The use of prosody in syntactic disambiguation

P. J. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong
SRI International, EJ 133, 333 Ravenswood Avenue, Menlo Park, California 94025

(Received 15 February 1991; accepted for publication 5 August 1991)

Prosodic structure and syntactic structure are not identical; neither are they unrelated. Knowing when and how the two correspond could yield better quality speech synthesis, could aid in the disambiguation of competing syntactic hypotheses in speech understanding, and could lead to a more comprehensive view of human speech processing. In a set of experiments involving 35 pairs of phonetically similar sentences representing seven types of structural contrasts, the perceptual evidence shows that some, but not all, of the pairs can be disambiguated on the basis of prosodic differences. The phonological evidence relates the disambiguation primarily to boundary phenomena, although prominences sometimes play a role. Finally, phonetic analyses describing the attributes of these phonological markers indicate the importance of both absolute and relative measures.

PACS numbers: 43.71.Es, 43.70.Hs

INTRODUCTION

The syntax of spoken utterances is frequently ambiguous, sometimes in more than one way. In spite of this, listeners usually arrive at something close to the speaker's intended meaning. Information that listeners might use to disambiguate structures includes knowledge of the world, context shared by the speaker and listener, and a further source of nonsyntactic information that is under-represented in written communication: the prosody of the utterance. By "prosody" we mean suprasegmental information in speech, such as phrasing and stress, which can alter perceived sentence meaning without changing the segmental identity of the components.

Prosody is an important component of spoken utterances. It serves to distinguish large classes of utterances (e.g., questions versus statements), and language learners appear to master its production and perception very early. There is evidence, for example, that infants only 4 days old can distinguish familiar from unfamiliar languages on the basis of prosody (see, e.g., Mehler *et al.*, 1988), and that 7- to 10-month-olds are sensitive to the prosodic appropriateness of pause location (Hirsh-Pasek *et al.*, 1987).

Since prosody plays an important role in speech communication, a clear understanding of the mapping between prosodic and syntactic structure would reveal significant aspects of the cognitive processes of speech production and perception. In addition, it would provide guidelines for the synthesis of more natural-sounding speech; Klatt (1987), for example, has cited prosody as a major requirement for more natural synthesis. Further, any contribution that prosody can make to the resolution of structural ambiguities will be particularly helpful in spoken-language understanding, where lexical and structural ambiguities of written forms are compounded by difficulties in finding word boundaries and in identifying words reliably in automatic speech recognition. Here, we use disambiguation experiments to study the mapping between prosody and syntax by minimizing the contribution of other possible cues to the resolution of ambiguity. The study described here forms the foundation for further work on modeling prosody by assessing a set of syntactic environments in which prosody alone might be used to disambiguate sentences, and by analyzing the correspondence between the phonological and phonetic attributes of the prosodic structure of utterances and their perceived meanings.

Specifically, in a series of perceptual experiments, listeners were presented with ambiguous utterances that had been produced in one of two contrasting context paragraphs, one appropriate for each of the two meanings, and were asked to choose the appropriate context from the two alternatives. The sentences were chosen to reflect seven classes of structural ambiguity, and the readers were professional FM radio announcers. Each spoken sentence was then hand labeled with prosodic phrase boundaries and prominences, and the duration and intonation correlates of the different markings were analyzed. The perceptual disambiguation scores were analyzed together with the prosodic markings to determine which structures were disambiguated, and which phonological cues were most often associated with those structures. This work extends the disambiguation experiments of Lehiste (1973) in ways that will be discussed below.

Our results offer support for the following hypotheses about resolving structural ambiguities via prosody.

(a) Different types of word strings with two competing syntactic structures differ in the degree to which they *can* be disambiguated by prosodic cues.

(b) Clause boundaries (i.e., boundaries of phrases that contain both a subject and a predicate) very often coincide with the boundaries of major prosodic constituents (as marked, e.g., by syllable-final lengthening, pause and boundary tone).

(c) Other syntactic constituents may be associated with any of several different levels of prosodic boundaries, i.e., speakers have more choice in phrasing, and prosodic boundaries need not correlate perfectly with syntactic ones, though they often do.

 0001-4966/91/122956-15$00.80

In our data, for the sentence pairs that were correctly disambiguated, the most frequent difference between the two spoken versions was the location and relative size of prosodic phrase boundaries. However, there were some structures for which phrasal prominence either was the only cue or played a supporting role in distinguishing the two versions.

We begin the paper by discussing previous work on the relationship between prosody and syntax, and on the use of prosodic cues for disambiguation. We then describe the recording of the corpus, and present results for the experimental studies which consider: (1) the accuracy and confidence of listeners in disambiguating different types of syntactic structures, (2) the phonological analysis of prosodic cues associated with the different structures and their relation to the disambiguation results, and (3) a phonetic analysis of the phonological markers. Finally, we discuss the implications of these results and raise some unresolved questions that suggest directions for future research.

I. BACKGROUND

With few exceptions (e.g., Geers, 1978), previous studies have focused either on relating phonological aspects of prosody to syntax (e.g., Gee and Grosjean, 1983; Nespor and Vogel, 1983; Bing, 1984; Ladd, 1986), or on relating phonetic/acoustic evidence to syntax (e.g., Scholes, 1971; Cooper and Sorensen, 1979; Cooper and Paccia-Cooper, 1980; Thorsen, 1980; Garro and Parker, 1982; Kutik *et al.*, 1983; Duez, 1985; Thorsen, 1985). A few studies, e.g., Pierrehumbert (1981), have considered the mapping from phonology to acoustics. The more phonetic/acoustic studies typically used a small number of minimal pairs of utterances in order to facilitate the acoustic measurements and to control parameters more precisely [exceptions include Klatt (1975) and Crystal and House (1990), where large data sets were used]. In contrast, the more phonological studies have focused either on "illustrative examples" or on text to which prosodic markers have been assigned on the basis of the syntax of the sentence. These studies have typically ignored the fact that there are several possible prosodic choices for a given syntactic structure. The focus in recent theoretical linguistics on human *competence* for language production, has resulted in neglect of actual language production and neglect of an area required for speech understanding (by human or by machine): the mapping from acoustics to meaning. Clearly, speech communication involves both production and perception, and it involves performance as well as competence.

A particularly interesting aspect of the syntax/prosody mapping involves the characterization of types of ambiguity that can be resolved by prosodic means and the types that cannot. Lieberman (1967) proposed that speakers can use prosody to disambiguate syntactically ambiguous strings of words that have several possible surface structures (i.e., the locations of syntactic bracketings differ, rather than just the labels on the brackets). For example, the sentence "I'll move on Saturday" could carry the message that the speaker will move personal household belongings to a different residence on Saturday, or that Saturday has been chosen as the date on which to "move on" to a new location or to a new topic. In the first case, "on" is structurally more closely linked to "Saturday;" in the second, it is linked to "move." By varying the location at which the utterance is broken into two prosodic units, the speaker can convey one or the other of these meanings because each corresponds to a different location for the edges of the main syntactic constituents:

"I'll move...on Saturday," versus

"I'll move on...Saturday."

In contrast, Lieberman proposed, such prosodic disambiguation is *not* possible when a string of words is syntactically ambiguous because of differing deep structures that do not affect the surface structure (i.e., where labels on brackets differ, but not the bracketing). The example most commonly cited for this claim involves structures like

"Flying planes can be dangerous,"

where "flying" can refer to the action of planes that fly, or to the action of another subject who flies the planes.

Lieberman's view that, of the three major kinds of ambiguity (lexical, surface-structure-related and deep-structure-related), only surface structure ambiguity can be resolved by prosodic (intonational) means was tested by Wales and Toner (1979). They used ten sentence pairs of each of the three types, and found that surface structure ambiguity was the only type that speakers and listeners resolved by prosodic means. Wales and Toner argued that prosody does *not* precisely signal the syntactic structure; rather, the means of disambiguation is a prosodic marking that warns the listener to look for a more unexpected meaning.

Lehiste (1973) conducted a perceptual study on prosodic disambiguation of different syntactic structures. Her experiments also showed disambiguation for sentence pairs with different surface-structure bracketings: of the 15 pairs tested, the ten with different bracketings were successfully disambiguated. However, one "deep structure" ambiguity was resolved by what appear to be intonational differences. An additional finding was that listener responses corresponded more closely to the intended meaning for versions where the speakers were consciously trying to distinguish between two possible meanings, compared to the condition in which they were reading sentences without necessarily being aware of the contrast.

Cooper and Paccia-Cooper (1980) investigated acoustic correlates of syntactic structure in a series of production studies focused on duration. Their data, based on sentence sets read by 10 to 20 talkers, support the claim that segments lengthen in constituent-final position. Their work focuses on durational cues to syntax at specific syntactic constituent boundaries and does not acknowledge the fact that the same syntactic structure could map to more than one prosodic structure.

Other investigators attempted to characterize the general relation between syntax and prosody, taking a more phonological approach. For example, Gee and Grosjean (1983) describe an algorithm that provides a mechanism for mapping from a syntactic parse to prosodic phrases. Application of their algorithm associates edges of major syntactic constituents (especially embedded sentences) with larger prosodic phrases. Selkirk (1984), on the other hand, suggests that the metrical structure of speech is determined by

the syntax, but that the intonational structure is determined by semantics.

In addition to studying which structures could be disambiguated by prosody, and how to map from syntactic to prosodic structure, researchers have investigated the acoustic cues associated with disambiguation and with boundaries, often without distinguishing explicitly between prosodic and syntactic constituents. In her original disambiguation study (1973), Lehiste noticed two strategies for marking constituent boundaries: pause or period of laryngealization, and duration lengthening in the sequence of words containing the boundary. She also suggested that intonation cues associated with disambiguation were less systematic, although intonation may play a role in "deep structure" disambiguation, as mentioned above. Klatt (1975) investigated duration lengthening at syntactic boundaries, by computing durations of both stressed and unstressed vowels in a connected discourse and determining where longer versions occurred. He found significant lengthening at several different types of syntactic junctures, including: noun phrase/verb phrase boundaries, sentence endings, before conjunctions, between nouns and prepositional phrases, and before embedded clauses. Crystal and House (1988) have more recently analyzed durations in connected discourse, confirming past findings of phrase-final duration lengthening.

Lehiste's disambiguation experiments motivated several further perceptual experiments on the relative importance of various acoustic cues to boundary locations. Two studies, Lehiste *et al.* (1976) and Scott (1982), investigated the role of duration and pausing, and found that lengthening the sequence of words containing a syntactic constituent can change the perceived meaning of a sentence. Others, however, e.g., Klatt (1975), claim that lengthening occurs only in the phrase-final syllable, which agrees with our own experimental observations (Wightman *et al.*, 1992). Streeter (1978) considered the relative importance of duration, intensity, and intonation. By varying each correlate independently, she found that duration and intonation both had significant (and additive) effects in changing perceived meaning, while intensity contributed to meaning change only in combination with other cues. In addition, she noticed that the two speakers studied used both duration and intonation cues in different ways.

The work presented in this paper extends previous work in several ways. First, focusing only on surface-structure ambiguities (since earlier work indicates that these are good candidates for disambiguation), we investigate the ability of listeners to disambiguate sentences for different types of syntactic structures, using several instances of each type. Second, our focus here is on both production and perception. We tried to avoid exaggeration of any disambiguating strategies on the part of speakers and listeners by separating the ambiguous pairs from each other in time (no two members of an ambiguous pair occurred in the same session either for speakers or for listeners). Third, to increase reliability without assessing a large pool of subjects, we used four professional FM radio announcers (these speakers have proved to be very consistent speakers in our pilot studies). Fourth, in analyzing the cues used in disambiguation, we have investigated the possible use of prominence associated with pitch accents, in addition to prosodic phrase boundary cues. Finally, to compare durational structures across the various sentences used, and to facilitate generalization beyond the specific sentences used, we present results in terms of relative, rather than absolute, durational patterns. By combining phonological analyses of prosodic elements such as boundary tones and prominences with investigation of their acoustic correlates and their perceptual effects, we hope to shed some light on both the mapping between syntactic and prosodic structure, and on the role of prosody in resolving various types of syntactic ambiguity.

II. CORPUS

The methodology used in this experiment involved (1) recording pairs of structurally ambiguous sentences, each preceded by a few sentences of disambiguating context, (2) presenting those recorded sentences to naive listeners for judgements of the appropriate context, and (3) comparing the phonological and phonetic characteristics of the spoken utterances with listeners' ability to disambiguate them. The recordings, which formed the basis for both perceptual experiments and phonetic and phonological analyses, are described below.

We used 35 pairs of sentences that were ambiguous in that the two members of each pair contained the same string of phones (in many cases the same string of words), and could be associated with two contrasting syntactic bracketings. The sentences manifested seven types of structural ambiguity: (1) parenthetical clauses versus nonparenthetical subordinate clauses, (2) appositions versus attached noun (or prepositional) phrases, (3) main clauses linked by coordinating conjunctions versus a main clause and a subordinate clause, (4) tag questions versus attached noun phrases, (5) far versus near attachment of final phrase,[1] (6) left versus right attachment of middle phrase, and (7) particles versus prepositions. In each category, there were five pairs of ambiguous sentences. In presentation, each sentence was preceded by a disambiguating context of one or two sentences. The target sentences were chosen to be fully voiced to facilitate pitch tracking for acoustic analysis. A list of the sentences with their disambiguating contexts is included in the Appendix.

We use the term *size of syntactic break* to reflect the number of syntactic brackets that would occur between two pairs of words: More brackets correspond to a larger syntactic break. The site with the largest number of brackets is referred to as the major syntactic break. For the structural categories (1)–(4), sentence A of the pair involved a larger syntactic break than sentence B. For the attachment ambiguities (5)–(7), sentence A of the pair had the larger syntactic break later in the sentence than did sentence B.

The sentences were recorded by four professional FM public radio newscasters, one male and three female, who were naive with respect to the purposes of the experiment. We will refer to the speakers using the codes: F1A, F2B, F3A, M1B. The newscasters were asked to read the sentences in context, using their standard radio style of speak-

ing. In a pilot study, we found the FM radio style to have more clearly and consistently marked prosodic cues than a nonprofessional speaking style (Price *et al.*, 1988), while sounding acceptably natural. Our hope was that this style would be easier to label prosodically, and therefore the contributions of specific phonological cues would be easier to identify. In addition, we believe the FM radio news style will be particularly useful for speech synthesis, since the style is a natural one, yet may be easier to model because the fewer, more clearly and consistently marked prosodic units may be predictable from a relatively shallow syntactic representation. However, care must be taken in generalizing our results to spontaneous speech, where prosodic cues may be less clearly marked.

The announcers were presented with the written sentences in context paragraphs, with the sentence types and A/B members of the pairs assigned to two recording sessions, so that the two contrasting members of a pair did not occur in the same session. The speakers were not told that there were special target sentences within the paragraphs. The recording sessions were separated by at least a few days and often several weeks, to minimize the possibility that the announcers would produce unnatural versions in an attempt to emphasize potential differences between the two members of a pair.

Our goal was to create sentence pairs that were segmentally identical but syntactically different, so that we could investigate the relationship between syntax and prosody independent of any differences contributed by the segments. Therefore, in an additional recording session, the speakers were asked to reread any problem sentences until an acceptable version was produced. For example, utterances which had a segmental cue that would disambiguate the pair (e.g., a reduced vowel in "an' Dewey" which would not occur in the corresponding "Anne Dewey," or an aspirated "h" in "would he" which would not occur in the corresponding "Woody") were rerecorded. For one talker, F2B, phonetic similarity was verified by examining phonetic labels automatically generated by the SRI Decipher speech recognition system (see Sec. V). Except for a few minor differences (such as schwa plus sonorant substitutions for syllabic sonorants), the sentences were found to be phonetically identical. If the experimenters judged that the prosody was inconsistent with the context (i.e., the speaker misinterpreted the intended context), the sentence was also rerecorded. An average of about 10% of the sentences were rerecorded because of phonetic differences, and an additional 10% were rerecorded to correct an incorrect prosody. The first speaker recorded (F2B) rerecorded 16 of her 70 sentences because of incorrect prosody; the sentence contexts were revised after this speaker to better elicit the intended meaning. For the three subsequent speakers, an average of six sentences per speaker needed to be rerecorded to correct the prosody. Except that all the speakers had trouble with at least one far versus near attachment sentence, there was very little in common among the prosodic errors that the different speakers made. Although they were not prosodically incorrect, tag sentences in which the tags were read as questions were rerecorded as statements (except for F2B) so that the question boundary tone cue would not confound the potential contribution of other prosodic cues.

Although the radio announcers were probably aware of the different contexts during the second recording, the similarity of sentence pairs was not pointed out to them. It is possible that asking the talkers to rerecord a sentence could lead them to search for alternate meanings and therefore to produce more exaggerated forms. However, the pattern of responses to the set of rerecorded sentences was very close to the pattern for those that were not rerecorded: The average accuracy of subject identification of the rerecorded sentences was 86%; for those that were not rerecorded, it was 85%.

Sentences might be prosodically distinct for reasons other than syntactic differences. The disambiguating contexts, for example, might have led to differences in focus. Therefore, to assess the relationship between prosody and syntax, it was essential to use several examples of each structural type. Semantic effects and response biases related to the particular sentence pairs or their contexts can then be "washed out" by looking for effects that occur throughout a class of sentences. Response biases can be assessed by looking at differences between the two members of the pair (see, e.g., Wales and Toner, 1979). It is important, however, to stress that the pairs never occurred in the same speaking session, in order to avoid productions with exaggerated contrast. The listeners were sensitive to the contrast, having the two written alternatives before them, but were presented with only one member of the pair per session, so they could not make contrastive judgments. Since normal speech perception involves the identification of structures, as opposed to the discrimination between two alternatives, it is important to look for cues to structure that are entirely contained in a given sentence, rather than those that simply distinguish it from the contrasting member of the pair. Again, we hope that using several members of each class can reveal cues that occur throughout the class, and that any particular contrast relating only to a particular pair will be just noise in the data.

III. PERCEPTUAL EXPERIMENTS

A. Methods

For the perceptual experiments, the spoken context sentences were edited out so that the target sentences could be presented in isolation. The 70 utterances (35 sentence pairs) produced by a single speaker were presented to listeners in two sessions; only one member of each pair was heard in each session using a mixed assignment of half type-A and half type-B sentences in each session (analogous to the strategy used for recording the sentences). The different syntactic types were interleaved, and, for each pair of sentences, A versions always appeared before B versions on the answer sheet. The listeners heard the sentences in a small conference room from a portable stereo. The tape player was stopped between sentences until subjects were ready to continue; the subjects were under no time constraints to make their judgements. Each listening session (35 sentences) took approximately 40 min, and was conducted without any additional breaks. Listening sessions were separated by at least 3 weeks

TABLE I. Perceptual experiment results, averaged over the four speakers, for ambiguous sentence interpretation. The version A/B figures are based on 285 total observations of each class. An asterisk marks the A and B version responses that had high accuracy in listener responses. ("High accuracy" was defined to be average accuracy minus the standard deviation greater than 50%.)

	Version A		Version B		Overall	
Type of ambiguity	% correct	% confident	% correct	% confident	% correct	% confident
(1) ± Parenthetical	77	53	96*	74	86	63
(2) ± Apposition	92*	70	91*	69	92	70
(3) M–M vs M–S	88*	46	54	28	71	37
(4) ± Tags	95*	69	81	46	88	57
(5) Far/near attach.	78	27	63	20	71	23
(6) Left/right attach.	94*	69	95*	67	95	68
(7) Particle/prep.	82*	39	81*	52	82	45
Average	87	53	80	51	84	52

to minimize listener recall of the previous session's sentences. Listeners were given an answer sheet with both disambiguating contexts written out for each sentence; the target sentence was printed in bold at the end of each context. They were asked to mark the context that they thought best matched what they heard, with an additional marker if they were confident of their decision. Subjects were rewarded with pizza and soft drinks after the session.

The subjects were all native speakers of American English, naive with respect to the purpose of the experiments. Most were engineering students, recruited through flyers advertising the free pizza. For the second two speakers, F3A and M1B, to attract more subjects, we increased the incentive by offering an additional $50 prize to the person who scored highest on this task. The number of listeners who heard both sessions for each of the different speakers was 13 for speaker F1A, 15 for F2B, 17 for F3A, and 12 for M1B. Different subjects participated in the experiments for the different speakers, although there was some overlap in the subject pool. Four subjects participated in all four experiments.

B. Results

For the analysis, we define a correct listener response to be one that matches the context in which it was produced. Accuracy is the percentage of correct listener responses. Confidence is the percent of the time that listeners indicated that they were confident of the response choice. Table I summarizes average subject accuracy and confidence for the different types of ambiguity. The averages are taken over the four speaker averages, so as not to more heavily weight the utterances that were heard by more listeners. The averages for each speaker are taken across five versions of each structural type, as well as across the various listeners (12–17 per talker).

Table I shows that subjects could reliably disambiguate many, but not all of the ambiguities. Subjects were rarely confident *and* incorrect, and the confidence is somewhat correlated (0.64) with the accuracy. On the average, subjects did well above chance (84% correct) in assigning the sentences to their appropriate contexts, although subjects were confident of their judgments only 52% of the time. Also on average, main-subordinate (3B) sentences and near attachments (5B) were close to the chance level ($\mu \approx$ chance); parentheticals (1A), far attachments (5A), and nontags (4B) were recognized at levels greater than chance but not reliably ($\mu - \sigma \approx$ chance); and all other sentence types were reliably disambiguated ($\mu - \sigma >$ chance). (Note that μ refers to the average identification accuracy, σ is the standard deviation, and "chance" is 50%.)

The sentence accuracy averages by talker differ significantly (accuracy figures for each sentence produced by each talker are included in the Appendix), indicating that the listener's ability to disambiguate a sentence via prosody is dependent on the speaker's particular choice of prosodic rendition for that sentence. However, by averaging across the talkers and across the listeners and across the five examples of each structural type, we should get an indication of the relative ease with which various sentence types can be disambiguated using prosody.

There was a fairly large standard deviation in subject response, ranging from 0.05 to 0.26 over the 14 different sentence types. For the most part, the sentence types with higher accuracy had lower variance, and the variance for different A/B sets of a pair could be quite different. The variability appeared to be associated mainly with the difference in particular productions of a sentence. There was little evidence for a systematic difference among speakers or among sentences, and the average sentence type variance for an individual speaker (0.14) was close to the four-speaker variance (0.15). We describe the variance of the results rather than a significance test, because a significance test would require the unreasonable assumption of *a priori* sentence probabilities. Wales and Toner (1979) suggest that biases toward one sentence or another may play a large role in perceptual studies such as these. However, the fact that for 21 of the 35 pairs, average accuracy was 70% or better for *both* members of the pair suggests that response biases are not playing a strong role in these results [as also noted in Lehiste (1973)], although response bias might explain the large discrepancy between main–main and main–subordinate results.

Listener accuracy as a function of sentence type and A/B structural distinction is illustrated in Fig. 1. Some of the sentence pairs differ in the size of the syntactic boundary (classes 1, 2, 3, 4) and others in the location of a major

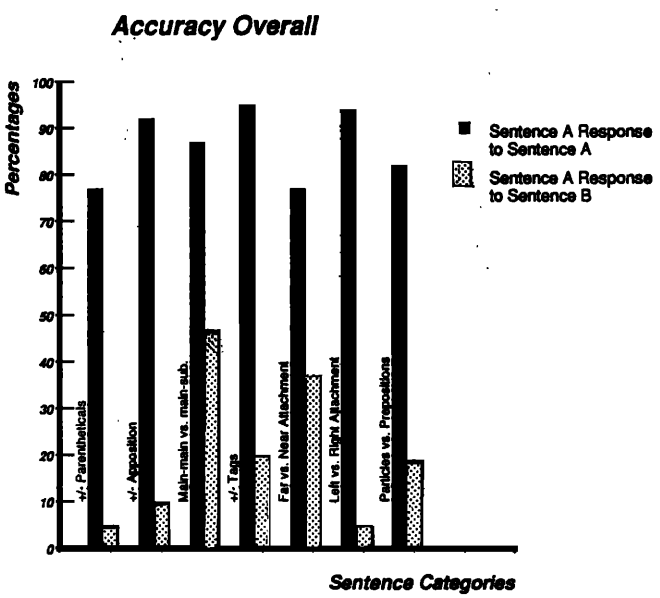

FIG. 1. Average listener percent A response to versions A and B. For categories 1–4: A is major versus B minor or no syntactic boundary. For categories 5–7: A is later versus B earlier syntactic boundary.

syntactic boundary (classes 5, 6, 7). For the first four classes, our initial hypothesis was that sentences with boundaries corresponding to higher level syntactic constituents (version A) would have greater prosodic breaks and, therefore, would be more reliably identified than their alternatives (version B). Our data did not confirm this hypothesis. In particular, parenthetical sentences were less reliably identified than their nonparenthetical alternatives. For the last three classes, our results also did not confirm the claim of Wales and Toner that disambiguation occurs only for sentences in which the preferred reading was associated with the earlier of two constituent boundaries. In fact, for the far versus near ambiguities, the sentence with the later boundary (far attachment) was somewhat more reliably disambiguated, although confidence in the significance of the difference is not high. Some of the parenthetical and appositive phrases are sentence-final and others are sentence-medial; no difference in identification accuracy was observed as a function of this difference in sentence position, but the data are rather sparse.

We found little difference among the four different radio announcers in overall average listener performance, as shown in Table II. Listener accuracy for the different announcers did vary somewhat as a function of sentence types, as illustrated in Fig. 2. For the four subjects who listened to all four speakers, there were no trends of increased accuracy or confidence in successive experiments (overall or for specific categories). Individual listeners were quite consistent in their average judgement accuracy, but varied significantly in the accuracy of their responses for specific sentence types.

Much previous work has focused on sentence pairs that are identical in their written form. Our study included ten such pairs. In addition, however, 22 of the pairs differed in placement of commas, and 14 involved homophones that were spelled differently (some sentence pairs involved both

TABLE II. Average ability of subjects to disambiguate sentences as read by different speakers.

	F1A	F2B	F3A	M1B	Average
% correct	81	85	85	82	83
% confident	58	57	49	44	52

punctuation and spelling differences). The differences in spelling could affect listener responses if the Wales and Toner (1979) claim is true that lexical ambiguity yields greater response biases. Particle/preposition examples could be viewed as lexical ambiguity, although they are spelled alike. However, whether the particle/preposition case is included in the lexical ambiguity (homophone) class or not, there is no evidence that response bias is more prevalent in one class or the other: In each class, there are just as many pairs where both members are reliably disambiguated.

Differences in punctuation also could have resulted in systematic differences in listener responses. For example, it may be that the punctuation in the written version caused the speakers to exaggerate these boundaries, which could lead to a greater accuracy for those sentences that differed in the placement of commas (22 of the 35 pairs). There is evidence for this effect: The average accuracy for pairs that differed in comma placement was 90% for the A version and 85% for the B version, while the average accuracy for pairs that did not differ in comma locations was only 81% for the A version and 71% for the B version. All of these accuracies, however, are above chance, and conventions for commas are probably related to a correspondence between syntactic and prosodic structure. In fact, Cooper and Paccia-Cooper (1980) found systematic durational lengthening at locations where commas would ordinarily have occurred, though in their stimuli the commas were omitted and had to be inferred by the reader from an accompanying paraphrase. The regularities of prosodic cues at sites where commas typically occur is good news for those understanding systems whose parsers take advantage of commas in written text, provided that the cues to "commas" can be reliably detected in speech. Our data suggest that this may not be impossible.

We hoped that by using multiple sentences in each structural class and by using multiple speakers, we could minimize the problem of sentence- and utterance-dependent results. This appears to have been successful. Although there were a few utterances that were outliers for their class (typically, particularly low accuracy) for an individual speaker, only two sentences were considered outliers generally, having atypically low identification accuracy for three of the four speakers. The preposition reading of "When I arrived, they were mulling over Andrea's annoying burner" was consistently misrecognized by listeners. Perhaps this was due to lexical bias of "mulling over" in the context of an "annoying burner:" "mulling over" (*pondering*) is much more likely than "mulling" wine. The far-attachment sentence "Andrea moved the bottle under the bridge," where the bottle was moved to a location under the bridge, also had atypically low performance, but we have no explanation for this result.

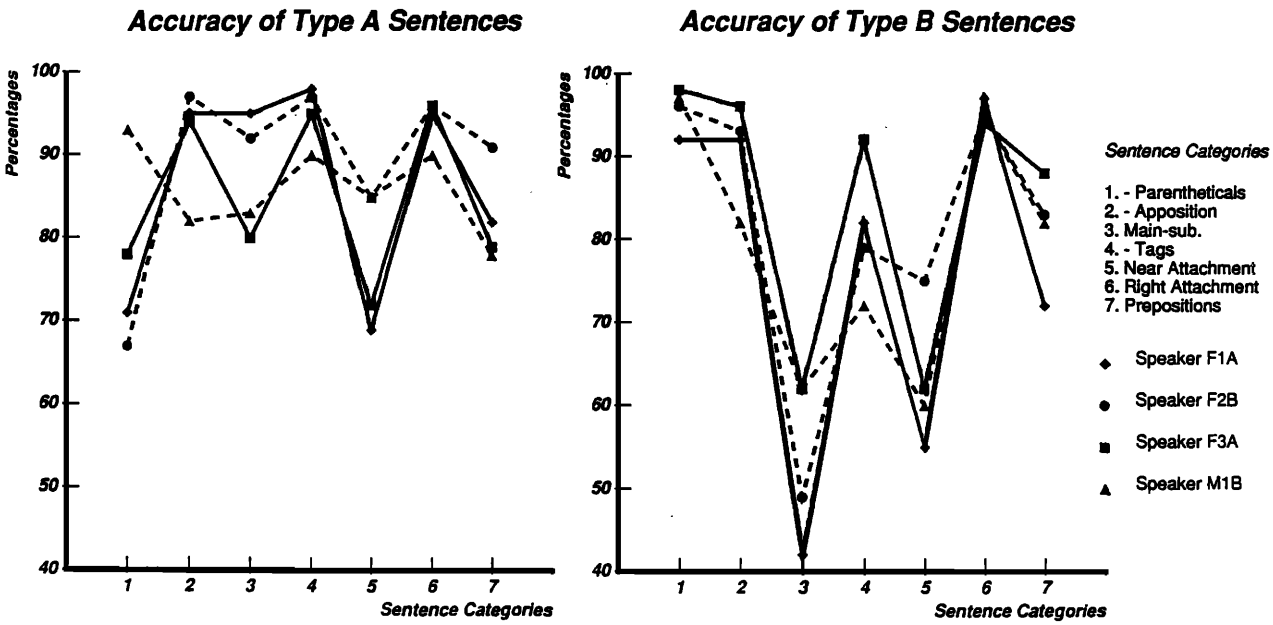

FIG. 2. Average listener accuracy in disambiguating sentences of different categories in A (at left) and B (at right) versions as a function of the different speakers (F1A, F2B, F3A, and M1B). Chance is 50%.

We have shown that naive listeners can, in general, reliably separate otherwise ambiguous sentences on the basis of prosody. To account for the patterns of listener accuracy, as well as the exceptions, we will compare, in the next section, these results using naive listener labelings to a phonological representation of the prosodic structures.

IV. PHONOLOGICAL ANALYSIS

The perceptual experiments described above clearly show that speakers can encode prosodic cues to structural ambiguities in ways that listeners can use reliably. This section attempts to find a phonological answer to the question: How do they do it? To approach this question, we labeled discrete, prosodic phenomena (specifically, prosodic phrase boundaries and prominences) that could mark structural contrasts phonologically. We then analyzed the relationship between these labels and the patterns in the perceptual accuracy study. There are other prosodic cues (e.g., the *type* of pitch accent), and other phonological correlates of the prosodic structure (e.g., phonological processes at prosodic boundaries) which can likely play a role in disambiguation. However, analysis of these phenomena was beyond the scope of the present study. In the following section, we describe our labeling system and analyze the associated constituents in terms of their relationship to the syntactic structures in our corpus, and the accuracy with which sentences are identified.

A. Perceptual labels

We chose labels based on three criteria: (1) they should be used consistently within and across labelers, (2) they should be rather close to surface forms (to make eventual automatic detection more tractable and to improve labeler consistency), and (3) they should provide a mechanism for communicating information to a parser. For these reasons, our notation differs somewhat from that of other systems, although it is similar in many respects.

We used seven levels to represent perceptual groupings (or, viewed another way, degrees of separation) between words as perceived by the trained labeler. These seven levels appeared adequate for our corpus and also reflected most the levels of prosodic constituents proposed in the literature (e.g., Liberman and Prince, 1977; Selkirk, 1984; Ladd, 1986; Nespor and Vogel, 1983; Beckman and Pierrehumbert, 1986). Researchers have not yet converged on a widely accepted set of notational conventions for representing prosodic phenomena. Since we found that many of our within and across labeler discrepancies were related to decisions about whether a break was large enough or not to warrant a label, we adopted a system that requires a prosodic break index between every pair of words. These break indices express the degree of decoupling perceived between each pair of words as follows: 0—boundary within a clitic group, 1—normal word boundary, 2—boundary marking a minor grouping of words, 3—intermediate phrase boundary, 4—intonational phrase boundary, 5—boundary marking a grouping of intonational phrases, and 6—sentence boundary. Break indices of 4, 5, and 6 are "major" prosodic boundaries; constituents defined by these boundaries are often referred to as "intonation phrases" (e.g., see Beckman and Pierrehumbert, 1986), and are marked by a boundary tone. Boundary tones were labeled using two types of falls (final fall and nonfinal fall), and two types of rises (continuation rise and question rise). A break index of 5 is typically found in long sentences and frequently coincides with a breath intake or long pause. The break label 5 was used rarely in this ambiguous sentence corpus, but was used more often when we labeled a different radio announcer corpus (consisting of

	The		men		won		over		their		enemies	
(a)	*s*	0	*P*0	3	*s*	0	*P*1 *s*	2	*s*	1	*P*0 *s FF*	6
(b)	*s*	1	*P*0	2	*P*1-*CR*	4	*s s*	1	*s*	1	*P*0 *s FF*	6.

news stories) which had much longer sentences. The break index 3 corresponds to the unit referred to as an "intermediate phrase" in Beckman and Pierrehumbert (1986), or a "phonological phrase" in Nespor and Vogel (1983). The "phrase accent" pitch marker theoretically associated with the intermediate phrase was not labeled.

Our conventions for labeling boundary tones correspond to those initiated by Pierrehumbert (1980) and elaborated by Beckman and Pierrehumbert (1986) as follows: final fall (LL%), continuation fall (HL%), continuation rise (LH%), and question rise (HH%). In a sample of about 450 words labeled in both conventions, we found about a 90% overlap in the locations of boundary tones.

Prominent syllables in the sentences were labeled using *P*1 for major phrasal prominence; *P*0 for a lesser prominence; *C* for contrastive stress, which occurred rarely in these sentences (marked on 1% of the total words for four speakers); and *s* for syllables with no prominence. A more detailed representation of prominence (e.g., the *type* of pitch accent) was beyond the scope of the present study.

The prosodic cues were labeled perceptually (i.e., with no visual display) by three listeners using multiple passes. The data were first labeled by the listeners individually; any differences in markings were then discussed; and then the sentence was replayed a few times to allow the labelers to revise their markings. Care was taken in the discussion to point out possible biases from the syntax in order to avoid such influences insofar as possible. Finally, a majority vote of the labels (which at this point had a correlation of 0.96 across labelers) was used as the final hand-marked label set. All labeling was perceptual. A sample of labelings for one sentence pair as produced by one of the talkers appears above.

B. Phonological analysis

To separate semantic effects (possible confounds for any particular pair of sentences) from effects that should occur throughout the syntactic class, we paid particular attention to those cues that reliably occurred in the A versions of one class, but never in the contrasting B versions, or vice versa. We also paid particular attention to those sentences that had high accuracy and confidence and to the outlier sentences. Below we mention some general results and then discuss briefly the individual classes investigated. In the discussion below, we use the term "major" or "large" prosodic break to indicate labels of 4 or higher; "small" breaks had labels of 3 or lower.

1. General observations

In summary, we found that prosodic boundary cues are associated with almost all reliably identified sentences. Presence of an intonational phrase boundary (break index 4 or 5) was often, but not always, a reliable cue and was most often observed at embedded or conjoined clause boundaries (marked by commas in the text). In addition, a systematic difference in the relative size of prosodic break indices, or in the location of the largest break regardless of size, was frequently the only disambiguating information in the labels for the smaller syntactic constituents that were reliably disambiguated. By and large, relatively larger break indices tended to mean that syntactic attachment was higher rather than lower. In contrast to the pervasive association of boundary cues with successful disambiguation, prominence seemed to play mainly a supporting role, and was the sole cue in only a few sentences. The data are too sparse to permit specific conclusions about structures where prominence is useful, other than to note that its presence on particles is much more likely than on prepositions.

a. Parenthetical (A) versus nonparentheticals (B): The A versions always have break indices larger than 3 surrounding the parenthetical, except for one talker's rendition of one sentence. The B members have break indices less than 4 at one or both of the corresponding sites. In all cases, the sentences with major prosodic breaks surrounding the parenthetical were identified as version A by 75% or more listeners, and sentences without the major prosodic breaks were identified as version B 80% of the time or more. This generalization includes an anomalous A version having a 3 at the parenthetical boundary, which was identified in accordance with the indices rather than in accordance with the speaker's intent.

b. Apposition (A) versus nonapposition (B): The A version of the pair, the appositive, always has a major prosodic break both before and immediately following the appositive. The B version of the pair, in contrast, typically has a small break index at one or both of the corresponding sites. One sentence pair, however, was difficult for both speakers and listeners. Two speakers produced a major break at the "wrong" location, i.e., after "are" in "Wherever you are in Romania or Bulgaria, remember me." This predicts that the sets should be clearly separable, except for this sentence, which is what we found: All were labeled by the naive listeners at 87% accuracy or higher, except for this sentence, which was 73% correct.

c. Main–main (A) versus main–subordinate sentences (B): The A versions of the pairs were typically well identified, whereas the B versions tended to be close to the chance level. This could be the result of a syntactic response bias if the conjunction constructions are preferred over the deleted "that" in the alternants. This is an interesting case since the bracketings differ for the two versions of the sentence, and yet the two versions are apparently not well separated perceptually. The prosodic transcriptions suggest a reason:

Both versions of the sentence have a major prosodic boundary in the same location, associated with the embedded (B) or conjoined (A) sentence.

d. Tags (A) versus nontags (B): The A members all have a major prosodic break before the tag, and these were all identified as A versions (92% or more or the time). One talker produced one B version with a major prosodic boundary before "Izzy," and 92% of the listeners identified this utterance as version A, in accordance with the prosody. Two other B versions were frequently misidentified; these sentences had no boundary tone, but did have a break index of 3 (the largest in these sentences) at the site corresponding to the boundary of the tag.

e. Far (A) versus near (B) attachment sentences: The A versions showed a tendency to have the largest break index in the sentence before the phrase to be attached to a "far" site (i.e., a site other than to a phrase ending in the immediately preceding word). This pattern occurred in 15 of the 20 A utterances and only one of the B utterances. One talker's production of one A version had a 2 at the site in question, and a majority of the listeners labeled this as version B, which happened with none of the other A versions. Thus the location of a relatively large break index at the site in question appears to block the "near" (low) attachment, and a relatively small index appears to enhance it.[2] The exception to this pattern, for all four talkers, is the sentence "I didn't wear them all day," which could be interpreted as a scope ambiguity rather than an attachment ambiguity. No rendition of this sentence contained a major prosodic boundary, although the pair was reliably separated. The apparent cue for disambiguation was a placement of prominence on "didn't" in the A versions and not the B versions.

f. Left (A) versus right (B) attachment sentences: For every rendition by every talker, there was a smaller break index at the attachment location than at the other end of the word or phrase to be attached. For the four sentence pairs that differed in comma location, the difference between the two break indices was large (2 or more), typically 0 or 1 in the location without a comma and 3, 4, or 5 in the location with the comma. These utterances were very reliably identified, with greater than 92% accuracy for all but one case. For the sentence pair with no commas ("They rose early in May"), only one talker produced a major prosodic boundary on "rose" for the B version, and this was the only talker whose version was correctly identified 100% of the time. The other versions of this sentence were identified with fairly high reliability; prominence placement, on "early" (A) versus on "May" (B), was likely the disambiguating cue.

g. Particles (A) versus prepositions (B): There is less frequently a major prosodic break before a prepositional phrase compared to conjoined or embedded sentences: 60% of the prepositional phrases in this class followed a major prosodic break, compared to 90% observed in the context of clauses. The real structural clue appears to be not the absolute size of the break index but its relative size. For all A versions, we observed a smaller break index between the verb and particle, compared to the indices before the verb or after the particle. For the B versions, the relations were reversed: There was a tendency to have a larger break between the verb and preposition, compared to those before the verb or after the preposition. This held in all but three of the 20 instances. For two of these, the break index was larger before the verb than after the verb; one of these was less often accurately identified than others in its group, but the other was not. The third, which appears to have an inappropriate prosodic pattern, was only identified correctly 69% of the time, but then other versions of this sentence ("mulling over") were also poorly identified (average accuracy was 50% for the preposition readings compared to 78% correct for the particle versions). This may well reflect a response bias related to the two meanings of "mulling over." In addition, 16 of the 20 particles had a prominence compared to only one of the 20 prepositions.

Overall, there was very little systematic difference in the speakers' use of prosodic cues. There were some differences in individual sentences which accounted for the variation in listener responses, but no consistent characteristics could be attributed to any one speaker. The correlation of break indices between pairs of speakers was 0.94–0.95, and the relative frequencies of prominences for the different speakers were also very similar. This result is consistent with the finding by Crystal and House (1990) of a high correlation in duration patterns between different versions of the same utterance read by nonprofessional speakers.

V. PHONETIC ANALYSIS

We have thus far presented evidence that naive listeners can reliably use prosody to separate structurally ambiguous sentences, and phonological evidence that suggests how listeners might use prosody to assign syntactic structure. Other studies have focused on syntactic differences associated with disambiguation. Our evidence shows that the prosodic structure can point to the syntactic differences in systematic ways: Sentences with certain correspondences between syntactic and prosodic structures are reliably disambiguated, whereas others are not. In this section, we investigate some of the phonetic evidence that might be responsible for the prosodic disambiguation. Since previous work suggests that the primary prosodic cues are duration and intonation, the present study is confined to these two cues. However, we acknowledge that other cues, such as the application or nonapplication of phonological rules, contribute to the perception of prosodic boundaries. We tried to minimize such effects by asking the speakers to reread sentences in which overt segmental cues were produced, i.e., where the gross phonetic transcription of the two versions of the sentence would differ.

In the results presented here, segment duration normalization is determined automatically using an HMM-based speech recognition system, the SRI Decipher system, which uses phonological rules to generate bushy pronunciation networks that enable more accurate phonetic transcription and alignment than single pronunciation speech recognizers (Weintraub *et al.*, 1989). On a sample of 21 sentences, including 643 segments, we found that 2/3 of the segment boundaries coincided or were off by less than 10 ms compared to hand labels; 95% were less than 50 ms off. Each phone duration was normalized according to speaker- and phone-dependent means μ_a and variances σ_a^2:

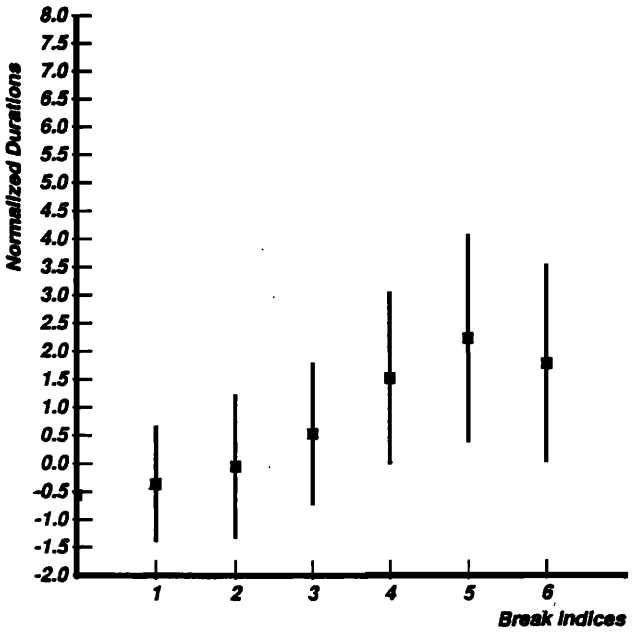

FIG. 3. Mean and variance of normalized duration in the rhyme of the final syllable in a word, as a function of the proceeding break index. The statistics are based on the following numbers of observations: 0—219, 1—918, 2—358, 3—128, 4—212, 5—25, 6—280.

$$\tilde{d} = (d - \mu_\alpha)/\sigma_\alpha,$$

where d is the duration of a segment, $\tilde{d}$ is the normalized duration given phone label α, and lexically stressed and unstressed vowels are modeled as separate phones. We have proposed normalized duration as a feature for automatically detecting prosodic boundaries (Ostendorf *et al.*, 1990), and normalized duration has also been used recently by Campbell (1990) for duration analysis. The variance of normalized duration in different contexts tends to be large, because the normalization has not accounted for effects such as syllable position, phonological and phonetic context, and speaking rate,[3] which also play a role in determining phoneme duration (Klatt, 1975; Crystal and House, 1988).

We observed longer normalized durations for phones preceding major phrase boundaries and for phones bearing major prominences compared to other contexts. As mentioned earlier, it has long been noted that syntactic breaks are often associated with duration lengthening in the phrase-final syllable. We measured average normalized duration in the rhyme of the final syllable of all words and found that higher break indices are generally associated with greater normalized duration, as shown in Fig. 3. The fact that duration is affected by constituents at many levels in the prosodic hierarchy is interesting, and consistent with our observations that relative break index size is meaningful even below the level of the intonational phrase (4,5). Although in this work we found only the difference between the groups 0–3 (without boundary tone) and 4–6 (with boundary tone) to be statistically significant, in recent related work (Wightman *et al.*, 1992), we found at least four statistically significant levels using duration alone, and possibly more when pauses and boundary tones are considered. Pauses are also associated with major prosodic boundaries, occurring at 48/212 (23%) boundaries marked with 4 and 17/25 (67%) boundaries marked with 5. The duration means and variances are 19.2 ± 10.4 ms for a pause following a 4 and 24.6 ± 15.6 ms for a pause following a 5. Sentence-final pauses could not be measured for these sentences, which were always the final sentence in a paragraph. In only one case did a pause occur after a 3.

Normalized duration of the vowel nucleus for the different prominence markings is illustrated in Fig. 4. The plots show that: (1) major prominences ($P1$, C) tend to be longer than unmarked or minor ($P0$) prominences, although the effect is small before major prosodic breaks;[4] (2) word-final syllables are longer than nonword-final syllables, except for syllables with contrastive stress before a small break (but there are only five instances in each case); (3) syllables tend to be longer in words before major breaks than before smaller breaks, though the effect is only significant for word-final syllables; and (4) the effects seem to be somewhat independent: The longest syllables are those with a major prominence, in word-final position, before a major break. The number of observations of each type of prominence is given in Table III.

Intonational cues observed included boundary tones, pitch range changes and pitch accents. Boundary tones are involved for the break indices 4, 5, and 6. Sentence-final (6) boundary tones are typically either final falls or question rises; level (5) boundary tones are usually perceived as incomplete falls; and intonational phrase (4) boundary tones are most often continuation rises but occasionally are perceived as partial falls. Tags were sometimes associated with a sentence-final question rise, though we tried to eliminate this cue as much as possible by asking the radio announcers to reread versions when this occurred. Another intonational cue was a perceived drop in pitch baseline and range in a parenthetical phrase, relative to the rest of the sentence. This pitch range change was not always perceived for appositives. In examining the associated fundamental frequency ($F0$) contours, we observed a region of reduced $F0$ excursion during the period of perceived range change, but this difference was difficult to quantify. Though intonation is an important cue, duration and pauses alone provide enough information to automatically label break indices with a high correlation (greater than 0.86) to hand-labeled break indices (Ostendorf *et al.*, 1990).

Since prominence was not consistently associated with specific syntactic structures in any systematic pattern we could discern (with the exception of particles), it appears that the disambiguating role of prominences (or pitch accents) differs from that of boundary phenomena, being associated more with the semantics rather than with the syntax of an utterance. In other words, we suspect that prominence is related more to the contextual focus of the sentence, although our data are sparse.

VI. DISCUSSION

We have confirmed that, for a variety of syntactic classes, but not all, naive listeners can reliably separate meanings on the basis of differences in prosodic information.

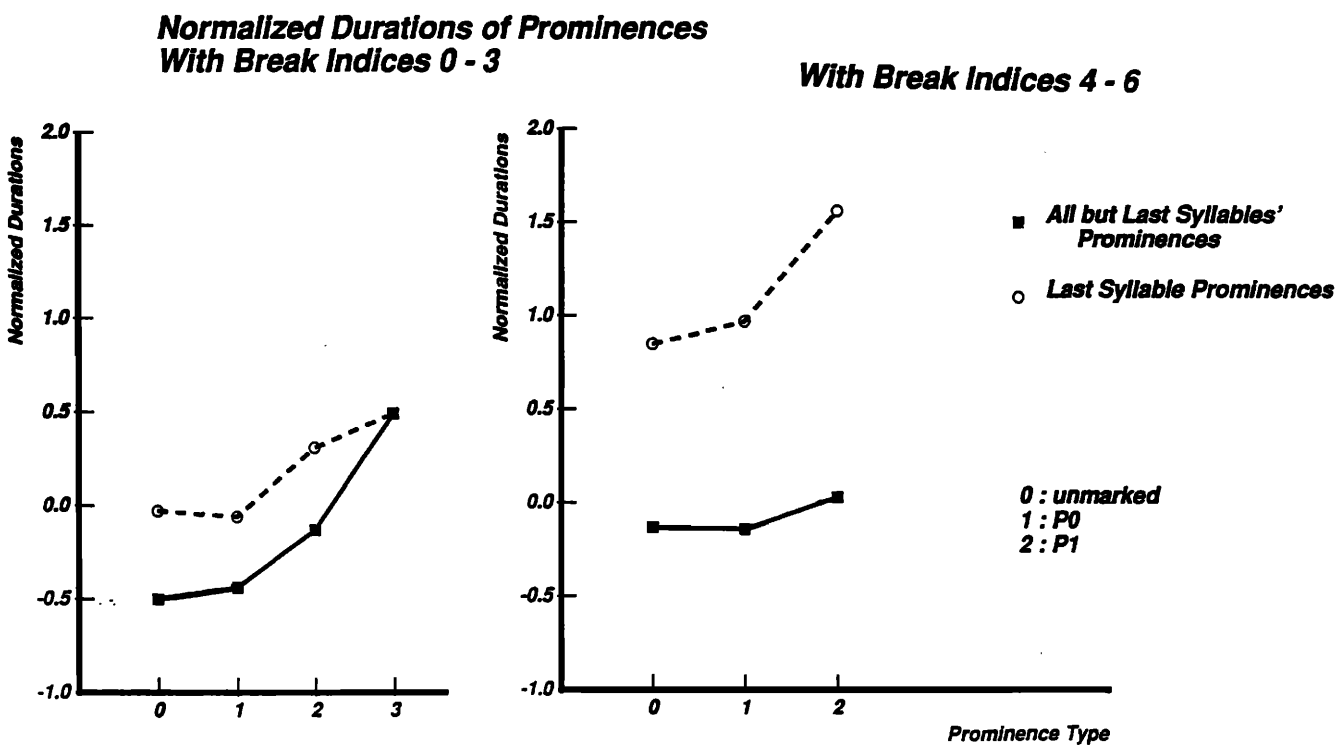

FIG. 4. Mean normalized duration of the vowel nucleus for different types of prominence markings, for syllables in words that precede a break index of 0–3 (at left), and for syllables in words that precede a major break (4–6) (at right).

We have further shown phonological and phonetic evidence bearing on how they might do this: by the tendency to associate relatively larger prosodic breaks with larger syntactic breaks. Further, syntactic boundaries of clauses that contain complete sentences nearly always coincide with the boundaries of major prosodic constituents (as marked, e.g., by syllable-final lengthening, a boundary tone and perhaps a pause). Syntactic constituents within these major constituents may be associated with any of several different levels of prosodic boundaries, i.e., speakers have more choice in phrasing, and prosodic boundaries need not correlate perfectly with syntactic ones, though they often do. We have also shown the importance of the relative size of prosodic breaks within a sentence. Though evidence relating to boundary phenomena appeared to be most important, there were some structures for which phrasal prominence either was the only cue or played a supporting role in distinguishing between the two versions.

Several aspects of the design of our experiment require comment involving the interpretation of our results. First, the disambiguation of some of the sentences may have been confounded by prosodic cues related to nonsyntactic factors, e.g., given versus new information, focus, contrastive stress, etc. However, the use of several sentences and of several speakers should minimize these effects, and should make it unlikely that there is a systematic correlation between such effects and the A and B versions of the sentences. Clearly, to fully elucidate the relationship between prosody and syntax will require the investigation of far more examples of far more syntactic constructions than we have been able to use in this study. Second, our finding of a correlation between the syntax and the phonological markers of prosody may have been corrupted by the fact that the labelers typically knew which version they were listening to. However, the labelers did not know the relative accuracy of the responses of the naive subjects. Therefore, these labels are relevant insofar as they account for both the accurate and the inaccurate responses. Third, we did not investigate the role of syntactic constituent length, which others have found to influence the placement of prosodic boundaries (Bachenko and Fitzpatrick, 1990). Fourth, the binary confident/not confident decision tended to reveal more about the persona-

TABLE III. Number of observations for different types of syllable markers, as a function of boundary location and word position.

Word-final syllable?	Word pre 4–6?	Unmarked	*P*0	*P*1	*C*
no	no	296	179	120	5
yes	no	1237	209	160	10
no	yes	160	83	106	5
yes	yes	267	84	162	1

lity of the rater than about the reliability of the judgment. In retrospect, we believe that a multivalued scale would provide a more useful measure of the relative confidence of listeners. Last, the use of read speech by professional radio announcers as speakers raises questions about generalizing the results to spontaneous speech by more average talkers. This is not a complicating factor for synthesis applications, but it is a complication for applications in speech understanding. We believe that the use of the professional speakers has allowed us to obtain initial results using far fewer speakers than would be needed using nonprofessionals. We hypothesize that the prosodic cues will be similar for nonprofessional speakers, although less consistently used and not as clearly marked.

One final observation of interest should be noted: the occurrence of apparently "neutral" prosodic patterns for some of the sentences in our corpus. That is, some utterances seemed appropriate for either of the possible interpretations of a sentence. This may arise when the prosodic breaks at different possible syntactic constituent boundaries are of equal size. The notion of a neutral prosodic pattern, if confirmed by further study, would be very useful in synthesis applications where sentence meaning may be quite difficult to compute.

Our results have both theoretical and empirical implications. We have shown that naive listeners can use prosody to separate structurally ambiguous sentence pairs, and we have further shown phonological and acoustic evidence of how they might do this. Prosodic cues may be particularly important in computer speech understanding applications, where the semantic rules available to the system are limited relative to the capabilities of human listeners. In addition, in these applications, prosodic cues can be used prior to semantic analysis, to reduce the number of syntactically acceptable parses by eliminating those that are inconsistent with the prosody (Ostendorf *et al.*, 1990). In speech generation applications, it is important to understand how the use of different prosodic markers will affect the interpretation of a sentence in order to improve comprehensibility of synthetic speech.

The results reported here provide evidence for some systematic relationships between prosody and syntax that should be explored further in several ways. First, a larger number of syntactic structures must be examined in order to make the prosody/syntax relationship more explicit. Second, we note that some sentences were successfully disambiguated with cues that were not represented in our labeling scheme. Since prominences were not differentiated as to type of pitch accent, a more detailed classification of intonation in such contexts could yield more information. Finally, for computer speech understanding applications, it will be important to investigate the extension of these results to spontaneous speech by nonprofessional speakers, where hesitation phenomena and speech errors will affect the prosodic structure.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IRI-8805680. The government has certain rights in this material. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. In addition, the authors wish to thank Andrea Levitt and Leah Larkey for their work with Patti Price in generating the ambiguous sentences; the radio announcers at WBUR in Boston who recorded the sentences; the subjects who participated in the perceptual experiments; Nanette Veilleux and Colin Wightman for many hours of prosodic labeling; and Gay Baldwin for verification of phonetic alignments. We thank John Bear for providing syntactic bracketings of the sentences and for many useful comments on the paper. We thank Julia Hirschberg for many useful discussions and for helpful comments on an earlier version of this paper. We thank Ivan Sag for syntactic advice on the sentences and sentence categories used.

APPENDIX: AMBIGUOUS SENTENCE PAIRS

For the contexts given below, the target sentences are given in italics and the percent accuracy in subject response is given in parentheses for F1A, F2B, F3A, and M1B, respectively.

1. Parenthetical main–main versus nonparenthetical main–subordinate

The A versions have parentheticals that are sentences (except number 4). The B members have an embedded sentence linked by a deleted subordinate conjunction to the previous word (except for 4). Sentences 4A and 4B did not seem to be outliers for their class based on the perceptual experiments.

1. A) Mary leaves on Tuesday. She will have no problem in Europe. *Mary knows many languages, you know.* (85, 93, 94, 100).

B) Mary and you have similar backgrounds and have both learned many languages. *Mary knows many languages you know.* (100, 93, 100, 100).

2. A) Don't think you can hide your plans from them. *They know, you realize, your goals.* (77, 80, 76, 92).

B) They think very highly of you and the way you get things done. *They know you realize your goals.* (92, 93, 100, 92).

3. A) Ian is so disorganized. I don't see how he could leave by Tuesday. *Does Ian know, I wonder, when he will leave*? (46, 60, 35, 100).

B) Ian has gotten his tickets but he has not told me what flight he is on. *Does Ian know I wonder when he will leave*? (85, 93, 100, 100).

4. A) Mel asked if you could give him a ride to the party next Saturday. *Mel knew, by the way, you were driving.* (54, 80, 88, 83).

B) It was obvious you had been drinking. Tom could smell it a mile away. *Mel knew by the way you were driving.* (85, 100, 94, 92).

5. A) For some reason it's Pat's job to tell her brother to move out. She's very upset about it, understandably. *The job, I mean, is never easy.* (92, 20, 94, 92).

B) You may be thinking of a job that is often difficult

but that has some moments of relief. I have another type of job in mind. *The job I mean is never easy.* (100, 100, 94, 100).

2. Apposition versus attached NP (or PP)

The A versions have NP apposition (except for 3 which is a PP adverbial). The B versions attach the NP or PP to the previous word, except for number 4, which is confounded by the Aull/all ambiguity. Neither 3A nor 4B seem to be outliers for their class based on the perceptual data.

1. A) The Smiths didn't know what to do with their time while their television was broken. *The neighbors who usually read, the Daleys, were amused.* (100, 93, 94, 75).

B) There was a funny Doonesbury today in all the local papers. *The neighbors who usually read the dailies were amused.* (100, 100, 100, 100).

2. A) Arthur took these pictures in a village where one caste does all the washing and another does all the wringing. *These are the ones who wring, the Belz.* (92, 100, 100, 100).

B) Here is the list of the people in work group *A* this week. *These are the ones who ring the bells.* (100, 100, 100, 100).

3. A) I can't remember whether you are going to Romania or Bulgaria. *Wherever you are, in Romania or Bulgaria, remember me.* (92, 100, 88, 67).

B) I hope you enjoy your tours of Bucharest, Sofia, and all those other towns you are visiting in Romania and Bulgaria. *Wherever you are in Romania or Bulgaria remember me.* (77, 80, 94, 42).

4. A) Your case would still be in court if not for May. *May gave Jane and Randy Aull, your lawyers, good ideas.* (92, 100, 88, 92).

B) Your lawyer's strategy would have worked if May had not been such a blabbermouth. *May gave Jane and Randy all your lawyer's good ideas.* (92, 87, 88, 100).

5. A) Most of the women had forgotten the strange event by the next week. *Only one remembered, the lady in red.* (100, 93, 100, 75).

B) Most of the people forgot about the strange visitor. *Only one remembered the lady in red.* (92, 100, 100, 67).

3. Main–main versus main–subordinate clauses

The A versions have conjunctions linking two sentences. The B versions have an embedded sentence attached by a deleted subordinate conjunction.

1. A) His mother and father did not have the same reaction when he announced he was going to become a hairdresser. *Mary was amazed and Dewey was angry.* (92, 93, 94, 92).

B) Mary couldn't believe anyone would object to such a harmless prank. *Mary was amazed Ann Dewey was angry.* (54, 53, 53, 58).

2. A) People get to work by various means. *Jane rides in the van and Della runs.* (85, 93, 65, 67).

B) Ann Della provides a commuter van service to the city. *Jane rides in the van Ann Della runs.* (31, 33, 94, 83).

3. A) I don't understand why Gary hadn't told the truth about where he'd been last night. *Gary knew we were worried, but he'd lied.* (100, 100, 71, 100).

B) Buddy had a bad habit of lying about his grades. *Gary knew we were worried Buddy'd lied.* (77, 80, 53, 17).

4. A) Lando's team won the last two games because we were arguing. *We'd better agree, or Lando may win again.* (100, 87, 76, 83).

B) Orlando is going to run for office again and we'd be foolish not to support him. *We'd better agree Orlando may win again.* (8, 13, 76, 75).

5. A) Do you think Mario will remember to hurry the visitors from Gannick Growers through the exhibit? *Mario will remember, or Gannick Growers may linger.* (100, 87, 94, 75).

B) He won't forget to schedule an extra long session when the organic farming association comes to see our farm. *Mario will remember organic growers may linger.* (38, 67, 35, 75).

4. Tags versus attached phrase

The A versions are all tag questions at the end of another sentence. The B versions contain at the site in question *V* plus *N* where the *N* is the direct object of the *V*, except for number 2, which contains an ADJ-*N*. This sentence does not appear to be anomalous based on the perceptual results.

1. A) Dave is always very angry, but it's futile to ask him why. *Dave will never know why he's enraged, will he?* (100, 93, 88, 100).

B) Dave can be obnoxious without realizing it. He just insulted Willy and is puzzled by his anger. *Dave will never know why he's enraged Willy.* (100, 47, 100, 75).

2. A) For years the Internationale was outlawed. Miles Davis could have played it after the socialists were elected. *Miles didn't know the melody was allowed, did he?* (100, 100, 100, 83).

B) When he turned on the music box in her hospital room he thought it would play a sweet little melody. *Miles didn't know the melody was a loud ditty.* (69, 100, 76, 100).

3. A) Most of Ben's friends and family had left but he remained. *Ben would never leave, would he?* (92, 93, 100, 83).

B) Woody was Ben's oldest friend. He took care of the invalid night and day. *Ben would never leave Woody.* (100, 53, 94, 100).

4. A) At the county fair Miriam and I were accosted by a revivalist asking if we believed. I thought I was very brave to say, "*Miriam and I don't believe, do we?*" (100, 100, 94, 92).

B) Dewey's lies are old hat to us now. *Miriam and I don't believe Dewey.* (100, 100, 94, 75).

5. A) I thought George would be home studying for his finals tonight, but he went dancing instead. *George isn't worrying, is he?* (100, 100, 94, 92).

B) George's threats affect some people, but Izzy still sells drugs. *George isn't worrying Izzy.* (38, 93, 94, 8).

5. Far versus near attachment of final phrase

The A versions have a final adverbial or PP attached nonlocally. The B versions have similar constituents attached to an immediately preceding word, except for number 3, which is scope ambiguity. This sentence was more often correctly identified than any other member of its class.

1. A) We each had to read technical articles in different languages. Jane read about bridge-building in Russian. *I read a review of nasality in German.* (46, 80, 65, 75).

B) For our linguistics projects, Jane read an article about negation in various French dialects. Sam read a series on Spanish consonants. *I read a review of nasality in German.* (69, 67, 71, 50).

2. A) Raoul's defense lawyer claimed that the murder weapon was a knife, but when we saw the body we knew they were wrong. *Raoul murdered the man with a gun.* (85, 93, 82, 92).

B) Raoul denied murdering either the man who carried a gun or his unarmed accomplice. We don't know yet who murdered the unarmed accomplice, but Raoul's blood-stained knife clinched the case for the other murder. *Raoul murdered the man with a gun.* (31, 80, 59, 33).

3. A) I never wear socks. Today was no exception. *I didn't wear them all day.* (85, 100, 65, 100).

B) I only wore these socks for an hour. *I didn't wear them all day.* (62, 67, 100, 58).

4. A) You'll never believe what she had on when she eloped. *Laura ran away with the man wearing a green robe.* (100, 100, 100, 67).

B) Which man did Laura run away with? *Laura ran away with the man wearing a green robe.* (62, 80, 29, 83).

5. A) Where did Andrea move the bottle? *Andrea moved the bottle under the bridge.* (31, 53, 47, 92).

B) There was one bottle under the bridge and another on the park bench. *Andrea moved the bottle under the bridge.* (54, 80, 53, 75).

6. Left versus right attachment of middle phrase

The A versions have a word (or, in one case, a prepositional phrase) attached to the left. The B versions members have the word or phrase attached to the right.

1. A) In spring there was always more work to do on the farm. May was the hardest month. *They rose early in May.* (85, 87, 82, 58).

B) Bears sleep all winter long, usually coming out of hibernation in late April, but this year they were a little slow. *They rose early in May.* (92, 87, 100, 92).

2. A) Rollo was terribly literal, often missing the forest for the trees. His approach to the satirical journal was no exception. *Rollo read the review literally, learning not an iota.* (92, 93, 100, 92).

B) He felt he had to read the journal, though it was poorly written and without content. *Rollo read the review literally learning not an iota.* (92, 100, 100, 92).

3. A) When I was a kid, I sneaked into an x-rated movie. *As I was eleven only, I knew my Dad would be angry.* (100, 100, 100, 100).

B) The other children were too young to know they were doing anything wrong. *As I was eleven, only I knew my Dad would be angry.* (100, 87, 94, 100).

4. A) John thought jogging in the woods would calm anyone down, but my nervous city cousins showed he was wrong. *Although they did run in the woods, they were uneasy.* (100, 100, 100, 100).

B) John and Jim liked running on a track. When they found out that part of the race went through the woods, they considered not running. *Although they did run, in the woods they were uneasy.* (100, 100, 100, 100).

5. A) My experience with slow learners has shown one thing. *When you learn gradually, you worry more.* (100, 100, 100, 100).

B) As you begin to study about nuclear war it becomes frightening. *When you learn, gradually you worry more.* (100, 100, 76, 92).

7. Particles and prepositions

The A versions have particle readings corresponding to the prepositional readings for the same words in the B versions.

1. A) There was mud on the floor but Marge was so busy she hadn't yet cleaned it up, thinking her kids would know enough to walk around it. But hearing the noise of little feet she turned around and asked in exasperation, "*Why are you grinding in the mud*?" (69, 87, 65, 42).

B) The scissors grinder had asked if he could use my yard to work in. I was surprised to see he had set up his wheel in the swampiest part of the yard, so I asked him, "*Why are you grinding in the mud*?" (92, 87, 94, 75).

2. A) The tactic of gentle persuasion led to a friendly settlement. *The men won over their enemies.* (92, 93, 82, 93).

B) Heartless violence led to a bloody victory. *The men won over their enemies.* (62, 93, 94, 100).

3. A) Tony and Ray were supposed to be painting Andrea's kitchen but were easily distracted. They loved to fix appliances and Andrea had mentioned how frustrated she was that one of the burners on the stove didn't get hot enough. *When I arrived, they were mulling over Andrea's annoying burner.* (69, 73, 88, 83).

B) Though Andrea's stove was cantankerous, Tony and Ray spent a lot of time getting it to work so they could mull the wine for the party. It had taken a long time, but a sweet smell was coming from the kitchen. *When I arrived, they were mulling over Andrea's annoying burner.* (69, 33, 59, 42).

4. A) Marge is a real card shark and adores dealing poker, but she will only play with women. We would sometimes try to get her to let one of our male friends into the game, but she always refused. *Marge would never deal in any guys.* (85, 100, 76, 92).

B) Marge loves cards but she refuses to deal. We would often try to trick her into doing it, but it never worked. *Marge would never deal in any guise.* (85, 100, 100, 100).

5. A) These new trucks are nice but they're too heavy for this residential-grade black-top. *They may wear down the road.* (92, 100, 82, 83).

B) That company claims that their ball bearings are the toughest made, but they haven't been submitted to the test of time and constant use. They look good now but who knows how long they will last? *They may wear down the road.* (54, 100, 94, 92).

[1] High versus low attachment is probably a more accurate syntactic description. However high versus low attachment could involve the same site in the string of words being parsed, and our instances of far (high) attachment all involve attachment to phrases ending in a word that is not neigh-

boring the word to be attached. Therefore, we instead use the more descriptive terms "far" and "near."

[2] Cooper and Paccia-Cooper (1980) investigated sentences analogous to those in this class (prepositional phrase attachment and scope of negation ambiguities) and found acoustic evidence (duration lengthening) at similar sites.

[3] In other work, we have found that variance can be reduced by adapting the phone means according to a local estimate of the speaking rate (Wightman *et al.*, 1992).

[4] This finding is similar to that of Beckman and Edwards (1991).

Bachenko, J., and Fitzpatrick, E. (**1990**). "A computational grammar of discourse-neutral prosodic phrasing in English," Comput. Linguist. **16**(3), 155–170.

Beckman, M., and Pierrehumbert, J. (**1986**). "Intonational structure in Japanese and English," edited by J. Ohala, Phonol. Yearbook **3**, 255–309.

Beckman, M., and Edwards, J. (**1991**). "Articulatory evidence for differentiating stress categories," in *Proceedings of the Third Conference on Laboratory Phonology*, edited by P. Keating (Cambridge U. P., Cambridge, MA).

Bing, J. (**1984**). "A discourse domain identified by intonation," in *Intonation, Accent and Rhythm: Studies in Discourse Phonology* (de Gruyter, New York), pp. 11–19.

Campbell, W. N. (**1990**). "Evidence for a syllable-based model of speech timing," in *Proceedings of the International Conference on Spoken Language Processing* (Acoust. Soc. of Japan, Japan), pp. 9–12.

Cooper, W., and Sorensen, J. (**1977**). "Fundamental frequency contours at syntactic boundaries," J. Acoust. Soc. Am. **62**, 683–692.

Cooper, W., and Paccia-Cooper, J. (**1980**). *Syntax and Speech* (Harvard U. P., Cambridge, MA).

Crystal, T. H., and House, A. S. (**1988**). "Segmental durations in connected-speech signals: Current results," J. Acoust. Soc. Am. **83**, 1553–1573.

Crystal, T. H., and House, A. S. (**1990**). "Articulation rate and the duration of syllables and stress groups in connected speech," J. Acoust. Soc. Am. **88**, 101–112.

Duez, D. (**1985**). "Perception of silent pauses in continuous speech," Language Speech **28**(4), 377–389.

Garro, L., and Parker, F. (**1982**). "Some suprasegmental characteristics of relative clauses in English," J. Phon. **10**, 149–161.

Gee, J. P., and Grosjean, F. (**1983**). "Performance structures: A psycholinguistic and linguistic appraisal," Cognitive Psychol. **15**, 411–458.

Geers, A. (**1978**). "Intonation contour and syntactic structure as predictors of apparent segmentation," J. Exp. Psych: Hum. Percept. Perform. **4**(3), 273–283.

Hirsh-Pasek, K., Kemler Nelson, D., Jusczyk, P., Wright Cassidy K., Druss, B., and Kennedy, L. (**1987**). "Clauses are perceptual units for young infants," Cognition **26**, 269–286.

Klatt, D. H. (**1975**). "Vowel lengthening is syntactically determined in a connected discourse," J. Phon. **3**, 129–140.

Klatt, D. H. (**1987**). "Review of text-to-speech conversion for English," J. Acoust. Soc. Am. **82**, 737–793.

Kutik, E., Cooper, W., and Boyce, S. (**1983**). "Declination of fundamental frequency in speakers' production of parenthetical and main clauses," J. Acoust. Soc. Am. **73**, 1731–1738.

Ladd, D. R. (**1986**). "Intonational phrasing: the case for recursive prosodic structure," Phonol. Yearbook **3**, 311–340.

Lehiste, I. (**1973**). "Phonetic disambiguation of syntactic ambiguity," Glossa **7**(2), 107–121.

Lehiste, I., Olive, J. P., and Streeter, L. A. (**1976**). "Role of duration in disambiguating syntactically ambiguous sentences," J. Acoust. Soc. Am. **60**, 1199–1202.

Liberman, M. Y., and Prince, A. S. (**1977**). "On stress and linguistic rhythm," Linguistic Inquiry **8**, 249–336.

Lieberman, P. (**1967**). *Intonation, Perception and Language* (MIT Research Monograph No. 38, MIT, Cambridge, MA).

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (**1988**). "A precursor of language acquisition in young infants," Cognition **29**, 143–178.

Nespor, M., and Vogel, I. (**1983**). "Prosodic structure above the word," in *Prosody: Models and Measurements*, edited by A. Cutler and D. R. Ladd (Springer-Verlag, New York), pp. 123–140.

Ostendorf, M., Price, P., Bear, J., and Wightman, C. (**1990**). "The use of relative duration in syntactic disambiguation," in *Proceedings of the 3rd DARPA Workshop on Speech and Natural Language* (Morgan Kaufman, San Mateo, CA).

Pierrehumbert, J. (**1980**). "The Phonology and Phonetics of English Intonation," Ph.D. Dissertation, MIT, Cambridge, MA.

Pierrehumbert, J. (**1981**). "Synthesizing intonation," J. Acoust. Soc. Am. **70**, 985–995.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., and Veilleux, N. (**1988**). "A methodology for analyzing prosody," J. Acoust. Soc. Am. Suppl. 1 **84**, S99.

Scholes, R. (**1971**). "On the spoken disambiguation of superficially ambiguous sentences," Language Speech **14**, 1–11.

Scott, D. (**1982**). "Duration as a cue to the perception of a phrase boundary," J. Acoust. Soc. Am. **71**, 996–1007.

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

Streeter, L. A. (**1978**). "Acoustic determinants of phrase boundary perception," J. Acoust. Soc. Am. **64**, 1582–1592.

Thorsen, N. (**1980**). "A study of the perception of sentence intonation—Evidence from Danish," J. Acoust. Soc. Am. **67**, 1014–1030.

Thorsen, N. (**1985**). "Intonation and text in standard Danish," J. Acoust. Soc. Am. **77**, 1205–1216.

Wales, R., and Toner, H. (**1979**). "Intonation and ambiguity," in *Sentence Processing*, edited by W. C. Cooper and E. C. T. Walker (Erlbaum, Hillsdale, NJ).

Wightman, C., and Ostendorf, M. (**1991**). "Automatic recognition of prosodic phrases," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*.

Weintraub, M., Murveit, H., Cohen, M., Price, P., Bernstein, J., Baldwin, G., and Bell, D. (**1989**). "Linguistic constraints in hidden Markov model based speech recognition," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* (Bell & Bain, Glasgow, Scotland), pp. 699–702.

Wightman, C., Shattuck-Hufnagel, S., Price, P., and Ostendorf, M. (**1992**). "Segmental durations in the vicinity of prosodic phrase boundaries," J. Acoust. Soc. Am. (to be published).