5p). In these experiments, the relatively few sentences with an "X" break were simply left out of the training set.

Another new problem introduced by the ATIS task was the definition of a "word", an important issue because prosodic break indices are labeled at each word boundary. The human labelers, the SUMMIT recognition system and the TINA natural language processing system all used different lexicons, differing on the definition of a "compound word" (e.g. *air-fare, what-is-the*). These differences were handled in training by: defining word boundaries according to the smallest unit marked in any of the three systems, using the MIT lexicons to associate the parse and recognition word boundaries, and assigning any hand-labeled "L" breaks to "1" where the recognizer or parser indicated a word boundary. In testing, only the mapping between the recognition and natural language components is needed, and again the smallest word units are chosen.

The main changes to the acoustic model in moving to the ATIS task were associated with the particular phone inventory used by the SUMMIT system. The differences in the phone inventory resulted in some minor changes to the syllabification algorithm (syllable boundaries are needed for acoustic feature extraction). In addition, the phone label set was grouped into classes for estimating robust duration means and variances. We also revised the pause duration feature to measure the total duration of all interword symbols.

The changes to the prosody/syntax model simply involved defining new questions for the decision tree design. The first change involved introducing new categories of parse tree bracketing labels, in part to handle the different naming conventions used in TINA and in part to take advantage of the semantic information provided by TINA. In addition, new types of questions were added to handle cases that included non-branching non-terminals, specifically, questions about the full level of bracketing and the bracketing defined only by binary branching non-terminals (i.e., using two definitions of the "bottom" of the syntactic tree) and questions about the non-terminal labels at multiple levels. Because of the differences in syntactic structure for word strings associated with a robust parse as opposed to a complete parse, we chose to model the prosody of breaks given a robust parse separately, which is equivalent to forcing the first branch of the tree to test for the use of the robust parser.

In summary, many changes were necessary in porting the algorithm to ATIS, some of which were required by the task of understanding spontaneous speech while others were specific to the particular recognizer and parser used here.

4. EXPERIMENTS

In the experimental evaluation of the prosody/parse scoring algorithm on ATIS, the acoustic and prosody/syntax models were trained on the subset of ATIS utterances that were hand-labeled with prosodic markers. The acoustic model was trained from phonetic alignments provided by the MIT recognizer, where the recognizer output was constrained to match the transcribed word sequence. The prosody/syntax model was trained from TINA parses of the transcribed word sequence.

For the parse scoring experiments, MIT provided the N

best recognition hypotheses and one parse per hypothesis for each utterance in the October 1991 test set. The sentence accuracy rate of the top recognition hypothesis, before any prosodic or natural language processing, was 32%. We rescored the top 10 hypotheses, choosing the same number used by the current version of the MIT ATIS system. 185 of 383 utterances (48%) included the correct word string in the top 10. Excluding a few other sentences because of processing difficulties, a total of 179 utterances were used in evaluating improvements in rank due to prosody. For each sentence hypothesis, we extracted a sequence of acoustic features from the phone alignments and F0 contours and a sequence of syntactic features from the associated parse. Thus, every utterance yielded ten sequences of acoustic observation vectors and ten associated sequences of parse features, one pair for each of the ten-best hypothesized word sequences. Each observation sequence was then scored according to the syntactic structure of the corresponding parse, yielding $p(\mathbf{x}_i|\text{parse}_i)$, $i = 1, \ldots, 10$ for each utterance.

The prosody/parse score was used as one component in a linear combination of scores, also including the MIT SUMMIT acoustic score and language model score, which was used to rerank the sentence hypotheses. We investigated the use of a combined prosody score and separate break and prominence scores, and separating the scores gave slightly better performance. The weights in the linear combination are estimated on the October 1991 data, using the method reported in [16]. (Although this is not a fair test in the sense that we are training the three weights on the test set, our experiments in recognition indicate that performance improvements obtained typically translate to improvements on independent test sets.) The acoustic scores were normalized by utterance length in frames, and the other scores by utterance length in words. We compared the rankings of the correct word string for the score combination using only the MIT acoustic and language scores with the rankings according to the score combination that also used the prosody/parse probability. The average rank of the correct utterance, for those in the top 10 to begin with, moved from 1.87 without the prosody score to 1.67 with the prosody score, a gain of about 23% given that the best rank is 1.0. A paired difference test indicates that the difference in performance is significant ($t_\alpha = 2.47$, $\alpha/2 < .005$). In addition, we noticed that incorporation of the prosody score rarely dropped the rank of the correct sentence by more than one, whereas it often improved the rank by more than one.

5. DISCUSSION

In summary, we have described a prosody/parse scoring criterion based on the probability of acoustic observations given a candidate parse. The model is general enough to handle a variety of prosodic labels, though we have focused here on prosodic breaks and prominences. Motivated by the good results in previous experiments with this algorithm on professionally read speech, the goal of this work was to extend the model to spontaneous speech and evaluate its usefulness in the context of an actual speech understanding system, i.e. the MIT ATIS system. Experimental results indicate that prosody can be used to improve the ranking of the correct sentence among the top N. We expect the improved ranking will translate to improved understanding accuracy, though clearly this needs to be confirmed in experiments with a spoken language system.

There are several alternatives for improving both the acoustic and prosody/syntax models. In particular, the current score uses a heuristic to account for differences in observation sequences, which could be better handled by explicitly representing $p(x|a)$ rather than the posterior probability $p(a|x)$ in the acoustic model. Other possible extensions include relaxation of independence assumptions, in particular the independence of breaks and prominences, since other work [9] has shown that breaks are useful for predicting prominence. Of course, this would require increased amounts of training data and somewhat more complex algorithms for computing the parse score. Finally, these experiments represent initial efforts in working with the MIT recognizer and parser, and new acoustic and syntactic features might take better advantage of the MIT system.

The parse scoring algorithm is trained automatically and is in principal easily extensible to other tasks and other speech understanding systems. However, our effort to evaluate the algorithm in the ATIS domain raised some issues associated with portability. New prosodic labels were added to accommodate hesitation and disfluency phenomena observed in spontaneous speech, a problem that we expect will diminish as prosodic labeling conventions converge. Problems arose due to the differences in the definition of a "word" among component modules in the system, which might be addressed by standardization of lexical representation and/or by additional changes to prosodic labeling conventions. Finally, the specific choice of questions used in the decision trees was determined in part by hand to accommodate the output "vocabulary" of the particular recognizer and parser used. Though this aspect could be completely automated by creating standards for parse trees and recognizer "phone" labels, the use of some hand-tuning of questions allows us to op-

timize performance by taking advantage of the features of different systems and knowledge of the task domain.

Clearly, performance in different spoken language systems will be affected by several factors, including the reliability and level of detail of the parser, the accuracy of the recognizer, the types of ambiguities in the task domain and the sophistication of other knowledge sources (e.g. semantic, discourse) in the system. We plan to explore these issues further by assessing performance of the algorithm in the SRI ATIS system. (Of course, it may be that the constrained semantics of the ATIS task make it difficult to assess the potential benefits of prosodic information.) Implementation and evaluation of prosody/parse scoring in the two systems should have implications for spoken language system design, and our initial work already raises some issues. In particular, there are cases where prosody could benefit speech understanding, but is not useful unless the natural language component provides more than one parse for a hypothesized word string, e.g. for lists of numbers and for utterances with possible disfluencies. In addition, it might be useful to have explicit filled pause models used in recognition (a capability available in some versions of the MIT system that was not used in this experiment), to help distinguish hesitations (marked by the "p" diacritic) from well-formed prosodic boundaries.

In conclusion, we emphasize that these experiments represent initial efforts at integrating prosody in speech understanding and there is clearly much more work to be done in this area. In addition to improving the basic components of the model and evaluating more parse hypotheses, there are many other possible architectures that might be investigated for integrating prosody in speech understanding.

ACKNOWLEDGMENTS

The authors gratefully acknowledge: C. Wightman for the use of his acoustic models; K. Ross for his prominence prediction model; E. Shriberg, K. Hunicke-Smith, C. Fong and M. Hendrix for help with prosodic labeling; and L. Hirschman, M. Phillips and S. Seneff for providing the MIT recognizer and parser outputs as well as many helpful discussions about the features/format of the MIT SUMMIT and TINA systems. This research was jointly funded by NSF and DARPA under NSF grant no. IRI-8905249.

References

1. P. Price, M. Ostendorf, S. Shattuck-Hufnagel, & C. Fong, "The Use of Prosody in Syntactic Disambiguation" *J. of the Acoust. Society of America 90,* 6, pp. 2956–2970, 1991.
2. M. Ostendorf, C. Wightman, and N. Veilleux, "Parse Scoring with Prosodic Information: An Analysis/Synthesis Approach," *Computer Speech and Language,* to appear 1993.
3. N. Veilleux and M. Ostendorf, "Probabilistic Parse Scoring with Prosodic Information," *Proc. of the Inter. Conf. on Acoustics, Speech and Signal Processing,* pp. II51–54, 1993.
4. L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees,* Wadsworth and Brooks/Cole Advanced Books and Software, Monterey, CA, 1984.
5. C. Wightman and M. Ostendorf, "Automatic Recognition of Prosodic Phrases," *Proc. of the Inter. Conf. on Acoustics, Speech and Signal Processing,* pp. 321–324, 1991.
6. C. Wightman and M. Ostendorf, "Automatic Recognition of Intonation Features," *Proc. of the Inter. Conf. on Acoustics, Speech and Signal Processing,* pp. 221–224, 1992.
7. M. Wang and J. Hirschberg, "Automatic classification of intonational phrase boundaries," *Computer Speech and Language,* 6-2, pp. 175–196, 1992.
8. J. Hirschberg, "Pitch Accent in Context: Predicting Prominence from Text," *Artificial Intelligence,* to appear.
9. K. Ross, M. Ostendorf and S. Shattuck-Hufnagel, "Factors Affecting Pitch Accent Placement," *Proc. of the Inter. Conf. on Spoken Language Processing,* pp. 365–368, 1992.
10. L. Hirschman *et al.,* "Multi-Site Data Collection for a Spoken Language Corpus," *Proc. of the DARPA Workshop on Speech and Natural Language,* pp. 7–14, 1992.
11. D. Pallett *et al.,* "DARPA February 1992 ATIS Benchmark Test Results," *Proc. of the DARPA Workshop on Speech and Natural Language,* pp. 15–27, 1992.
12. V. Zue *et al.,* "The MIT ATIS System: February 1992 Progress Report," *Proc. of the DARPA Workshop on Speech and Natural Language,* pp. 84–88, 1992.
13. S. Seneff, "TINA: A Natural Language System for Spoken Language Applications," *J. Association for Computational Linguistics,* pp. 61–86, March 1992.
14. S. Seneff, "A Relaxation Method for Understanding Spontaneous Speech Utterances," *Proc. of the DARPA Workshop on Speech and Natural Language,* pp. 299–304, February 1992.
15. K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg, "TOBI: A Standard Scheme for Labeling Prosody," *Proc. of the Inter. Conf. on Spoken Language Processing,* pp. 867–870, Banff, October 1992.
16. M. Ostendorf, A. Kannan, S. Austin, O. Kimball, R. Schwartz and J. R. Rohlicek, "Integration of Diverse Recognition Methodologies Through Reevaluation of N-Best Sentence Hypotheses," *Proc. of the DARPA Workshop on Speech and Natural Language,* February 1991, pp. 83–87.