Psychological Review
1993, Vol. 100, No. 2, 233–253

0033-295X/93/$3.00

Creation of Prosody During Sentence Production

Fernanda Ferreira

Phrase-final words tend to be lengthened and followed by a pause. The dominant view of prosodic production is that word lengthening and pausing reflect the syntax of a sentence. The author demonstrates that, instead, lengthening and pausing reflect a distinctly prosodic representation, in which phonological constituents are arranged in a hierarchical, nonrecursive structure. Prosodic structure is created without knowledge of words' phonemic content. As a result, within a single sentential position, greater word lengthening necessitates shorter pauses, but across positions, word and pause durations show a positive correlation. The author presents a model of prosodic production that describes the process of prosodic encoding and provides a quantitative specification of the relation between word lengthening and pausing. This model has implications for studies of language production, comprehension, and development.

Before the publication of Karl Lashley's (1951) classic paper, "The Problem of Serial Order in Behavior," many psychologists thought of language as a linear system. Words in a sentence were strung together like beads on a string, and sentence production consisted of generating words in the correct sequence. The tasks of sentence production, on this view, amounted simply to locating words in the mental dictionary and using the current word to trigger production of the next. Lashley argued that the speed with which many complex motor tasks could be performed made it unlikely that the elements of those tasks were planned sequentially. The existence of anticipatory speech errors such as "breakfast is the Wheaties of champions" (Garrett, 1980) also indicated that speech was not planned one word at a time. Lashley suggested instead that groups of words were produced under the guidance of a hierarchical plan.

A decade later, theorists in linguistics and psychology proposed that language was organized into hierarchical linguistic representations. Chomsky (1957, 1965), following Harris (1951), developed a theory of the syntactic component of the grammar that made use of the concepts of hierarchy and recursion. Yngve (1960) designed a complexity metric for sentence production that was based on a node count of a syntactic tree. By the end of the 1960's, the notion that sentences were organized into hierarchical syntactic structures at some level of description was widely accepted (see Fodor, Bever, & Garrett, 1974, for a review of relevant arguments). Garrett (1975, 1976, 1980) used speech-error data to develop a psychological model of the way syntactic structures are generated during sentence production. More recently, the processes responsible for syntactic production have been explored experimentally (Bock, 1986, 1987b; Bock & Loebell, 1990; Bock & Miller, 1991; Ferreira, 1991).

Although the field of psycholinguistics has progressed in accounting for syntactic regularities in production, little is known about the processes responsible for generating a sentence's sound pattern. For example, some words in a sentence are louder and longer than others. Semantically important content words are acoustically prominent, whereas words such as *the* are virtually inaudible outside their sentential context. These are not mere tendencies; no speaker of English would say "the cat" so that both words were equal in duration or stress. The widely accepted view of how these prosodic regularities are generated has been that the speaker produces a syntactic structure for a sentence and then uses this hierarchical structure to assign phonetic values such as duration and degree of stress to particular words (Cooper, 1980; Cooper & Danly, 1981; Cooper & Paccia-Cooper, 1980; Kaisse, 1985; Klatt, 1975; Selkirk, 1984).

The purpose of this article is to examine how syntax affects sentence prosody. The issue, I believe, is not whether syntax affects prosody; it is clear that many syntactic variables can be shown to influence the prosodic form of a sentence. The issue is whether the syntactic representation of a sentence directly influences phonological variables (as argued recently by Chen, 1990, and Odden, 1990) or whether syntax is used to construct an intermediate representation, a prosodic representation, which

Parts of this research were presented at the Annual Meeting of the Psychonomic Society, Seattle, Washington; the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge; the Annual City University of New York Sentence Processing Conference, New York; the Department of Linguistics, University of California, Los Angeles; and the Max Planck Institut für Psycholinguistik, Nijmegen, The Netherlands. I thank those audiences for lively discussions and helpful comments. Experiment 3 was part of a dissertation submitted to the University of Massachusetts, Amherst; I thank Chuck Clifton, Lyn Frazier, Keith Rayner, David Rosenbaum, and Elisabeth Selkirk for their assistance with that project. I thank Chuck Clifton, Pete Dixon, Vic Ferreira, John Henderson, and three anonymous reviewers for comments on an earlier version of this article and Ramona Clifton, Lucille Lavender, and Andrea Kross for their assistance with the experiments. Experiments 1, 2, and 4 were supported by Natural Sciences and Engineering Research Council of Canada Grant OGP-37323 to Fernanda Ferreira; Experiment 3 was supported by National Institutes of Health Grant HD-18708 to Chuck Clifton and Lyn Frazier.

Correspondence concerning this article should be addressed to Fernanda Ferreira, Department of Psychology, Michigan State University, East Lansing, Michigan 48824-1117. Electronic mail may be sent to fernanda@msu.edu.

in turn influences variables such as word duration and stress (Inkelas & Zec, 1990; Selkirk, 1986). In this article, I present experimental evidence for the psychological reality of a prosodic representation in sentence production (see also Bock & Loebell, 1990; Ferreira, 1991; Gee & Grosjean, 1983; Levelt, 1989). I also provide evidence that a prosodic structure is created from a sentence's syntactic structure but without knowledge of its phonemic content. In the General Discussion section, I describe a model of the relation between word lengthening and pausing and relate it to a larger theory of prosodic encoding (Levelt, 1989). In addition, I describe the differences between the view I am advocating and the one promoted by Gee and Grosjean (1983), who also argued for a distinctly prosodic structure but with radically different characteristics. Finally, I address the implications of this work for psycholinguistics in general.

Phrase-Final Lengthening

Previous research has demonstrated that a word and its following pause[1] tend to have a longer duration at the end of a phrase than in any other phrasal position (Boomer, 1965; Cooper & Paccia-Cooper, 1980; Klatt, 1975; Nakatani, O'Connor, & Aston, 1981; Oller, 1973). This phenomenon is referred to as *phrase-final lengthening,* although the term *prepausal lengthening* is also sometimes used, to highlight the tendency of a pause to follow a lengthened word. To illustrate, consider the sentences in Example 1.

Example 1

a. The table that I thought was black tempted me.
b. The black table tempted me.

The word *black* and the pause after it would tend to be longer in 1a than in 1b, because in 1a, *black* is the last word of the subject-noun phrase, whereas in 1b, *black* occurs in the middle of a phrase.

Cooper (1980; Cooper & Paccia-Cooper, 1980) has claimed that a sentence's timing pattern is highly correlated with its syntactic phrasal bracketing. Therefore, research on word duration and pausing may reveal the characteristics of syntactic representations constructed during sentence production. However, does an effect of phrasal position necessarily implicate syntactic structure? Before the 1980's, there was little choice but to answer "yes," for no other candidate representation was available to account for these effects. However, the past decade has seen an explosion of work in linguistics on the prosodic structure of language (for a recent anthology, see Inkelas & Zec, 1990). In the same way that linguistic argumentation has motivated the existence of syntactic constituents such as noun phrases, verb phrases, and sentences (e.g., Chomsky, 1957, 1965, 1986), similar arguments have been used to motivate the existence of prosodic constituents—units such as syllables, prosodic words, phonological phrases, and intonational phrases. Furthermore, a great deal of experimental work has demonstrated the psychological reality of syntax in language comprehension and production (e.g., Bock, 1986; Bock & Loebell, 1990; Ferreira, 1991; Fodor et al., 1974; Frazier, Clifton, & Randall, 1983; Frazier & Rayner, 1982; Garrett, 1975, 1980, 1988; MacDonald, 1989; Nicol & Swinney, 1989). The goal of the work described here is to demonstrate the psychological reality of prosodic structure in sentence production.

The work that I present here is part of a larger trend within psycholinguistics toward examining the role of prosody and phonological information generally in language processing (Ferreira & Anes, in press). In the area of comprehension, it has been shown that variations in word and pause duration can be used to assist the parser in assigning a syntactic structure to a sentence (Beach, 1991; Slowiaczek, 1981) and that variations in the pronunciation of words (because of processes such as flapping and palatalization) can be used to assist in spoken word identification (Church, 1987; see also Frauenfelder & Lahiri, 1989). In a similar vein, Kelly (1992) has presented evidence that phonological information can be used to resolve grammatical category assignments. In the area of language acquisition, a number of researchers have recently explored children's use of prosodic information to learn the syntactic organization of their language (e.g., Cassidy & Kelly, 1991; Fernald & Mazzie, 1991; Gerken, 1991; Jusczyk et al., 1992; Morgan, Meier, & Newport, 1987). Researchers in language production have also been giving a great deal of attention to the question of how language is phonologically encoded (Levelt, 1989; Meyer, 1990, 1991, 1992; Schriefers, Meyer, & Levelt, 1990). Finally, even in the field of reading, researchers have begun to emphasize the influences of phonological information on basic reading processes (Coltheart, Avons, Masterson, & Laxon, 1991; Daneman & Stainton, 1991; Perfetti & Zhang, 1991; Van Orden, Johnston, & Hale, 1988).

Theory of Prosodic Structure

A number of theories of prosodic structure have been proposed in the linguistic literature (for a discussion, see Levelt, 1989). The theory I describe here is based largely on the work of Elisabeth Selkirk (1984, 1986). This work has been extremely influential in both linguistics and psychology. According to Selkirk and others, prosodic structure consists of distinctly prosodic constituents arranged in a hierarchy, as shown in Figure 1 (Nespor & Vogel, 1987; Selkirk, 1986). Branches in a prosodic tree cannot cross, so that a constituent low in the hierarchy (e.g., the prosodic word) cannot be part of more than one higher level constituent (e.g., the phonological phrase). The utterance is the highest unit in the hierarchy, and therefore it constitutes the root node in any prosodic tree. Thus, the entire sentence shown in Figure 1 is one utterance.

The constituent immediately below the utterance is the intonational phrase, which Nespor and Vogel (1987) defined as the unit over which pitch in a declarative sentence typically falls and that tends to be followed by a pause. For example, the sentence in Figure 1 consists of two intonational phrases. The

[1] The type of pause I discuss in this article may be termed *timing based* (Ferreira, 1988, 1991), to contrast with other pauses that reflect factors such as word-finding difficulty. In earlier articles, (Ferreira, 1988, 1991), I provided evidence that these two types of pauses are distinct and should be treated separately in any theory of sentence production. This issue is taken up briefly in the General Discussion section.

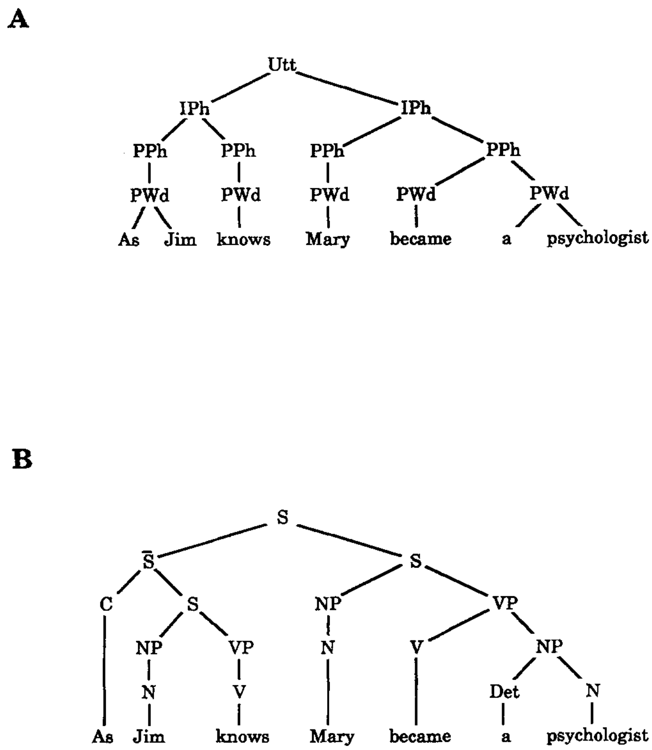

Figure 1. Prosodic and syntactic structures for the sentence "As Jim knows, Mary became a psychologist." (Utt = utterance; IPh = intonational phrase; PPh = phonological phrase; PWd = prosodic word; S̄ = sentence-bar; S = sentence; NP = noun phrase; VP = verb phrase; N = noun; V = verb; Det = determiner; C = complementizer.)

first one includes the sequence *as Jim knows,* because preposed phrases of this type obligatorily constitute a single intonational unit (Nespor & Vogel, 1987; Pierrehumbert, 1980; Selkirk, 1984). As a result, the remaining sequence *Mary became a psychologist* must also be an intonational phrase, because any level in the prosodic hierarchy must be exhaustively divided into the constituents appropriate for that level (Selkirk, 1986).

Below the level of the intonational phrase is the phonological phrase (Hayes, 1984; Kanerva, 1990; Nespor & Vogel, 1987; Selkirk, 1986; Selkirk & Shen, 1990). According to Selkirk (1986), phonological phrases are created through an operation she terms the *X-max algorithm* (described in Ferreira, 1991). This algorithm scans the syntactic structure of a sentence (shown in the lower half of Figure 1) looking for the right boundary of a full phrasal constituent. All the material up to the syntactic boundary constitutes a single phonological phrase. Thus, in Figure 1, *Mary* would constitute the first phonological phrase, because *Mary* occurs at the end of the subject-noun phrase. The sequence *became a psychologist* constitutes the second phonological phrase, because the next syntactic phrase remains open until after the word *psychologist.*[2]

Below the phonological phrase is the prosodic word. Prosodic words are based on the distinction between content and function words. Content words are members of the major lexical categories, including nouns, verbs, adjectives, and adverbs, and tend to convey much of the semantic content of a sentence. Function words are members of the minor lexical categories, including prepositions, complementizers, determiners, and conjunctions, and tend to convey the syntactic roles played by content words. Selkirk (1984) suggested that, phonologically, function words are not really words at all; they are more like clitics that attach to an appropriate lexical host. For example, the word *the* tends to be spoken in close temporal proximity to the following adjective or noun; the particle *up* in a phrase such as *throw up* tends to attach to the preceding word. A prosodic word, then, is a content word together with any attached function word or words.[3]

According to such models of prosodic structure, syntax influences the prosodic characteristics of a sentence, but not directly. Syntax has an indirect influence, because the syntactic structure of a sentence influences the sentence's organization into prosodic units. For example, a clause boundary virtually forces an intonational-phrase boundary (Nespor & Vogel, 1987), and a syntactic-phrase boundary, at least on Selkirk's (1986) model, forces a phonological-phrase boundary. As a result, these syntactic boundaries will sometimes be the locus of phrase-final lengthening, pausing, and pitch fluctuation. Of course, given that there is some correlation between syntactic and prosodic structures, it is important to investigate the differences between them to justify the need to postulate a prosodic level of representation in language processing.

One major difference between syntactic and prosodic structures is in the way constituents are defined; for example, a prosodic word is defined differently from a syntactic word. Another crucial difference is that a prosodic constituent cannot be embedded inside an identical or higher level constituent; that is, prosodic structures are not recursive (Nespor & Vogel, 1987; Selkirk, 1984, 1986; for a different view see Ladd, 1986). For example, the intonational phrases shown in Figure 1 are made up of phonological phrases, the unit immediately below in the prosodic hierarchy. Syntactic structures, in contrast, can be recursive. For example, in Figure 1, the topmost sentence node in the syntactic structure dominates two lower level sentence nodes.

The lack of recursion in prosodic structures makes them less complex than syntactic structures. This decrease in complexity has implications for global models of sentence production. Bock (1982, 1987a) and Levelt (1989) have argued that the goal

[2] It should be noted that Selkirk's (1986) model of phonological-phrase construction is somewhat different from that proposed by others, particularly Nespor and Vogel (1987). According to Nespor and Vogel, phonological phrases typically included material up to and including a lexical head; thus, their definition of a phonological phrase is quite similar to that of Gee and Grosjean (1983). The sequence *Mary became a psychologist* would therefore consist of three phonological phrases as follows: (*Mary*) (*became*) (*a psychologist*). One problem with the Nespor and Vogel and Gee and Grosjean model is that it can produce structures in which the main verb patterns with the sentential subject, a prosodic pattern that speakers rarely produce (Ferreira, 1991). For this and other reasons, I have chosen to work with the Selkirk model of phonological-phrase construction.

[3] Again, substantial differences exist between Selkirk's (1984, 1986) definition and Nespor and Vogel's (1987) definition of *prosodic word.* For consistency, I have adopted Selkirk's approach.

of the sentence-production system is to map thoughts and ideas, which are inherently unordered, onto a sequential speech channel. Pinker and Bloom (1990) made a similar point, arguing that one purpose of language is to enable communication of propositional structures over a serial channel (the vocal apparatus). It is unlikely that this translation happens in a single step. Instead, intermediate representations are constructed that make the propositional structure progressively more linear. A propositional structure, which is multidimensional, hierarchical, and recursive (Anderson, 1990), is translated into a syntactic structure. A syntactic structure, which is hierarchical and recursive, but only two-dimensional (or possibly three-dimensional; Espinal, 1991; Williams, 1978), is translated into a prosodic structure; a prosodic structure, which is hierarchical and two-dimensional, but not recursive, is translated into a linear phonetic structure. Thus, prosodic structure serves as one of the representations that effect the translation between thought and speech.

The phonological processes and structures discussed thus far fall within what is termed *autosegmental phonology*. Autosegmental phonology concerns timing, stress, and tone and units such as syllables, phonological phrases, and intonational phrases; segmental phonology concerns the behavior of phonemes (also termed *segments*) and articulatory features. These two aspects of phonology are presumed to operate on different planes or tiers (Prince, 1984), and therefore they are formally independent. As noted by Levelt (1989), linguistic independence suggests a possible psychological independence between the processes that create the prosodic form of a sentence and those that access the segmental forms of words. In Levelt's model of language production, the procedures that access the segmental composition of a word are independent both of the procedures that access a word's segmental structure and those that construct the prosodic representation of a sentence. The results of experiments described in the following sections are consistent with this model.

Separating Syntactic and Prosodic Structure

According to a model in which syntax directly influences phrase-final lengthening (Cooper, 1980; Cooper & Paccia-Cooper, 1980), the position a word occupies within a syntactic phrase affects that word's duration. According to a prosodic model, lengthening occurs because of the position a (prosodic) word occupies within an intonational or phonological phrase. Sentences such as 1a and 1b illustrate the problem of establishing whether lengthening is a syntactic or prosodic phenomenon. The two versions of the sentence differ in both syntactic and prosodic structure, and so the critical word *black* occurs in different syntactic and prosodic contexts. At the syntactic level, the critical word occurs either in the middle or at the end of the subject-noun phrase. At the prosodic level, the critical word occurs either in the middle or at the end of a phonological phrase. To unconfound the two types of structure and to show the need for a distinctly prosodic structure, one type of structure must be manipulated independently of the other, and the two must be shown to produce different effects. There are two logical possibilities: One is to vary prosodic structure and not syntactic structure; the other is to vary syntactic structure and not prosodic structure. The first possibility is shown in Example 2:

Example 2

a. The girl left the room.
b. The GIRL left the room.

The capitalization of *girl* in 2b indicates that *girl* receives contrastive prominence; 2b might be an appropriate statement in a context in which the speaker incorrectly believed a boy had left the room (Rochemont, 1986). As a result of the extra stress added to *girl* in 2b, an intonational-phrase boundary at that position is more likely than in 2a (Rochemont, 1986; Selkirk, 1984). The sentences in Example 2, then, may be produced with the following intonational structures, where all the material inside parentheses is inside one intonational phrase (IPh)[4]:

Example 3

a. (The girl left the room)$_{\mathrm{IPh}}$.
b. (The GIRL)$_{\mathrm{IPh}}$ (left the room)$_{\mathrm{IPh}}$

According to a direct syntactic account of phrase-final lengthening, the duration of *girl* should be equal in the two cases because their syntactic structures do not differ. The prosodic account states that the amount of lengthening is a function of the number of prosodic constituents that end on a particular word. The word *girl* should be longer in 3b than in 3a because of the potential intonational-phrase boundary after *girl* in 3b.

However, this prediction might seem unfair to the syntactic hypothesis. One acoustic correlate of stress is not only increased amplitude or loudness but also increased duration (Crystal & House, 1988; Fry, 1955). Therefore, it is possible that *girl* must be longer to be stressed. There are two ways to deal with this objection. One is to measure the pause after *girl*. It may be necessary to lengthen a word to give it stress, but it is not necessary to pause after the word. If an intonational-phrase boundary occurs after *girl* because of contrastive prominence, then according to Nespor and Vogel (1987), a pause is likely, because pauses typically follow intonational phrases. Therefore, a pause after *girl* in 3b but not in 3a would tend to support the prosodic view. These predictions are examined in Experiment 1.

A second way to deal with the objection is to examine the other logical possibility: vary the syntactic structure but hold the prosodic structure constant. Two sentences with the same

[4] It may seem somewhat weak to claim merely that the sentences may have the indicated intonational structures. However, as many workers in the field of sentence phonology have noted, intonational variability must be accounted for in any model of language production (see the discussion in Levelt, 1989). A model such as Gee and Grosjean's (1983), for instance, which algorithmically generates a single intonational structure for any given sentence, cannot account for the available options in intonational phrasings. Instead, a model of sentence production should specify the factors that may influence the probability of an intonational-phrase boundary at a particular point in a sentence. In the case of the sentences in Example 2, the presence of contrastive prominence on a word increases the likelihood of a following intonational-phrase boundary.

prosodic structure but different syntactic structures are shown in Example 4:

Example 4

a. The friendliest cop saw the enterprising girls.
b. The guy who's a cop saw the enterprising girls.

As spelled out in more detail later, these two versions have the same prosodic structure; intuitively, the rhythm, timing, and intonational characteristics of these two sentences are similar.[5] However, the subject-noun phrase in 4a is less complex than in 4b (see Example 9 for the relevant syntactic structures). In 4a, the word *cop* is followed by the right bracket of one syntactic constituent, and in 4b, the word *cop* is followed by five such brackets. (See Ferreira, 1991, for more details about these structures.) According to Cooper's syntactic model (1980; Cooper & Paccia-Cooper, 1980), the more right brackets following a word, the more that word will be lengthened. Cooper and Paccia-Cooper suggested that this effect may be due to a relaxation response on the part of the sentence-production system. They propose that constructing syntactic phrases is resource intensive, and therefore the system takes a short "break" at the end of a phrase by lengthening the phrase-final word and sometimes pausing. The more phrases that have been constructed, the more breaks that will be required.

According to the syntactic model, then, the duration of *cop* should be longer in 4b than in 4a. According to the prosodic model, their durations should be equal, because the two sentences do not differ in their prosodic structures. A problem with the comparison illustrated in Example 4, however, is that evidence for the prosodic model would be based on a null result. To deal with this problem, a sentence such as the one in Example 5 will be compared with the ones shown in Example 4.

Example 5

The cop who's my friend saw the enterprising girls.

Because the word *cop* is now phrase medial in both a syntactic and prosodic phrase, both types of models predict that its duration should be shorter than in either of the two conditions shown in Example 4. These predictions are examined in Experiment 2.

Establishing the psychological reality of prosodic structure would have important consequences for current theories of sentence production. Until recently, these theories have not included explicit procedures for the creation of a distinctly prosodic level of representation (e.g., Garrett, 1975, 1976, 1988), perhaps because the need for such a representation was not apparent. One might be tempted to argue that linguistic evidence has already established the need for such a level and that experimental studies of the type to be presented here are unnecessary. However, there are two problems with this argument. First, there is still disagreement even among linguists about the need for prosodic structure (see the discussion in Inkelas & Zec, 1990), although it appears that the proponents are gaining ground. Second, much of the linguistic evidence is based on subjective intuition about a word's duration in a particular context and whether a pause follows a word. This evidence is suspect both because linguistic intuitions generally can be unreliable and biased (Gerken & Bever, 1986) because it is difficult to rely on intuition to determine whether one word's duration is longer than another's and because it is virtually impossible to distinguish perceptually between word lengthening and pausing (Martin, 1970). The first two experiments in this article, I believe, make the case convincingly for the psychological reality of prosodic structure in sentence production.

In the remaining two experiments, I address a more specific question: Is the production of sentence prosody independent of the selection of words' segmental content? On the basis of the results of Experiments 3 and 4 in this article, I argue that the production of prosody precedes the selection of word forms, implying a separation between segmental and autosegmental processing. In the General Discussion section, I relate these findings to Levelt's (1989) model of prosodic encoding. A second issue that I address on the basis of Experiments 3 and 4 is the relation between word lengthening and pausing. I examine this issue by varying both the prosodic and segmental properties of sentences and then measuring word and pause durations. Many researchers have found that lengthening and pausing tend to co-occur (e.g., Cooper & Paccia-Cooper, 1980; Klatt, 1975) but have not provided an explanation of this co-occurrence. The model that I describe in the General Discussion section includes a quantitative mechanism to account for the relation between lengthening and pausing.

Finally, before turning to the experimental portion of this article, I address one potential methodological concern. In any study of language production, the researcher must decide how sentences will be elicited from speakers. The task chosen will balance the desire to mimic normal speaking conditions as much as possible with the need to control the linguistic characteristics of the spoken sentences. In the field of timing, almost invariably the task used is a reading task (described in detail in Cooper & Paccia-Cooper, 1980): Speakers silently read a sentence typed on a card until they believe they can produce the sentence without error, and then they simply read the sentence out loud. I also used this task, despite some of its obvious limitations, for a number of reasons. First, to compare the results obtained here with those obtained in the literature on timing more generally, it seemed advisable to use the same task. Second, distinguishing between the prosodic and syntactic models of timing requires careful construction of materials and precise control over the syntactic, prosodic, and segmental characteristics of the speakers' sentences. Third, Anderson and Cooper (1986) have found that intonation and timing patterns obtained in reading tasks are quite similar to those observed in more natural speaking situations. Consistent with this finding, Experiments 3 and 4 in the present article produced remarkably similar results, yet the former used a sentence-memorization task (Ferreira, 1991) while the latter used the more standard reading task. Thus, although in the long run it will be useful to examine the prosodic characteristics of sentences spoken in a wide variety of circumstances, the reading and sentence-me-

[5] It is possible to produce 4b with an intonational-phrase boundary after *guy*, thus resulting in different prosodic structures for 4a and 4b. However, as I show later, a boundary after *guy* would not affect the prosodic structure around the critical word *cop*.

morization tasks were deemed appropriate for the purposes of this article.

Experiments Separating Prosodic and Syntactic Structure

The purpose of the first two experiments was to determine whether word lengthening at the ends of phrases is due to the effects of prosodic or syntactic structure. In Experiment 1, only the prosodic structure was varied; in Experiment 2, only the syntactic structure was varied.

Experiment 1

In this experiment, sentences were varied so that their prosodic structure changed but their syntactic structure remained constant. Consider Example 6:

Example 6

The crate contains the missing book.

The final word of the subject in this sentence (*crate*) was spoken either with contrastive prominence or in a neutral manner. In what I term the *prominent* condition, the existence of contrastive prominence on *crate* increases the likelihood of an intonational boundary after the word (Rochemont, 1986; Selkirk, 1984). Thus, the sentence may sometimes be spoken as two intonational phrases, one including the subject of the sentence and the other including the verb phrase (*contains the missing book*).

The condition that I term the *neutral* condition requires more comment. Nespor and Vogel (1987), among others (Levelt, 1989; Selkirk, 1984), pointed out that there will be variability in how a given sentence is intonationally phrased. At least three factors determine whether a sentence will be spoken as a single intonational phrase or as more than one. The first is the existence of what Selkirk (following Pierrehumbert, 1980) termed a *pitch accent* on a word. A pitch accent is a general term to describe the presence of some sort of prosodic prominence on an element of a sentence. A pitch accent affects the likelihood of an intonational-phrase boundary. In a normal or neutral rendition of a sentence, the most prominent word will tend to be the last word of the sentence (the Nuclear Stress Rule; Chomsky & Halle, 1968; Liberman & Prince, 1977). Thus, in a neutral rendition, the intonational-phrase boundary will occur at the end of the sentence, and therefore the sentence will consist of only one intonational phrase. If a word in the middle of a sentence is prominent, it may be followed by an intonational-phrase boundary, so that the sentence would consist of two intonational phrases.

The second factor is speech rate. The more slowly a sentence is spoken, the greater the tendency to divide it into more than one intonational phrase. This factor was controlled in the present experiment by having all subjects speak at what they judged to be their normal or typical speech rate.

The third factor concerns the length of the sentence (Gee & Grosjean, 1983; Nespor & Vogel, 1987). The longer a sentence, the greater the likelihood that it will be spoken as more than one intonational phrase. In the present experiment, this factor was controlled by constructing sentences that were relatively short. For example, the sentence in Example 6 consists of eight syllables and six words.

These considerations lead to the expectation that, in the neutral condition, the sentence in Example 6 will be spoken as one intonational phrase. In the prominent condition, however, the existence of prominence on the final word of the subject will increase the likelihood that the sentence will be spoken as two intonational phrases. According to an extreme version of the syntactic account, word and pause durations will not differ in these two conditions. A more modified view might predict longer word duration in the prominent condition because of the extra stress on the word but would not necessarily predict a longer following pause. According to the prosodic account, word and pause durations will both be longer in the prominent condition than in the neutral condition because of the increased likelihood of an intonational-phrase break after the prominent word.

Method

Subjects. Ten University of Alberta undergraduates participated in the experiment in exchange for course credit. All subjects were native speakers of Canadian English.

Materials. Each sentence appeared in one of two versions, shown in Example 7:

Example 7

a. The crate contains the missing book. (neutral condition)
b. The CRATE contains the missing book. (prominent condition)

In the neutral condition, the critical sentence occurred by itself and with neutral punctuation. In the prominent condition, the head noun of the subject-noun phrase (the word *crate* in 7b) appeared in all capital letters, and subjects were told to place heavy emphasis on any such word. To induce contrastive prominence, sentences were preceded by a biasing question. In 7b, the appropriate question was "Why do you keep looking in those boxes?" As a result of these two features of the materials, subjects were expected to place contrastive prominence on *crate* and consequently to produce 7b as two intonational phrases.[6]

Ten sets of materials such as the one shown in Example 7 were created. All materials were constructed so that the target word (i.e., *crate*) began and ended with a stop consonant, to facilitate acoustic analysis. The word after the target word also began with a stop, to allow accurate measurements of the pause after the target word.

Subjects saw a total of 20 items. Ten of the items were single sentences and constituted the neutral condition, and 10 items were the same sentences paired with an appropriately biasing question and constituted the prominent condition. The 20 items were typed onto 5×8 in. (12.7×20.3 cm) index cards, 1 item per card. In the neutral condition, the sentence appeared by itself centered on the card. In the prominent condition, the sentence appeared below the biasing question, and both were centered on the card.

Procedure. Subjects were run individually. Before a subject arrived, the cards were shuffled to produce a pseudorandom order of the 20

[6] It may not be possible, given the way the conditions in this experiment were set up, to determine whether it is the preceding question, the capitalization of the critical word, or both that differentiate the neutral and prominent conditions. However, this distinction is not important given the purpose of this study, which is to show that these two versions will be produced with different prosodic structures.

cards, constrained so that the same item would not appear consecutively in its two conditions. Subjects were asked to produce each sentence as follows: First, they were to read the sentence over to themselves until they felt confident that they could say the sentence flawlessly. Second, subjects were to read the biasing question (if it occurred) and the target sentence in a natural fashion and at a normal speech rate, and their productions were tape-recorded. If the subject made an error producing the sentence, he or she simply tried again. No subject required more than three attempts for any sentence. The subject then turned over the card and went on to the next one. (This procedure is essentially the same as that described in Cooper & Paccia-Cooper, 1980.) The experimental session took about 15 min.

Data analysis. Acoustic analyses were performed as described in Cooper and Paccia-Cooper (1980). The tape-recorded sentences were digitized at a 10 kHz rate and then analyzed using a waveform editor. Word duration was measured from the onset of visible activity associated with a particular word to its offset. Pauses were measured between the critical word and the final word and could take any value greater than 0 ms. Measurements were taken independently by two assistants. Accuracy to within 5 ms was achieved on 93% of the measurements, and the remaining cases were resolved by discussion and remeasurement. Analyses of variance (ANOVAs) were performed with both subjects (F_1) and items (F_2) as random effects.

Results and Discussion

The results of this experiment were straightforward: The duration of the target word was longer when it received contrastive prominence than when it did not (376 ms vs. 345 ms), $F_1(1, 10) = 37.03$, $MS_e = 262$, $p < .001$; $F_2(1, 9) = 8.77$, $MS_e = 1{,}060$, $p < .05$. The pause after the target word was also longer in the prominent condition (106 ms vs. 26 ms), $F_1(1, 10) = 20.89$, $MS_e = 1{,}676$, $p < .005$; $F_2(1, 9) = 18.51$, $MS_e = 1{,}694$, $p < .005$. These results demonstrate that it is not necessary to vary the syntactic structure of a sentence to obtain differences in word and pause durations. Variations in prosodic structure alone are sufficient to affect these temporal variables. Thus, it appears that a level of representation containing information about the prosodic structure of an utterance, the existence of contrastive prominence and the sentence's division into prosodic constituents such as intonational phrases, must be postulated to account for timing variations.

Experiment 2

The second experiment was conducted to make the complementary point: Variations in syntactic structure alone, without corresponding changes in prosodic structure, are not sufficient to affect word duration. To make this point, the syntactic characteristics of a sentence must be varied without changing its prosodic form. Consider the sentences in Example 8:

Example 8

a. The *cop* who's a friend infuriated the boyfriend of the girls. (control condition)
b. The friendliest *cop* infuriated the boyfriend of the girls. (adjective condition)
c. The friend of the *cop* infuriated the boyfriend of the girls. (prepositional-phrase condition)
d. The man who's a *cop* infuriated the boyfriend of the girls. (relative-clause condition)

The critical word in each sentence is the word *cop,* which is followed by differing numbers of syntactic right brackets across the four conditions. The labeled bracketings for the sentential subject of each sentence are shown in Example 9 (NP = noun phrase; VP = verb phrase; PP = prepositional phrase; AdjP = adjective phrase; S = sentence; $\bar{S}$ = S-bar):

Example 9

a. [The cop [who's [[[a friend]NP]VP]S]$\bar{S}$]NP
b. [The [friendliest]AdjP cop]NP
c. [The friend [of [the cop]NP]PP]NP
d. [The man [who's [[[a cop]NP]VP]S]$\bar{S}$]NP

The word *cop* is followed by zero, one, three, and five right brackets in the control, adjective, prepositional-phrase, and relative-clause conditions, respectively. The syntactic model thus predicts the following ordering of conditions: control condition < adjective condition < prepositional-phrase condition < relative-clause condition.

The following are the corresponding prosodic structures for these conditions (PWd = prosodic word; PPh = phonological phrase; IPh = intonational phrase):

Example 10

a. (((The cop)PWd (who's a friend)PWd)PPh)IPh
b. (((The friendliest)PWd (cop)PWd)PPh)IPh
c. (((The friend)PWd (of the cop)PWd)PPh)IPh
d. (((The man)PWd (who's a cop)PWd)PPh)IPh

The prosodic word and phonological-phrase boundaries shown in Example 10 are straightforward, because these units are generated algorithmically, at least on Selkirk's (1986) theory. The intonational-phrase boundary at the end of the subject, however, is optional, because of the variability inherent in intonational phrasing. However, if it is present in one condition, it is likely to be present in all of them, because none of the factors listed earlier in the introduction to Experiment 1 varies across these conditions. The results show that this assumption is indeed correct.

In the control condition (10a), *cop* is followed by only one prosodic boundary. In the adjective, prepositional-phrase, and relative-clause conditions (10b, 10c, and 10d), the word *cop* is followed by three prosodic constituent boundaries. If it is assumed, as in Experiment 1, that the more prosodic constituent boundaries, the greater the amount of word lengthening, thus, the following pattern of results is predicted: control condition < adjective condition = prepositional-phrase condition = relative-clause condition. Another potential concern about these prosodic structures relevant to the pattern predicted is whether there might be an intonational-phrase boundary following the head noun of the subject in 10a, 10c, and 10d, corresponding to a nonrestrictive reading of the following modifier. This factor might appear to be confounded among the conditions, because such a reading is virtually impossible for 10b. However, even if it is assumed that some speakers will adopt nonrestrictive readings of the modifiers in 10a, 10c, and 10d but not in 10b, this assumption will not affect the predictions for the duration of *cop* in each of the noncontrol conditions (10b–10d), as shown in the following list of structures:

a. ((((The cop)PWd)IPh (who's a friend)PWd)PPh)IPh
b. ((((The friendliest)PWd (cop)PWd)PPh)IPh
c. ((((The friend)PWd)IPh (of the cop)PWd)PPh)IPh
d. (((The man)PWd)IPh (who's a cop)PWd)PPh)IPh

The predictions for the duration of *cop* in 10b through 10d are not affected because the presence of an intonational-phrase boundary at an early point in a sentence does not affect a later portion of the sentence. (This is one of the consequences of the assumption that prosodic structures are not recursive.) Of course, the word *cop* in 10a is affected, because I have now added an extra prosodic boundary after it. Even still, more prosodic boundaries follow the word *cop* in 10b through 10d than in 10a, and so the pattern of predictions given earlier still holds; that is, the duration of *cop* should be equal in 10b through 10d and longer than in 10a.

Method

Subjects. Eleven subjects from the same pool as in Experiment 1 participated in this experiment. None of the subjects had participated in Experiment 1.

Materials. Ten items such as in Example 8 were constructed. For each item, a monosyllabic word beginning and ending with a stop consonant was designated to be the critical word. The word occurred in the four contexts shown in Example 8. The three noncontrol conditions (adjective, prepositional phrase, and relative clause) were constructed so as to differ only in syntactic structure and not in the number of syllables, the stress pattern, or the number of prosodic words.[7] The control condition was constructed by taking the relative-clause condition and reversing the position of the two nouns in the subject-noun phrase.

Each subject produced every version of every item and thus produced 40 sentences in total. The 40 sentences were typed on index cards and presented in a different pseudorandom order to each subject, constrained so that two versions of the same sentence did not appear consecutively.

Procedure and analysis. The same experimental procedure was used as in Experiment 1, as well as the same methods of waveform analysis for word and pause durations. The experiment was analyzed as a one-way ANOVA with four levels (adjective, prepositional-phrase, relative-clause, and control conditions).

Results and Discussion

Mean word durations of the critical word in the four conditions are shown in Table 1. An ANOVA revealed a significant effect of condition, $F_1(3, 10) = 14.21$, $MS_e = 283$, $p < .001$; $F_2(3, 9) = 8.21$, $MS_e = 441$, $p < .001$. To analyze this pattern, the following planned comparisons were made: between the adjective and prepositional-phrase conditions, between the prepositional phrase and relative-clause conditions, and between the control condition and the shortest of the noncontrol conditions.

The adjective and prepositional-phrase conditions differed significantly, $F_1(1, 10) = 7.76$, $MS_e = 253$, $p < .025$; $F_2(1, 9) = 8.77$, $MS_e = 201$, $p < .025$. Notice that the difference goes in the wrong direction from that predicted by the syntactic hypothesis: Greater lengthening was associated with the version having fewer syntactic right brackets. The prepositional-phrase and relative-clause conditions did not differ significantly, $F_1(1, 10) = 2.59$, $MS_e = 165$, $p > .10$; $F_2(1, 9) = 1.12$, $MS_e = 3{,}637$, $p > .30$. Of the three noncontrol conditions, the shortest duration was found for the prepositional-phrase condition. This condition was significantly longer than the control condition, $F_1(1, 10) = 11.04$, $MS_e = 332$, $p < .01$; $F_2(1, 9) = 10.18$, $MS_e = 321$, $p < .025$.

Table 1
Word and Pause Durations (in ms) in Experiment 2

	Condition			
Duration	Adjective	Prepositional phrase	Relative clause	Control
Word	335	316	325	291
Pause	55	55	57	33

Pause durations are also shown in Table 1. There was a significant effect of condition, $F_1(1, 10) = 3.66$, $MS_e = 413$, $p < .05$; $F_2(1, 9) = 3.79$, $MS_e = 333$, $p < .05$. However, only the control condition differed from the other three ($p < .01$).

The results from this experiment indicate that variations in syntactic structure alone do not affect word and pause duration. Comparing the adjective, prepositional-phrase, and relative-clause conditions, it is clear that an increase in the number of syntactic right brackets was not correlated with an increase in either word or pause duration. In an earlier article (Ferreira, 1991), I examined a similar set of conditions and found the same result: Variations in syntactic structure alone did not yield differential amounts of word lengthening. That study used a different experimental task and a slightly different design: subjects read and produced the sentences from memory in response to a visual cue, and no subject saw an item in more than one of its experimental conditions. Yet, even with these differences, the same general results were obtained as in the present experiment.[8]

The two experiments described in this section support the prosodic view of word lengthening and pausing. The first experiment demonstrated that variations in syntactic structure were not necessary to affect word and pause duration. The second experiment demonstrated that variations in syntactic structure alone were not sufficient to affect word and pause duration. These results indicate that word lengthening and pausing are not a direct effect of syntactic structure, but instead they reflect the influence of prosodic structure. The next step is to specify how that prosodic level of representation is created during sentence production.

[7] Word length was not equated across the three noncontrol conditions: The adjective condition consisted of three or five words, and the remaining conditions generally consisted of five words. It was impossible to control both prosodic and standard word length given the constraints of the language. Because the purpose of these three conditions was to examine the effect of varying syntactic structure while holding prosodic structure constant, it seemed more appropriate to control prosodic word length.

[8] One discrepancy was that in Ferreira (1991), the adjective condition was not significantly longer than the prepositional-phrase or relative-clause conditions.

Experiments Examining the Creation of Prosodic Structures

I begin by addressing the issue of when a prosodic structure is created by describing Garrett's (1975, 1976, 1988) speech-error model, mainly because it has been a point of departure for most other models of production (e.g., Bock, 1987a; Dell, 1986; Lapointe, 1985). According to Garrett's model, a message-level (conceptual) representation is converted into what he terms a *functional-level* representation. The functional-level representation includes the semantic content of lexical items and the overall syntactic organization of the sentence. The next level created during production is a positional-level representation. Here, phrasal frames containing bound and free grammatical morphemes are accessed, and the phonological forms of words are retrieved and inserted into the phrasal frames. The positional-level representation is then converted into a phonetic motor program, which (among other tasks) computes the absolute durations to be assigned to segments in the utterance.

Given that the creation of intonational phrases is dependent on semantic information (Selkirk, 1984), and the creation of phonological phrases is dependent on syntactic information about phrasal boundaries, it follows that the creation of a prosodic representation requires semantic and syntactic information. Does the prosodic representation require information about the segmental content of words? At this point, there is no experimental evidence bearing on this question. However, Garrett (1980) has described a suggestive finding from the speech-error literature. Consider a typical word-exchange error such as "I left the briefcase in my cigar." The stress pattern of the sentence is not disrupted by the error: The heavy stress that normally occurs at the end of a sentence remains at that position, rather than moving with the word *briefcase.* Word lengthening tends to behave the same way: The de facto final word of the sentence is lengthened, not the word that was intended for sentence-final position. Finally, notice that although sentence stress does not move with the exchanged words, lexical stress does; that is, *cigar* and *briefcase* have their correct lexical-stress pattern (see also Cutler & Isard, 1980). These observations suggest that a sentence is assigned a stress pattern first, and then phonological word forms (including information about segmental content and lexical-stress pattern) are inserted into their (usually) appropriate phrasal positions.

One way to investigate the question of whether prosodic representations can be constructed independently of segmental information is to vary both the prosodic structure of a sentence and the segmental content of a critical word within it and then measure the amount of word lengthening and pausing. Consider, for example, the following sentences:

Example 11

a. The table that I thought was *black* tempted me.
b. The table that I thought was *green* tempted me.
c. The *black* table tempted me.
d. The *green* table tempted me.

The prosodic structures surrounding the critical word *black* or *green* differ in 11a and 11b versus 11c and 11d, as shown in Example 12.

Example 12

a. ((The table)PWd (that I thought)PWd (was *black*)PWd) PPh)IPh?
b. ((The table)PWd (that I thought)PWd (was *green*)PWd) PPh)IPh?
c. (The *black*)PWd (table)PWd
d. (The *green*)PWd (table)PWd

In 12a and 12b, the critical word occurs at the end of a prosodic word, phonological phrase, and possibly also an intonational phrase; in 12c and 12d, the critical word occurs only at the end of a prosodic word. Therefore, its duration and the duration of any following pause should be longer in 12a and 12b than in 12c and 12d. (Of course, the syntactic positions of the words differ as well, but because Experiments 1 and 2 indicate that prosodic, not syntactic, structure is responsible for word lengthening and pausing, I refer only to the characteristics of these sentence's prosodic structures.)

The segmental characteristics of the critical word differ also. In 12a and 12c, the critical word *black* has a short intrinsic duration, and in 12b and 12d, the critical word *green* has a longer intrinsic duration. The operational definition of intrinsic length used in these experiments was short versus long vowel for the short and long intrinsic-duration conditions, respectively (Peterson & Lehiste, 1960). The consonants varied as well; for example, the final segment of *black* is a voiceless stop, and the final segment of *green* is a nasal stop. In general, I attempted to use consonants higher on the sonority hierarchy (Hooper, 1976) for the long intrinsic-length condition and consonants lower on the hierarchy for the short intrinsic-length condition. However, I could not impose this restriction for all sets of items and still use real English words as stimuli. As the results show, the words I have designated a priori as belonging to the two conditions do indeed differ in intrinsic length.

According to the view that segmental and autosegmental phonological processes are independent at an early stage of sentence production, a particular abstract timing interval is assigned to a position in a sentence on the basis of prosodic structure, without benefit of segmental information about the word eventually to occupy that position. I use the term *abstract* for two reasons: First, by hypothesis, the timing interval may be assigned without information about the segmental content of the word occupying that interval. Second, durations must be specified relatively, because speech rate affects absolute timing. The timing interval will then be filled by the lexical item occupying that sentential position. For example, a position at the end of a phonological phrase would be allocated a larger timing interval than one in the middle of a phonological phrase, because of the effect of prosodic constituency on duration. Therefore, at the phonetic level of processing, a word such as *black* with a short intrinsic duration would be followed by a long pause to fill the entire timing interval; a word such as *green* with a longer intrinsic duration would be followed by a much shorter pause. In both cases, word and pause durations adjust themselves at the phonetic level to fit the timing interval created for the position by the prosodic structure and to adapt to the constraints imposed by the occupying word's segmental content.

The pattern predicted for the two types of words in phrase-medial position is shown in Panel A of Figure 2. A certain total

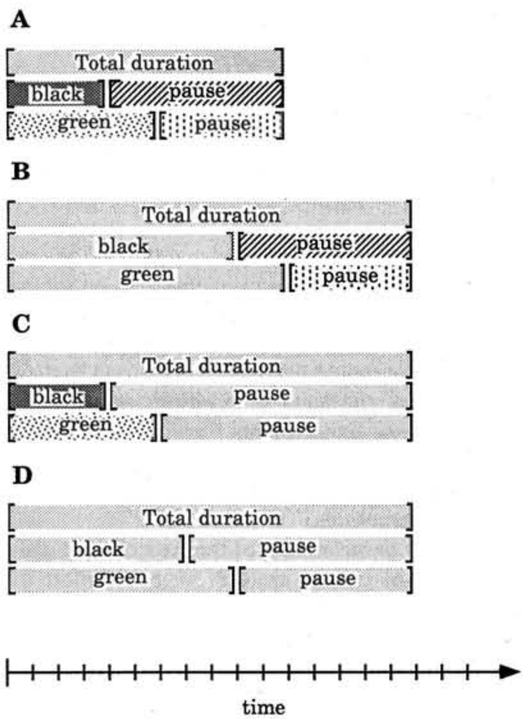

Figure 2. Predicted patterns for word and pause durations in Experiment 3. (Panel A represents the phrase-medial condition; Panels B, C, and D represent the phrase-final condition. In Panel B, only word durations increase; in Panel C, only pause durations increase; and in Panel D, both increase.)

duration for that position is specified. With an intrinsically short critical word (*black*), a relatively long pause will occur; with an intrinsically long critical word (*green*), a shorter pause is predicted. In both cases, the size of the timing interval remains constant.

As shown in Figure 2, the phrase-final position would be associated with a longer total duration, because of the effect of prosodic constituency on the size of the timing interval. The larger total duration could be accommodated in three different ways, as illustrated in Figure 2. One would be to increase only word duration (Panel B); another would be to increase only pause duration (Panel C); and another possibility would be to increase both word and pause duration roughly proportionally (Panel D). The first possibility is a priori implausible, because of the reported correlation between word lengthening and pausing (Cooper & Paccia-Cooper, 1980). The second possibility is also implausible for the same reason, but was proposed by Selkirk (1984) as the stretchability hypothesis. According to this hypothesis, a pause never occurs until a word has reached its limit of stretchability, the point beyond which the word, for phonetic reasons, can be lengthened no further. A pause will occur when the limit of stretchability is reached but time still remains in the timing interval. If a pause should occur in the phrase-medial position (as shown in Panel A of Figure 2), then the preceding word must have reached the limit of its stretchability. Therefore, in the phrase-final position, word lengthening cannot be used to expand the timing interval; only pausing can occur. This hypothesis thus predicts an interaction between interval size and intrinsic word duration—the pattern shown in Panel C of Figure 2.

The third possibility, that both word and pause durations increase to accommodate the larger interval, is shown in Panel D of Figure 2. This pattern of results would indicate that both word lengthening and pausing are used in any sentential position to fill the timing interval. Furthermore, as the total duration of the timing interval becomes larger, lengthening and pausing both increase to fill the interval.

These possibilities are examined in the next two experiments. In Experiment 3, the phrasal position and segmental content of a critical word were varied, and word and pause durations were measured. In Experiment 4, the length of the timing interval assigned to a particular position in a sentence was varied by either placing or not placing contrastive prominence on that word. Experiment 1 demonstrated that contrastive prominence on a word increases both the word's duration and the duration of the following pause, possibly because of the presence of an intonational-phrase boundary after the word. Therefore, it should be possible to increase the size of the timing interval assigned to a word by making it prominent.

Experiment 3

My first goal in Experiment 3 was to determine whether prosodic structure is created during sentence production without knowledge of the sentence's segmental content. My second goal was to examine the relation between word lengthening and pausing. Both the position of a word within a prosodic structure and the word's intrinsic duration were varied, and word and pause durations were measured.

Method

Subjects. Twelve subjects from the University of Massachusetts subject pool participated in the experiment in exchange for course credit. All subjects were native speakers of American English.

Materials. Each subject was shown 56 sentences; 32 of these were experimental sentences, and the remainder were filler items. The 32 experimental sentences consisted of two different sets of 16 items each. The first set consisted of items such as the one shown in Example 11, where the position of an adjective was varied so that it occurred either in the middle or at the end of a phrase. The intrinsic length of the adjective was varied as well. Words with short vowels made up the short intrinsic-length condition, and words with long vowels made up the long intrinsic-length condition. The second set consisted of items such as the following:

Example 13

a. The chauffeur thought he could *stop* the car. (phrase medial, short intrinsic length)
b. The chauffeur thought he could *drive* the car. (phrase medial, long intrinsic length)
c. Even though the chauffeur thought he could *stop*, the passengers were worried. (phrase final, short intrinsic length)
d. Even though the chauffeur thought he could *drive*, the passengers were worried. (phrase final, long intrinsic length)

The phrase-medial and phrase-final conditions were created by varying whether the verb (*stop* or *drive* in Example 13) was used with or without a direct object, respectively.

Each sentence occurred in one of the four experimental conditions. Each subject saw a single sentence in only one of its versions but saw all conditions of the experiment. The experimental and filler sentences were presented in a different random order for each subject.

Procedure. On each trial, the following events took place. First, a sentence appeared on a computer screen positioned in front of the subject. Subjects were told to learn the sentence so they could produce it from memory. Once the subjects were confident they could say the sentence from memory, they pushed a button on a button panel also positioned in front of them. This event caused the screen to go blank, followed by presentation of the question, "What happened?" This question was the subject's cue to begin to say the sentence. The onset of the subjects' vocalization triggered a voice-activated relay, which in turn turned on a tape recorder. The experimenter listened to the subject's rendition of the sentence. If an error occurred, the experimenter hit a key on the keyboard, which turned off the tape recorder and caused the same trial to repeat from the beginning. If an error did not occur, the experimenter hit a different key, which signaled the beginning of the next trial and turned off the tape recorder. The experimental session lasted approximately $\frac{1}{2}$ hour and was preceded by six practice trials.

Sentences were digitized and waveform measurements were taken as in the previous experiments, except that only a single set of measures was taken by one laboratory assistant.

Results

Results for word, pause, and total duration (word + pause duration) are shown in Figure 3. Results were analyzed as a $2 \times 2 \times 2$ factorial (critical-word adjective vs. verb; phrase-medial vs. phrase-final position; and short vs. long intrinsic length). There was no main effect of the adjective–verb factor (both Fs < 1), nor did it interact with any other variable. Therefore, this variable is not discussed further.

Words had a longer duration in the phrase-final condition than in the phrase-medial condition (285 ms vs. 239 ms), $F_1(1, 11) = 20.07$, $MS_e = 2{,}477$, $p < .005$; $F_2(1, 30) = 85.36$, $MS_e = 840$, $p < .001$. Words in the long-intrinsic-duration condition were longer than words in the short-intrinsic-duration condition (278 ms vs. 246 ms), $F_1(1, 11) = 40.68$, $MS_e = 585$, $p < .001$; $F_2(1, 30) = 7.09$, $MS_e = 4{,}813$, $p < .01$. As is apparent in Figure 3, these two variables did not interact (both Fs < 1).

Like word durations, pause durations were longer in the phrase-final condition than in the phrase-medial condition (67 ms vs. 35 ms), $F_1(1, 11) = 11.87$, $MS_e = 2{,}050$, $p < .01$; $F_2(1, 30) = 19.18$, $MS_e = 1{,}245$, $p < .01$. In contrast with word durations, pause durations were shorter in the long-intrinsic-duration condition than in the short-intrinsic-duration condition (34 ms vs. 68 ms), $F_1(1, 11) = 29.12$, $MS_e = 960$, $p < .01$; $F_2(1, 30) = 25.06$, $MS_e = 1{,}376$, $p < .01$. Again, the two variables did not interact, $F_1(1, 11) = 1.50$, *ns;* $F_2(1, 30) = 1.05$, *ns.* The final panel of Figure 3 reveals that when word and pause durations are added, the sum is approximately equal in the two intrinsic-length conditions (Fs < 1).

Discussion

The results of this experiment provide support for the hypothesis that a prosodic structure is created without knowledge of the segmental forms of words occupying the structure. First, the results show that, as expected, words in phrase-final position had a longer duration than those in phrase-medial position. Because the first two experiments demonstrated that it is prosodic and not syntactic structure that affects word lengthening and pausing, this effect can be attributed to the position occupied by a word in a prosodic structure. Second, words in the short-intrinsic-length condition had a shorter duration than words in the long-intrinsic-length condition. This result simply indicates that the operational definition chosen for this variable, short versus long vowel, was appropriate.

Most important, the results show that when a word's duration in a particular position is relatively short, the corresponding pause must be long; when a word's duration is long, the corresponding pause must be short. As predicted in Panels B, C, and D of Figure 2, when word and pause durations are added, they

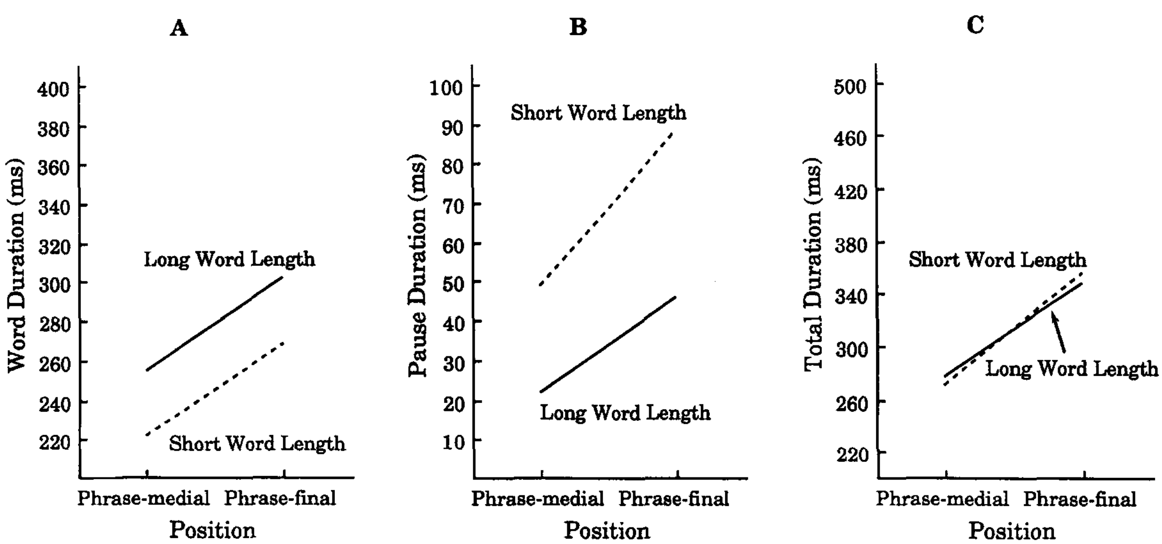

Figure 3. Results for word, pause, and total duration in Experiment 3.

equal the entire size of the timing interval specified for the word in that position, regardless of intrinsic word length. This result indicates that time is assigned to a position in a sentence without knowledge of the segmental content of the word to occupy that position; an interval of some size is created, and the next stage in sentence production must sort out how to fill the interval given the segmental content of the word in that position.

Furthermore, the pattern of results corresponds to Panel D of Figure 2: Both word and pause durations increase to accommodate the larger interval allocated to the phrase-final position. This result is contrary to the stretchability hypothesis proposed by Selkirk (1984). According to Selkirk's hypothesis, a pause occurs only once a word reaches its limit of stretchability. However, Figure 3 shows that although a pause occurred after the critical word in the phrase-medial condition, the word was still lengthened further in the phrase-final condition. It appears that at any position within a sentence, both word and pause durations are capable of filling the timing interval associated with that position.

The relation, then, between lengthening and pausing is complex. To understand the relation, it is important to distinguish what can happen within a particular timing interval versus across intervals. Within an interval, word lengthening and pausing are inversely related: The more a word is lengthened, the shorter the pause must be. This relation must hold to prevent more time from being allocated to a word than is specified in the abstract timing interval. Across intervals (e.g., from the phrase-medial to the phrase-final position), word lengthening and pausing are positively related: To the extent that a word's duration increases across intervals, any pause after the word will tend to increase as well.

Experiment 4

To strengthen the conclusions drawn from Experiment 3, I conducted a final experiment. Both the size of the timing interval and the intrinsic duration of the word occupying that interval were varied, and word and pause durations were measured. The last word of a sentential subject was spoken either with contrastive prominence or in a neutral manner. Recall that contrastive prominence on a word increases the likelihood of an intonational-phrase boundary after it, and so a word receiving prominence should on average be assigned a larger abstract duration than one lacking such prominence. If, in the final experiment, contrastive prominence works like phrasal position (i.e., both create a prosodic constituent boundary), the conclusion would be strengthened that the factor determining a word's duration is simply the number of prosodic constituent boundaries following that word, regardless of the source of those boundaries (i.e., whether the boundary was created through syntactic structure or through a phonological process such as pitch accent). Experiment 4 also provided an opportunity to replicate the main findings of Experiment 3.

Method

Subjects. Ten subjects from the University of Alberta subject pool participated in the experiment in exchange for a $6 payment.

Materials. Each item was varied along two dimensions: A word could receive contrastive prominence (the prominent condition) or not (the neutral condition) and had either a short or long intrinsic duration (defined as in the previous experiment). The resulting four conditions are shown here:

Example 14

a. The cat/CAT crossed the busy street. (neutral/prominent, short intrinsic length)
b. The mouse/MOUSE crossed the busy street. (neutral/prominent, long intrinsic length)

As in Experiment 1, contrastive prominence was induced by typing the head noun of the subject all in uppercase letters and by preceding the target sentence with a question such as "Why are you worried about the mouse/cat?"

Ten items like Example 14 were created, for a total of 40 items. Each subject saw all 40 items, and therefore, each subject saw each item in all its conditions. No filler items were used.

Procedure. The experimental procedure used was the same as in Experiments 1 and 2.

Results

Results for word, pause, and total duration (word + pause duration) are shown in Figure 4. Words had a longer duration in the prominent condition than in the neutral condition (389 ms vs. 248 ms), $F_1(1, 9) = 29.09$, $MS_e = 6{,}823$, $p < .001$; $F_2(1, 9) = 661.61$, $MS_e = 294$, $p < .001$. Words in the long-intrinsic-duration condition were longer than in the short-intrinsic-duration condition (338 ms vs. 299 ms), $F_1(1, 9) = 21.47$, $MS_e = 682$, $p < .005$; $F_2(1, 9) = 19.54$, $MS_e = 754$, $p < .005$. As is clear from the figure, these two factors did not interact, $F_1(1, 9) = 1.93$, $MS_e = 253$, *ns;* $F_2 < 1$.

Pause durations were longer in the prominent than in the neutral condition (76 ms vs. 22 ms), $F_1(1, 9) = 11.29$, $MS_e = 2{,}275$, $p < .01$; $F_2(1, 9) = 175.38$, $MS_e = 146$, $p < .001$. In contrast with word durations, pause durations were shorter in the long-intrinsic-duration condition than in the short-intrinsic-duration condition (34 ms vs. 63 ms). This effect was significant by subjects, $F_1(1, 9) = 13.72$, $MS_e = 98$, $p < .005$, but marginal by items, $F_2(1, 9) = 3.05$, $MS_e = 368$, $p < .15$. There was no interaction between the two variables, $F_1(1, 9) = 3.39$, $MS_e = 32$, *ns;* $F_2 < 1$.

As can be seen in the final panel of Figure 4, when word and pause durations are added, the effect of intrinsic duration is virtually eliminated (Fs < 1), but the effect of prominence remains. This pattern replicates Experiment 3.

Discussion

Experiment 4 reinforces the conclusions drawn from Experiment 3. First, the size of the timing interval associated with a particular word in an utterance is established independently of that word's segmental content. Second, word and pause durations both increase as the size of the timing interval increases, contrary to the predictions of Selkirk's (1984) stretchability hypothesis. In addition, this experiment adds weight to the conclusions drawn from Experiments 1 and 2. It appears that the size of a timing interval is determined by the number of prosodic constituents that terminate at a particular point in an

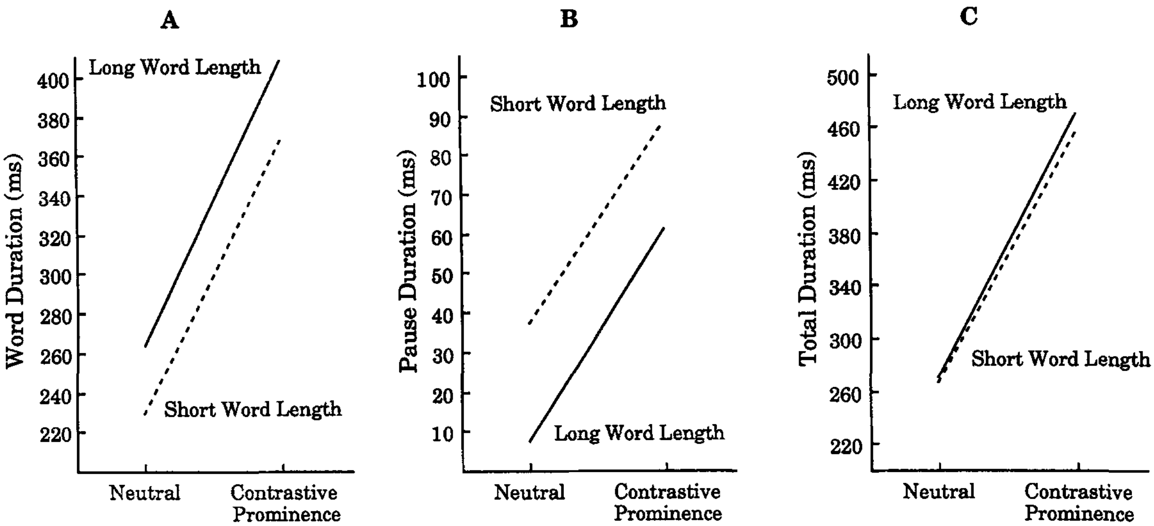

Figure 4. Results for word, pause, and total duration in Experiment 4.

utterance; the greater the number of right boundaries, the longer the timing interval. These prosodic constituents may be created in two ways. One method is mediated by syntactic structure: The right boundary of a syntactic constituent is used to mark the right boundary of a prosodic constituent, as outlined in the introductory paragraphs. The second method is through specifically prosodic processes such as contrastive stress, which increase the likelihood of a following intonational-phrase boundary, as in Experiments 1 and 4. Constituent boundaries appear to have the same effect on the size of a timing interval regardless of the method by which they were created.

General Discussion

The experiments described here demonstrate that syntax is not the only type of hierarchical structure generated during sentence production. I showed in the first two experiments that syntactic structure by itself is neither necessary nor sufficient to explain the duration of a word and its following pause. Word and pause durations could be explained more successfully from a hierarchical prosodic representation of the type described in the introductory paragraphs. I then argued that prosodic constituent structure is used to assign abstract timing intervals to positions in a sentence. The third and fourth experiments demonstrated that the size of an interval is determined without any knowledge of the segmental content of the word that will ultimately occupy it. Eventually, of course, segmental factors must be taken into account in producing a sentence. At the phonetic level of production, a word with a relatively long intrinsic length will take up more space in the timing interval, but that greater length also necessitates a shorter pause. Thus, word and pause durations trade off so as to respect the total duration of the interval allocated to a word.

In this section, I give an overview of the creation of prosody during sentence production. The description will consist of two parts. In the first part, I describe and elaborate on Levelt's (1989) model of prosodic encoding and show how it accounts for the creation of a prosodic constituent structure, including the assignment of timing intervals to words, and the insertion of word phonological forms, including their segmental and suprasegmental content. I also point out how the model accounts for the results of the experiments reported here. In the second part, I provide a mathematical model of how intrinsic word duration, word lengthening, and pausing are used at any given interval to fill the total duration assigned to that interval. This section serves the purpose of integrating some of what is known about the production of prosody with what is known about sentence production generally, and it delineates some directions for further research.

Description and Elaboration of Levelt's (1989) Model of Prosodic Encoding

This description of Levelt's model of language production will necessarily be incomplete; the interested reader is referred to Levelt (1989) for further details. The model consists of three main components: (a) a *conceptualizer,* whose task is to take a communicative intention and translate it into a semantic representation that can be used by the linguistic system (a *message,* a term also used by Garrett, 1975, 1976); (b) a *formulator,* which takes the message as input and constructs a grammatical and phonological representation and ultimately a phonetic plan; and (c) an *articulator,* which translates the phonetic plan into commands to the musculature of the vocal apparatus. These components are assumed to operate incrementally: One component can operate on the incomplete output of another, so that all components operate in parallel during the construction of a spoken sentence. I focus mainly on the formulator, because that is the component that deals with both syntactic and prosodic planning and that accesses the phonological representations of words (including their syllabic, stress, and segmental characteristics). However, first I briefly describe the conceptualizer, because some of its activities involve information that may be relevant to the intonational pattern of a sentence.

The conceptualizer creates a preverbal semantic representation. The form this representation takes is similar to a proposition (Kintsch, 1974). A message includes information about the type of message being conveyed (e.g., an event versus a state; Jackendoff, 1983, 1987), the action, and the thematic roles of the entities involved in the action (e.g., the action of *giving* involves an agent, an object undergoing transfer, and a recipient). In addition, a message must include information about the meanings of the words used in the utterance, but information about their phonological forms is not necessary. Levelt (1989), following Kempen and Huijbers (1983), separated a lexical entry into two components: its lemma, corresponding to its meaning and syntactic characteristics, and its lexeme, corresponding to its morphological and phonological form. In addition, the preverbal semantic representation specifies unusually prominent entities in the utterance. For example, if an entity in a discourse needs to be contrasted with another (as in "Why do you keep looking in those boxes? The CRATE contains the missing book"), that entity will be marked as "+ prominent." This marking will eventually translate phonologically into a pitch accent, which in turn could influence the eventual utterance's intonational phrasing. Other kinds of information contained in a message include tense and modality, among others.

The output of the conceptualizer (which, according to the assumption of incremental production, may be a fragment of the complete utterance) is then input to the formulator, which translates the fragment into a linguistic representation. First, the *grammatical encoder* uses the syntactic-category information contained in a lemma to access an appropriate syntactic procedure, one in which the lemma can act as head of a phrase. For example, the lemma for *give* will include the information that the word is a verb, which will in turn call the phrase-structure rule stating that verbs are heads of verb phrases. This procedure is followed for words that can serve as heads of phrases—verbs, nouns, prepositions, adjectives, and adverbs. The lemmas corresponding to function words (e.g., determiners and complementizers) are accessed during this syntactic-encoding stage. For example, the lemma for the noun *cat* would access the syntactic procedure for a noun phrase, and the noun-phrase procedure would in turn access the lemma of an appropriate determiner. The output of the grammatical encoder is a syntactic structure for a sentence. The syntactic structure is input incrementally to the next stage, prosodic encoding.

Prosodic encoding involves a number of processes, some operating over the domain of the word and others operating over the domain of the sentence. It will be important to specify the timing of these processes correctly, to account for the results of the experiments presented here. As the syntactic structure of a sentence unfolds, phonetic plans for words are generated. The lemma is used to access its corresponding morphological and metrical structure, a procedure Levelt (1989) termed *morphological-metrical spellout*. For example, for a word such as *giving*, morphological spellout would indicate the word's division into a stem (*give*) and affix (*ing*), and metrical spellout would indicate the number of syllables in the word (two) and its stress pattern (main word stress on the first syllable). This morphological and metrical information is then used to access the word's segmental content, a procedure termed *segmental spellout*, which specifies the phonemes of the word and their locations within each syllable. This division of labor between the procedures responsible for generating the morphological and metrical pattern of the word on the one hand, and its segmental content on the other hand, is reminiscent of the results obtained in Experiments 3 and 4, which indicated that metrical timing intervals were created without knowledge of segmental information.

Before discussing this point in more detail, it is necessary to discuss the procedures responsible for generating the phonetic plans for utterances (procedures executed by the *prosody generator*). These procedures generate the sentence's rhythmic and tonal pattern, with the former being more relevant to the purposes here. The generation of rhythm consists of two main steps: (a) the generation of prosodic constituents such as prosodic words, phonological phrases, and intonational phrases and (b) the creation of a metrical grid from the prosodic constituent structure. The grid will ultimately represent the stress and timing pattern of the utterance. Prosodic constituents are created as follows: Prosodic words are produced from lemmas as a result of morphological–metrical spellout, together with procedures for grouping clitics to hosts. Phonological phrases are created by scanning the syntactic structure of a sentence and incorporating all information up to and including the right bracket of a syntactic phrase into a single phonological phrase (the X-max algorithm proposed by Selkirk, 1986). (It should be noted that Levelt, 1989, described a different procedure that is based on Selkirk, 1984.) Intonational-phrase boundaries result from a speaker's decision to break at a certain point in an utterance, a decision that will be affected by factors such as the ones given by Nespor and Vogel (1987) and described earlier in the introductory paragraphs, as well as others. For example, a speaker may decide to create an intonational-phrase break to take a breath. A consequence of requiring the break will then be that an intonational-phrase boundary is placed at that location by the prosody generator. Finally, these prosodic constituents will be grouped into the single dominating node, *utterance*, completing the prosodic constituency of the utterance. The result will be a tree such as the one shown in Figure 1, Panel A, with one important difference: The terminal nodes in the tree are not simple words as shown in the figure, but rather they are the words' metrical representations.

Recall that the metrical information contained with a word includes its number of syllables and stress pattern. Linguists have proposed metrical grids to represent this information (Halle & Vergnaud, 1987; Prince, 1983; Selkirk, 1984). A grid is a graphlike structure in which the horizontal dimension represents time, and the vertical dimension represents degrees of stress. For example, a metrical grid for the word *Manitoba* is shown in Example 15:

Example 15

```
          x
x         x
x    x    x    x
Ma   ni   to   ba
```

On the horizontal dimension, each x represents a syllable or beat; on the vertical dimension, the higher the column of xs, the greater the amount of stress on the syllable. It should be noted that the letters symbolizing the word *Manitoba* represented in the grid are there for expository purposes only. It is assumed that at this point segmental spellout has not yet occurred, and the word is still represented in terms of its lemma and its morphological and metrical composition.

Of course, a spoken sentence will not simply be a series of concatenated metrical grids for individual words. Instead, the prosody generator uses the prosodic structure of the sentence and the metrical grids for each word to generate a metrical grid for the entire utterance. (At this point, my proposals differ in detail from the one offered by Levelt, 1989.) A grid is constructed for the sentence to represent its overall stress and timing pattern and to reflect the changes in its metrical pattern brought about by the sentential context. To illustrate, I use the sentence "The girls left." The three words making up the sentence would have the following individual metrical grids:

Example 16

```
                x          x
 x              x          x
 x              x          x
The           girls       left
```

The two content words, *girls* and *left,* are each monosyllabic and so contain only a single column of *x*s. To represent the main word stress, the column contains three *x*s (that is, goes up three levels). The function word *the* is also one syllable, but because it does not receive main word stress, the column corresponding to its syllable only goes up two levels (Selkirk, 1984).

A prosodic structure for this sentence is shown in Example 17 (assuming that the sentence is spoken as one intonational phrase; Utt = utterance):

Example 17

(((((The girls)PWd)PPh (left)PWd)PPh)IPh)Utt

The metrical grid for the sentence should represent its stress and timing pattern. As was discussed earlier, words at the ends of prosodic constituents tend to receive greater amounts of stress (Inkelas & Zec, 1990). To capture this regularity, I modify the grid for each word so that an extra level of stress is added for the boundary of each prosodic constituent. For example, after *the,* no prosodic constituent boundary follows; after *girls,* two follow; and after *left,* four follow. The grid for the entire sentence would at this point look like this:

Example 18

```
                           x
                           x
                x          x
                x          x
                x          x
 x              x          x
 x              x          x
The           girls       left
```

This grid now indicates that *left* is the most highly stressed word in the sentence and that *the* receives less stress than *girls.*

It was also seen in Experiments 1 and 2 that the more prosodic boundaries that end on a word, the longer its duration. Selkirk (1984) proposed that the metrical grid be used to capture syllable duration across an utterance. She suggested that an *x* be added horizontally immediately next to each syllable for every right boundary of a syntactic constituent—syntactic because her 1984 model was one in which syntactic structure directly determined word duration and pausing. Selkirk termed these added *x*s *silent demibeats.* Their effect is to lengthen the syllable with which they are associated; the more silent demibeats associated with a syllable, the longer its duration. I modify her suggestion by adding an *x* horizontally immediately to the right of each word for every right boundary of a prosodic constituent. The following is the result:

Example 19

```
                              x
                              x
                x             x
                x             x
                x             x
 x              x             x
 x              xxx           xxxxx
The           girls          left
```

This metrical grid indicates that the final word of the utterance is both the most stressed and has the longest duration.

Recall that earlier I added one *x* vertically for every prosodic constituent; here I have added one *x* horizontally for every prosodic constituent. A more compact rule to describe the modification of metrical grids for individual words because of prosodic structure is as follows: Add one *x* vertically and horizontally for each right boundary of a prosodic constituent. This rule is not only economical, but it also captures the empirical generalization that the more a word is stressed, the longer its duration tends to be (Crystal & House, 1988; Fry, 1955). A word such as *the,* which tends to receive little stress, also receives no silent demibeats; the word *left,* which would tend to be the loudest and longest of the utterance, receives a number of vertical *x*s and silent demibeats.

Modifying the grid for sentences that contain multisyllabic words is only slightly more complicated. For example, if our sentence was "The girls argued," the final grid would look like this:

Example 20

```
                             x
                             x
                x            x
                x            x
                x            x
 x              x            x  x
 x              xxx          x  xxxxx
The           girls          ar gued
```

The word *argued* has two syllables, with main stress on the first. The syllable with main stress is the one to which additional stress will be attracted because of prosodic constituency (Selkirk, 1984), and therefore, the syllable *ar* receives a column of *x*s up to level seven—three levels for main word stress and one level for the right boundary of each prosodic constituent. The silent demibeats are added to the final syllable of the word, because that is the one that tends to be lengthened (Klatt, 1975). Four silent demibeats are associated with the syllable *gued,* one for every boundary of a prosodic constituent. As can be seen, the number of syllables in a word does not affect how the grid is modified to accommodate its prosodic structure. However, a two-syllable word will generally have a longer duration than a one-syllable word, because it takes some amount of time to produce a syllable, whether lengthened (associated with silent demibeats) or not.

I have now set up a metrical grid for the complete utterance by modifying the grid for each individual word. The procedures I have described account for the results of the first two experiments. First, because the model assigns extra *x*s on the basis of the sentence's prosodic, not syntactic, constituency, it can account for the finding that word lengthening and pausing are a consequence of prosodic, not syntactic, structure. The more prosodic constituent boundaries that end on a word, the longer the timing interval for that word will be. The number of horizontally aligned *x*s assigned to a syllable represents the size of its timing interval. If a syllable receives no silent demibeats and therefore has only one *x*, it will be assigned a small timing interval; if the syllable has many silent demibeats, it will be assigned a larger timing interval.

In addition, the model can account for the findings of Experiments 3 and 4, which indicated that timing intervals are created without regard for the segmental content of the words eventually to occupy those intervals. Recall that in Levelt's (1989) model of prosodic encoding for words, morphological–metrical spellout precedes segmental spellout; that is, words' metrical structures are available independently of their segmental content. Earlier I modified each word's metrical grid in accordance with the overall prosodic structure of the utterance, and again, these procedures did not refer to any segmental information. Thus, the procedures that create the timing intervals for words are created without knowledge of the words' segmental content.

The timing interval assigned to each word in a sentence, then, is its sequence of horizontal grid marks. Segmental information is then inserted into the slot corresponding to that interval. The size of the interval is respected by adjusting word and pause durations to fill the interval. In the next section, I provide a model of how this adjustment takes place.

Relation Between Word Lengthening and Pausing

To begin, the total duration (dur) of an interval is the sum of word and pause durations:

$$I_{dur} = W_{dur} + P. \qquad (21)$$

The terms W_{dur}, P, and I_{dur} refer to word, pause, and total duration, respectively. Examination of the data from Experiments 3 and 4 suggests that Equation 21 should be revised to account for the finding that intrinsic word length affects word duration; long words have longer durations than short words. Think of intrinsic duration as being due only to a word's segmental content and number of syllables and including no lengthening. (I say more about this notion later.) I refer to this minimal (min) word duration as W_{min}. W_{dur} thus has two components: one is W_{min} and the other is the lengthened (len) component, W_{len}. That is,

$$W_{dur} = W_{min} + W_{len}, \qquad (22)$$

and Equation 21 is expanded accordingly:

$$I_{dur} = W_{min} + W_{len} + P. \qquad (23)$$

Given that W_{dur} increases as the size of the interval increases, in principle either W_{min} or W_{len} could vary with I_{dur}. By definition, however, only W_{len} and not W_{min} is dependent on the value of I_{dur} because W_{min} represents minimal duration due only to its segmental content and number of syllables. Therefore, W_{len} must be the factor that changes to accommodate the total duration specified by the prosodic structure. I stipulate that W_{len} is some constant proportion α of the total interval duration:

$$W_{len} = \alpha I_{dur}. \qquad (24)$$

Thus, Equation 23 takes the following final form:

$$I_{dur} = W_{min} + \alpha I_{dur} + P. \qquad (25)$$

I_{dur} is determined by prosodic structure; the more prosodic constituent boundaries that end on a word, the larger I_{dur} will be. W_{min} differs from word to word and is unaffected by I_{dur}. W_{len} is the amount a word is lengthened to fill out I_{dur}. Finally, P compensates for the varying length of W_{min} across words and fills any time remaining after W_{len} has exerted its effect.

The next step is to assign some values to these parameters, using the data from Experiment 3, shown in Table 2. Values for W_{dur} and I_{dur} can be obtained directly from the data (word duration and total duration, respectively). By modifying Equation 22 in accordance with Equation 25, a word-duration equation is derived that allows one to estimate W_{min} and α:

$$W_{dur} = W_{min} + \alpha I_{dur}. \qquad (26)$$

W_{min} is assumed to be equal in the short-length, small-interval condition and in the short-length, long-interval condition, because W_{min} differs only for different words. Applying Equation 26 to the short-length, small-interval condition, the following equation is derived:

$$222 = W_{min} + \alpha 274; \qquad (27)$$

that is, the word duration value of 222 ms is equal to W_{min} plus some proportion of the total interval duration. I have used the

Table 2
Predicted and Obtained Word and Pause Durations and Obtained Total Durations (in ms) in Experiment 3

	Small interval[a]		Large interval[b]	
Duration	Short word length	Long word length	Short word length	Long word length
Word				
Obtained	222	255	268	302
Predicted	221.7	254.6	267.7	301.2
Pause				
Obtained	49	22	88	46
Predicted	52.3	19.3	84.3	51.3
Total obtained	271	277	356	348

[a] The mean total obtained duration for small interval was 274. [b] The mean total obtained duration for large interval was 352.

value of 274, the average total duration in the two small-interval conditions, as representative of the total duration assigned due to interval size in the small-interval condition. Equation 26 can also be applied to the short-length, large-interval condition, yielding

$$268 = W_{min} + \alpha 352. \tag{28}$$

Equations 27 and 28 represent two equations with two unknowns. Solving for the unknowns, a value of 60.34 is obtained for W_{min} and .59 for α. These values imply that the minimal duration of the words in the short-intrinsic-length condition is approximately 60 ms and that the amount a word is lengthened is 59% of the total interval size.

The value of $\alpha = .59$ can now be used to calculate the appropriate value of W_{min} (which I label $W_{min'}$) for the long-length condition. The result takes the form in Equation 29:

$$255 = W_{min'} + .59(274), \tag{29}$$

yielding a value of 93.34 for $W_{min'}$. Finally, I use the values of $W_{min'} = 93.34$ and $\alpha = .59$ to predict word duration in the long-length, large-interval condition:

$$\begin{aligned} W_{dur} &= W_{min'} + \alpha I_{dur} \\ &= 301.2. \end{aligned} \tag{30}$$

The predicted value of 301.2 is remarkably close to the obtained value of 302 ms.

In the present model, W_{min} and $W_{min'}$ represent the duration of a word outside any prosodic context. Because it is virtually impossible to produce an utterance of any type (whether a word list, phrasal fragment, or full sentence) without prosody, minimal durations can never be produced by human speakers. However, the equations given here can be used to generate estimates of minimal spoken word durations (that is, in the smallest possible prosodic context), which will include W_{min} and the minimal amount of word lengthening possible for a given word and speech rate.

The smallest value of I_{dur} occurs when P is equal to zero. In this situation, I_{dur} equals W_{dur}, and because W_{len} is a constant proportion of I_{dur}, W_{len} is also at its minimal value. I have estimated W_{min} in the short-length condition to be 60 ms and α to be .59. Therefore, using Equation 25, Equation 31 is derived:

$$\begin{aligned} I_{dur} &= 60 + .59(I_{dur}) + 0 \\ &= 146, \end{aligned} \tag{31}$$

and because I_{dur} without a pause equals W_{dur}, W_{dur} also equals 146. The minimal spoken duration (not to be confused with W_{min}) for a word such as *black* is 146 ms. The same procedure can be used to estimate the minimal I_{dur} and W_{dur} in the long-intrinsic-length condition:

$$\begin{aligned} I_{dur} &= 93 + .59(I_{dur}) + 0 \\ &= 227; \end{aligned} \tag{32}$$

that is, the minimal spoken duration for a word such as *green* is 227 ms.

These estimates of minimal spoken word durations are close to those obtained by Allen, Hunnicutt, and Klatt (1987). Allen et al. developed a text-to-speech system, which includes a specification of the minimal durations of segments to be produced by their system. Short vowels have a minimal duration between 40 and 80 ($M = 60$) ms, long vowels between 100 and 150 ($M = 125$) ms, and stop consonants between 20 and 60 ($M = 40$) ms. Using the means, a word such as *black* would have a minimal duration of about 180 ms and *green* about 245 ms. These values are similar to the estimates of minimal spoken word duration obtained here.

Thus far, I have used the data from Experiment 3 to estimate α, W_{min}, and $W_{min'}$. The estimates for these parameters can be used to generate predictions for word duration in all four experimental conditions, and these are shown in Table 2. Pause durations can also be estimated, using Equation 33;

$$P = I_{dur} - W_{min} = \alpha I_{dur}. \tag{33}$$

The predicted values are also shown in Table 2. As can be seen from the table, the fit between predicted and obtained values is close (standard error of the difference between predicted and obtained values for word and pause duration is 1.04 ms).

Word and pause durations in Experiment 4 can also be predicted in a similar fashion. I begin by using the estimates of α, W_{min}, and $W_{min'}$, generated from Experiment 3. This procedure is conservative, because Experiment 4 used different subjects and items than Experiment 3. I again predict word durations using Equation 26. For example, the equation for the short-word, neutral condition would be

$$W_{dur} = 60 + .59(273), \tag{34}$$

where 60 and .59 represent the estimates of W_{min} and α, and 273 is the mean total duration in the neutral condition. The same procedure can be used for the remaining three conditions. The resulting predictions for W_{dur} are given in Table 3, labeled *Predicted* ($\alpha = .59$).

Pause durations were also predicted, using Equation 33. For example, the equation for the short-word, neutral condition would take the following form:

$$P = 273 - 60 - .59(273), \tag{35}$$

and $P = 51.9$. The same procedure was followed for the remaining three conditions. The predicted values for P, Predicted ($\alpha = .59$), are given in Table 3.

The fit for the data in Experiment 4 is not as close as in Experiment 3 ($SE = 9.78$ ms). This result is not surprising, given that the same parameter values for α, W_{min}, and $W_{min'}$ were used. W_{min} and $W_{min'}$ could be expected to differ from Experiment 3 to Experiment 4 because the two experiments used entirely different sets of items (that is, different critical words). However, examination of the predicted values, Predicted ($\alpha = .59$), for Experiment 4 in Table 3 indicates that the predictions are worse in the two prominent conditions compared with the two neutral conditions; word durations are underestimated, and pause durations are overestimated. This pattern suggests that α should be adjusted for the two prominent conditions, on the assumption that placing contrastive prominence on a word not only alters its prosodic context but it also affects the trade-off between word lengthening and pausing. (Recall that increased word duration is an acoustic correlate of stress; Crystal &

Table 3
Predicted and Obtained Word and Pause Durations and Obtained Total Durations (in ms) in Experiment 4

	Neutral[a]		Prominent[b]	
Duration	Short word length	Long word length	Short word length	Long word length
Word				
Obtained	232	264	366	412
Predicted (α = .59)	223.8	256.8	339.0	372.0
Predicted (α = .66)	NA	NA	366.9	399.9
Pause				
Obtained	29	21	83	68
Predicted (α = .59)	51.9	16.2	126.0	93.0
Predicted (α = .66)	NA	NA	98.1	65.1
Total obtained	261	285	449	480

Note. NA = not applicable.
[a] The mean total obtained duration for the neutral condition was 273. [b] The mean total obtained duration for the prominent condition was 465.

House, 1988; Fry, 1955). Therefore, I reestimated α only for the prominent conditions and then recalculated predicted values for word and pause durations.

To recalculate α, Equation 26 was used as follows:

$$366 = 60 + \alpha 465, \quad (36)$$

yielding a value of .66. This value of α was then used to calculate predicted word and pause durations in the two prominent conditions, and these predictions are shown in Table 3, that is, Predicted (α = .66). As can be seen, the fit is much closer, as reflected in the standard error (*SE* = 4.08).

Thus, the model I have presented here (summarized as Equation 25) provides a good description of the relation between word lengthening and pausing. A number of issues arise from the model. First, speech rate may affect the relation between lengthening and pausing, by means of minimal spoken word durations. (Recall that this concept is not the same as W_{min}; minimal spoken word durations are made up of W_{min} plus the smallest amount of word lengthening possible in a prosodic context.) If words have a minimal duration due solely to their segmental content and number of syllables, then the fastest speech rate possible will be determined by the minimal spoken word duration. Investigation of this question could also reveal whether the estimates of minimal spoken duration obtained here correspond to the minimal durations that human speakers can produce.

Another issue concerns how the value of α is affected across different prosodic contexts. For example, would α stay constant (as in Experiment 3) across a variety of phrasal and sentential positions? Furthermore, the model of prosodic encoding that I described in the previous section does not specify that prominence affects the size of the timing interval; the size of the interval is determined simply by the prosodic constituency of the sentence. However, contrastive prominence does appear to change α, that is, the relation between lengthening and pausing. Further research must investigate whether contrastive prominence does indeed have the effect of changing the value of α, as suggested in Experiment 4.

Finally, the model as it stands explicitly addresses only monosyllabic words. Obviously, a multisyllabic word will have a longer duration than a monosyllabic word in any prosodic context. The effect of number of syllables on duration is similar to the effect of segmental content: The phonological characteristics of the word affect duration above and beyond any potential lengthening. Therefore, intrinsic duration has two phonological components, one that is due to the word's segmental content and the other that is due to its own metrical structure. I have assumed that both of these components may be included in W_{min}, so that Equation 25 can apply both to mono- and multisyllabic words. Verifying whether this assumption is accurate is yet another project for future research.

Comparison to Gee and Grosjean (1983)

I have argued that sentence timing is determined by prosodic, not syntactic, structure. Gee and Grosjean (1983) made similar arguments, claiming that pause duration within a sentence could be accurately predicted from a prosodic representation. They noted that function words almost always have a short duration and are rarely followed by a pause. This simple fact cannot be accounted for on a purely syntactic model, because syntax does not distinguish between function and content words. Second, Gee and Grosjean found that sentence length affects the pause pattern of a sentence: The longer a sentence, the more likely it is to be divided into more than one intonational phrase (Nespor & Vogel, 1987) and thus to contain a significant pause. This finding is troublesome for a purely syntactic model because the syntactic structure of a sentence does not change depending on the length of its constituents. To account for these findings, Gee and Grosjean proposed an algorithm to generate a pause structure for a sentence, together with a method for computing pause duration from the pause structure.

However, the Gee and Grosjean (1983) model differs significantly from the one I have proposed here. First, their model generates highly recursive pause structures, because intonational phrases are permitted inside of intonational phrases. Furthermore, the model allows only binary branching, with the result that often the prosodic structure for a sentence is more complex than its corresponding syntactic structure. Second, their model deals only with pauses and not with word duration. If the model of the relation between lengthening and pausing that I have proposed is correct, then pausing should not be treated in this manner, because pause duration is dependent on the segmental characteristics of the preceding word. Furthermore, a model that handles both word lengthening and pausing is preferable to one that simply handles pausing.

Third, the Gee and Grosjean (1983) model calculates pause duration on the basis of the complexity of the material both preceding and following a potential pause location. Psychologically, this amounts to saying that a pause occurs both because the system is taking a rest after executing a certain amount of prosodic structure and also because the system is planning the upcoming stretch of speech. In earlier work (Ferreira, 1988), I argued that it is useful to distinguish between pauses resulting from these two sources. Pauses that are based on already-spoken material I termed *timing-based pauses,* and pauses that are based on material to be produced I termed *planning-based pauses.* In another earlier work (Ferreira, 1991), I demonstrated that planning-based pauses are affected by the complexity of the upcoming syntactic structure of a sentence and reflect the process of translating the syntax into phonological phrases using Selkirk's (1986) X-max algorithm. Timing-based pauses, in contrast, are dependent on prosodic structure, as shown in this article. A pause occurs simply to fill whatever time is left in a timing interval after intrinsic word length and word lengthening have exerted their effect. Thus, as assumed by Gee and Grosjean (1983), pause durations are affected both by preceding and upcoming material. However, contrary to their model, the same type of representation does not affect both sides of the pause site. The effect of preceding material is due to prosodic structure, and the effect of upcoming material is due to syntactic structure. (Further criticisms of the Gee and Grosjean model can be found in Fowler, 1985; Levelt, 1989; and Van Wijk, 1987).

Implications

The work presented here has implications for language comprehension and development as well as production. In the area of language comprehension, a large body of research has examined topics such as word recognition (e.g., Burgess, Tanenhaus, & Seidenberg, 1989), syntactic parsing (e.g., Holmes, Stowe, & Cupples, 1989), and text comprehension (e.g., Fletcher & Bloom, 1988). However, because visually presented materials are the norm, less is known about spoken than visual language comprehension (Ferreira & Anes, in press). One reason for this state of affairs is that auditory studies are technically demanding. Stimulus materials are difficult to prepare, and no on-line task comparable to the moving-window paradigm (Just, Carpenter, & Woolley, 1982) or eye-movement-monitoring technique (Frazier & Rayner, 1982) has been developed. However, a more theoretical reason for the relative neglect of auditory language comprehension may be that spoken sentences necessarily have a prosodic structure, and without an understanding of its characteristics, it is unclear how to control and manipulate the relevant prosodic variables in a comprehension experiment. Because this article contributes to researchers' understanding of prosodic structure, it may be useful to researchers considering working in auditory language comprehension.

The view of prosodic structure I have presented here may have implications for the study of language development as well. For example, on the basis of Cooper and Paccia-Cooper's (1980) work on timing, Morgan and his associates (Morgan, 1986; Morgan et al., 1987; Morgan, Meier, & Newport, 1989) have claimed that the timing pattern of a sentence provides reliable cues to a sentence's syntactic structure and that these cues are used by children to bootstrap them into the syntax. The research presented in this article suggests that there is some relationship between a sentence's syntactic structure and timing pattern, but it is not transparent. Syntactic constituents do not have one-to-one durational correlates, and the presence of features such as contrastive prominence can distort a sentence's prosodic structure relative to its syntactic structure. Therefore, prosodic cues are not highly diagnostic of syntactic structure.

Similarly, Hirsh-Pasek et al. (1987) have found that infants prefer to listen to speech in which pauses occur at clause boundaries compared with other randomly selected locations within a sentence. Hirsh-Pasek et al. argued that this finding shows a sensitivity to the existence of clauses and therefore to the syntactic structure of language, present as early as 7 months of age. However, clause boundaries typically also coincide with intonational-phrase boundaries (Nespor & Vogel, 1987; Selkirk, 1984), and therefore, the infants may have been sensitive to the latter rather than to the former. Furthermore, intonational-phrase boundaries are the site of pitch changes, word lengthening, and pausing. Removal of any one of these normally correlated features (e.g., pausing) could result in unnatural speech, which infants may find unappealing.

Finally, this article has implications for the study of sentence production. In the course of speaking, one not only generates a semantic intention and a syntactic structure, but one also creates a prosodic representation. I have examined only the processes of translating a fixed semantic–syntactic representation into its corresponding prosodic structure. However, just as one makes choices about which lexical items and syntactic forms will best convey one's intended message, one also makes decisions about how best to use prosody to achieve one's communicative goals. For example, a speaker must decide on a sentence's intonational phrasing and must select a word to receive semantic and phonological prominence. It is not surprising that the word at the end of a sentence normally receives this prominence, given Haviland and Clark's (1974) evidence that information held in common between conversational participants typically precedes information that is new to the discourse. However, on the basis of semantic considerations, the speaker can also decide to place prominence on a word located early in a sentence. An understanding of the circumstances under which prosodic decisions such as these are made would greatly advance researchers' understanding of the creation of prosody and the process of language production more generally.

References

Allen, J., Hunnicutt, M. S., & Klatt, D. (1987). *From text to speech: The MITalk system.* Cambridge, England: Cambridge University Press.

Anderson, J. R. (1990). *Cognitive psychology and its implications* (3rd ed.). New York: Freeman.

Anderson, S. W., & Cooper, W. E. (1986). Fundamental frequency during spontaneous picture description. *Journal of the Acoustical Society of America, 79,* 1172–1174.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30,* 644–663.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 72,* 467–482.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18,* 355–387.

Bock, J. K. (1987a). Coordinating words and syntax in speech plans. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 337–390). Hillsdale, NJ: Erlbaum.

Bock, J. K. (1987b). An effect of the accessibility of word forms on sentence structure. *Journal of Memory and Language, 26,* 119–137.

Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition, 35,* 1–40.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23,* 45–93.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech, 8,* 148–158.

Burgess, C., Tanenhaus, M. K., & Seidenberg, M. S. (1989). Context and lexical access: Implications of nonword interference for lexical ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 620–632.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language, 30,* 348–369.

Chen, M. Y. (1990). What must phonology know about syntax? In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 19–46). Chicago: University of Chicago Press.

Chomsky, N. (1957). *Syntactic structures.* The Hague, Netherlands: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: M.I.T. Press.

Chomsky, N. (1986). *Barriers.* Cambridge, MA: MIT Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Church, K. W. (1987). Phonological parsing and lexical retrieval. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 53–69). Cambridge, MA: MIT Press.

Coltheart, V., Avons, S. E., Masterson, J., & Laxon, V. J. (1991). The role of assembled phonology in reading comprehension. *Memory & Cognition, 19,* 387–400.

Cooper, W. E. (1980). Syntactic-to-phonetic coding. In B. Butterworth (Ed.), *Language production. Vol. 1: Speech and talk* (pp. 297–334). San Diego, CA: Academic Press.

Cooper, W. E., & Danly, M. (1981). Segmental and temporal aspects of utterance-final lengthening. *Phonetica, 38,* 106–115.

Cooper, W. E., & Paccia-Cooper, J. M. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

Crystal, T., & House, A. (1988). The duration of American English vowels: An overview. *Journal of Phonetics, 16,* 263–284.

Cutler, A., & Isard, S. D. (1980). The production of prosody. In B. Butterworth (Ed.), *Language production. Vol. 1: Speech and talk* (pp. 245–270). San Diego, CA: Academic Press.

Daneman, M., & Stainton, M. (1991). Phonological recoding in silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 618–632.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Espinal, M. T. (1991). The representation of disjunct constituents. *Language, 67,* 726–762.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech in infants and adults. *Developmental Psychology, 27,* 209–221.

Ferreira, F. (1988). *Planning and timing in sentence production: The syntax-to-phonology conversion.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30,* 210–233.

Ferreira, F., & Anes, M. (in press). Why study spoken language processing. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics.* San Diego, CA: Academic Press.

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language, 27,* 235–244.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar.* New York: McGraw-Hill.

Fowler, C. (1985). Current perspectives on language and speech production: A critical overview. In R. Daniloff (Ed.), *Speech science* (pp. 193–278). London: Taylor & Francis.

Frauenfelder, U. H., & Lahiri, A. (1989). Understanding words and word recognition: Can phonology help? In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 319–341). Cambridge, MA: MIT Press.

Frazier, L., Clifton, C., & Randall, J. (1983). Filling gaps: Decision principles and structure in sentence comprehension. *Cognition, 13,* 187–222.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178–210.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 35,* 765–769.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133–177). San Diego, CA: Academic Press.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–256). Amsterdam: North-Holland.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production. Vol. 1: Speech and talk* (pp. 177–220). San Diego, CA: Academic Press.

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey. Vol. III: Language: Psychological and biological aspects* (pp. 69–96). Cambridge, England: Cambridge University Press.

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

Gerken, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language, 30,* 431–451.

Gerken, L., & Bever, T. G. (1986). *Linguistic intuitions are the result of interactions between perceptual processes and linguistic universals* (Tech. Rep. No. 35). Rochester, NY: University of Rochester, Cognitive Science Program.

Halle, M., & Vergnaud, J.-R. (1987). *An essay on stress.* Cambridge, MA: MIT Press.

Harris, Z. S. (1951). *Methods in structural linguistics.* Chicago: University of Chicago Press.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512–521.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry, 15,* 33–74.

Hirsh-Pasek, H., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269–286.

Holmes, V. M., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language, 28,* 668–689.

Hooper, J. B. (1976). *An introduction to natural generative phonology.* San Diego, CA: Academic Press.

Inkelas, S., & Zec, D. (1990). *The phonology-syntax connection.* Chicago: University of Chicago Press.

Jackendoff, R. (1983). *Semantics and cognition.* Cambridge, MA: MIT Press.

Jackendoff, R. (1987). *Consciousness and the computational mind.* Cambridge, MA: MIT Press.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24,* 252–293.

Just, M. A., Carpenter, P. A., & Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 10,* 833–849.

Kaisse, E. M. (1985). *Connected speech: The interaction of syntax and phonology.* San Diego, CA: Academic Press.

Kanerva, J. M. (1990). Focusing on phonological phrases in Chichewa. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 145–162). Chicago: University of Chicago Press.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99,* 349–364.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185–209.

Kintsch, W. (1974). *The representation of meaning in memory.* Hillsdale, NJ: Erlbaum.

Klatt, H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3,* 129–140.

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

Lapointe, S. G. (1985). A theory of verb form use in the speech of agrammatic aphasics. *Brain and Language, 24,* 100–155.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

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

MacDonald, M. C. (1989). Priming effects from gaps to antecedents. *Language and Cognitive Processes, 4,* 35–56.

Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior, 9,* 75–78.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29,* 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30,* 69–89.

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42,* 181–212.

Morgan, J. L. (1986). *From simple input to complex grammar.* Cambridge, MA: MIT Press.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19,* 498–550.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language, 28,* 360–374.

Nakatani, L. H., O'Connor, K. D., & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica, 38,* 84–106.

Nespor, M., & Vogel, I. (1987). *Prosodic phonology.* Dordrecht, The Netherlands: Foris.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18,* 5–19.

Odden, D. (1990). Syntax, lexical rules and postlexical rules in Kimatuumbi. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 259–278). Chicago: University of Chicago Press.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America, 54,* 1235–1247.

Perfetti, C. A., & Zhang, S. (1991). Phonological processes in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 633–643.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32,* 693–703.

Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation.* Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707–784.

Prince, A. (1983). Relating to the grid. *Linguistic Inquiry, 14,* 19–100.

Prince, A. (1984). Phonology with tiers. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure* (pp. 234–244). Cambridge, MA: MIT Press.

Rochemont, (1986). *Focus in generative grammar.* Amsterdam: John Benjamins.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production picture-word interference studies. *Journal of Memory and Language, 29,* 86–102.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook, 3,* 371–405.

Selkirk, E. O., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 313–338). Chicago: University of Chicago Press.

Slowiaczek, M. L. (1981). *Prosodic units as language processing units.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 341–386.

Van Wijk, C. (1987). The PSY behind PHI: A psycholinguistic model for performance structures. *Journal of Psycholinguistic Research, 16,* 185–199.

Williams, E. (1978). Across-the-board rule application. *Linguistic Inquiry, 9,* 31–43.

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society, 104,* 444–466.

Received July 23, 1991
Revision received June 15, 1992
Accepted September 14, 1992 ■