SYNTACTIC CONTROL OF TIMING IN SPEECH PRODUCTION

by

WILLIAM EDWIN COOPER

A.B., Brown University

1973

A.M., Brown University

1973

SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF

PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

February, 1976

Signature of Author.....

Department of Psychology,

January 12, 1976

Certified by.....

Thesis Supervisor

Accepted by.....

Chairman, Departmental Committee

on Graduate Students

SYNTACTIC CONTROL OF TIMING IN SPEECH PRODUCTION

by

William Edwin Cooper

Submitted to the Department of Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

A sentence-reading procedure was used to study the influence of syntactic and semantic structure on the timing of syllables in speech production. In each sentence, easily segmentable key words were placed in the environment of a putative syntactic boundary. The immediate phonetic environments of the key words were held fixed or independently assessed for their effect on syllable timing.

Chapter 1 includes an introduction placing the study within a larger context of studies on complex motor skills that are segmented and hierarchically organized. The chapter contains a discussion of the study's principal motives, outcomes, and implications for theory and practical applications.

Chapters 2 through 7 include descriptions of experimental studies and their relation to issues discussed Chapter 1. The experiments of Chapter 2 showed that speakers shortened the durations of segments just prior to the locations of two syntactic deletions which erased material from the beginning of a clause, whereas speakers lengthened the durations of segments just prior to the location of a deletion erasing material from the middle of a clause. The results provided support for the notion that a syntactic level of processing controls timing and that this level computes a syntactic representation that corresponds more closely to linguistic surface structure than underlying structure.

Chapter 3 includes a study of speech timing at the clause boundaries of complement structures. Two types of complement were studied, triggered in the main experiment by the verbs expect and persuade. The results provided further support that a syntactic level of processing controls speech timing and that this level computes a clausal representation similar to linguistic surface structure. An auxiliary experiment showed that a phonetic influence on syllable timing, conditioned by the presence of a following voiced vs. voiceless segment, extends across a verb-noun phrase boundary.

Chapter 4 contains a report of experiments on the possible effects of syntactic rules of preposing on timing. The rules Passive, Topicalization, Adverb Preposing, and Prepositional Phrase Proposing were considered. The results provided no support for the possibility that morphologically null traces, as defined in current linguistic theory, are accompanied in speech production by changes in the timing of immediately surrounding words.

Chapters 5 and 6 include experiments showing that the semantic relation of coreference plays a role in the control of timing. In the experiment reported in Chapter 5, speakers lengthened the duration of a noun when it was referred to again later in the same sentence by a pronoun. This effect was observed for the very first word of an utterance and for non-adjacent coreferents that spanned a major clause boundary. A second experiment, reported in Chapter 6, added further support to the notion that coreference relations control timing and that such control is not restricted to the domain of a single clause.

Chapter 7 includes a sketch of additional experiments for which a relatively small data base was obtained. Three major issues in timing research are discussed in conclusion with reference to the present study and plans for further work.

Thesis Supervisor: Merrill F. Garrett

Title: Associate Professor of Psychology

TABLE OF CONTENTS

Acknowledgements

I wish to thank the following individuals who have been particularly helpful to me during the course of this thesis: Dumont Billings, Irene Bonner, Noam Chomsky, Ann Cutler, Thomas Felton, Jerry Fodor, Merrill Garrett, John Groopman, Richard Held, Dennis Klatt, Steven Lapointe, Cornelia Parkes, John Robert Ross, Kenneth Stevens, Hans-Lukas Teuber, Luis Tiant, and Robin Tuckerman. [see individual chapters for specific acknowledgements].

Biographical Note

The author was born on March 20, 1951 in Baltimore, Maryland. He attended public schools in Lancaster, Pennsylvania from 1956 to 1969. During the 9th grade, he received local recognition for placing three root beer Fizzies in the water bucket of a certain algebra teacher, just prior to the teacher's routine washing of the blackboard (thereby ending a promising career in mathematics). From 1969 to 1973, he attended Brown University. From 1973 to 1975, he attended graduate school at M.I.T., where he divided his TV viewing time between Star Trek and You Bet Your Life. He has published numerous articles, one of which (co-authored by Professor J. R. Ross) was discussed in a recent column of the Christian Science Monitor. His professional ambition, presently elusive, is to learn to do science with pride and without arrogance.

Note to the Reader

Some of the chapters in this thesis were written as self-contained papers, intended for journal publication. For this reason, the format differs slightly from chapter to chapter, and there exist a few instances of unnecessary repetition. All references appear at the ends of individual chapters.

Chapter 1

Introduction

The aim of this chapter is to provide an overview of the study by describing its motivation, outcome, and applications. The chapter is intended for anyone who might be curious about how people speak, as well as for some who, by their own admission, are not at all curious about this topic, but who are interested nonetheless in the more general problem of information-processing by organisms.

Section 1 includes a statement of a major assumption of the study, as well as some discussion of the possibility that this assumption represents a general property of information-processing in motor skills which exhibit two particular properties -- segmentation and hierarchical organization. Section 2 contains remarks on previous theory and research on speech production insofar as such work has guided the present venture. Section 3 includes a description of the study's principal aims, methods, and results. Finally, Section 4 includes an outline of a revised theory of speech production and the implications of the study for further research and health- and engineering-related applications. At worst, a look at this chapter should allow the reader to make a relatively informed decision about whether his interests would be served better by reading the study in detail or by turning attention instead to a nearby journal, game of squash, or any of a number of other attractions.

1. Aside from general assumptions concerning the orderliness of nature, much of this study was based on one major assumption about a relation between centralized motor commands and motor output. In its general form, the assumption can be stated as follows:

(1) Given a domain of information processed by the central nervous system with an ordered sequence of segmented contents $\{s_1, s_2, s_3 \ldots s_n\}$, the motor output controlled by $\{s_n\}$ will be longer in duration than the average of the segments controlled by the processing domain as a whole, ceteris paribus.

We may refer to this assumption as domain-final or phrase-final lengthening. The assumption is simple, intuitively appealing, and, as we shall see, can be used to pursue an understanding of processing domains in speech production, an area about which very little is currently known.

The intuitive appeal of the assumption can be pinpointed for certain types of behavior, including speech. If a listener attends to the durations of speech sounds during a conversation, he notices that syllables which occur at the ends of sentences and at the ends of certain clauses are lengthened in comparison with their typical durations. In addition, the lengthened segments are often followed by a pause. These impressions of the unaided ear have been confirmed by actual measurements of acoustic duration (Martin, 1970; Lindblom and Rapp, 1973; Klatt, 1975a; Kloker, 1975).

Before considering speech in detail, however, it should be noted that the appeal of the phrase-final lengthening assumption extends to other forms of behavior. A good example is musical composition, for which, like speech, we have some idea on independent grounds about the boundaries of domains of processing, including movements, stanzas, phrases, and the like.[1] The independent grounds are provided by the theory of musical composition as by the theory of grammar. The unaided ear tells us that musical notes are generally longer at the ends of phrases than the average duration of notes within such phrases. This lengthening effect occurs with remarkable regularity in classical and modern music of the Western world with which I am familiar. The reader may wish to test this claim by listening to the radio for a half-hour or so. Although a number of other factors may influence the durations of individual musical notes (see below), the phrase-final lengthening effect appears consistently throughout most compositions.

One of the more famous examples of the principle is taken from the score of Beethoven's 5th Symphony, consisting of a phrase of four notes -- three short followed by one long. If the ordering of the short and long notes were reversed, the phrase would sound not only unfamiliar, but, I would argue, unnatural as well. According to this view, a musical phrase beginning with a long note and ending with a short one is an unnatural output for a composer, other factors being equal.

The general principle of phrase-final lengthening appears to be ingrained in the listener as well as the composer. A simple experiment in auditory perception confirms this notion. If a listener is presented a tape loop containing two tones of the same frequency and amplitude, but of different durations, the listener hears the sequence of tones as a pair of short-long tones, not as a pair of long-short tones (Allen, 1975). It seems, then, that the auditory system naturally processes information in terms of a domain containing short-long rather than long-short sequences, paralleling the constraint on the motor output of both musical composer and speaker.[2]

Two further points should be made about the nature of phrase-final lengthening for speech and music. First, it seems that the general principle stated in (1) might be represented more precisely by taking into account different magnitudes of lengthening for processing domains of different hierarchical status. In both speech and music, there exists a general trend for the lengthening at the ends of minor phrases to be less pronounced than the lengthening at the ends of successively more inclusive domains. Thus, the end of a multi-clause sentence is generally longer than the end of an individual clause within that sentence, which is in turn longer than the end of an individual phrase within that clause, according to impressions of the unaided ear. Similarly, in a symphony, the end of a major movement is generally longer than the end of a lesser domain.

A second point concerning speech and music, already hinted at, is that the principle in (1) should hold with absolute certainty only when other factors that might influence segment durations are held equal, which they of course seldom are. Such factors may include changes in overall rate of movement, fatigue,[3] trade off between amplitude and duration,[4] as well as factors that rely on relatively complex constraints on information-processing. Unfortunately, too little is known about most of these factors to provide a flawless, independently motivated account of any cases of segment length which violate the general principle proposed above. Nevertheless, the available impressions of the listener for both speech and music are sufficiently supportive of the principle that it seems justifiable to claim that the principle is not merely intuitively appealing but valid, at least within the realms of speech and music.

Speech and music provide the best available sources of information about the general principle because in these cases there exist independent grounds for defining domains of processing, based on the theories of grammar and musical composition. Since the domains can be independently defined, any possible circularity in testing the principle can be avoided. For other forms of complex motor behavior involving segmented information, there typically exist weaker, sometimes meager, grounds for defining the bounds of processing domains. However, for the few cases where such

grounds are relatively well-established, the evidence provides support for the principle of phrase-final lengthening. We turn now to consider these cases in an attempt to determine the necessary and sufficient conditions under which the general principle applies.

The communication systems of other species provide a few tests of the general principle, and in most cases where I have been able to find a documented difference in lengthening among phrase segments, the differences are in support of the general principle.[5] In acoustic traces of the song phrases of the chaffinch (fringilla coelebs), for example, phrase-final segments are typically longer than segments at the beginning or middle of a phrase (see Figure 1). This claim is based on an examination of the acoustic traces published by Thorpe (1961) and Nottebohm (1970). Unfortunately for present purposes, however, these and other researchers of birdsong have been generally concerned with analyzing acoustic traces in terms of the frequency domain rather than in terms of duration, and to my knowledge no highly systematic

---

Insert Figure 1 about here

---

treatment of segment durations in birdsong has yet been published.

If the spectrographic traces published by Nottebohm (1970) are representative of the durational characteristics of chaffinch song, a second major point about birdsong duration is worth noting. Phrase-final

lengthening is found not only in traces of the adult song but in the earliest stage of song development as well, even for birds reared in the absence of an auditory model. If the latter condition applies to phrase-final lengthening generally, it would provide evidence that the lengthening effect is genetically pre-determined.

A better documented example of phrase-final lengthening in animal communication involves chirps emitted by the insect Amblycorypha oblongifolia, a member of the family Orthoptera (du Mortier, 1963, p. 359). This insect produces chirp phrases consisting of four segments, having durations averaging 8, 13, 21, and 38 msec, successively. The example illustrates another possible refinement of the general principle in (1), namely that gradations in length occur within a processing domain, making the lengthening effect observed in absolute phrase-final position a special case of a duration principle that applies to each segment of a phrase, other factors being equal.

A consideration of acoustic communication in other species has suggested that phrase-final lengthening is not restricted to our own species. In addition, more systematic study of animal communication may confirm the possibility, suggested by an examination of published birdsong traces, that phrase-final lengthening is a genetically pre-determined characteristic of acoustic communication.

In addition to birdsong and insect chirps, one might expect to find tests of the general lengthening principle in other types of animal communication, in particular, those that have been studied in some acoustic detail, including the sound emissions of the dolphin (Lilly, 1963, 1965), squirrel monkey (Winter, Ploog, and Latta, 1966), bullfrog (Capranica, 1965, 1968), and others (see Busnell, 1963, for a review). Yet, as in the case of birdsong, research on these species has not treated segment timing systematically, but has been primarily concerned with the frequency domain. The results of this study on speech timing suggest, however, that a similar research strategy might be applied with profit in other areas of animal communication, where, as in the case of speech, very little is known about the hierarchical structure of processing.

In search of other properties of the general principle, we can consider non-acoustic forms of motor output as well as non-communicative outputs. In our species, non-acoustic forms of communicative behavior include handwriting, telegraphy, sign language, and dance, all of which are segmentable to some extent. My intuition suggests that phrase-final lengthening exists for these behaviors, independent of other factors that may influence movement duration (such as the need for precision in reaching a goal). However, I have been able to find no documented evidence either in support of, or in opposition to, the phrase-final lengthening principle for these behaviors. In other animals, relevant

activities include communicative insect dancing (von Frisch, 1950) and perhaps certain non-communicative behaviors as well, including pecking in chicks and swimming in fish, to the extent that such behaviors can be shown on independent grounds to be hierarchically organized and segmentable. So far, all evidence in support of phrase-final lengthening comes from communicative behavior, and it remains a moot question whether communicative function is a necessary condition for the operation of the principle.

The phrase-final lengthening principle is used in this study primarily as a starting point for probing the control of timing in speech, and we need not know why the principle holds to undertake this effort. Yet, the question will ultimately become an important one for speech theorists, and we will consider here three possible answers that apply generally to the various behaviors that exhibit such lengthening.

Two types of explanation can be classified as transmitter-oriented, while a third can be classified as receiver-oriented. One of the transmitter-oriented accounts states that organisms produce phrase-final lengthening in order to permit an extra fraction of timing during which to compose a following phrase. This account rests on the reasonable assumptions that (a) motor output is planned on a phrase-by-phrase basis to some extent and (b) the planning of an upcoming phrase occurs primarily near the end of the currently produced phrase. This

account may be naturally extended to provide an explanation of the high frequency of pauses occurring between phrases, on the assumption that planning continues during the pause interval. The planning account, however, cannot explain another related finding, namely that segments are usually lengthened at the ends of discourses, symphonies, etc., where no further planning is required. It is conceivable, of course, that the planning account provides a correct explanation of phrase-final lengthening but that an independent principle of timing accounts for the lengthening observed at the termination of discourses and musical compositions. The latter principle might take the form of a receiver-oriented principle, which operates under the assumption that exaggerated lengthening at the end of the entire behavior pattern serves as a cue to the listener that the behavior has in fact ended.

The second transmitter-oriented account of phrase-final lengthening, like the planning account, relies on the assumption that the timing of a phrase is planned in a unitary fashion at some processing level. According to this second account, however, lengthening is produced at the ends of phrases not to aid the planning of upcoming phrases but as a consequence of the manner in which the previously planned unit was timed and stored before segment output. For example, segments of a processing domain might be stored in a buffer which operates like a push-down store (Simon and Kotovsky, 1963)[6], containing a "spring"-like mechanism. The force on the spring is directly proportional to the

number of segments currently in storage. As a consequence, successive segments are emitted from the buffer at successively slower rates, and hence, possess successively longer durations. This account provides an explanation of the graded lengthening effect noted for insect chirps, although the planning account can also be modified to handle this case. Discourse-final lengthening is not handled well by either the planning or buffer accounts, and there is generally little or no a priori reason to favor one of these accounts over the other.

A third general account of phrase-final lengthening is based on the assumption that lengthening is produced as a cue for the receiver, a cue which aids the receiver in recovering structural properties of the message. Unlike the transmitter-oriented accounts, the receiver-oriented explanation is limited to communicative behavior, but, as we have noted above, all known instances of phrase-final lengthening belong to this category.

The three general types of explanation for phrase-final lengthening are not incompatible with one another. This fact makes it possible that more than one of these accounts is responsible for the lengthening effect observed in any particular case, increasing the degrees of theoretical freedom and making it very difficult to provide tests that distinguish among the three main alternatives. To complicate matters further, it should be pointed out that, in addition to the three general alternatives, there exist certain specific accounts which may

apply to particular forms of phrase-final lengthening. An example of such an account occurs with speech, where it has been suggested that phrase-final lengthening is produced to permit a longer interval during which to produce intonational changes in phrase-final position (see Klatt, 1975a). Another possibility for speech, suggested by Stevens (personal communication), is that phrase-final lengthening is produced to allow the speaker time to reset laryngeal and articulatory postures for the upcoming phrase. This view is based on the notion that such postures "run down" during the production of a phrase and thus require resetting. Work has not proceeded to the point where this hypothesis can be formulated in precise terms, however. Eventually, it should be possible to provide critical tests of the alternative accounts noted here, but a number of other facts need to be tacked down beforehand. The present results are accountable in terms of any of the explanations cited above, with one important class of exceptions. Some of the consistently observed effects on duration were very small in average magnitude ($\sim$10 msec), too small to be reliably detected by listeners. Accordingly, such effects cannot be accounted for in terms of the receiver-oriented hypothesis noted above.

2. We turn now to consider speech timing in some detail[7]. Of the various behaviors discussed above, speech is believed to be the best area in which to undertake an in-depth study of timing control. The reason is that speech is our most vital communicative activity, and the study of timing control in this area should allow us to apply proper constraints on a theory of speech production, a matter of both intellectual and practical value. On one hand, providing a refined theory of speech production would bring us closer to understanding the essence of mental operations. On the other hand, such a theory can be applied directly to problems in communications engineering and speech pathology. These considerations, discussed more concretely below, provided the general motivation for the present study.

It was noted at the outset of this chapter that speech segments tend to be longer at the ends of phrases, clauses, and sentences than in non-final positions. For some types of sentence structures, this lengthening effect is clearly audible. It is not very surprising, then, that acoustic analyses of speech signals have confirmed this impression. And yet, such studies have brought to our attention in a particularly direct manner the fact that the control of speech timing is a rather intricate process, involving not only relatively low-level properties of the articulators, but syntactic properties of the utterance as well.

Martin (1970) discovered that grammatical boundaries were typically accompanied by lengthening of the preceding syllable in a spectrographic study of English spontaneous speech. The speech corpus was not very large (only 60 utterances) and the measurement technique was relatively crude, yet Martin's results did show the presence of a general clause-final lengthening effect. Kloker (1975) has corroborated and extended this finding for spontaneous speech, and Lindblom and Rapp (1973) and Klatt (1975a) have found similar effects for practiced reading, using Swedish and English, respectively. The combined results of the studies lend strong support to the notion that the ends of major grammatical domains are accompanied by segment lengthening.

Each of these studies contained two limitations which are important from the standpoint of the present study. First, with the exception of one aspect of the study conducted by Lindblom and Rapp (1973), none of the studies involved tests of syntactic effects that controlled for other influences known to affect lengthening, in particular influences of sound structure and sound environment (see Lehiste, 1970, for a review of some of these factors). Klatt (1975a) took such factors into account in a post hoc analysis, but experiments were not conducted to test comparisons between phonetically-matched sentences differing only in syntactic variables. Second, none of the previous studies included direct tests of a variety of syntactic boundaries, or cases in which the

locations of syntactic boundaries were controversial according to competing linguistic theories or according to distinct levels of representation within a given theory.

The present study was designed to provide a high degree of phonetic control and at the same time provide a testing ground for a number of hypotheses about the control of speech timing by syntactic (and semantic) factors. The information available about the influence of other factors on timing control is fairly extensive, enabling us to control for such factors wherever possible or to take their influence into account when their presence must be tolerated to test a syntactic hypothesis.

The additional influences on speech timing include gross factors such as overall speaking rate (Malecot, 1969; Gay, Ushijima, Hirose, and Cooper, 1974), word emphasis (Lieberman, 1967; Bolinger, 1972), word frequency (Coker, Umeda, and Browman, 1973), as well as detailed properties of sound structure, including the inherent duration of segments (Klatt, 1975a) and effects of immediate phonetic environment. An example of the last type of effect is the influence of a following voiced vs. voiceless consonant (e.g. [b] vs. [p]) on the duration of an immediately preceding vowel -- the vowel duration is longer when followed by a voiced consonant (Peterson and Lehiste, 1960; House, 1961; Delattre, 1966; Lisker, 1974). And, in addition to such

structural effects, the overall phonological stress pattern (Fry, 1955; Chomsky and Halle, 1968) plays a major role in the control of timing.

In normal speech, the interplay of these various factors may obliterate the audibility of phrase-final lengthening, although not to the extent that such lengthening is not generally detectable during the course of a conversation. From a practical standpoint, the presence of such a great number of factors necessitates that a detailed study of syntactic control of timing neutralize these other factors wherever possible. From a theoretical standpoint, the presence of so many extra factors suggests that either (a) speech timing and its possible perceptual relevance make use of extremely complex processing machinery, or (b) a much smaller number of control factors is accountable for the wide range of influences observed in speech behavior.

A version of a currently popular model of speech production is presented in Figure 2. The model represents a slightly more detailed

---

Insert Figure 2 about here

---

version of one described by Liberman (1970), among others. The principal features of the model are its strictly serial stages of processing and the close correspondence between such stages and components of a generative grammar (Chomsky, 1965). The model is presented here

primarily because it provides a useful framework for introducing aspects of the present study, not because it holds any special claim to validity.

The first stage of the model represents the conversion of thought into some form of linguistic representation. It is typically assumed that the linguistic representation is in essence a semantic representation, although it need not be (Fodor, 1975). Next, an underlying syntactic structure is formed, analogous to the level of deep structure proposed in generative grammar (see Postal, 1964 for introductory motivation for the linguistic distinction between deep and surface structures). Lexical items are inserted after the underlying structure has been formed, resulting in an output roughly corresponding to a terminal string in the grammar of Chomsky (1965). A system of transformations operates to move, add, or delete elements from the structure. The output of the transformational stage is a syntactic surface structure. This structure is transmitted to the phonological component of the speech processor, where word- and syllable-level rules are applied to produce the phonetic output, including the desired phonetic structure as well as proper stress and timing relations.

If the general outline of this serial model is correct, then the control of timing relations in the phonological stages of processing should be influenced by syntactic and/or semantic information only to the extent that such information is preserved at the final stage of syntactic processing. That is, information available in the surface structure representation should be capable of influencing the control

of timing, but not information about semantic structure, underlying syntactic structure, or transformational derivation, which does not also appear in surface structure.

The model is speculative and rests on no more than a few strands of experimental evidence. The prediction of the model concerning timing control has not been tested, nor have any of the previous studies on phrase-final lengthening provided information that has direct bearing on the prediction. The previous evidence which does concern the general form of the model involves analyses of errors in spontaneous speech (Fromkin, 1971; MacKay, 1972, 1973; Shattuck, 1974; Garrett, 1975), a few studies of hesitation phenomena, as well as some experimental attacks on the problem (see Fodor, Bever, and Garrett, 1974, Chapter 7, for review). The data from speech errors suggest that more than a single syntactic level of processing may exist in speech production, as required by the present model (Garrett, 1975). However, error analyses and other tests have so far failed to uncover any detailed properties of levels of syntactic processing other than properties of a level that corresponds to a surface structure representation.

A major drawback facing someone interested in discovering properties of the speech production system is the lack of an experimental paradigm that can be applied validly and efficiently to a large sub-set of the problems at hand. Those who analyze speech errors have

properly laid claim to the face validity of their enterprise, but while the analysis of errors has provided a very useful source of information, the work involved in obtaining a sufficiently large number of errors relevant to testing a particular hypothesis is at best arduous and, in some cases, impossible.

In addition to the general model of speech production reviewed above, a number of models have been proposed to account for speech timing. One model relies on the notion of isochrony, which states that speakers attempt to produce the onsets of stressed syllables at approximately equal intervals (Kozhevnikov and Chistovich, 1965; Lehiste, 1970). Both methodological and theoretical objections have been raised against the isochrony principle (Ohala, 1970; Klatt, 1975a). A fairly elaborate theory of stress timing, based in part on the isochrony principle, has been advanced by Martin (1972) to account for timing relations in both speech and musical composition. Other models of timing have been concerned primarily with two major processing distinctions, the difference between linear vs. hierarchical planning (Lashley, 1951) and between central vs. peripheral feedback control (MacNeilage, 1970; Ohala, 1970). These distinctions will be of marginal use in providing an account of the results of this study.

3. The principal aims of this study were:

   (1) to determine whether surface, transformational, and/or underlying syntactic stages of processing control speech timing

   (2) to determine whether semantic relations control speech timing

   (3) to determine the domain of processing at any stage of processing for which timing control can be shown.

In order to provide adequate tests of these questions, the control of timing was studied for a number of different syntactic structures with a single experimental design. The basic intent of the design was to provide a means of assessing the timing effects of syntactic variables while holding phonetic variables fixed as much as possible. To achieve this goal, a sentence-reading task was employed in which speakers were asked to read sentences as if they were uttering the sentences spontaneously. Ideally, analyses of spontaneous speech itself would provide the most fitting tests of a speech production model, but such analyses are strictly impossible in cases like the present where tight control over phonetic and situational variables is required. The sentence-reading task allows the experimenter to control for these variables, and yet provides a relatively natural speech situation. Since previous data on phrase-final lengthening showed similar effects for both spontaneous speech

and practiced reading (Martin, 1970; Lindblom and Rapp, 1973; Klatt, 1975a; Kloker, 1975), there is some reason to believe that the syntactic effects observed in practiced reading will be applicable to spontaneous speech.

Each hypothesis in this study was tested using one or more lists of sentences. Each list contained from 2 to 39 sentences, and each sentence in a list contained one or more key words. The key words, to be measured for duration, were placed at locations just before or after a putative syntactic boundary (a different procedure was used in tests of semantic variables; see Chapter 5). Wherever possible, the key words appeared in the middle of each sentence string, in order to minimize any possible effects of changes in subglottal pressure on timing (Lindblom and Rapp, 1973). In addition, the key words were selected on the basis of their phonetic structure, so that each word was readily segmentable in the speech waveform.

The sentences of a given list were closely matched for total number of words and syllables. In addition, the lexical material and sentence stress contour of the sentences were matched wherever possible. Finally, the sentences of a list were equally plausible, to a first approximation[8].

Speakers were tested individually in a sound-insulated room. A typical experimental session lasted about 45 minutes. During this time, the speaker was given from 5 to 7 sentence lists. At the outset of the session, the speaker was told that the general purpose of the experiments

was to study the control of speech timing in a relatively natural speech setting. The speaker was then provided general instructions that concerned the format of all tests in the session. These instructions included the following main points: (a) a number of different sentence lists will be presented; each list should be treated as a separate experiment, and each sentence within a list should be treated independently of any other sentences; (b) when a sentence list is presented, the speaker should first practice saying the sentence until he is comfortable with the utterance and is satisfied that he is able to utter the sentence with normal rhythm as a unitary whole, not word-by-word as in unpracticed reading; (c) emphatic or contrastive stress should not be placed on any words or syllables in a sentence (this instruction did not apply to the semantic test in Chapter 5); (d) after the speaker finishes practice, he should inform the experimenter and then utter one more practice utterance to allow the experimenter to check for undesirable emphatic or contrastive stress and to check recording levels; (e) on cue from the experimenter, the speaker should then say the sentence 6 times (or 2 times, as noted for particular experiments) in succession; (f) if the speaker departs from his normal practiced rhythm or utters a mispronunciation during recording, he should utter the word "repeat", pause, and then say the utterance again as necessary until the appropriate number of correct occurrences of the sentence is obtained. For most tests, the speaker

practiced and read a given sentence for recording before practicing any other sentences in the list; in some tests, as noted in later chapters, the speaker instead practiced each sentence in the list before reading any of the sentences for recording purposes.

In addition to these general instructions, each speaker was given particular instructions for certain lists, as indicated in later chapters. After the completion of each list, the speaker was usually encouraged to take a drink of water, and a longer rest period was provided about halfway through the test session.

For tests in which 6 occurrences of each sentence were recorded, the first 5 occurrences were digitized at a sampling rate of 10 kHz and analyzed for duration with the aid of a computer controlled cursor (Huggins, 1969). The reliability of the duration measurements varied slightly from experiment to experiment, as indicated later, depending on the phonetic structure and environment of the key word or segment. In all cases, however, the reliability was estimated to be within $\pm$ 5 msec and in most cases within $\pm$3 msec (see individual chapters for reliability estimates).

The sentence-reading technique outlined here provided a fairly efficient means of testing a variety of hypotheses about the organizational structure of speech timing control. Further efforts to automate aspects of the measurement procedure should make the technique more desirable as a tool for studying sentence production.

We now turn to consider the main findings of the study. These can be summarized as follows:

(1) evidence indicates rather strongly that a surface level of syntactic structure controls speech timing; preliminary evidence indicates that transformational and underlying syntactic structure may also control timing

(2) the magnitude of clause-final lengthening differs as a function of the particular type of clause

(3) semantic relations of coreference control timing

(4) the domain of semantic processing which controls timing is not restricted to a single clause; however, at least one domain of syntactic processing which controls timing probably is so restricted

A number of specific findings were obtained in the study, but they will not be reviewed here because they rely on linguistic constructions whose properties have not yet been discussed.

4. The principal findings of the study indicate the need for important revision and extension of the theoretical model outlines in Figure 2. In particular, the model must now be re-constructed so that higher level semantic information plays a role in the control of timing, either by a direct route or by some other route not presently available in the model.

In addition, the results suggest that semantic and syntactic control operate over different domains. Semantic processing is not limited to the domain of a single clause, whereas certain syntactic processing does appear to be so limited. Currently, work is being directed at specifying other details of the model (see Cooper, in preparation).

It was noted earlier in this chapter that one of the prime motives for undertaking this study was the expectation that it would be useful in guiding work in related areas of research. The outcome of the research has confirmed this expectation, and we note below three specific areas in which the results may play a guiding role.

One area, quite directly related to the work on speech production, involves studies on the perception of duration. In a recent report, Klatt and Cooper (1975) showed that listeners could detect differences in the durations of speech sounds of the same order of magnitude as the differences in duration produced by speakers, as a function of the syntactic environment. The results of the current study and further work using the sentence-reading paradigm provide important information about the magnitudes of speech production differences, information that can be used to guide the perceptual work. For example, cases can be distinguished from the production data in which lengthening effects are either clearly too small and inconsistent to be of perceptual relevance or are well within the range of a listener's detectability.

Perceptual research can thus be directed at studying the latter cases, in order to find out whether listeners can not only detect but utilize duration information to recover structural properties during sentence perception.

In addition to guiding work on sentence perception, the present results provide useful input to programs designed to synthesize speech from a phonetic transcription (Coker, Umeda and Browman, 1973; Klatt, 1975b). Currently, such programs include durational rules to some extent in order to produce more natural-sounding speech, but these rules typically do not involve syntactic or semantic variables or else treat such variables at only a very general level. In Klatt's (1975b) program, for example, a clause-final lengthening rule is included to increase the duration of clause-final segments by a small amount. However, the results of the present study indicate that the magnitude of the lengthening effect observed for clauses of different types may differ substantially. These results can be used to implement a more refined set of clause-final lengthening rules which take these differences into account. The results of similar studies should permit the implementation of sufficiently precise lengthening rules whose combined effect will be to noticeably enhance the naturalness and understandability of synthetic speech.

Improvements in speech synthesis have social significance because synthesis programs are implemented as a component of reading machines for the blind (Allen, 1973) and as part of an overall effort to achieve full man-machine communication by speech. The latter effort represents a much-desired but very remote goal which, if attained, would revolutionize information-processing as we know it in daily life. Typing would become obsolete, and many interactions within and among businesses, education, and government would take place over telephone lines between man and computer. Although this goal is well beyond the reach of current understanding and technology (see Reddy, 1974), results such as those provided here should play a small part in directing the effort.

A third application of the present study concerns the teaching of speech rhythm in the deaf. Boothroyd, Nickerson, and Stevens (1975) have noted that improper rhythm is one of the more important drawbacks in the speech of congenitally deaf children, and teachers of these children are often unaware of the importance of rhythmic structure in speech training. Work on phrase-final lengthening and other aspects of speech timing may culminate in a system of durational rules that could be taught to deaf children. However, like the engineering goal of man-machine intercommunication by speech, this application can only be attained after a much greater portion of the systematic research effort on speech timing is completed.

Footnotes

1. Note that the theories of grammar and musical composition provide an independent means of determining domains of structure but not domains of processing in a psychological sense. We adopt as a working assumption the principle that an isomorphic relation exists between domains of structure and domains of processing at some stage of central motor activity.

2. It seems that the poet should also be included in this group, since iambic meter is more common than trochaic. Of related interest is the observation by Allen (1975) that languages having accent based primarily on duration (e.g. French) have stress on final syllables of words, whereas languages having strong tonic accent (e.g. English, German) have syllable-initial stress primarily.

3. In most cases, general fatigue can be discounted as an explanation for lengthening by showing that a segment at the end of a phrase is longer than the average duration of segments in an immediately following phrase.

4. It is expected that in cases where the phrase-final segment is short, it will be of high amplitude. This effect is observed in some modern symphonies.

5. The only documented counter-example that I have found so far concerns the bird song of Pileated Tinamou, cited by Thorpe (1961: 3). The song phrase of this species consists of note groupings that possess rising gradations in pitch and gradations of decreasing length.

6. The assumption that the buffer operates like a push-down store can be applied to this situation only if another, counterintuitive, assumption is made, namely that speech segments of a domain enter the store such that the domain-final segment enters the store first.

7. In addition to phrase-final lengthening, there exists another speech phenomenon which supports the general principle in (1). Cooper and Ross (1975) have noted that, in pairs of conjoined words for which the linear ordering of the words is rigidly fixed (e.g. kit and caboodle/ *caboodle and kit), words which are fixed in second position have a vowel with an inherently longer duration than the vowel of the word fixed in first position (e.g. stress and strain, hem and haw), other factors being equal or taken into account. The vowel length principle for fixed conjuncts suggests an instance in which the principle in (1) has become frozen in the structure of the language itself, in addition to operating

to influence the lengths of speech segments during on-line speech production.

8. Since the reading of test sentences is practiced prior to recording, it is anticipated that small differences in plausibility will play a negligible role in determining segment length.

References

Allen, G. D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. Journal of Phonetics 3, 75-86.

Allen, J. (1973). Speech synthesis from unrestricted text. In J.L. Flanagan and L. R. Rabiner (Eds.) Speech Synthesis. Stroudsburg, Pa.: Dowden, Hutchinson, and Ross, Inc.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). Language 48, 633-644.

Boothroyd, A., Nickerson, R.S., and Stevens, K.N. (1975). Temporal patterns in the speech of the deaf: a study in remedial training. Submitted for publication.

Busnel, R.-G. (1963). Acoustic Behavior of Animals. Amsterdam: Elsevier.

Capranica, R.R. (1965). The Evoked Vocal Response of the Bullfrog. Cambridge: M.I.T. Press.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge: M.I.T. Press.

Chomsky, N. and Halle, M. (1968). The Sound Pattern of English. New York: Harper and Row.

Coker, C.H., Umeda, N., and Browman, C.P. (1973). Automatic synthesis from ordinary English text. IEEE Audio and Electroacoustics AU-21, 293-297.

Cooper, W.E. (in preparation). Syntactic Control of Speech Timing.

Cooper,W.E. and Ross, J.R. (1975). World order. In R.E. Grossman, L.J. San, and T.J. Vance (Eds.) Papers from the Parasession on Functionalism. Chicago: Chicago Linguistic Society. Pp. 63-111.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. International Review of Applied Linguistics 4, 183-198.

Du Mortier, B. (1963). The physical characteristics of sound emissions in arthropoda. In R.-G. Busnel (Ed.) Acoustic Behavior of Animals. Amsterdam: Elsevier. Pp. 346-373.

Fodor, J.A. (1975). The Language of Thought. New York: Thomas Y. Crowell.

Fodor, J.A., Bever, T.G., and Garrett, M.F. (1974). The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. New York: McGraw Hill

Fromkin, V.A. (1971). The non-anomalous nature of anomalous utterances. Language 47, 27-52.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America 27, 765-768.

Gay, T., Ushijima, T., Hirose, H., and Cooper, F. S. (1974). Effect of speaking rate on labial consonant-vowel articulation. Journal of Phonetics 2, 47-63.

Garrett, M.F. (1975). The analysis of sentence production. In G. Bower (Ed.) Advances in Learning Theory and Motivation: Vol. 9. New York: Academic Press.

House, A. (1961). On vowel duration in English. Journal of the Acoustical Society of America 33, 1174-1178.

Huggins, A.W.F. (1969). A facility for studying perception of timing in natural speech. Quarterly Progress Report of the M.I.T. Research Laboratory of Electronics 95, 81-83.

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

Klatt, D.H. (1975b). Structure of a phonological component for a synthesis-by-rule program. Paper presented at the 90th Meeting of the Acoustical Society of America, November, 1975.

Klatt, D.H. and Cooper, W.E. (1975). Perception of segment duration in sentence contexts. In A. Cohen and S.G. Nooteboom (Eds.) Structure and Process in Speech Perception. Heidelberg: Springer-Verlag.

Kloker, D. (1975). Vowel and sonorant lengthening as cues to phonological phrase boundaries. Paper presented at the 89th Meeting of the Acoustical Society of America, April, 1975.

Kozhevnikov, V.A. and Chistovich, L.A. (1965). Speech: Articulation and Perception. Joint Publications Research Service 30, 543, U.S. Department of Commerce, Washington, D.C.

Lashley, K.S. (1951). The problem of serial order in behavior. In L.A. Jeffress (Ed.) Cerebral Mechanisms in Behavior: The Hixon Symposium. New York: Wiley.

Lehiste, I. (1970). Suprasegmentals. Cambridge: M.I.T. Press.

Liberman, A.M. (1970). The grammars of speech and language. Cognitive Psychology 1, 301-323.

Lieberman, P.H. (1967). Intonation, Perception, and Language. Cambridge: M.I.T. Press.

Lilly, J.C. (1963). Distress call of the bottlenose dolphin: stimuli and evoked behavioral responses. Science 139, 116-118.

Lilly, J.C. (1965). Vocal mimicry in Tursiops: ability to match number and durations of human vocal bursts. Science 147, 300-301.

Lindblom, B. and Rapp, K. (1973). Some temporal regularities of spoken Swedish. Papers from the Institute of Linguistics, University of Stockholm, Publication 21.

Lisker, L. (1974). On "explaining" vowel duration variation. Glossa 8, 233-246.

MacKay, D.G. (1972). The structure of words and syllables: evidence from errors in speech. Cognitive Psychology 3, 210-227.

MacKay, D.G. (1973). Complexity in output systems: evidence from behavioral hybrids. American Journal of Psychology 86, 785-806.

MacNeilage, P.F. (1970). Motor control of serial ordering of speech. Psychological Review 77, 182-196.

Malecot, A. (1969). The effect of syllabic rate and loudness on the force of articulation of American stops and fricatives. Phonetica 19, 205-216.

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

Martin, J.G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behaviors. Psychological Review 79, 487-509.

Nottebohm, F. (1970). Ontogeny of bird song. Science 167, 950-956.

Ohala, J.J. (1970). Aspects of the control and production of speech. Unpublished Ph.D. Dissertation, U.C.L.A., Los Angeles, Ca.

Peterson, G.E. and Lehiste, I. (1960). Duration of syllabic nuclei in English. Journal of the Acoustical Society of America 32, 693-703.

Postal, P.M. (1964). Underlying and superficial linguistic structure. Harvard Educational Review 34, 246-266.

Reddy, D.R. (Ed.) (1974). Speech Understanding Systems. New York: Academic Press.

Shattuck, S.R. (1974). Unpublished Ph.D. Dissertation. M.I.T., Cambridge, Mass.

Simon, H.A. and Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. Psychological Review 70, 534-546.

Thorpe, W.H. (1961). Bird-song. Cambridge: Cambridge University Press.

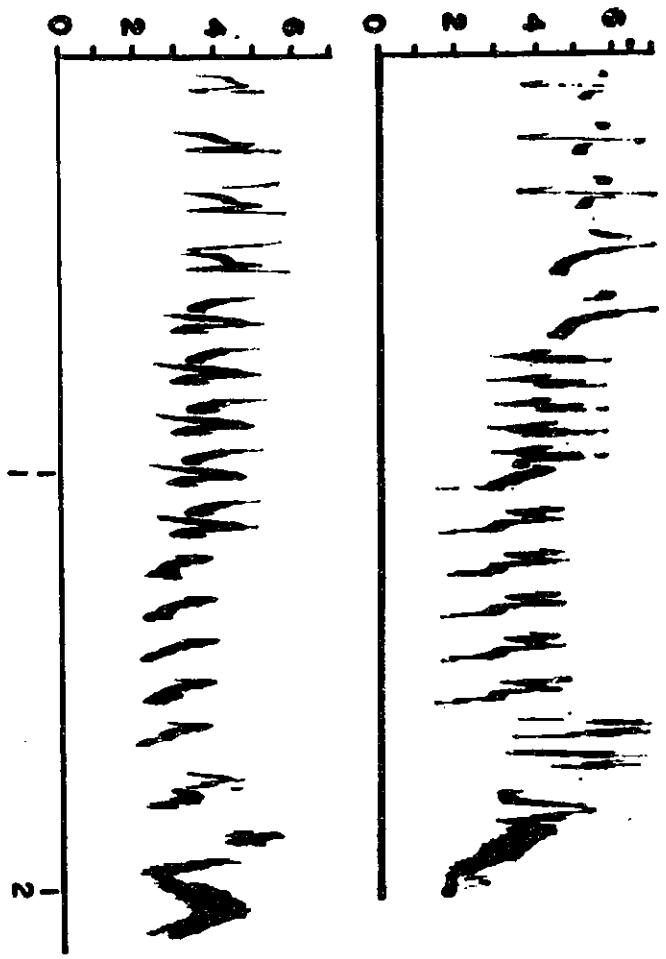
von Frisch, K. (1950). Bees, Their Vision, Chemical Senses, and Language. Ithaca, New York: Cornell University Press.

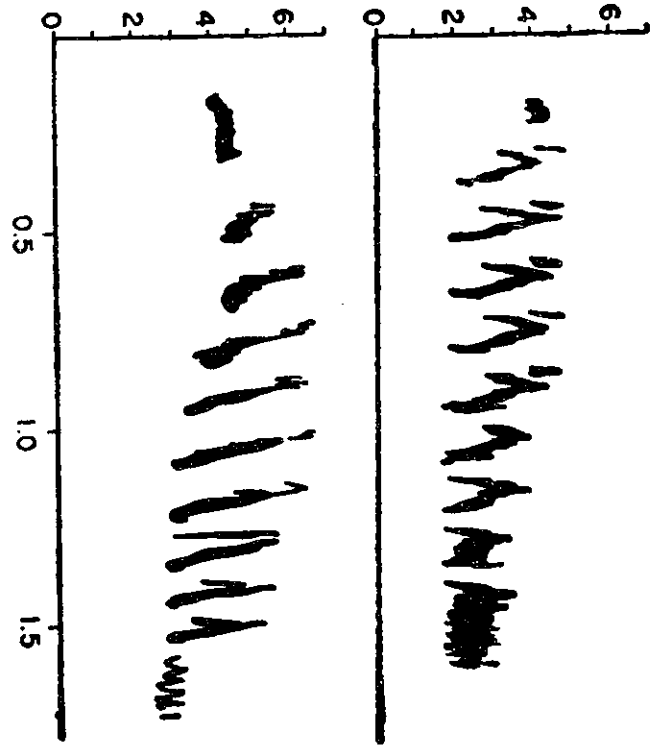
Winter, P., Ploog, D., and Latt, J. (1966). Vocal repertoire of the squirrel monkey (Saimiri sciureus). Experimental Brain Research 1, 359-384.

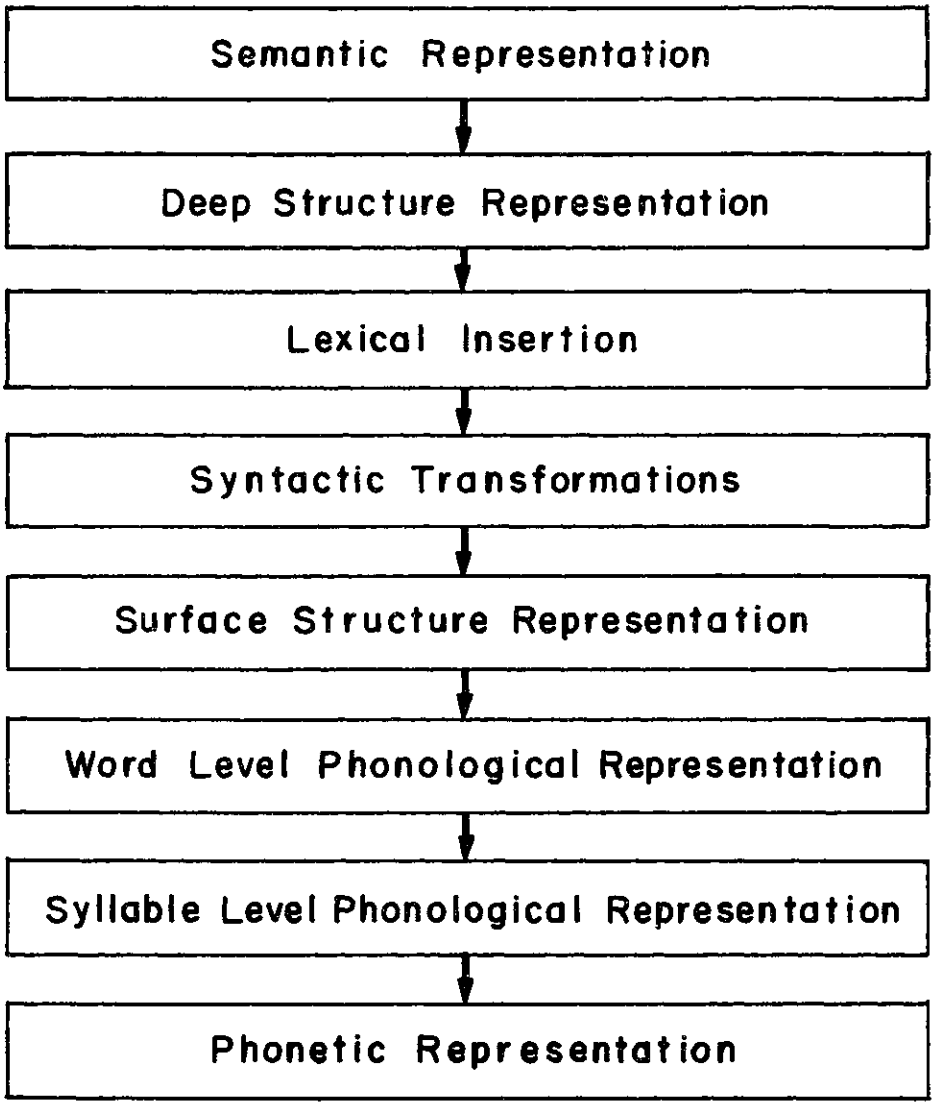
Figure Captions

Figure 1: Spectrographic traces of birdsong in the chaffinch (from Nottebohm, 1970).

Figure 2: A currently popular model of information-processing in speech production.


0
2
4
6
1
2


0
2
4
6
0.5
1.0
1.5


Semantic Representation
Deep Structure Representation
Lexical Insertion
Syntactic Transformations
Surface Structure Representation
Word Level Phonological Representation
Syllable Level Phonological Representation
Phonetic Representation

CHAPTER 2

Syntactic Control of Segment Length in Speech

Abstract

The durations of segments in spoken sentences were measured to determine whether speakers alter durations depending on syntactic structure. The speakers read aloud sentences for which the effects of three syntactic deletions were assessed in a relatively fixed phonetic environment. Speakers shortened the durations of segments just prior to the locations of two deletions which erase material from the beginning of a clause, whereas speakers lengthened the durations of segments just prior to the location of a deletion that erases material from the middle of a clause.

The control of timing in speech production is governed by a variety of factors, including syntactic and phonetic influences. For example, the duration of a stressed vowel is modified according to its position in a surface clause (1) and whether it is followed by a voiced or voiceless consonant (2).

The work to date on syntactic influences of timing has suggested that major clause boundaries produce lengthening of segments in a preceding word (1). However, the syntactic effects have not been assessed independently of phonetic influences on timing, nor has work been directed at the question of whether one or another type of linguistic representation corresponds to the level of syntactic processing that mediates legthening.

A control for phonetic influences was provided here by requiring speakers to read sentences in which a key word or other segment was inserted in a relatively fixed phonetic environment. Using this procedure, it was possible to test two linguistic hypotheses about the kind of clausal analysis that a speaker computes during sentence production.

According to one hypothesis, pre-clausal lengthening is determined at a level of computation comparable to the level of linguistic representation known as deep, or underlying structure (3). Alternatively, lengthening could be controlled at a level of processing comparable to a representation of surface structure. The linguistic rationale for distinguishing deep and surface syntactic descriptions is well-motivated (3) and forms a central thesis of generative grammar as well as most other current grammars that attempt to provide an adequate account of sentence well-formedness. According to generative grammar, the deep and surface

levels of description are mediated by a set of transformations which permute, add, or delete elements from the underlying structure in order to derive a surface form. The experiments here were designed to test whether pre-clausal lengthening is determined by a level of speech computation corresponding to deep vs. surface linguistic structure.

Experiment I involved sentences like (a) I have the tape (that) the officer erased, in which a relative pronoun that can be deleted optionally by a transformation known as Relative Clause Reduction (4). The deletion of a relative pronoun in Sentence (a) is accompanied by the deletion of the sentence-node dominating the subordinate clause the officer erased the tape in underlying structure (5). Thus, if pre-clausal lengthening is determined by surface rather than underlying structure, lengthening should be observed for the word tape in Sentence (a) only when this sentence contains the relative pronoun in surface structure.

The experiment involved two sentences, namely the full (no deletion) and reduced (deletion) versions of Sentence (a). The key word tape was bound on the left by the same word the in both sentences and on the right by the words the and that, containing the same initial phonetic segment [ð]and stress (6).

Sixteen speakers read each of these two sentences 6 times in succession, beginning with either the full or reduced versions, according to random assignment. The speakers were seated in a sound-insulated chamber and were told at the beginning of the test to practice the sentences so as to be able to read the sentences "as unitary wholes" rather than "word-by-word" as in unpracticed reading. In addition, the speakers were instructed not to place contrastive or emphatic stress on any words in the sentences.

All utterances were recorded onto magnetic tape via a Neumann U87 microphone and a Revox A77 tape recorder. The duration of the key

word tape was measured from digitized oscillographic traces (sampling rate = 10 kHz) of the first five occurrences of each sentence for each speaker. The segment duration for tape excluded the [t] and [p] closure intervals and a [p] release (if any existed) in order to minimize variability. The segment boundaries were marked with the aid of a computer controlled cursor (7), and the accuracy of each measurement was estimated to be within $\pm$ 2 msec.

The results for the 16 speakers showed a small but statistically significant difference in the duration of the key word tape ($p<.02$, $t = 2.755$, $df = 15$; two-tailed t-test for matched pairs). The average duration of the sentence containing the relative pronoun was 6.1 msec longer than the duration of the reduced sentence. The results support the hypothesis that pre-clausal lengthening is controlled in part by a level of computation that corresponds to a surface rather than deep level of linguistic description. The results are paralleled by a recent study of speech perception which showed a stronger clause-boundary influence on click location in the case of full vs. reduced relative clauses (8).

Experiment II was designed to test the generality of the above effect with a second deletion that also operates to delete material from the beginning of a clause as well as deleting a sentence-node. This deletion, Conjunction Reduction, deletes material optionally under identity with material in another clause, as in Dave rehearsed the show on Tuesday and (Dave) taped it the same day. If the results of Experiment I depended on surface structure relations, rather than on the particular form of the deletion rule (identity vs. non-identity deletion), then the

effect obtained in Experiment I should also be observed with Conjunction Reduction.

Experiment II involved 18 speakers, 16 of whom served in Experiment I. Three sentences were used: (a) Dave rehearsed the show on Tuesday and Ted taped it the same day; (b) Dave rehearsed the show on Tuesday and taped it the same day; and (c) Dave rehearsed the show on Tuesday and taped it the very same day. Sentence (a) represented a full coordinate structure, whereas Sentences (b) and (c) were derived by Conjunction Reduction. Sentence (c) was included to test the possibility that the lengthening effect for full vs. deleted versions is conditioned by the total length of the sentences. The segment chosen for measurement in each of the three sentences included the final syllable of Tuesday, the word and, and any pause existing in between.

The results showed that the durations of this segment were significantly longer for the full coordinate structure of Sentence (a) than for either of the reduced Sentences (b) and (c), as predicted on the basis of the surface structure hypothesis [(a) vs. (b): $p<.05$, $t = 1.982$, $df = 17$; (a) vs. (c): $p<.05$, $t = 1.926$, $df = 17$; one-tailed t-tests for matched pairs]. The average duration of the key segment in Sentence (a) was 27.6 msec longer than in Sentence (b) and 39.8 msec longer than in Sentence (c). No significant difference was observed between the two reduced Sentences (b) and (c) ($p>.20$, $t = 0.938$, $df = 17$). The effects obtained in this experiment add further support to the notion that pre-clausal lengthening is determined at least in part by clausal relations that exist at a surface rather than underlying level of description. The results of this experiment, however, are also compatible

with an account of timing in terms of rhythmic stress patern (9).

Experiment III was conducted to test the possibility that a deletion rule which erases material from the middle of a clause and is not accompanied by sentence-node deletion would also produce lengthening at the site of deletion. If so, then the previous effects would need to be regarded as being determined by the process of syntatic deletion per se rather than by the particular surface structure relations that hold as a consequence of deletions that erase material from the beginning (or presumably, the end (10) of a clause and delete an underlying sentence-node.

Gapping (11) is a rule that optionally deletes a verb from the middle of a clause under identity with the verb of another clause, leaving the general structure of both clauses intact. Two sentences were used in Experiment III: (a) The lecture began before lunch and the test began before three; and (b) The lecture began before lunch and the test before three o'clock. Sentence (a) involved no deletion, whereas Sentence (b) was derived via Gapping. Both sentences contained the same number of words and syllables. Measurements of duration were made on the key word test.

The results showed that the average duration of the key word in the reduced Sentence (b) was 8.6 msec longer than the duration of this same word in Sentence (a). This trend was opposite in direction from the lengthening effects observed for the non-deletion versions of sentences in Experiments I and II. The effect observed in Experiment III did not reach statistical significance ($.20 > p > .10$, $t = -1.492$, $df = 15$; two-tailed t-test for matched pairs). The results of this experiment show that the effects observed in the previous experiments

must be accounted for by a level of processing that is sensitive to the clause structure relations that exist in surface structure. Other experiments using the sentence-reading procedure also support this claim (12).

The magnitude and consistency of the lengthening effects for some speakers was sufficiently great to warrant the implementation of such effect in programs to synthesize speech by machine from a phonetic transcription (13).

References and Notes

1. J.G. Martin, J. Verb. Lrn. Verb. Beh. 9, 75 (1970); B. Lindblom and K. Rapp, Papers from Inst. Ling., Univ. Stockholm, Publ. 21 (1973); D.H. Klatt, J. Phonetics 3, 129 (1975).

2. G.E. Peterson and I. Lehiste, J. Acoust. Soc. Amer. 32, 693 (1960); A. House, J. Acoust. Soc. Amer. 33, 1174 (1961).

3. N. Chomsky, Syntactic Structures (Mouton, The Hague, 1957); P.M. Postal, Harvard Ed. Rev. 34, 246 (1964); N. Chomsky, Aspects of the Theory of Syntax (MIT Press, Cambridge, 1965); N. Chomsky, in A Festschrift for Morris Halle, S.R. Anderson and P. Kiparsky, Eds. (Holt, Rinehart, and Winston, New York, 1973), pp. 232-285.

4. C.S. Smith, Lang. 40, 37 (1964).

5. J.R. Ross, in Modern Studies in English, D.A. Reibel and S.A. Schane, Eds. (Prentice-Hall, Englewood Cliffs, N.J., 1969), pp. 288-299. Ross's original formulation of sentence-node pruning was motivated for the form of relative reduction known as whiz deletion (4), in which the relative pronoun and verb are deleted. However, the argument involving extraposition which Ross used as motivation for sentence-node pruning in the case of whiz deletion also applies to cases in which the relative pronoun alone is deleted (J. Hankamer, Constraints on Deletion in Syntax, Ph.D. Dissertation, Yale Univ., 1971).

6. To the extent that slightly more stress was placed on that vs. the in the two sentences, a phonetic influence of lengthening on the key word would be predicted in the direction opposite that predicted on the basis of the syntactic hypothesis (A.W.F. Huggins, Quart. Prog. Rpt. M.I.T. Res. Lab. Electr. 114, 179 (1974).

7. A.W.F. Huggins, Quart. Prog. Rpt. M.I.T. Res. Lab. Electr. 95, 81 (1969).

8. J.A. Fodor, J.D. Fodor, M.F. Garrett, and J.R. Lackner, Quart. Prog. Rpt. M.I.T. Res. Lab. Electr. (1976), in press. The reduced relatives in their work differed from those of the present study in that they were derived by whiz deletion (see Note 5).

9. Note that the key segment of the undeleted sentence was followed immediately by two strongly stressed words, whereas the corresponding segment of the deleted sentences was followed by a single stressed word. It is conceivable that speakers lengthened the duration of the key segment in the undeleted sentence because this segment terminated a rhythmic phrase group. Such an explanation, however, cannot account for the difference obtained in Experiment I and wrongly predicts other results obtained with the sentence-reading paradigm (12).

10. Conjunction Reduction can operate to delete material at either the beginning or end of a clause.

11. J.R. Ross, in Progress in Linguistics, M. Bierwisch and K.E. Heidolph, Eds. (Mouton, The Hague, 1970); R.S. Jackendoff, Ling. Inq. 2, 21 (1971).

12. W.E. Cooper, Syntactic Control of Timing in Speech Production, Ph.D. Dissertation, M.I.T., 1976).

13. C.H. Coker, N. Umeda, and C.P. Browman, IEEE Aud. and Electroacoust. AU-21, 293 (1973).

14. Supported by a National Science Foundation Graduate Fellowship and NIH Grant HD-05168.

CHAPTER 3

Syntactic Control of Timing in Speech Production: a Study of Complement Clauses

Abstract

A sentence-reading procedure was used to study the influence of syntactic structure on the timing of syllables in speech production. In each sentence, easily segmentable key words were placed in the environment of a putative syntactic boundary. The immediate phonetic environments of the key words were held fixed or independently assessed for their effect on syllable timing. In the main experiment, the effects of complement clause boundaries were tested in sentences containing the verbs expected and persuaded. These verbs were used because they trigger structurally distinct complement types. Each of a group of 15 speakers read aloud a set of 5 sentences 6 times each. The waveforms of the first 5 occurrences of each sentence were digitized at a sampling rate of 10 kHz and measured for the duration of the last two syllables of the verb and for the duration of the immediately following noun. The results showed that the last two syllables of the verb expected but not of persuaded were significantly shorter, by a small amount averaging about 12 msec, when the verb was followed by a complement clause appearing in the surface structure of the sentence as compared with when the verb was followed by a non-surface complement. The results support the notion that a level of syntactic computation controls timing relations in speech production and that this level computes a clausal representation similar to a variant of linguistic surface structure. Other experiments showed that (1) a phonetic effect on syllable timing, conditioned by the presence of a following voiced vs. voiceless segment, extends across a verb-noun phrase boundary and (2) the difference noted for the verbs expected and persuaded generalizes at least in part to other verbs in sentences having the same structural distinction.

1. Introduction

The control of timing in speech production is governed by a variety of factors, including overall speaking rate as well as phonetic and syntactic influences. For example, the duration of a stressed vowel differs depending on whether a following consonant is voiced vs. voiceless (Peterson and Lehiste, 1960; House, 1961; Delattre, 1966), whether the vowel is contained in an emphasized word (Lieberman, 1967; Bolinger, 1972), or whether the vowel is contained in a clause-final vs. non-clause-final syllable (Martin, 1970; Lindblom and Rapp, 1973; Huggins, 1974; Klatt, 1975; Kloker, 1975).

The syntactic influences on speech timing may provide an opportunity to examine the kinds of computations that a speaker performs during the formation of sentence structure. Currently, very little is known about such computations because sentence production is not easily amenable to experimental manipulation. Evidence from speech errors (see Garrett, 1975 and references cited therein) has been the most useful source of information heretofore, but an experimental setting is needed to test hypotheses generated either by speech error analyses or by linguistic theory.

In this paper, a sentence-reading procedure is used which allows the investigator to examine some of the effects of syntactic structure on syllable timing in a controlled yet relatively natural speaking situation. The sentence-reading procedure has been applied to a number of syntactic constructions (Cooper, 1975a), and the data obtained with the technique are used to distinguish possible computational models of sentence production as well as to guide two related research

areas--speech synthesis by rule (Coker, Umeda, and Browman, 1973) and the study of perception of duration and its possible role in aiding listeners to recover structural information (Klatt and Cooper, 1975).

1.1 Underlying vs. Surface Structure

The research on syntactic influences of timing to date has shown that syllable lengthening occurs primarily at the ends of major clause and phrase boundaries,[1] where boundary locations are determined by parsing methods similar to those taught in grade school. However, modern grammatical work typically distinguishes between two or more levels of syntactic representation for a given sentence, and studies of syllable lengthening have not yet been designed to test whether the lengthening effect is produced by one or another level of syntactic processing.

Regardless of whether one chooses to adopt a transformational grammar or one of a number of alternatives, a property common to virtually all current formulations of grammar is the distinction between a surface representation of word order and another level or levels of syntactic, logical, and semantic relations. A variant of this distinction appears to underlie the computational processes that mediate speech production, according to a recent analysis of speech errors by Garrett (1975).

The present study was designed to test whether syllable lengthening might be sensitive to the distinction between surface and underlying structure. Complement clauses were chosen for sentence materials, since

linguistic hypotheses distinguishing underlying and surface representations have been particularly explicit in the case of complements (Rosenbaum, 1967; Chomsky, 1973; Postal, 1974); moreover, as recent discussions by Chomsky and Postal indicate, the analysis of complement structures is pivotal in current controversy over the general form that grammatical rules and constraints on such rules should take (for a review, see Cooper, 1975b). Thus, in addition to providing a means of testing whether syllable timing is primarily controlled at an underlying or surface level of syntactic representation, the study of complementation affords an opportunity to test the relevance of competing linguistic analyses for developing a performance model of the kinds of computations carried out by these processing levels.

Although the clausal analysis of complements does provide a strong linguistic backdrop to the present study, there is no assurance a priori that complement clauses are accompanied by syllable lengthening at all in speech production, unlike other clause types (Cooper, 1975a). Coordinate clauses, non-restrictive relatives, and conditionals are marked by a comma in written English and are bounded by perceptible syllable lengthening and pauses in spontaneous speech; by contrast, complement clauses are neither accompanied by a comma in writing nor by a perceptibly obvious terminal lengthening in speech.

1.2 EXPECT vs. PERSUADE: underlying structure

The structure of the complement clauses used in this study will now be reviewed, beginning with an analysis of underlying structure. A major distinction between the underlying structures of two types of

complement was pointed out by Chomsky (1965) and is illustrated by sentences (1) and (2), containing the verbs expect and persuade.

(1) The host expected Kate to be at breakfast.

(2) The host persuaded Kate to be at breakfast.

Although these two sentences are quite similar superficially in their word order, they have quite different logical structures. This difference is brought out clearly when one considers the meanings of passive sentences containing expect vs. persuade, as in (3) vs. (4):

(3) The host expected Kate to be brought by an escort.

(4) The host persuaded Kate to be brought by an escort.

Sentence (3) is synonymous with the active sentence The host expected an escort to bring Kate, whereas sentence (4) is not synonymous with the corresponding The host persuaded an escort to bring Kate. The logical distinction underlying this difference is that expect Kate X does not entail expect Kate, whereas persuade Kate X does entail persuade Kate.

The question of how this underlying distinction between the complements of expect and persuade is to be represented (i.e., in the underlying structure of the syntactic component of grammar or in a separate component marking logical relations) remains to some extent problematic. If the distinction is represented in the syntactic component, according to the proposal of Rosenbaum (1967), then the complments of the two verbs are assigned different clause status. Thus, to the extent that the underlying clause representation of complements determines syllable lengthening, the complements of expect vs. persuade should produce an observable difference in speech timing.

In particular, Rosenbaum proposed that sentences like (1) and (2) have underlying structures like those shown in Figure 1.

Insert Figure 1 about here

Under this proposal, the critical distinction between the complements is that the expect complement is immediately dominated by a Noun Phrase node in underlying structure, whereas the persuade complement is immediately dominated by a Verb Phrase node. As a consequence, Kate is a member of the superordinate clause in the structure of the persuade sentence but is a member of the subordinate clause in the structure of the expect sentence. Put another way, the major clause boundary for the persuade sentence occurs immediately after Kate, whereas the major boundary for the expect sentence occurs immediately before Kate, just after the verb expect. If the underlying clause structure of complements determines syllable lengthening, then lengthening should be observed for Kate in sentence (2) and for expect in sentence (1), relative to some reference duration.

1.3 EXPECT vs. PERSUADE: surface structure

The underlying structures shown in Figure 1 can be converted into surface structures by application of transformational rules--rules which move, add, or delete elements (Chomsky, 1965). For a verb like expect, the underlying structure can be converted into two types of surface complement--an infinitival complement (introduced by to), as in sentence (1), or a that complement (introduced optionally by that), as in sentence (5) below.

(5) The host expected (that) Kate would be at breakfast.

Generative grammarians agree that the surface structure of that complements

like (5) preserves the constituent structure of the underlying representation shown in Figure 1. In bracketed form, the constituent structure of (5) can thus be represented by (5'):

(5') $[_S[_S[_{NP}$the host$][_{VP}$expect$]][_S[_{NP}$Kate$][_{VP}$be at breakfast$]]]$

According to this analysis, there is a major clause boundary immediately after the verb expected in sentence (5) both in surface and underlying structure.

For infinitival complements with expected, however, there exist two major alternative ways of describing the surface representation. According to one alternative, advocated principally by Rosenbaum (1967) and Postal (1974), the surface structure of infinitival complements differs from its underlying structure, and hence, differs from the surface structure of that complements as well. Rosenbaum and Postal propose that the noun Kate in infinitival complements like (1) has been moved ("raised") from its position as the subject of the subordinate clause in underlying structure (see Figure 1) into the object position of the higher clause in surface structure. The transformational rule proposed to account for this movement is termed Raising (from subject to object).[2]

Note that according to the Raising analysis of infinitival complements, the surface structure of (1) contains a major clause boundary immediately after the noun Kate; in contrast, the surface structure of that complements contains a major clause boundary after the main verb, just prior to Kate. The bracketed surface form of (1) is represented by (1') below, assuming the Raising analysis:

(1') $[_S[_S[_{NP}$the host$][_{VP}$expect Kate$]][_S[_{NP}\emptyset][_{VP}$be at breakfast$]]]$[3]

The Raising analysis of infinitivals was motivated by the need

to account for a number of differences in grammaticality between infinitival vs. that complements under the application of certain transformations (e.g., Passive) and constraints on transformations (e.g., Inclusion Constraint--see Postal, 1974 and Cooper, 1975c for discussion). However, another plausible alternative account of infinitival complements has been proposed by Chomsky (1973) to account for the same range of facts. According to Chomsky's analysis, the surface structure of (1) is the same as (5) insofar as major constituency relations are concerned. To account for the variety of differences in grammaticality between infinitival and that complements, Chomsky proposes that a distinction be made between finite and infinitival clauses, and that a universal condition on transformations, the Tensed-S Condition, be used to account for the differences between the complement types.

At present, the linguistic controversy surrounding the two alternative accounts of infinitival complements is not settled. However, it is of independent interest whether either of these two proposals provides a better model of performance in sentence production. This question can be tested in this study of syllabel timing because the two proposals make conflicting predictions about the location of syllable lengthening, under the assumption (requiring independent verification) that a surface as opposed to underlying level of representation controls such lengthening, at least in part. According to the Raising analysis, lengthening should appear on the noun Kate of an infinitival complement with expect but on the main verb of the corresponding that complement; in contrast, according to Chomsky's Tensed-S account, lengthening should appear on the main verb for both types of complement.

We turn now to consider the surface structure of persuade complements, since the difference between the Raising and Tensed-S proposals is also represented here. Unlike expect, the verb persuade takes only infinitival complements with an underlying structure like that shown in Figure 1. The infinitival sentence in (2), repeated below for convenience, is the relevant form.

(2) The host persuaded Kate to be at breakfast.

Given the underlying representation of Figure 1, a transformational rule must be postulated to delete one of the two occurrences of the noun Kate to convert the underlying structure into a surface form. Generative grammarians agree for the most part that the correct formulation involves deleting the subordinate occurrence of Kate, under identity with the superordinate one, by a rule known as Equi-NP Deletion. After this rule has applied, the surface structure of (2) takes the bracketed form shown in (2').[4]

(2') $[_S[_S[_{NP}$the host$][_{VP}$persuaded Kate$]][_S[_{NP}\emptyset][_{VP}$be at breakfast$]]]$

The major clause break of this structure appears immediately after the noun Kate. Thus, syllable lengthening should be observed for this noun to the extent that such lengthening is determined by either underlying or surface structure.

Note that according the the Raising analysis, the major clause boundary for the persuade infinitival occurs at the same location as for the expect infinitival in surface structure, namely after Kate. But, according to Chomsky's analysis, a difference in the boundary locations for expect vs. persuade infinitivals exists, with the boundary occuring after the main verb for expect but after Kate for the persuade complement. By including persuade infinitivals in this study, it was thus possible to

provide another test of the merits of the two proposals in accounting for syllable lengthening in speech.

A summary of the clausal analysis of expect and persuade complements is provided in Table 1, indicating the major clause boundaries predicted by the Raising and Tensed-S proposals.

Insert Table I about here

1.4 TO BE deletion

Since the Raising and Tensed-S analyses make conflicting predictions about the locations of clause boundaries for the surface but not the underlying representations of complements, it is necessary to try to provide an independent test of whether syllable lengthening is primarily determined by one or the other of these two levels of representation. Fortunately, evidence from a variety of other experiments using the same testing procedure as the present study indicates the presence of syntactic effects which can be attributed to surface but not underlying structural relations (Cooper, 1975a). However, none of these other experiments involved complement structures, so a further test of surface structure effects was desired here.

Consider sentence (6), which contains no complement clause in surface structure. This sentence is nearly synonymous with sentence (1) [The host expected Kate to be at breakfast]. Until recently, it was

(6) The host expected Kate at breakfast.

assumed that (6) and (1) contained identical underlying structures, with (6)

being derived by application of a rule which deletes to be. Under such an analysis, a comparison of sentences like (1) and (6) would provide the desired independent test of the role of surface vs. underlying structure as a determinant of syllable lengthening. Borkin (1973) has shown that a slight difference in meaning is usually associated with sentences (1) vs. (6), however, such that the version in which to be has been deleted has a greater tendency to denote personal experience on the part of the subject. Sentences (7) and (8) appear to bring out Borkin's point clearly.

(7) I find this chair to be uncomfortable. = Borkin's (10b)

(8) I find this chair uncomfortable. = Borkin's (10c)

Borkin notes that either (7) or (8) would be appropriate for a speaker who is reporting on his personal experience with the chair, whereas only (7) would be approrpiate for a speaker who is reporting the results of a consumer reaction test in which he himself did not have experience with the chair. Since both (7) and (8) can be used in the former circumstance, however, it is still maintainable that sentence pairs like (1)-(6) and (7)-(8) are derived from the same underlying structure on their most common reading. Assuming this analysis,[5] if syllable lengthening is controlled primarily by surface as opposed to underlying structure in the case of complements, lengthening should be observed for the verb expected in (1) vs. (6), since only in the former sentence does the underlying complement clause exist in surface structure. For (6), the subordinate sentence node in underlying structure is presumably deleted by the condition of S-node pruning (Ross, 1969; Hankamer, 1971; Reis, 1973), a convention which, according to its original formulation, deletes all S-nodes from surface structure

which do not branch into a verb phrase and some other constituent.

Unlike expect, the optional deletion of to be is not permitted for the complement of a verb like persuade, as evidenced by the major difference in meaning between a sentence like (2) and (7) below.

(7) The host persuaded Kate at breakfast.

Unlike (2) [The host persuaded Kate to be at breakfast], (7) is assumed to contain a single clause in underlying as well as surface structure. In addition to the surface structure contrast between expect sentences like (1) and (6), it was decided to include the contrast between persuade sentences like (2) and (7) in the present study. Because (2) and (7) differ greatly in their meaning as well as in their constituent structure at both underlying and surface levels, however, it was not possible to make any firm predictions about the effects of this comparison on syllable timing. The absence of prediction in this case resulted from a lack of prior knowledge about any effects that semantic representation might have on timing.

2. A Sentence-Reading Paradigm

In order to study a speaker's computational processes during speech production, it would be most desirable to study a corpus of spontaneous speech. While certain effects, including very general effects of syntactic structure on syllable timing, can indeed by studied in this manner (Kloker, 1975), work on specific hypotheses cannot yet be conducted with proper control or efficiency using such a data base. In the case of complement clauses, the need for a controlled experimental setting is particularly acute, since the unaided ear

suggests that, to the extent that syllable lengthening effects exist at all for complements, these effects must be quite small compared with the lengthening effects exhibited for some other clause types, such as coordinates and non-restrictive relatives. The small magnitude of the complement effects would presumably be impossible to ascertain from a corpus of spontaneous speech, since the interplay of other influences on syllable timing (particularly speaking rate and phonetic environment) would mask any systematic effects of complement structure.

Accordingly, a sentence-reading procedure was used in this and similar studies, in whcih the phonetic form of the utterances was tightly controlled. In each test, a short list of sentences was read by a speaker after a period of practice during which he was familiarized with the task and the particular sentence materials. Each sentence contained one or more key words, placed at the location(s) of putative syntactic boundaries. The choice of key words was influenced by the degree to which their phonetic representations could be readily segmented from a digitized oscillographic trace of the speech waveform. In most of the studies conducted with the procedure, the key words were CVC monosyllables containing a long vowel (Cooper, 1975a),[6] bounded by stop consonants or fricatives (e.g., Kate, tape, case). In some instances, including the present study, key words were required which were slightly more difficult to segment (e.g., expected, persuaded), although in each case the additional measurement error incurred with such words was small compared with the magnitude of the effects observed.

Wherever possible, the key words were placed in the middle of the sentence string in order to neutralize any possible influence of differences in subglottal pressure on segmental timing (Lindblom and

Rapp, 1973). The sentences themselves typically contained the same number of words and syllables and shared as many words as possible with one another, compatible with signalling the structural differences under study.

Each speaker was tested individually in a sound-insulated chamber for approximately 40 minutes per session. During this time, the speaker was typically tested on from 5 to 7 sentence lists, each representing a separate experiment. Each of the experiments with complements reported here was conducted as part of a different session. At the beginning of a session, the speaker was told that the general purpose of the experiments was to study the syntactic control of syllable timing. The speaker was then told that he would be given practice in reading lists of sentences in order to train him to utter each sentence "as a unitary whole, as if it were spoken spontaneously", rather than "word-by-word", as in unpracticed reading.[7] The speaker was encouraged to speak as naturally as possible but to avoid placing contrastive or emphatic stress on any word or syllable in a sentence.

Following these preliminary instructions, the speaker was given the first of a series of lists of sentences to practice until he figured out how he intended to utter each sentence in the list. The speaker was told to consider each sentence in the list independently of the others. Following practice, the speaker was asked to utter each sentence aloud once, providing the experimenter with a final opportunity to check for undesirable contrastive stress or emphasis and to check recording levels. The speaker then uttered each sentence in the list 6 times in succession for recording. The speaker was

asked to say each token of the sentence at the same overall rate and with the same rhythm.

The speaker was told to expect to produce a few mispronunciations during the test. For mispronunciations, or for any changes from the normal intonation or timing which the speaker chose to adopt, the speaker was instructed to pause, utter the word "repeat", and then say the sentence token again, as often as needed until 6 appropriate occurrences of the sentence were produced. After reading a given list, the speaker was provided a short rest period, encouraged to take a drink of water, and then asked to begin the practice procedure with a new list.

3. Experiment I

In this experiment, the durations of syllables in 5 sentences were measured to determine whether differences in the syntactic structure of complements following the verbs expected and persuaded would produce differences in syllable timing. The verb and the following noun of each sentence were chosen as the key words for measurement, since the locations of the major syntactic boundaries postulated according to both major linguistic analyses of complementation occurred immediately after one of these two words. The words immediately surrounding the verb in each sentence were held fixed, so that any durational effects observed for the verb could be attributed directly to syntactic structure. For the noun, the following word, of necessity, covaried with the syntactic structure of the sentence. For this reason, an independent test (Experiment II) was required to determine whether any effects of syllable duration observed for the noun in this

experiment were due to phonetic as opposed to syntactic influences.

3.1 Method

Subjects

Fifteen M.I.T. students and employees served as volunteers in the experiment. The subjects were native speakers of English with no history of speech or hearing impairment.

Sentence Materials

The sentences used in the experiment are listed below, along with their descriptive labels (written in capital letters):

(a) The host expected Kate to be at breakfast. (EXPECT-INF = expect infinitival complement)

(b) The host expected Kate would be at breakfast. (EXPECT-THAT = expect that complement)

(c) The host expected Kate at the big breakfast. (EXPECT-SIMPLE = expect single surface clause)

(d) The host persuaded Kate to be at breakfast. (PERSUADE-INF = persuade infinitival complement)

(e) The host persuaded Kate at the big breakfast. (PERSUADE-SIMPLE = persuade single surface clause)

Each of these 5 sentences contained 8 words and 11 syllables. When read without contrastive or emphatic stress, the sentences also had the same approximate sentence stress countour, with primary stress on Kate.[8]

Sentences (a) and (d) were equivalent to sentences (1) and (2) discussed in the Introduction. Sentences (c) and (e) had the same structure

as (6) and (7), while sentence (b) represented a version of (5) in which the that complementizer had been deleted. The complementizer was deleted in (b) so that the immediate phonetic environment of the verb would be identical for all 5 sentences.

Procedure

The testing procedure was described in Section 2. Each speaker read each sentence in the list (a)-(e) 6 times, beginning with sentence (a) or (e), according to a randomized assignment. The first 5 occurrences of each sentence, excluding false starts, mispronunciations, and sentences containing contrastive or emphatic stress, were digitized at a sampling rate of 10 kHz on a PDP-9 computer at the M.I.T. Research Laboratory of Electronics. The waveforms were analyzed for segment durations with the aid of a computer controlled cursor (Huggins, 1969). This cursor was maneuvered by velocity and position dials to mark the onset and offset of the bisyllabic segments [spɛktəd] and [swedəd] of the verbs expected and persuaded as well as of the monosyllable [ket] (Kate). The time difference between the segment onset and offset was displayed on an oscilloscope screen to the nearest 100 μsec.

The onset of visible frication following the closure gap of the first [k] in expected was taken as the onset of the final two syllables of this word.[9] For some speakers, a [k] release burst was discernible in the waveform and in such cases this burst was not included as part of the measured bisyllabic segment, since including the burst duration would have introduced greater variability into the data.

The onset of visible frication following the [r] in persuaded was similarly taken as the onset of the bisyllabic segment for this verb. The offset of the bisyllabic segments for both verbs was measured to be the termination of visible glottal pulsing in the unstressed syllable [ə d], not including the following closure interval or any [d] release burst. As above, the decision not to include portions of the waveform as part of the bisyllabic segment was based upon the undesirability of increasing the measurement variability.

The reliability of each measurement of the verb segments was estimated to be within ± 3 msec, based on a duplicate set of measurements for a small sample of 20 utterances made without the experimenter's remembrance of, or reference to, the original measurements taken for these utterances. All waveform measurements were made by the author.

The onset of the monosyllable Kate was measured at the release burst of [k], and the offset of the syllable was measured at the termination of visible glottal pulsing. The onset release burst was included in the measurements of Kate because this burst was clearly discernible in the waveforms of all speakers and did not increase the variability of the data appreciably. The closure interval of the [t] or any [t] offset burst was not included as part of the duration of Kate. The reliability of each measurement for Kate was estimated to be within ± 2 msec, based on duplicate measurements of the same utterances cited above.

3.2 Results and Discussion

Verb Duration

The mean durations of the verb segments for expected and persuaded are presented for the individual speakers in Table II. Two-tailed

---

Insert Table II about here

---

t-tests for correlated observations were applied to the mean durations for the set of speakers to determine whether any differences in these durations were statistically significant. The analysis for the bisyllabic segment of the verb expected showed that the segment in both Sentence (a) EXPECT-INF and Sentence (b) EXPECT-THAT were significantly longer than the segment in Sentence (c) EXPECT-SIMPLE (EXPECT-INF vs. EXPECT-SIMPLE: $t = 2.675$, $df = 14$, $p < .01$; EXPECT-THAT vs. EXPECT-SIMPLE: $t = 2.487$, $df = 14$, $p < .02$). The average duration of the bisyllabic segment for EXPECT-INF was 13.3 msec longer than for EXPECT-SIMPLE, while the average duration of the segment for EXPECT-THAT was 9.8 msec longer than for EXPECT-SIMPLE.

The significant lengthening of the verb segment for the sentences (a) and (b) containing surface complement clauses vs. sentence (c), which had the same meaning as the other two sentences and presumably the same underlying syntactic structure as well, suggests that complement clause boundaries in surface structure are among the types of syntactic boundaries which help to determine segmental lengthening. As expected, however, the lengthening observed for surface complements was small compared with the effects observed at the boundaries of

other clause types such as coordinates and conditionals (Cooper, (1975a).

No significant difference was found between the duration of the verb segment for expected between the two complement types, EXPECT-INF and EXPECT-THAT ($t = 0.856$, $df = 14$, $p > .20$). The average duration of the segment in EXPECT-INF was longer than the duration in EXPECT-THAT by 3.3 msec. According to the Raising analysis of complementation, whereby a major clause boundary exists immediately after expected for EXPECT-THAT but not EXPECT-INF, the duration of the verb segment for EXPECT-THAT should have been longer. Since the data in fact show a slight trend in the opposite direction, the results provide no support for a syntactic level of computation in speech production that corresponds to the surface representations proposed by the Raising analysis. This finding is paralleled in speech perception by a recent study of click location in which a test for Raising also failed to show support for such an analysis (Fodor, Fodor, Garrett, and Lackner, 1975).

The lack of a significant difference between EXPECT-INF and EXPECT-THAT is, on the other hand, consistent with the analysis of complement clause structure proposed by Chomsky (1973), whereby infinitival and that complements have the same surface as well as underlying constituent structure. However, the present data cannot be taken as very strong support for a performance analog of Chomsky's analysis, given the possibility of a Type II statistical error.

We now turn to consider the bisyllabic segment durations for the verb persuaded. These durations were generally about 10-20 msec shorter than the durations of the corresponding segment of the verb expected.[10]

The average duration of the verb segment in PERSUADE-INF was 3.3 msec longer than the average duration of the segment in PERSUADE-SIMPLE. The difference in duration for the set of subjects was not statistically significant ($t = 1.149$, $df = 14$, $p > .20$). Acceptance of the null hypothesis here is consistent with the predictions of both major analyses of complementation, since no major clause boundary occurred immediately after persuaded in either the underlying or surface structure of these two sentences.

We conclude the discussion of the verb duration data by noting that, according to either of the linguistic analysis of complementation, it was anticipated that some difference in the duration of the EXPECT sentences should have occurred, whereas no such differences should have occurred for the PERSUADE sentences. In fact, the only statistically significant effect obtained in the experiment was for the EXPECT sentences, providing support for the notion that syllable timing is conditioned in part by complement clause structure in a manner corresponding at a general level to a syntactic clause analysis.

The difference obtained between EXPECT-INF and EXPECT-THAT vs. EXPECT-SIMPLE provides some evidence that the site of syntactic computation which controls syllable timing computes a clausal representation corresponding approximately to linguistic surface as opposed to underlying structure. Since a valid test of the Raising account of the possible difference between EXPECT-INF and EXPECT-THAT rested on the assumption that syllable lengthening was controlled in part at a surface as opposed to underlying level of representation, and since this assumption received some independent support, the absence of a difference between infinitival and that complements predicted by

the Raising analysis in this experiment must be considered problematic for the view that speakers generally compute a syntactic representation of complements corresponding to such an analysis.

Noun Duration

The results of the duration measurements for the noun Kate are presented in Table III. Unlike the durations for the verb

---

Insert Table III about here

---

segments, the data for the noun can be compared for a phonetically fixed environment in the case of sentences (a) vs. (d) only, the infinitival complements for expected and persuaded. In both sentences, Kate was preceded by the unstressed syllable [ɚd] and was followed by the infinitive to. The results showed that the average duration of Kate for PERSUADE-INF was 2.5 msec longer than the average duration of Kate for EXPECT-INF. The difference in duration between these two test conditions was not statistically significant ($t = 1.134$, $df = 14$, $p > .20$). The slight trend for the noun following persuaded to be longer than the noun following expected is consistent with the notion that speakers compute the structural representations of the two complements according to Chomsky's (1973) account of infinitivals, whereby a major clause boundary exists immediately after the noun for persuaded but after the verb for expected. On the other hand, the absence of a statsitically significant difference between the two complement types is also consistent with a Raising analysis, whereby a

major clause boundary exists immediately after the verb for both persuaded and expected. In summary, then, the data for the noun durations of EXPECT-INF vs. PERSUADE-INF provide no independent basis for distinguishing the merits of the Chomsky and Rosenbaum-Postal proposals as models of speech performance.

In addition to the single comparison between phonetically stable nouns in EXPECT-INF vs. PERSUADE-INF, other comparisons were made for the noun duration data. These comparisons revealed some significant differences, potentially attributable to the effects of the immediate phonetic environment. In particular, the durations of Kate before at in EXPECT-SIMPLE and PERSUADE-SIMPLE were significantly longer than the durations of Kate before to in EXPECT-INF and PERSUADE-INF and longer than the duration of Kate before would in EXPECT-THAT ($p < .05$ in each case). The average duration of Kate before at in the SIMPLE sentences was about 15 msec longer than before to and would in the complements.

4. Experiment II

It is known that the duration of a vowel is longer when it is followed immediately by a voiced vs. voiceless phonetic segment occurring within the same word (Peterson and Lehiste, 1960; House, 1961; Delattre, 1966). In addition, Barnwell (1971) has shown that this effect of phonetic environment is stronger within a word than across a word boundary, although his data base was not sufficiently large, nor were his sentence materials sufficiently well-matched, to test the possibility that the phonetic effect does operate across a

word boundary to some extent. The data for the noun Kate in Experiment I, regarding the difference between simple and complement sentences, could be accounted for by a similar effect, as opposed to some unexpected difference in syntactic structure. Of necessity, the structural differences of interest in Experiment I made it impossible to control for the phonetic environment of the noun Kate, unlike the verb. In this experiment, an independent test was thus carried out to examine the possibility of a phonetic effect across word boundaries of the type that would directly account for the results for Kate in the previous experiment.

4.1 Method

Subjects

Ten subjects, one of whom (I.B.) served in Experiment I, participated as volunteers in this experiment. All new subjects had the same qualifications as the subjects of Experiment I.

Sentence Materials

The following three sentences were used in the experiment.

(a) We skate to the farm.

(b) We skate with the crowd.

(c) We skate at the pond.

Each of these sentences contained 5 monosyllabic words and the same approximate sentence stress countour in non-emphatic speech. In addition, the underlying and surface structure representations of the sentences were identical with regard to major constituent relations,

consisting of a subject noun phrase and a verb phrase dominating a verb and a prepositional noun phrase.

Each sentence contained the key word skate, immediately followed by a word beginning with [t], [w], or [æ], These three segments corresponded to the three segments immediately following the word Kate in Experiment I. It was decided to use the above sentences for the test rather than simple phrase like Kate to, Kate would, and Kate at (taken directly from the sentences of Experiment I) so as to preserve a sentence context.

Procedure

The procedure for testing described previously was used for this experiment. Each speaker read each sentence in the list (a) through (c) 6 times, beginning with sentence (a) or (c), according to a randomized assignment. The durations of the monosyllabic words we and skate were measured using the same general procedure as Experiment I.

4.2 Results and Discussion

The mean durations for the words we and skate are presented in Table IV. The average durations of we in the three sentences were

---

Insert Table IV about here

---

all within 3.5 msec of one another, and the differences among these durations for the set of subjects were not statistically significant

($p > .20$ in each case). These results are consistent with the notion that the duration of a word is not significantly influenced by differences in the phonetic structure of another word that is two words removed from it.

The average duration of skate in the three sentences showed systematic differences. The word skate was longest preceding at, somewhat shorter preceding with, and shortest preceding to, covering an average range of more than 24 msec. Statistical tests showed that the duration of skate before at was significantly longer than the duration of skate before to ($t = 3.683$, $df = 9$, $p < .01$) and that the duration of skate before with was also significantly longer than the duration of skate before to ($t = 7.043$, $df = 9$, $p < .001$). On the other hand, the duration of skate before at was not significantly longer than the duration of skate before with ($t = 1.246$, $df = 9$, $p > .20$).

The results of this experiment indicate that the duration of skate is significantly longer when the following word begins with either of the voiced segments [æ] or [w], in comparison with the voiceless segment [t]. This pattern of results demonstrates that the conditioning effect of following voiced vs. voiceless segments extends across a word boundary, and, furthermore, that it extends across a relatively major surface phrase boundary separating the verb and prepositional noun phrase. The magnitude of this effect is the same order of magnitude as the effects observed in Experiment I. We can thus conclude that most, if not all, of the segmental lengthening observed for that noun was attributable to the phonetic structure of the following segment rather than to the distinction between simple and complement sentences.[11]

5. Post-Hoc Analysis of Individual Speakers' Data For Experiment I

Based on the results of Experiment II, a post-hoc analysis was carried out on the data of Experiment I to determine whether any individual speakers showed a pattern of results that was strikingly consistent with either a Raising or Tensed-S analysis of complementation. If a speaker computed syntactic representations for infinitival vs. *that* complements according to a Raising analysis, then the duration of the verb segment of *expected* should have been longer for EXPECT-THAT than for EXPECT-INF. In addition, based on a consideration of the results of Experiment II, the duration of the noun following *expected* in EXPECT-INF should have been longer than the duration of the noun in EXPECT-THAT, whereas the duration of the noun in EXPECT-INF should have been about equal to the duration of the noun in PERSUADE-INF. Two of the 15 speakers of Experiment I, R.F. and M.J., showed this pattern of results. Thus, while a Raising analysis appears incapable of accounting for the results of the speakers as a group, it is possible that this analysis is represented as a level of speech computation for two of the 15 speakers. Further research is required to test the possibility that the particular pattern of results obtained for these two speakers was not coincidental.

A similar analysis of individual speakers' data was conducted in search of speakers who showed a pattern of results corresponding to Chomsky's Tensed-S proposal. As noted earlier, speakers who computed a syntactic representation of infinitival and *that* complements according to the Tensed-S account should have produced approximately

equal durations for the verb segment of expected in EXPECT-INF and EXPECT-THAT. Furthermore, taking into account the results of Experiment II, the speakers should have produced a noun duration for EXPECT-THAT that was longer than for PERSUADE-INF, while the noun durations for EXPECT-THAT and PERSUADE-INF should have been about equal. None of the individual speakers of Experiment I showed this particular pattern of results. It must thus be concluded that the post-hoc analysis of individual speakers' data provides no support for the notion that speakers represent the structure of complements in a manner like that proposed under a Tensed-S analysis and only very marginal support for the notion that some speakers represent the complements in a manner corresponding to a Raising analysis.

6. Experiment III

The single result thus far providing any strong support for the notion that a surface structure representation of complements affects syllable timing was the significant lengthening of the verb segment in Experiment I for the complements EXPECT-INF and EXPECT-THAT, in comparison with the duration of EXPECT-SIMPLE. To test whether this difference reflected an improbable chance result or some idiosyncratic property of the verb expected, it was decided to conduct a third experiment, similar to Experiment I but using different verbs.

Unlike the verb expect, there exist some verbs which trigger the same complement structures as expect but which can also occur as main verbs in sentences which contain a single underlying clause.

Believe is such a verb, and it was used in this experiment in both single-clause and complement sentences.

6.1 Method

Subjects

Seven M.I.T. students and employees, 5 of whom served in Experiment I, participated in this experiment. The two new subjects had the same qualifications as the subjects of previous experiments.

Sentence Materials

The sentences used in this experiment were divided into three categories, Expect-type 1, Expect-type 2, and Persuade-type 1. The Expect-type 1 sentences contained comparisons between infinitival complements and single clause structures from which to be had presumably been deleted (cf. Borkin, 1973). Four different verbs were used, each triggering complements in the same manner as expect with regard to the major constituency relations of relevance (Bresnan, 1972). The resulting list of 8 sentences appears below:

Expect-type 1

(a) We believed Kate to be crazy.

(b) We believed Kate crazy at times.

(c) We considered Kate to be crazy.

(d) We considered Kate crazy at times.

(e) The boss wants Ted to be at the station by 3 o'clock.

(f) The boss wants Ted at the old train station by 3 o'clock.

(g) The boss needs Ted to be at the station by 3 o'clock.

(h) The boss needs Ted at the old train station by 3 o'clock.

The list of Expect-type 2 sentences included three-way comparisons among infinitival complements, that complements, and single clause structures considered to contain a single clause in underlying as well as surface structure. Three different verbs triggering complements like expect were used, making a total of 9 sentences listed below.

Expect-type 2

(i) Kate believed John to be at the trial.

(j) Kate believed John was at the trial.

(k) Kate believed John at the trial last week.

(l) Kate understood John to be at the trial.

(m) Kate understood John was at the last trial.

(n) Kate understood John at the trial last week.

(o) John proved Bayes' Theorem to be applicable to my sampling problem.

(p) John proved Bayes' Theorem was applicable to my sampling procedure.

(q) John proved Bayes' Theorem at the conference on statistical procedures.

Finally, the list of Persuade-type 1 sentences included two-way comparisons between infinitival complements and single-clause sentences considered to contain a single clause in both underlying and surface structure. These sentences used two verbs which trigger complements like persuade (Bresnan, 1972).

Persuade-type 1

(r) We convinced Kate to be at breakfast.

(s) We convinced Kate at the big breakfast.

(t) Ted challenged Kate to be at breakfast.

(u) Ted challenged Kate at the big breakfast.

In summary, the Expect-type 1 and Persuade-type 1 sentences were comparable to sentences used for expect and persuade in Experiment I. The Expect-type II sentences, however, contained verbs of the expect type in single clause sentences of the same kind as the sentences with persuade type verbs.

Procedure

The same testing procedure was used as in previous experiments. The sentences appeared on three separate sentence lists. The durations of the verbs were measured using the general technique described earlier. Unlike Experiment I, the duration of the entire verb segment, excluding terminal gaps and offset bursts, were measured.

6.2 Results and Discussion

Expect-type 1

For the 4 Expect-type 1 verbs, 3 verbs showed an average lengthening effect in the infinitival complement vs. single-clause sentences for the set of 7 speakers. The average lengthening effect for believe was 20.1 msec, for want 5.1 msec, and for need only 0.2 msec. For the verb consider, an average shortening of 7.9 msec was obtained, although, unlike the date for the other verbs, the average effect for

consider was heavily influenced by the effect for a single speaker. We conclude from these results that the effect observed in Experiment I generalizes to some extent to other verbs in the context of synonymous complement vs. single-clause sentences, although a much larger data base would be required to test generality in a definitive manner.[12]

Expect-type 2

In contrast to the trend observed for the data of Expect-type 1 sentences, the duration of the main verb for Expect-type 2 sentences was either shorter for the complement than single-clause structures or was about equal. For the verb believe, which showed a sizable lengthening effect for the complement in Expect-type 1 sentences, an equally sizable lengthening effect was observed for the single-clause sentence when this verb occurred in an Expect-type 2 context. The average duration of the verb in the single-clause sentence for believe in the latter context was 27.0 msec longer than the duration of the verb with an infinitival complement and 27.4 msec longer than the duration of the verb with a that complement. An average lengthening effect for the single-clause sentence was also observed for understood, amounting to slightly more than 12 msec in comparison to each complement. For the verb proved, however, the verb of the single-clause was shorter than the verb in the infinitival complement by 11.3 msec and was longer than the verb in the that complement by 3.4 msec. In summary, it appears that the verb duration data for Expect-type 2 sentences differs from that of Expect-type 1 sentences in the predicted direction. The lengthening for the single-clause sentences of Expect-type 2 is unaccountable in terms of clausal analysis but may

well reflect semantic differences of focus that will need to be studied further. Since the present data base is small and subject to idiosyncrasies, we will not pursue a discussion of the possible significance of the single-clause lengthening of Expect-type 2 sentences here.[13]

Persuade-type 1

The data for the two Persuade-type 1 verbs, convinced and challenged, showed approximately equal average durations for the single-clause and complement sentences. The average difference in duration was within 2 msec for both verbs. These results indicate that the results for persuaded in Experiment I generalize to these other verbs of the same structural classification.

7. General Discussion

The present experiments have provided evidence for the existence of both syntactic and phonetic effects on syllable timing. At a syntactic level, speakers lengthened the duration of the last two syllables of the verb expected when this verb was followed by a complement clause in the surface structure of the sentence, as compared with when it was followed by a simple phrase in surface structure which, according to one linguistic analysis, was derived from a full complement in underlying structure. This finding suggested that a speaker's surface structure representation of an utterance is the primary level of syntactic representation that exercises control over syllable timing.

The effect observed for full vs. reduced complements is noteworthy also because the verb in both sentences occurred two words prior to the point in the sentences at which the distinction between the surface structures became apparent. This fact suggests that the speakers in this study were computing the durations of syllables in part as a function of the hierarchical constituent structure of the sentences rather than on a linear word-by-word basis.

The major negative finding of the present study was the failure to find evidence in strong support of a system of syntactic computation that corresponded to either the Raising or Tensed-S analyses of complementation. As in the case of previous work in psycholinguistics, it is perhaps not surprising that this study failed to show a close correspondence between the kinds of structural analyses employed by a speaker during sentence production and the kinds of analysis proposed on independent linguistic grounds (see discussion in Fodor and Garrett, 1966 and in Fodor, Bever, and Garrett, 1974). Although it remains a reasonable research strategy to test the relevance of linguistic theory to a performance model of sentence production, greater reliance must be placed in constructing a model of performance based primarily upon available psycholinguistic data and other psychological considerations. The main problem facing this approach is that the data required are for the most part unavailable in order to verify even the most basic of assumptions needed as a foundation for further hypothesis testing.

Aside from the long-range goal of providing a model of the computational processes involved in speech production, the results of studies like the present serve an immediate function as a guide for

research on speech syntehsis by rule and on the perception of segmental timing. The present results indicate that a speech synthesis by rule program can comfortably avoid any specialized rules for complement clause boundaries, since the average effects are comparable to the perceptual tolerance of listeners (Huggins, 1972; Klatt and Cooper, 1975). At most, a 4-5% lengthening of syllables in the verb preceding a complement clause boundary might be implemented. Since it appears unlikely that the lengthening found in speech production could be detected reliably by listeners in normal speech situations, there is no need to conduct a systematic study of the perceptual relevance of segmental lengthening with complements, unlike the case for some other types of lengthening which are syntactically determined (Cooper, 1975a).

It is tempting, finally, to speculate on the precise cause of the syllable lengthening produced by the presence of a surface clause boundary. Perhaps, if speakers compute articulatory commands on a clause-by-clause basis to some degree, lengthening is produced to provide the speaker with an extra fraction of time during which to plan the articulatory commands for the upcoming clause. On the other hand, it is conceivable that speakers lengthen syllables to provide a slightly longer interval over which to place a rise or fall in intonation associated with the clausal structure. This possibility is not considered very likely, but a systematic study of fundamental frequency contours remains to be conducted.

Acknowledgements

This article is a version of a chapter of my Ph.D. dissertation (Cooper, 1975a). The work was supported by a National Science Foundation Graduate Fellowship and NIH Grant HD-05168. I am irretrievably indebted to Merrill Garrett for guidance throughout the course of this work. In addition, I thank Dumont Billings, Ann Cutler, Jerry Fodor, A. W. F. Huggins, Dennis Klatt, Steven Lapointe, Cornelia Parkes, and John Robert Ross for advice and assistance.

References

Akmajian, A. and Jackendoff, R.S. (1970). Coreferentiality and stress. Linguistic Inquiry 1, 124-126.

Barnwell, T.P. (1971). An algorithm for segment durations in a reading machine context. Technical Report No. 479. Research Laboratory of Electronics, M.I.T., Cambridge, Ma.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). Language 48, 633-44.

Borkin, A. (1973). To be and not to be. In Papers from the Ninth Regional Meeting of the Chicago Linguistic Society (C. Corum, T.C. Smith-Stark, and A. Weiser, Eds.). Chicago: Chicago Linguistic Society. Pp. 44-56.

Bresnan, J. (1971). Sentence stress and syntactic transformations. Language 47, 257-81.

Bresnan, J. (1972). The Theory of Complementation in English Syntax. Ph.D. Dissertation, M.I.T., Cambridge, Ma.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge: MIT Press.

Chomsky, N. (1973). Conditions on transformations. In A Festschrift for Morris Halle (S.R. Anderson and P. Kiparsky, Eds.). New York: Holt, Rinehart, and Winston. Pp. 232-285.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior 12, 335-359.

Coker, C.H., Umeda, N., and Browman, C.P. (1973). Automatic synthesis from ordinary English text. IEEE Audio and Electroacoustics AU-21, 293-7.

Cooper, W.E. (1975a). Syntactic Control of Timing in Speech Production. Ph.D. Dissertation, M.I.T., Cambridge, Ma.

Cooper, W.E. (1975b). Why reordering transformations? a review of Paul M. Postal's On Raising. Forthcoming.

Cooper, W.E. (1975c). Inclusion rleations in language and perception. Forthcoming.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. International Review of Applied Linguistics 4, 183-98.

Fodor, J.A., Bever, T.G., and Garrett, M.F. (1974). The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. New York: McGraw-Hill.

Fodor, J.A., Fodor, J.D., Garrett, M.F., and Lackner, J.R. (1975). Effects of surface and underlying clausal structure on click location. Quarterly Progress Report of the M.I.T. Research Laboratory of Electronics, in press.

Fodor, J.A. and Garrett, M.F. (1966). Some reflections on competence and performance. In Psycholinguistic Papers (J. Lyons and R.J. Wales, Eds.). Edinburgh: Edinburgh University Press. Pp. 135-179.

Garrett, M.F. (1975). The analysis of sentence production. In Advances in Learning Theory and Motivation, Vol. 9 (G. Bower, Ed.). New York: Academic Press, in press.

House, A. (1961). On vowel duration in English. Journal of the Acoustical Society of America 33, 1174-1178.

Huggins, A.W.F. (1969). A facility for studying perception of timing in natural speech. Quarterly Progress Report of the M.I.T. Research Laboratory of Electronics 95, 81-83.

Huggins, A.W.F. (1972). Just-noticable difference for segment duration in natural speech. Journal of the Acoustical Society of America 51, 1270-1278.

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

Klatt, D.H. and Cooper, W.E. (1975) Perception of segment duration in sentence contexts. In Structure and Process in Speech Perception (A. Cohen and S.G. Nooteboom, Eds.). Heidelberg: Springer-Verlag.

Huggins, A.W.F. (1974). An effect of syntax on syllable timing. Quarterly Progress Report of the M.I.T. Research Laboratory of Electronics 114, 179-185.

Kloker, D. (1975). Vowel and sonorant lengthening as cues to phonological phrase boundaries. Paper presented at the 89th Meeting of the Acoustical Society of America, April, 1975.

Lieberman, P. (1967). Intonation, Perception, and Language. Cambridge: MIT Press.

Lindlbom, B. and Rapp, K. (1973). Some temporal regularities of spoken Swedish. Papers from the Institute of Linguistics, University of Stockholm, Publication 21.

Martin, J.G. (1970). On judging pauses in spontaneous speech. Journal of Verbal Learning and Verbal Behavior 9, 75-8.

Peterson, G.E. and Lehiste, I. (1960). Duration of syllabic nuclei in English. Journal of the Acoustical Society of America 32, 693-703.

Postal, P.M. (1974). On Raising: One Rule of English Grammar and Its Theoretical Implications. Cambridge: MIT Press.

Reis, M. (1973). Is there a rule of subject-to-object raising in German? In Papers from the Ninth Regional Meeting of the Chicago Linguistic Society (C. Coru, T.C. Smith-Stark, and A. Weiser, Eds.). Chicago: Chicago Linguistic Society.

Rosenbaum, P.S. (1967). The Grammar of English Predicate Complement Constructions. Cambridge: MIT Press.

Ross, J.R. (1969). A proposed rule of tree-pruning. In Modern Studies in English (D.A. Reibel and S.A. Schane, Eds.). Englewood Cliffs, N.J.: Prentice-Hall. Pp. 288-99.

Footnotes

[1]Syllable lengthening has also been found at the beginnings of major phrase boundaries in Swedish speech (Lindblom and Rapp, 1973). However, this effect is small compared with the effect at the ends of major phrase boundaries in Swedish, and the phrase-initial effect has not been found for English (Klatt, 1975). The possibility of a phrase-initial or clause-initial lengthening cannot be ruled out for English, but for the arguments presented in this paper it is necessary to assume only that such effects, if they exist, are small compared with the effects in clause-final position.

[2]This rule was originally included under the rule It-Replacement (Rosenbaum, 1967). The rule must be kept distinct from a similar rule proposed to raise a subordinate subject into superordinate subject position. The latter rule converts the underlying structure of sentences like It appeared that John was sick into John appeared to be sick, with John as the raised element.

[3]Under an alternative possibility, (1') would contain a single surface clause, with the sentence node dominating be at breakfast deleted by the convention of S-node Pruning (Ross, 1969; Hankamer, 1971; Reis, 1973). However, it appears that a Raising analysis, which preserves a two-clause structure by selectively raising the subordinate subject, is sufficient without pruning in English (see Reis, 1973). If, conversely, S-node pruning is assumed, then the major result of Experiment I to be discussed seems to have occurred at a level of processing prior to the level at which pruning occurs.

[4]An alternative to the Equi-NP analysis has recently been proposed by Postal (1974). However, the location of the major clause boundary under his analysis remains the same as in the Equi analysis.

[5]Although this analysis is assumed throughout the rest of the text, the linguistic evidence favoring at least one aspect of the analysis is less than compelling. John Robert Ross (personal communication) has pointed out that while the underlying representation of (6) may in fact take an underlying complement (the critical assumption for present purposes), the underlying complement verb may be a verb of motion rather than to be. A main verb like want, used in Experiment III, does appear to require an underlying to be in sentences like (6), however. Ross cites as syntactic evidence of an underlying to be for want complements the existence of idioms like I want headway made on this by Friday and I want tabs kept on John. The argument applied to headway goes as follows: (a) terms like headway can occur in underlying object position but not underlying subject position (*Headway is fun); (b) in order to derive a sentence like I want headway made on this by Friday, it is necessary to postulate a rule moving headway to the left of make; (c) the relevant rule, Passive, requires to be; (d) thus, the structure of I want headway made on this by Friday must include an underlying to be which is present when Passive applies. For my speech, the same argument applies to the main verb expect, lending support for an underlying to be for the complement of this main verb also.

[6]Klatt (1975) has noted that the modifications of vowel length are more pronounced for vowels with inherently long durations.

[7]Speakers were not, however, asked to close their eyes while they spoke.

[8]Although speakers did not place contrastive stress on any syllables in the sentences, some difference in stress pattern existed between the simple sentences (c) and (e) and the other sentences because of the presence of the adjective big in the simple sentences. This adjective was inserted to control for total number of words across sentences. An account of the present results in terms of rhythmic groupings of stressed syllables, however, would run counter to the stress timing prediction needed to account for the results of the second experiment in Chapter 2 (see especially Chapter 2, footnote 9). Thus, in the absence of a more refined theory of stress timing, the syntactic account to be proposed here should be favored.

[9]This and similar locations in the waveform differed to some extent depending on the recorded amplitude of the speaker's voice. An effort was made to keep the overall amplitude the same across speakers, but in cases where the amplitude was perceptibly lower for a given speaker, the amplitude display of each of the speaker's waveforms was doubled to facilitate marking of the segment boundaries. Since all comparisons of interest were across sentences within speakers, the doubling of amplitude for a few speakers allowed for a more reliable segment measuring procedure without introducing experimenter bias.

[10]Recall that Chomsky predicts a difference in clause boundaries between the infinitival complements of expect and persuade, whereas Rosenbaum and Postal do not (see Table 1, surface structure predictions). Assuming that the typical duration of [swedəd] is longer than that of

[spɛktəd] on purely phonetic grounds (an assumption that is plausible but has not yet been tested), the finding of longer durations for [spɛktə̀d] must be accounted for on syntactic grounds and would provide a strong point in favor of a performance analog of Chomsky's analysis, since on his account [spɛctəd] is clause-final, whereas [swedəd] is not. The difference in verb durations between [spɛktəd] and [swedəd] for simple sentences could also be accounted for on syntactic grounds, by phrase-final lengthening, according to both Chomsky and Rosenbaum-Postal.

[11]The only difference found for the noun Kate in Experiment I which is not easily accounted for by the phonetic factor demonstrated in Experiment III was the significant difference between EXPECT-THAT and EXPECT-SIMPLE, for which the duration of Kate was 13.0 msec longer before at than before would. In Experiment III, a difference for skate before [a] vs. [w] (at vs. with) in the same direction was observed, averaging 8.0 msec, but this difference failed to reach statistical significance with the relatively small N = 10.

[12]Since a major clause boundary existed in the complement sentences of Expect-type 2, an account of the single-clause lengthening for believed and understood must include some reference to a non-syntactic factor (e.g., focus) whose effectiveness overrides the syntactic clause effect. Although some study of the relation between sentence stress and focus has been conducted within a transformational framework (see Akmajian and Jackendoff, 1970; Bresnan, 1971), work on this general topic has not proceeded far enough to provide any good hints about the precise account of the present cases of single-clause lengthening.

[13]Clark (1973) has suggested the use of variance statistics like $\underline{\min}\ \underline{F}'$, where $\underline{\min}\ \underline{F}'(\underline{i},\underline{j}) = \underline{F}_1\underline{F}_2/(\underline{F}_1 + \underline{F}_2)$, for testing the generality of an effect across different sentence materials and different speakers using a design similar to that of Experiment III. However, because of the lack of a large data base in this experiment, it was decided to limit the analysis to computing the average durations across speakers for the sentences of each verb.

Table I

Syntactic Clause Boundaries Predicted on the Basis of Two Linguistic Analyses[a]

Linguistic Analysis	Complement Type		
Rosenbaum-Postal	expect to	expect that	persuade to
a. underlying structure	expect$_{\wedge}$	expect$_{\wedge}$	Kate$_{\wedge}$
b. surface structure	Kate$_{\wedge}$	expect$_{\wedge}$	Kate$_{\wedge}$
Chomsky			
a. underlying structure	expect$_{\wedge}$	expect$_{\wedge}$	Kate$_{\wedge}$
b. surface structure	expect$_{\wedge}$	expect$_{\wedge}$	Kate$_{\wedge}$

[a]Based on the sentences The host expected Kate to be at breakfast, The host expected that Kate would be at breakfast, and The host persuaded Kate to be at breakfast.

Table II

Mean Durations of the Last Two Syllables of the Main Verbs in Experiment I

	Sentence				
Subject	(a)	(b)	(c)	(d)	(e)
I.B.	407.6	402.4	396.2	391.8	392.4
R.B.	326.5	317.6	310.2	305.3	309.0
L.C.	376.2	363.2	351.7	371.6	328.3
B.F.	415.6	381.2	367.8	380.1	432.7
P.F.	358.3	335.9	330.9	342.2	325.8
R.F.	378.7	399.0	368.5	349.0	325.5
J.G.	381.3	384.5	363.5	354.8	328.1
R.G.	422.3	435.0	411.7	349.7	358.5
L.H.	366.3	365.6	366.9	368.8	343.1
M.J.	357.9	378.0	341.3	349.2	323.1
S.K.	353.5	353.5	355.5	314.3	364.2
B.P.	350.4	349.8	352.8	345.0	350.5
J.P.	369.7	362.4	370.9	357.8	350.1
R.T.	339.6	317.3	335.1	316.7	334.8
P.V.	342.9	348.9	323.8	326.5	306.9
Grand Mean	369.8	366.3	356.5	348.2	344.9

(a) = EXPECT-INF

(b) = EXPECT-THAT

(c) = EXPECT-SIMPLE

(d) = PERSUADE-INF

(e) = PERSUADE-SIMPLE

Table III

Mean Durations for the noun Kate in Experiment I

	Sentence				
Subject	(a)	(b)	(c)	(d)	(e)
I.B.	194.5	210.9	208.4	192.7	190.4
R.B.	169.7	163.1	181.5	161.8	167.3
L.C.	171.9	184.8	178.6	159.2	187.3
B.F.	196.0	214.0	227.6	211.0	261.4
P.F.	212.4	211.9	228.1	223.2	244.5
R.F.	156.8	149.2	214.3	160.5	179.8
J.G.	168.6	164.3	200.5	175.3	184.7
R.G.	177.2	194.9	194.0	187.5	223.1
L.H.	168.7	167.5	172.4	162.1	158.2
M.J.	191.9	187.9	176.2	180.8	179.1
S.K.	172.4	174.9	194.7	182.6	209.3
B.P.	161.0	164.1	197.0	172.8	210.9
J.P.	157.7	163.9	153.1	161.6	147.9
R.T.	143.3	149.7	172.5	143.1	167.7
P.V.	167.2	172.8	170.9	172.9	177.5
Grand Mean	174.0	178.3	191.3	176.5	192.6

(a) = EXPECT-INF

(b) = EXPECT-THAT

(c) = EXPECT-SIMPLE

(d) = PERSUADE-INF

(e) = PERSUADE-SIMPLE

Table IV

Mean Durations of the Pronoun *we* and the Verb *skate* in Experiment II

	Sentence			Sentence		
	(a)	(b)	(c)	(a)	(b)	(c)
Subject		*we*			*skate*	
A.B.	165.1	175.2	162.6	391.0	419.1	401.9
D.B.	99.6	57.2	63.9	283.9	301.1	324.4
I.B.	127.6	135.8	129.1	366.1	384.5	403.9
J.B.	127.5	96.3	91.1	290.3	308.2	281.3
M.I.	61.9	66.5	70.7	254.9	267.1	274.1
C.K.	99.1	108.0	126.8	361.7	365.6	363.3
J.L.	59.7	96.7	83.0	282.2	293.5	325.7
S.L.	98.6	91.3	61.3	295.5	316.1	313.9
C.P.	128.3	115.5	144.3	390.4	399.2	411.3
J.P.	77.5	77.2	77.5	395.1	419.3	454.6
Grand Mean	104.5	102.0	101.0	331.1	347.4	355.4

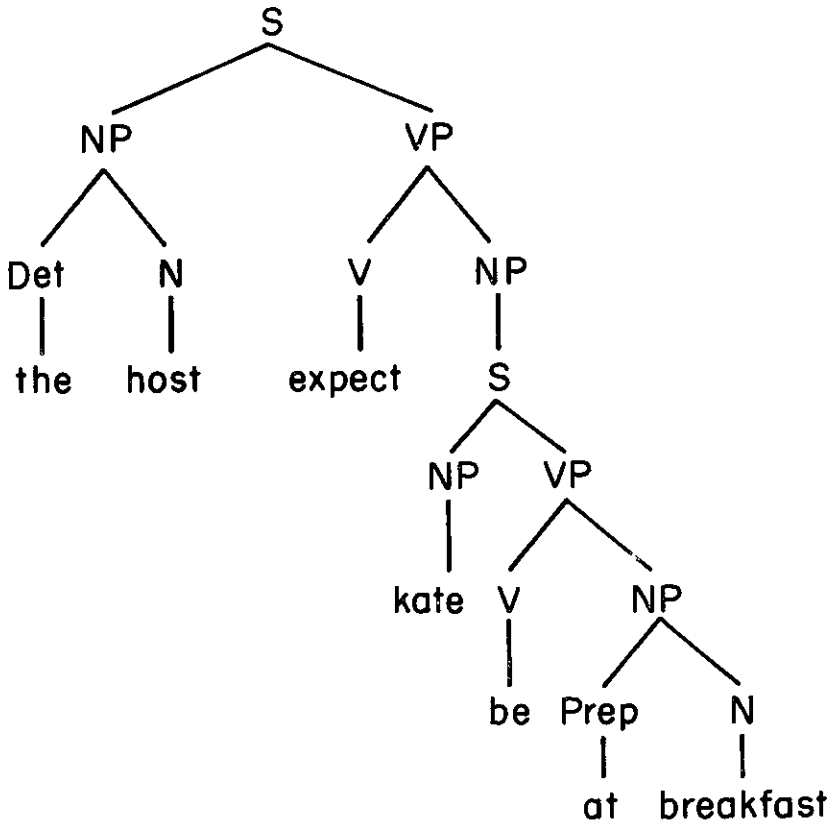
(a) = *We skate to the farm.*

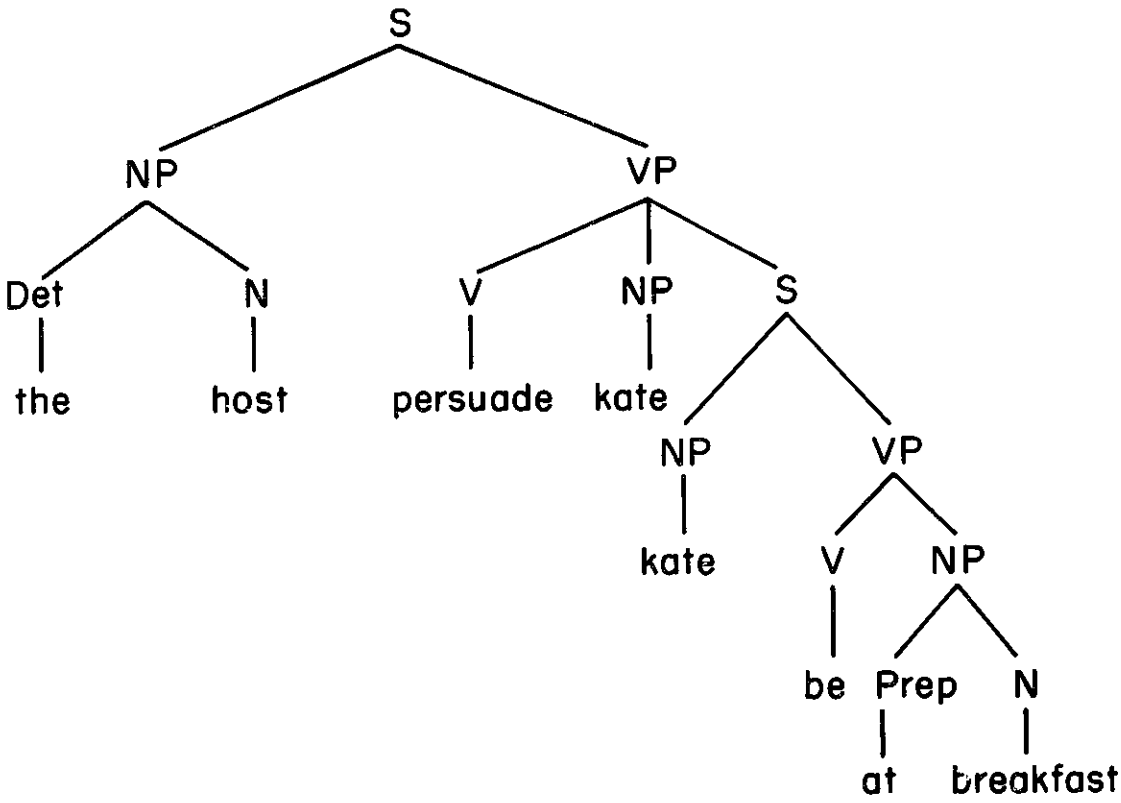
(b) = *We skate with the crowd.*

(c) = *We skate at the pond.*

Figure Caption

Figure 1. Hierarchical tree diagram representing the underlying structure of sentences (1) [upper tree] and (2) [lower tree], according to the analysis of complements provided by Rosenbaum (1967).


S
NP
VP
Det
N
V
NP
the
host
expect
S
NP
VP
kate
V
NP
be
Prep
N
at
breakfast


S
NP
VP
Det
N
V
NP
S
the
host
persuade
kate
NP
VP
kate
V
NP
be
Prep
N
at
breakfast

CHAPTER 4

Preposing Rules

Preface to Chapter 4

"Naval historians think it worth while to argue about Nelson's tactical plan at Trafalgar because he won the battle. It is not worth while arguing about Villeneuve's plan. He did not succeed in carrying it out, and therefore no one will ever know what it was." [from R.G. Collingwood, An Autobiography, London: Oxford, 1939].

The above quotation applies by analogy to most scientific work--reports and theoretical discussions center around the outcomes of successful experiments, where "successful" is typically defined in terms of data which permit the rejection of a statistical null hypothesis. But null results may be informative too, particularly in cases where (a) the likelihood of a Type II statistical error (by which the null hypothesis is incorrectly accepted) is small, and (b) the null result can be contrasted with positive results obtained with the same experimental method.[1] Because these conditions appear to hold for some of the results reported in this chapter, it was decided to include the report as part of this thesis. The report should be of interest to anyone who wishes to pursue work using the present method, as well as for those who would simply have liked to know what Villeneuve's plan really was.

---

[1]When these conditions are satisfied, I take it that the null result is "valid", in the sense used in the M.I.T. Graduate School Manual (p. 40), and is therefore not merely desirable to report by required under M.I.T.'s thesis regulations.

Linguistic Background

In addition to putative transformations like Raising, there exist a number of other transformations which move NP or other constituents.

Among these are the rules Passive, Topicalization, Adverb Preposing, and Prepositional Phrase Preposing (hereafter, PP Preposing). This chapter is a report of experiments designed to test the possibility that the application of such movement rules influences speech timing, either directly or indirectly via the particular form of surface structure they produce.

We begin with the Passive transformation, a rule which has received much attention in linguistic work. This rule has undergone a number of revisions since its earliest formulation in generative grammar by Chomsky (1957).

Chomsky originally argued that sentences like (1) and (2) should be related transformationally by an optionally-applied rule.

(1) John ate the apple.

(2) The apple was eaten by John.

According to this formulation of the rule, Passive operated to move the underlying NP John to the right, attaching it to a by-phrase, and also moved the underlying NP the apple to the left, to the position originally occupied by John.

Later, however, Chomsky reformulated the Passive rule as an obligatory transformation which operated on structures like (3):

(3) NP-Aux-V-...-NP---by passive (where passive is a dummy element in deep structure (Chomsky, 1965: 104).

Under this formulation, the transformation substituted the first NP for the dummy element passive and placed the second NP in the original position of the first.

The reformulation was designed to account for a number of facts not handled well under the earlier optional formulation, including the fact that passivization is normally restricted to sentences whose verbs take manner adverbials (e.g., "with enthusiasm", "slowly"). Passive is blocked for a verb like resemble, for example (*Mary is resembled by Jane), a verb which cannot take manner adverbials (*Mary resembled Jane enthusiastically). By adding a base rule to the grammar like (4), combined with the formulation of Passive in (3), the restriction of Passive to verbs that can take manner adverbials was

(4) Manner→ by passive (Chomsky, 1965: 104).

captured by the grammar. In addition, the new formulation enabled Chomsky to provide a somewhat more natural account of the existence of "pseudo"-passives (e.g., The issues were enthusiastically argued against).

Still further revision in the Passive rule was made when Chomsky (1970) discussed this rule in relation to certain nominalizations under the lexicalist hypothesis. He noted that the two operations which comprise Passive, namely NP Postposing of the underlying subject and NP Preposing of the underlying object, could be viewed as independently applicable operations, assuming a relation between passive and certain nominalizations whose existence could be accounted for by the application of one or the other (but not both) of these two operations (Chomsky, 1970: 203-204).

A final revision of Passive has been made which relies on the recent "trace" theory of movement rules (see Chomsky, 1973, to appear; Fiengo, 1974). Under this formulation, Passive and other NP movement rules are assumed to leave a trace of their application. The trace is morphologically null and acts like a bound variable, in the same sense as an anaphoric element (Fiengo, 1974). The existence of such morphologically null traces allows some generalizations to be stated in a manner that appears to be more accurate and elegant than that of any currently competing alternatives.

In the case of Passive, according to the new formulation, a trace t is left behind in the place of both the underlying subject and object after each is moved. However, the trace left by postposing the underlying subject is erased by preposing the underlying object to the position formerly held by this trace. Hence, only one trace appears in the derived surface structure, occupying the position of the underlying object, as shown in the string below:

(5) the apple was eaten t by John

An apparent advantage of the trace formulation of Passive is that it allows the statement of a constraint on why idiomatic phrases like kick the bucket can occur in only a non-idiomatic reading when they are passivized (compare (6) and (7)):

(6) John kicked the bucket.

(7) The bucket was kicked by John.

Doubtless there exist other ways to state the idiomatic restriction, but the trace theory as currently being developed appears to provide a good way of stating this constraint as well as many others,

and the theory appears to be a highly promising approach to an account of movement transformations.[1]

Objective

The goal of this preliminary study was to determine the timing of words in a passive sentence, in comparison with a control sentence containing a by-prepositional phrase. Because no previous studies of timing have been concerned with passives, it was decided to provide date on a number of candidate hypotheses, in particular, the three listed below:

(a) the trace left by passive, according to Chomsky's most recent formulation of movement rules, is accompanied by a durational change in the immediately preceding or following word--this hypothesis states that the trace is accompanied by a phonetically observable change in segment duration, even though the trace itself is morphologically and phonologically null.

(b) the duration of a noun moved by Passive differs from the duration of a corresponding stationary noun in the control sentence

(c) the duration of the by in Passive differs from the duration of the prepositional by

Because this study represented the first attempt to test timing relations in passive sentences, the above hypotheses and the experimental design itself were intentionally somewhat scattershot. If either Hypothesis (a) or (b) turned out to be supported, then a rather profound

revision of current thinking about the control of speech timing would seem to be in order. And from the linguist's vantage point, any evidence favoring Hypothesis (a) would constitute a performance analog to trace theory of a kind that could be very useful in guiding further development of the theory itself.

Experiment I

Method

Subjects

Sixteen M.I.T. students and employees served voluntarily in this experiment. All subjects were native speakers of English and had no history of speech or hearing impariment.

Sentence Materials

Two sentences were used in the experiment:

(a) The students were seated by the widow. (Passive by)

(b) The students were seated by the window. (Prep. by)

The sentences were matched for total number of words and syllables, typical stress contour, and, indeed, differed from each other phonetically only by the presence or absence of a nasal [n] in the first syllable of the last word of the sentence.

The surface structure bracketing of both sentences was as follows:

$[_S [_{NP}$the students$] [_{VP} [_{AUX}$were$] [_V$seated$] [_{NP}$by the wi(n)dow$]]]$

According to the trace theory, the passive version (Sentence (a)) also contains a morphologically null trace preceding by.

Sentences (a) and (b) can also be said to differ in other grammatical ways. The first NP of (a) is a patient whereas the first NP of (b) is agentive; the verb in (a) is transitive whereas the verb in (b) is intransitive; and the last NP in (a) is agentive whereas the last NP in (b) is a locative prepositional phrase.

Procedure

The general testing procedure described in Chapter 1 was used here. Speakers were told at the outset of the experiment to read Sentence (a) as a passive sentence rather than on its possible reading as a sentence containing a locative prepositional phrase (e.g., The students were seated by (=beside) the widow.) Each sentence was practiced individually before it was read for recording purposes 6 times in succession. Speakers began with Sentence (a) or (b) according to random assignment. The first 5 occurrences of each sentence (excluding mispronunciations) were digitized at 10 kHz and analyzed for the durations of each of the first 5 words of the sentence. The measurements of each word included all phonetic segments of the word, with the sole exception that any pre-voicing for by was excluded from the measurement of this word (in this case, pre-voicing was noticed in a few utterances, distinguishable from the voicing of the previous word by a change in the amplitude of voicing). The occasional pre-voicing for by was excluded to minimize the duration variability of this word.

Results and Discussion

The mean duration of each of the 5 words for each speaker is shown in Table I. Two-tailed t-tests for matched pairs were

---

Insert Table I about here

---

applied to the mean durations for the group of speakers and revealed no statistically significant differences between the two test conditions for each word ($p > .20$ in each case). The absence of any significant effects indicates that the data provide no support for any of the three hypotheses outlined in the Objective. Thus, there is no evidence to support the notion that morphologically null traces are accompanied by a phonetic change in duration, that a moved NP differs in duration from a stationary NP (in terms of a sentence derivation), or that the duration of passive _by_ differs significantly from prepositional _by_.

Of the 16 speakers, two (J.L. and P.V.) produced an average lengthening of all 5 words in the passive sentence in comparison with the prepositional sentence, while one speaker (A.B.) produced an equally systematic effect for all 5 words in the opposite direction. It is quite possible that these individual trends represent changes in overall speaking rate rather than changes in duration as a direct function of the grammatical structure. Further work is required, however, to test the latter possibility.

Other Preposing Rules

A second experiment was designed to provide another test of a phonetic manifestation for morphologically null traces (Chomsky, 1973, to appear). Three preposing rules were studied--Topicalization, Adverb Preposing, and PP Preposing.

The first of these rules, Topicalization, serves to move an NP to the beginning of its clause, transforming the structure of sentences like (8) into (9):

(8) We taped the concert with the Ampex recorder.

(9) The concert we taped with the Ampex recorder.

This construction rarely appears in written text, but it occurs commonly in normal conversation. In my speech, the moved NP is lengthened, and a noticeable pause occurs immediately afterwards. The concern here, however, is not with any lengthening that may occur for the moved NP but with any change in duration that may occur just prior to its underlying position, where a morphologically null trace appears in surface structure. This trace occurs immediately after the verb (Fiengo, 1974).

The second transformation, Adverb Preposing, moves an adverb to the beginning of its clause, accounting for the relation between (10) and (11):

(10) We taped adeptly with the Ampex recorder.

(11) Adeptly we taped with the Ampex recorder.

Here, as in the case of Topicalization, the application of the preposing rule results in a construction that rarely appears in written English. In my speech, Sentence (11) is accompanied by lengthening of the

preposed adverb and a puase between this word and the rest of the sentence, similar to the case for topicalized sentences. However, unlike Topicalization, there is no currently available linguistic basis for postulating a trace at the site of adverb movement. Trace theory, as developed so far, pertains only to rules that move NPs.

A third preposing rule, PP Preposing, moves a prepositional phrase to the beginning of its clause, deriving the structure of (13) from (12):

(12) We taped with the Ampex recorder at sundown.

(13) At sundown we taped with the Ampex recorder.

This construction appears more commonly in written and spoken English than either of the two preceding preposed constructions. Like the others, lengthening of the moved NP is noticeable in my speech, as is the presence of a pause following this NP. But unlike the rule of Topicalization, the location of the trace left by PP Preposing is believed to occur after the direct object NP, not immediately after the verb. Thus, according to the hypothesis that traces are accompanied by a change in the duration of an immediately preceding word, the durations of the verb in topicalized sentences should differ from the verb durations in control sentences (no preposing), whereas such a difference should not be found for the duration of the verb in a PP-preposed sentence compared with its control.

Experiment II

Method

Subjects

Nineteen M.I.T. students and employees served as volunteers in the experiment, having the same qualifications as the subjects who served in Experiment I.

Sentence Materials

Five sentences were used in the experiment:

(a) I sang before we taped with the Ampex recorder. CONTROL

(b) At lectures we taped with the Ampex recorder. PP PREPOSING (LOCATIVE)

(c) At sundown we taped with the Ampex recorder. PP PREPOSING (TEMPORAL)

(d) Adeptly we taped with the Ampex recorder. (ADVERB PREPOSING)

(e) The concert we taped with the Ampex recorder. (TOPICALIZATION)

Each sentence contained from 7 to 9 words and from 12 to 13 syllables. Each sentence contained the same approximate stress contour in the vicinity of the key word.

Procedure

The general testing procedure described in Chapter 1 was used here. Each speaker read each sentence 6 times, beginning with (a) or (e) according to random assignment. The duration of the key word tape was measured. The measured segment included the [t] release burst and the vowel [e], but did not include the [p] closure

interval or any [p] release burst. The decision not to include the latter segments was made to minimize the duration variability.

Results and Discussion

The results appear in Table II. The only consistent trend emerging from the data is for the verb of both locative and temporal PP preposing sentences (b) and (c). The verbs in both sentences were shorter in duration than the verb duration of the control sentence (a).

Insert Table II about here

The results of the experiment provide no support whatsoever for the hypothesis that morphologically null traces are accompanied by a change in the duration of a preceding word. According to this hypothesis, a comparison between the verb durations of the preposed and control sentences should have revealed consistent differences for Topicalization only, yet the only difference observed was found for PP Preposing, where no trace is postulated immediately after the verb.

These results, taken together with the null results of Experiment I on passives, suggest that, to the extent that morphologically null traces are processed during sentence production, such traces are accompanied by no changes in the duration of an immediately preceding word. The data from Experiment I suggest that no changes in the duration of an immediately following word occur either.

This particular set of tests was conducted primarily in search of support for an a priori very improbable, but theoretically exciting, hypothesis concerning a phonetic manifestation of traces. No support was found for this hypothesis, but further work using preposed constructions may still be of value toward developing a theory of speech timing. In particular, it was noted above that, for my speech, preposed constituents appear to be lengthened and accompanied by a pause in comparison with non-preposed constructions. Although duration measurements were not made for preposed constituents in the two experiments, I noticed the same trend as found in my speech while listening to the utterances of nearly all of the 19 speakers. On the basis of such listening, it appears that lengthening of the final word of a preposed constituent in the cases of Topicalization, Adverb Preposing, and PP Preposing averages at least 25 msec, and perhaps as much as 75 msec. If acoustic measurements verify these estimates, then a fairly interesting question can be raised, namely, why is lengthening so pronounced for these preposing rules when it is altogether absent in the case of passivization (cf. Experiment I)? One possibility is that Passive, unlike the other rules, incorporates both NP Postposing and NP Preposing, such that the surface structure contains NPs in roughly the same positions as they appear in underlying structure. The other preposing rules produce a more drastic alteration of the normal sentence structure. In linguistic terminology, this impression has been given a formal representation by Emonds (1970). Emonds defines two types of transformational rules, root transfromations and structure-preserving transformations. Root transformations move constituents into positions which could not be

generated by the phrase structure rules of the base component of grammar, while structure-preserving transformations may move constituents only into positions which could be so generated. According to this formulation, Topicalization, Adverb Preposing, and PP Presposing are all root transformations, whereas Passive is structure-preserving. It is possible that this difference in transformational type is associated with a major difference in the control of speech timing; i.e., that root transformations but not structure-preserving transformations produce changes in word duration. Work directed at testing this principle may lead to a firmer test of whether speech timing is controlled by a level of processing comparable to the transformational component of generative grammar.

References

Chomsky, N. (1957). Syntactic Structures. The Hague: Mouton.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge: MIT Press.

Chomsky, N. (1973). Conditions on transformations. In S.R. Anderson and P. Kiparsky (Eds.) A Festschrift for Morris Halle. New York: Holt, Rinehart, and Winston. Pp. 232-286.

Chomsky, N. (to appear). Trace theory. Linguistic Inquiry.

Emonds, J.E. (1970). Root and Structure-Preserving Transformations. Unpublished Ph.D. Dissertation, M.I.T., Cambridge, Ma.

Fiengo, R. (1974). Semantic Conditions on Surface Structure. Unpublished Ph.D. Dissertation, M.I.T., Cambridge, Ma.

Footnote

[1]In Chomsky's most recent formulation (class lectures, 1975), the application of trace theory may result in an elimination of the Passive rule as such, replacing it and other NP movement rules by a general rule Move NP, whose application is contrained by metaconditions on transformational theory.

Table I

Mean Word Durations of Each Speaker for Sentences (a) and (b) of Experiment I

Sentence (a): The students were seated by the widow.

Sentence (b): The students were seated by the window.

Word / Sentence / Speaker	the (a)	the (b)	students (a)	students (b)	were (a)	were (b)	seated (a)	seated (b)	by (a)	by (b)
A.B.	89.4	89.8	519.9	565.7	111.2	141.5	448.9	487.7	199.7	233.0
L.C.	80.4	65.3	383.7	394.9	108.6	128.4	346.4	358.6	155.9	167.4
B.F.	66.1	66.3	577.3	564.2	105.6	109.1	384.3	417.0	160.8	161.3
P.F.	45.2	59.9	492.3	404.9	114.1	110.4	382.4	391.6	168.5	171.5
R.F.	64.5	61.3	424.4	407.6	109.7	118.9	362.0	308.7	154.0	137.4
J.G.	55.2	45.8	414.5	445.6	109.4	107.1	355.0	347.3	133.3	138.3
R.G.	64.7	70.6	525.1	542.5	114.9	143.9	403.1	397.1	156.7	155.4
M.I.	51.5	56.5	408.8	377.9	83.4	69.1	335.9	322.4	145.2	142.0
C.K.	70.5	67.9	471.9	466.0	84.8	81.8	376.3	394.7	147.3	168.0
S.K.	49.6	48.8	387.5	418.0	115.6	118.8	383.4	409.4	152.4	171.6
J.L.	54.5	49.9	496.7	468.9	109.4	103.3	462.0	445.8	255.7	235.4
S.L.	51.2	51.8	454.8	463.4	101.8	112.5	377.9	371.5	152.8	156.1
E.M.	62.1	69.0	435.1	423.1	96.9	89.7	344.7	335.4	167.3	164.4
J.P.	63.9	69.1	451.5	425.2	78.2	84.7	376.1	368.6	136.7	148.7
R.T.	65.5	62.9	497.0	497.1	127.6	132.9	331.0	332.6	153.9	134.7
P.V.	41.7	38.7	454.4	438.0	141.9	124.8	371.2	361.3	131.9	120.9
Grand Mean	61.0	60.9	462.2	456.4	107.1	111.1	377.5	378.1	160.8	162.9

Table II

Mean Verb Durations of Each Speaker for Sentences (a)-(e) of Experiment II

Speaker	Sentence (a)	(b)	(c)	(d)	(e)
A.B.	209.2	205.8	209.9	214.5	227.8
A.Br.	228.0	228.1	215.1	211.7	195.5
D.B.	193.9	196.7	197.5	196.0	200.8
I.B.	184.6	178.7	184.0	189.4	192.8
J.B.	184.6	172.5	175.2	156.5	169.1
S.B.	184.1	172.7	175.8	187.2	176.4
D.D.	137.2	128.7	129.3	149.2	140.7
J.D.	177.9	179.4	184.4	175.7	177.2
L.D.	177.6	179.0	181.3	169.8	175.5
M.I.	184.1	170.7	161.2	174.5	170.3
C.K.	175.5	181.5	176.7	196.7	186.4
D.K.	191.4	188.1	190.8	184.8	180.9
J.L.	197.1	184.8	188.8	205.6	198.1
S.L.	181.4	168.4	169.0	162.2	158.9
S.Lo.	145.9	152.4	150.8	155.0	163.6
E.M.	176.6	160.0	170.9	165.3	158.8
E.Mc.	178.4	185.3	183.9	184.5	200.9
C.P.	184.5	188.2	171.8	197.6	188.5
J.P.	193.4	194.0	169.4	184.2	190.8
Grand Mean	183.4	179.7	178.2	182.1	181.7

Sentence (a): I sang before we taped with the Ampex recorder.
(b): At lectures we taped with the Ampex recorder.
(c): At sundown we taped with the Ampex recorder.
(d): Adeptly we taped with the Ampex recorder.
(e): The concert we taped with the Ampex recorder.

CHAPTER 5

Speech Timing of Coreference

Abstract

Speakers lengthened the duration of a noun when it was referred to again later in the same sentence by a pronoun. The effect was observed for the very first word of an utterance and for coreferents that spanned a major syntactic clause boundary.

CONSIDER the following ambiguous sentence:

(1) Kate told Jane that she didn't need to pass French.

This sentence contains two clauses, whose underlying representations, according to generative grammar, are roughly [Kate told Jane X] and [Kate (or Jane) didn't need to pass French.] The second clause is subordinate to the main clause.

The pronoun she in the subordinate clause can refer back to either Kate or Jane. Linguists have noted that a speaker may place emphatic stress on a pronoun in ambiguous sentences like (1) in case the speakre intends the pronoun to be a coreferent of the more distant antecedent (e.g., Kate).[1,2] Since duration is a major concomitant of stress,[3] it was expected that in sentences like (1) the duration of she would be much longer when this pronoun refers back to Kate.

The present experiment shows, however, that speakers more consistently lengthen the duration of the antecedent noun itself, by a small amount averaging about 8 msec. The result indicates that a semantic level of processing partly controls the timing of words in speech. Furthermore, the result suggests that the timing of coreferents may take place at a different stage of processing than one limited to the domain of a single clause.[4]

The experiment included 12 speakers from the M.I.T. community. Each speaker was presented Sentence (1) and was instructed to consider each of its two possible meanings. The speaker was instructed to practice saying the sentence according to each meaning and to try to make the intended meaning understandable to a listener. Following practice without feedback from the experimenter, each speaker uttered each of the two versions of Sentence (1) 6 times in succession for

purpose of recording, beginning with either version according to random assignment. The waveforms of the first 5 occurrences of each sentence for each speaker were digitized at a sampling rate of 10 kHz and measured for the duration of the words Kate, Jane, and she. The measurements were made from digitized oscillographic traces of the speech waveforms with the aid of a computer controlled cursor.[5] The offset of visible waveform periodicity was taken as the end of each word segment; hence, the [t] closure interval for Kate or any following [t] release burst was not included as part of the duration of Kate. The decision not to include such segements was made in order to minimize duration variability. The reliability of each measurement was estimated to be within ± 2 msec for Kate and ± 3 msec for both Jane and she. The measurements for Jane for one speaker were excluded from the data analysis because the reliability of the measurements exceeded ± 3 msec.

The results for the version of Sentence (1) in which she referred to Kate showed the following average durations for the group of speakers: 182.8 msec for Kate, 318.5 msec for Jane, and 232.1 msec for she. For the version of Sentence (1) in which she referred to Jane, the average durations were: 174.8 msec for Kate, 325.8 msec for Jane, and 196.8 msec for she. The average duration of Kate was thus 8.0 msec longer when she referred back to Kate, and the average duration of Jane was 7.3 msec longer when she referred back to Jane. This lengthening effect was obtained with 9 of the 12 speakers for Kate and with 9 of 11 speakers for Jane. In addition, the average duration of she was 35.3 msec longer when she referred back to Kate, its more distant antecedent. This effect, however, was observed for

only 7 of the 12 speakers. For 4 of the speakers, the effect averaged more than 65 msec and represented a clearly perceptible indication of emphatic stress. In summary, while the duration of she for some speakers was much longer when it referred to the more distant antecedent, a smaller but more consistent effect of lengthening was obtained for the duration of the antecedents themselves.

Antecedent lengthening may be controlled by the relation of coreference per se or by a relation of focus, whereby the antecedent is lengthened because it happens to serve as the main information-bearing referent of the sentence. Regardless of how this distinction is resolved,[4] the present effect indicates that speakers time the durations of words with reference to their semantic relation to other words in a sentence. This semantic influence can be observed even for the first word of an utterance. In addition, the influence is not blocked by the presence of a major syntactic clause boundary. The latter condition suggests that the level of processing at which semantic control of timing takes place occurs prior to a level of processing that is limited to clause-by-clause computation.[4,6]

Further work should be directed at the question of whether antecedent lengthening is observed under conditions that more closely approximate a natural speaking situation. For example, antecedent lengthening should be studied for non-ambiguous sentences otherwise similar to (1) [e.g., Kate told Don that she didn't need to pass French] in which the speaker is asked to read the sentences in conversational context.

This work was supported by a National Science Foundation Graduate Fellowship and by the National Institutes of Health. I thank Dr. Merrill F. Garrett for advice.

References

[1]Akmajian, A. and Jackendoff, R.S., Ling. Inq., 1, 124-126 (1970).

[2]Cantrall, W.R. in Papers from the parasession on functionalism, (edit. by Grossman, R.E., San, L.J., and Vance, T.J.), 36-46 (Chicago Linguistic Society, Chicago, 1975).

[3]Fry, D.B., J. Acoust. Soc. Amer., 27, 765-768 (1955).

[4]Cooper, W.E. Syntactic control of timing in speech production (Ph.D. Dissertation, M.I.T., 1976).

[5]Huggins, A.W.F. Q. Prog. Rep. Res. Lab. Electr., M.I.T., 95, 81-83 (1969).

[6]Garrett, M.F. in Advances in learning theory and motivation (edit. by Bower, G.) (Academic Press, New York, in the press).

CHAPTER 6

Speech Timing of Coreference. II. *Precede* and *Command* Relations

Introduction

This chapter includes the report of a second experiment on coreference. The results provide a further demonstration that the control of timing is not necessarily restricted to the domain of a single clause.

The present study concerns two syntactic conditions on pronominalization. One of these conditions, termed the precede condition (Lanacker, 1969), involves the linear relation between a pronoun and its antecedent. The other condition, termed command, involves a hierarchical relation between two coreferents. By comparing word durations in sentences in which one or the other of these two conditions is violated, one can determine whether the linear and hierarchical syntactic constraints on pronominalization play different roles in the control of speech timing. In addition, the comparison between these two conditions allows one to provide a further test of whether coreference relations play a role in the control of timing and whether such control, if it exists, is restricted to the domain of a single clause (see the preceding chapter for discussion).

Langacker (1969) formulated a general constraint on pronominalization as follows: a pronoun cannot both precede and command its antecedent (the term antecedent as used by linguists refers to the definite noun phrase that is coreferential with a pronoun--the antecedent need not precede the pronoun). The hierarchical notion command was defined by Langacker in terms of structural nodes A and B as follows: A commands B if the sentence-node most immediately dominating A also dominates B, and if A and B do not dominate each other. By applying

the definition of command to the example structure in Figure 1, we note that A commands B, C, and D; B commands A, C, and D; C commands only D; and D commands only C.

Insert Figure 1 about here

Langacker's formulation of the constraint on pronominalization in terms of the notions precede and command was designed to account for the ungrammaticality of sentences like (1) below, in which the indexed nouns are intended to be coreferential:

(1) *He$_i$ will leave if Heath$_i$ is in real trouble.

Ross (1969) and Postal (1970) also noted and discussed this constraint.[1]

Both of the normal precede and command relations between pronoun and antecedent must be violated in order for pronominalization to be blocked. If either of the relations is violated singly, fully grammatical sentences are obtained, as in (2) and (3):

(2) If he$_i$ leaves the state, Heath$_i$ will be in real trouble.

(3) If Heath$_i$ leaves the state, he$_i$ will be in real trouble.

Sentence (2) contains a violation of the normal precede relation, resulting in so-called backwards pronominalization. Sentence (3), on the other hand, contains a violation of the normal command relation, with the definite noun phrase Heath occurring in a conditional clause dominated by the main clause containing the pronoun.

Sentences (2) and (3) were used here as materials in a sentence-reading experiment, similar in design to the experiment reported in the previous chapter. Instead of measuring the durations of the pronoun and

antecedent as in the earlier experiment, it was decided to measure the duration of the key word state. This word was chosen because there exists no conceivable difference in the semantic focus assigned to this word in Sentence (2) vs. (3). Thus, any duration effects obtained must be attributed either directly to a difference between the coreference relations of precede and command or to another factor that depends on this difference.

Experiment Method

Subjects

Twelve M.I.T. students and employees served voluntarily in the experiment. All subjects were native speakers of English with no history of speech or hearing impairment. All 12 subjects served in the experiment described in the previous chapter during the same test session.

Sentence Materials

Sentences (2) and (3) were used in the experiment. These sentences are repeated below for convenience:

(2) If he leaves the state, Heath will be in real trouble.

(3) If Heath leaves the state, he will be in real trouble.

Each of the two sentences contained 11 words and 12 syllables. The sentences differed from each other solely in the location of the coreferential nouns Heath and he, whose positions were interchanged in the two sentences. In (2), the pronoun preceded but did not command its antecedent, while in (3), the pronoun commanded but did not precede its antecedent. To control for any possible effects of phonetic environment on the duration of the key word state, this word was bounded

on the left in both sentences by the same word, the, and on the right by either Heath or he, containing the same initial phonetic segments. The latter words were also separated from the key word by a major syntactic clause boundary, making it unlikely that any difference in the stress (with greater stress on Heath) or phonetic structure of the two words systematically influenced the duration of the key word (cf. Huggins, 1974).[2]

Procedure

The subjects were tested individually according to the general protocol used in the experiment reported in the previous chapter. In this experiment, however, the speaker was told not to place contrastive or emphatic stress on any words in order to minimize variability and avoid any systematic stress differences on the coreferential nouns Heath and he that might influence the duration of the key word (see Footnote 2). Each speaker was instructed to practice reading each of the two sentences in a natural speaking voice and was encouraged to read the sentences as unitary utterances rather than word-by-word as in unpracticed reading. Following practice, each speaker read each of the two sentences 6 times in succession. The speaker began with Sentence (2) or (3) according to random assignment. The first 5 occurrences of each sentence were digitized at a sampling rate of 10 kHz and analyzed for the duration of the key word state. The onset of this word was taken as the beginning of visible frication in the digitized oscillographic trace. The offset of the word was taken as the end of visible periodicity in the vowel. The closure interval of the syllable-final [t] and any closure release burst was not included as part of the word duration in order to minimize

duration variability. The reliability of each measurement was estimated to be within ± 3 msec.

Results and Discussion

The average durations of the word state in each sentence for each speaker are presented in Table 1. The average duration for the 12 speakers in Sentence (2) was 13.3 msec longer than in Sentence (3). The longer duration in (2) was found for 9 of the 12 speakers,

---

Insert Table 1 about here

---

and the lengthening effect for the speakers as a set was statistically significant ($t = 2.636$, $p<.05$, $df = 11$; two-tailed t-test for matched pairs).

The present effect provides a further demonstration that coreference relations play a direct or indirect role in the control of speech timing and that this influence is not restricted to the domain of a single clause. The only surface difference between the two sentences in the experiment was the location of the two coreferents, which were members of separate clauses in both sentences.

The experiment of the previous chapter also showed a relation between coreference and speech timing. That experiment was conducted with a structurally ambiguous sentence and speakers were instructed to read each version of the sentence so as to make the intended meaning

understandable to a listener. The previous results could have reflected a capability on the speaker's part to take account of coreference relations in speech timing in a special circumstance and not in a setting approximating normal speech. In the present experiment, however, the sentence materials were not structurally ambiguous, and speakers were simply instructed to read the sentences in their normal speaking voice. Thus, the results indicate that coreference relations play a role in timing for non-ambiguous as well as ambiguous sentences and suggest that such relations may play a role in the timing of natural speech.

In addition to these general considerations, it may be possible to provide a specific account of the direction of effect observed in the present experiment. One possibility is that the duration of state was longer in Sentence (2) because violation of the precede relation resulted in a structure that is less typical than the structure in (3). At the word level, it has been shown by Coker, Umeda, and Browman (1973) that the duration of atypical or low-frequency words is longer than the duration of more typical words. If this principle can be extended to the level of sentence structure, and if violation of the precede relation indeed produced a less typical structure than violation of the command relation, then an account would be provided for the greater duration of the clause-final word in (2).

Footnotes

This work was supported by a National Science Foundation Graduate Fellowship, a grant from the Sloan Foundation to the Department of Psychology, M.I.T., and by NIH Grant HD-05168. I thank Dumont Billings for assistance.

[1]Postal (1970: 20ff.) notes that Langacker's precise formulation of the constraint may require modification. In addition, Postal shows that this major constraint on proniminalization must be accompanied by other constraints (in particular, a cross-over restriction). For present purposes, however, it is sufficient to assume that a hierarchical notion similar to Langacker's command is required to account for the major constraint on pronominalization under discussion.

[2]In order to test this possibility directly in future work, a control experiment should be conducted using sentences like (1') and (2'):

(1') If Joan leaves the state, Heath will be in real trouble.

(2') If Joan leaves the state, he will be in real trouble.

References

Coker, C.H., Umeda, N., and Browman, C.P. (1973) Automatic synthesis from ordinary English text. IEEE Audio and Electroacoustics AU-21, 293-297.

Huggins, A.W.F. (1974) An effect of syntaz on syllable timing. Quarterly Progress Report of the M.I.T. Research Laboratory of Electronics 114, 179-185.

Langacker, R.W. (1969) Pronominalization and the chain of command. In (D.A. Reibel and S.A. Schane, Eds.) Modern Studies in English. Prentice-Hall, Englewood Cliffs, N.J. Pp. 160-186.

Postal, P.M. (1970) Cross-Over Phenomena. Holt, Rinehart, and Winston, New York.

Ross, J.R. (1969) On the cyclic nature of English pronominalization. In (D.A. Reibel and S.A. Schane, Eds.) Modern Studies in English. Prentice-Hall, Englewood Cliffs, N.J. Pp 187-200.

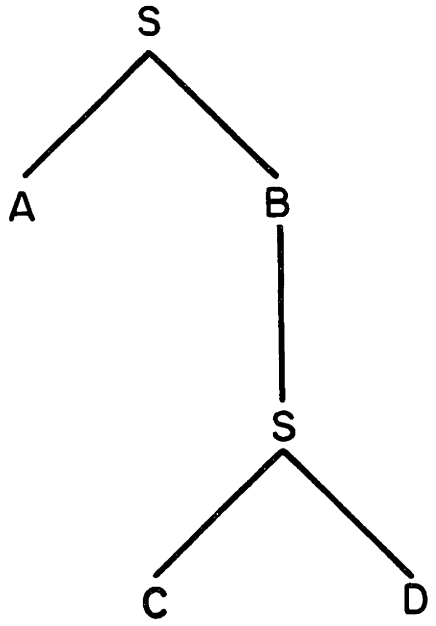
Table I

Mean Durations of the Key Word state for Each Speaker in Sentences (2) and (3)

Speaker	Sentence (2)	Sentence (3)
R.B.	303.7	306.6
L.H.	328.5	337.8
M.I.	354.4	348.8
M.J.	341.4	326.9
C.K.	348.5	319.8
J.L.	381.3	394.3
S.L.	348.3	322.6
E.M.	325.3	310.8
B.P.	367.7	341.8
C.P.	391.0	361.5
J.P.	360.8	339.2
K.P.	435.7	416.8
Grand Mean	357.2	343.9

Figure Caption

Figure 1. Structural respresentation to illustrate command relations.


S
A
B
S
C
D

CHAPTER 7

Conclusion

In addition to the experiments described in preceding chapters, a number of other experiments were conducted using the same testing method, typically using only from 3 to 7 speakers in each experiment. Because of the small data base for these experiments, few if any firm conclusions can be drawn from the results. Current work is aimed at building larger data bases for some of these experiments, and the final results of this work will be reported elsewhere (Cooper, in preparation). However, it seems useful to provide a preliminary sketch of these additional studies here in order to give some indication about the direction which the work is currently taking.

The experiments can be classified into four main areas, concerning (1) deep structure ambiguities, (2) deletion rules, (3) movement rules, and (4) clause types. The work in each of these areas is discussed briefly below, followed by concluding remarks summarizing the direction of the work as a whole.

1. Deep Structure Ambiguities

The experiments in the preceding chapters were primarily designed to uncover possible effects controlled by transformational or surface levels of syntactic processing. However, some preliminary work was also aimed at the possibility that underlying syntactic structure plays a role in the control of speech timing. To test the relevance of this syntactic level of processing (if it indeed exists), the timing of words in two ambiguous sentences was studied. These sentences, presented below in (1) and (2), represent so-called "deep structure ambiguities" because they have two possible underlying syntactic representations but only a single surface structure.

(1) Which criminals have the police surrounded?[1]

(2) The shooting of the criminals was appalling.

Sentence (1) can question either which criminals surround the police or which police surround the criminals. In essence, the ambiguity concerns the underlying grammatical roles of subject and object. Sentence (2) concerns a statement about either the criminals' act of shooting or someone else's act of shooting the criminals.

The duration of each word in both Sentences (1) and (2) were measured for a group of 5 speakers. At the outset of the test, each speaker was told that the sentences were ambiguous and was informed about the nature of the ambiguity. The speaker was then told to practice saying each of the intended meanings of each sentence so as to make the meaning understandable to a listener if possible. Each version of each sentence was then recorded 6 times.

The word durations showed some sizable differences between the two meanings for individual speakers, in one case averaging more than 100 msec. However, these effects were not very systematic in direction across different speakers, and a larger data base is clearly needed to test for population effects. The presence of large and consistent effects within the data from each subject, however, provided rather convincing support for the notion that deep structure differences can influence speech timing. It should be noted, however, that the evidence provided here is also compatible with a model of timing in which logical relations control timing directly rather than via an underlying syntactic representation.

2. Deletion Rules

Additional work on deletion rules has centered on the putative rules of to be deletion (Borkin, 1973) and will deletion (Vetter, 1973). The work on to be deletion, designed to test for evidence of underlying and/or transformational structure effects on timing, involved the two sentences below.

(3) John appeared abruptly at the meeting.

(4) John appeared abrupt at the meeting.

While both (3) and (4) contain a single clause in surface structure, (4) contains a putative two-clause underlying structure, represented roughly by the strings John appeared to be X and John was abrupt at the meeting. If the underlying clause boundary occurring after the verb appeared plays a role in word lengthening, then it would be expected that the duration of this verb should be longer in (4) than in (3). This hypothesis was tested with 6 speakers, one of whom showed the predicted effect by an average of 39.6 msec. However, only two other speakers of the group showed a trend in the same direction, indicating that a larger data base is required before any firm conclusions cam be reached about the role of underlying clause structure for this sentence construction. Work on a putative rule of will deletion involved the following sentences:

(5) Tomorrow I will face the students cheerfully.

(6) Next week I face the committee on academic standing.

(7) I will face the committee tomorrow.

(8) On Tuesday John will face the new students.

(9) Next month I face the prospect of hard times.

(10) I will face the dean on Monday morning.

(11) On Tuesday John will face the new students.

(12) Next week I will face the president's envoy.

(13) Tom will face the crowd on Thursday.

These sentences were included as part of a 39-sentence list used in the earliest stages of pilot work. Notice that, in sentences (6) and (9), future tense is indicated by the presence of appropriate lexical information, but a future tense marker will is absent, unlike the remaining sentences in the list. No durational effect of this deletion was found for the key word face in these sentences in data for three speakers. In addition, no consistent differences in the duration of the key word were obtained depending on whether the temporal prepositional phrase in the sentences occurred in clause-final position or in a preposed position (see Chapter 4) (compare, for example, Sentences (5) and (7)).

The work on to be deletion is currently being extended because of its importance in testing for effects of underlying clause structure. The work on will deletion is not being continued at present because no major theoretical issues appear to depend on the outcome of such work.

3. Movement Rules

Additional studies of movement rules have included sentences involving A-Verb Raising (Postal, 1974) and Particle Movement. The former rule converts the structure of a sentence like (14) into (15), while the latter rule converts the struct of (16) into (17) (or 17 into (16)--the direction of movement being unclear in the case of this

rule--see Jacendoff, 1975 for discussion favoring the derivation of (16) from (17)).

(14) It seems that John is angry.

(15) John seems to be angry.

(16) John phoned up Martha.

(17) John phoned Martha up.

The preliminary results obtained with these constructions indicate that some speakers produce systematic differences for the sentences of each of the two types. However, no firm conclusions can be reached concerning the presence or absence of population effects.

4. Clause Types

Of major interest in current timing research are differences in the magnitude of clause-final lengthening obtained with different types of clauses, e.g. complement, relative, conditional, causal. It was pointed out earlier that audible differences in magnitude exist in the lengthening produced by different clause types, yet no systematic account of these differences or of their theoretical basis exists. To provide information on the empirical issue, a test using 6 speakers was conducted with sentences containing two clauses, one main clause and one subordinate clause, the latter being either a conditional clause (introduced by _if_ or _unless_) or a temporal clause (introduced by _before_ or _after_). The following four sentences were used:

(18) If we reveal Kate's hopes the Doctor will see her.

(19) Unless we reveal Kate's hopes the Doctor will see her.

(20) Before we reveal Kate's hopes the Doctor will see her.

(21) After we reveal Kate's hopes the Doctor will see her.

The duration of the vowel [o] of the key word hopes was measured for duration in each sentence. The average durations for the key word of the two conditional clauses (18) and (19) were 151.6 msec and 159.8 msec, respectively, while the average durations for the key word of the two temporal clauses (20) and (21) were 147.0 msec and 147.5 msec, respectively. On the basis of these preliminary data, it appears that clause-final lengthening is more pronounced for conditional than for temporal subordinate clauses when followed by a main clause. However, this difference is small compared to the differences expected to exist amoung some other clause types. For example, it is expected, based on impressions of listening to conversational speech, that clause-final lengthening at the end of a complement clause is very small in magnitude compared with the lengthening effects observed at the ends of other subordinate clauses, as well as at the end of a conjoined clause. Further experiments are required to verify and quantify this impression. As noted in the introduction, such differences in magnitude should probably be implemented in programs to synthesize speech by rule. A theoretical account of the differences may rely on the degree to which two clauses separated by a clause boundary are semantically related to one another (the closer the relation, the less lengthening). However, the latter possibility requires a much more precise definition.

Another area of current study on clause types for which some preliminary data have been obtained concerns a comparison between the timing of a word preceding a complement vs. relative clause.[2] Four

sentences were used in this experiment:

(22) I regretted the fact that John killed Sam.

(23) I regretted the fact that John reported.

(24) I regretted the fact that John died.

(25) I regretted the fact that John announced.

Sentences (22) and (24) contain a main clause followed by a complement. On the other hand, Sentences (23) and (25) are ambiguous, containing a main clause followed by either a complement or a relative. The ambiguity in (23) involves the distinction, for example, of whether John's reporting was regretted (complement) or whether the fact which he happened to report was regretted (relative). Speakers were instructed to read sentences (23) and (25) on their relative clause meanings.

Measurements of the duration of the key word fact were made for these sentences using 7 speakers. The average durations for sentences (22)-(25) were, respectively, 261.6 msec, 269.6 msec, 250.6 msec, and 273.0 msec. All 7 speakers produced a longer duration in the relative-containing (23) than in the complement-containing (24). However, remaining comparisons between complement- and relative-containing sentences failed to uphold this highly systematic lengthening effect for relatives. If the greater lengthening for relative-containing sentences turns out to be significant on the basis of further experimentation, this effect can be accounted for by the notion that a grammatical boundary of greater hierarchical status occurs immediately after fact in the case of relatives.

5. Concluding Remarks

If this chapter serves its purpose, it leaves the unmistakable

impression of a job unfinished. A major reason for undertaking this work on speech timing was the suspicion that the work would not, and should not, end with the completion of this thesis. The study overall is viewed as a programmatic series of steps, only a few of which have been taken.

In addition to the experiments described above, there exist at least three major issues in timing research on which a great deal of theoretical and empirical work is required. By way of concluding the discussion, these issues are sketched below with reference to the findings of previous chapters.

a. Syntax vs. Rhythm. One of the issues that has cropped up from time to time in the current work on speech timing is the problem of determining whether a specific durational effect should be attributed to a level of syntactic processing or to a level of phonological processing that is concerned with rhythmic factors such as stress timing. In some cases, it is possible to construct experiments that adequately test the effects of a syntactic variable independently of rhythmic variables. However, in other cases, it is difficult to achieve this controlled situation; a change in the value of a syntactic variable may be typically or invariably accompanied by a perceptible change in speech rhythm. An example of a case in which the two variables typically covary concerns the deletion Conjunction Reduction as used in Chapter 2. In the relevant experiment, a comparison was attempted between a two-clause sentence in which the subject of the second clause was either present or deleted from surface structure. Deletion of the subject (a change in a syntactic variable) was accompanied by a change in the number of stressed syllables in the vicinity of the major clause

boundary, and it is conceivable that the latter change produced a reorganization of rhythmic stress affecting the duration of the key segment, occurring just prior to the clause boundary. In order to determine whether syntactic or rhythmic factors (or both) were responsible for the obtained effect, two types of test need to be conducted. First, a test could be constructed in which the difference in speech rhythm between the comparison sentences is minimized, in this case, by substituting an unstressed pronoun for a stressed noun in the subject position of the second clause in the non-deletion sentence. Second, an independent test could be constructed in order to test the effects of rhythmic stress pattern independently of syntactic variables. The latter type of test is of course particularly important in cases where the first type of test cannot be constructed, as may well be the case for some syntactic constructions. Work along these lines should help in the development of separate theories of syntactic and rhythmic phonological influences on speech timing.

b. Information-Flow. A concern related to the issue about a possible rhythmic level of processing is the more general question of what distinct processing stages actually influence speech timing, and how is the flow of information among these stages organized. So far, two distinct types of non-phonetic timing effects have been studied, including clause-final lengthening and antecedent lengthening. As noted earlier, it is believed that these two types of lengthening can be distinguished empirically on the basis of their domain of control. Clause-final lengthening, according to any of its major alternative explanations (see Chapter 1), is considered to be a reflection of the speaker's clause-by-clause computation (see Cooper, in preparation,

for detailed assumptions of this viewpoint). Antecedent lengthening, on the other hand, can be controlled by a coreferent that belongs to a different syntactic clause, eliminating the possibility that non-phonetic timing control takes place exclusively at a level of computation restricted to the scope of a single clause.[3]

Up to this point, we have at least two distinct levels of non-phonetic timing control. There may be more levels, but for the present, it is appropriate to ask whether these two levels operate in series or in parallel. Intuitively, it is appealing to hypothesize that semantic effects like antecedent lengthening (as well, perhaps, as effects like emphatic and contrastive stress assignment), are programmed before the speaker's computation is restricted to clause-by-clause scope. A behavioral test of this hypothesis is currently difficult but not impossible to construct. The essential ingredients for such a test would be two separate lengthening effects, one akin to antecedent lengthening, the other clause-final lengthening, with each producing a large lengthening effect, say 30%. If the two 30% effects are combined in the same sentence, and if these effects add linearly, then a parallel model of lengthening predicts a total increase in length of 60%, whereas a serial model predicts a somewhat larger increase, amounting to 69%. Further work on timing effects and their interaction might allow us to conduct an experiment of this sort to test the seriality hypothesis. In any event, it appears that further work on timing should be directed at providing a better model of information-flow during sentence production.

c. Explanations of Clause-Final Lengthening. Finally, we return to an issue raised in Chapter 1 concerning possible explanations of clause-final lengthening, the effect that has occupied most of our experimental concern thus far. An explanation of the effect is desirable primarily because it might provide information about the operation of the level of timing constrained by clause-clause computation. Three major types of explanation were offered earlier, concerning the speaker's planning strategy, the speaker's storage and retrieval of clauses, and a listener-oriented account, whereby clause-final lengthening is produced in order to aid the listener in recovering structural information. Based on the results of the present study, it is possible to reject a least one of these major accounts for certain constructions. Recall that the average magnitude of lengthening was no more than 10 msec for some structural distinctions. In these cases, a listener-oriented account can be ruled out, since listeners are unable to reliably detect such small differences in duration during sentence perception (Klatt and Cooper, 1975).[4] We are left then, with the planning and buffer accounts. Tentatively, it seems that the buffer account should be favored, since it is far from clear that a 10 msec-interval would be sufficiently large to be of any advantage to clause planning. But only when much more is known will we be in a position to test the merits of the planning and buffer accounts adequately.

References

Borkin, A. (1973). To be and not to be. in C. Corum, T.C. Smith-Stark, and A. Weiser (eds.) Papers from the Ninth Regional Meeting of the Chicago Linguistic Society. Chicago: Chicago Linguistic Society. Pp. 44-56.

Cooper, W.E. (in preparation). Syntactic Control of Speech Timing.

Jackendoff, R.S. (1975). Morphological and Semantic Regularities in the Lexicon. Language 51, 639-671.

Postal, P.M. (1974). On Raising: One Rule of English Grammar and Its Theoretical Implications. Cambridge: M.I.T. Press.

Vetter, D.C. (1973). Someone solves this problem tomorrow. Linguistic Inquiry 4, 104-108.

Footnotes

[1]I am grateful to John Robert Ross for providing this example.

[2]The comparison between complement and relative clauses was suggested to me by Merrill Garrett.

[3]The results of Experiment 6 are not easily handled by the distinction drawn here. For further work on this problem, see Cooper (in preparation).

[4]It is important to note that the smallness in magnitude of some of the present effects does not detract from their potential value in aiding in an understanding of the speech production system. It is true, however, that such small-magnitude effects are of less interest than large-magnitude effects from the standpoint of practical applications, including speech synthesis.