7816707

TUSHER, ALAN LEWIS
THE UTILIZATION OF SILENT PAUSES IN SPONTANEOUS SPEECH BY THE LISTENER.

CITY UNIVERSITY OF NEW YORK, PH.D., 1978

University Microfilms International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

THE UTILIZATION OF SILENT PAUSES
IN SPONTANEOUS SPEECH BY THE LISTENER

by

ALAN L. TUSHER

A dissertation submitted to the Graduate
Faculty in Psychology in partial fulfillment
of the requirements for the degree of Doctor of
Philosophy, The City University of New York.

1978

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

May 4, 1978
date

Chairman of Examining Committee

May 8, 1978
date

Executive Officer

Dr. Bernard Seidenberg

Dr. Dorothy Murgatroyd

Dr. Harold Webster

Supervisory Committee

The City University of New York

Abstract

THE UTILIZATION OF SILENT PAUSES IN SPONTANEOUS SPEECH BY THE LISTENER

by

Alan L. Tusher

Adviser: Professor Bernard Seidenberg

The role of silent pauses for the listener in spontaneous speech was investigated in this series of studies. First, a review of the major background research in spontaneous speech was presented with a delineation of a proposed method for extending the data base into the area of listener extraction of information.

In essence, quasi-expert listeners were presented with a tape recording of a spontaneous monologue. Their task was to indicate on a verbatim transcript those groups of words or non-lexical sounds which together conveyed a single thought or "unit." The results indicated that the "units" were very likely to have been preceded by either a silent pause, a major syntactic boundary, or their combination. Further analyses reveal that silent pauses had been highly associated with the underlying syntactic structure. However, even in those cases where a "unit" had not been preceded by this linguistic information, silent pauses were still found to be highly associated with the beginning of these "units."

In order to clarify this relationship, an experimental tape was constructed in which the pausal structure had been interfered with. Listeners given a similar task of "unit" isolation were also found to

be making use of the silent pause as was a Control group that had not received this version of the tape. The results obtained with a third group which had only received transcript information verified that the presence of written information was not responsible for the results obtained. In addition, a test of comprehension (comprehension here measured by how confident subjects were of their answers) revealed a generalized deficit in the experimental group. A third study utilizing a modified procedure ruled out the possibility that the deficit observed in the experimental group was due to stimulus novelty effects.

The results of these studies were interpreted in terms of a model that views the listener as actively engaged in extracting information from the speaker by making use of the pausal information in determining the underlying propositions of the communication. The generalized comprehension deficit observed in the experimental subjects of the second study, then, was the result of providing them with erroneous pausal information.

ACKNOWLEDGEMENTS

I would like to express my thanks to all those who were involved in my work. I especially wish to thank Dr. Florence Denmark, Executive Officer of the Ph.D. program in Psychology, for her warm and continued support. I also want to thank Dr. Bernard Seidenberg who acted as my dissertation adviser. Much thanks go to Dr. Arthur Reber for his help in the guidance of the work to be presented here. Dr. Dorothy Murgatroyd and Dr. Harold Webster are to be thanked for the giving of their time to serve as members of the dissertation committee. Dr. Lois Hood and Dr. Ray McDermott were most generous with their time in their capacity as outside readers and I deeply appreciate it. I also wish to express my thanks to Prof. Morey J. Wantman and Dr. Solomon Weinstock for their advice with the statistical procedures in the dissertation. Several others were very kind in the giving of their time and assistance: Joseph Geliebter who assisted in the analyses and running of part of the research; Roy Nemerson who helped in the running of some of the research; Steven Bloom whose audio-technical expertise was of great assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In normal, unrehearsed (e.g. spontaneous) speech, periods of silence abound plentifully. Goldman-Eisler (1968) has found that on the average as much as 50% of total speaking time (defined as periods of speech added to periods of silence) is composed of pauses. Given this plethora of hesitation pausing, the question naturally arises as to its functional significance. Exactly what is the relationship between speech and pausing? In order to provide a reference point to the literature and the research to be presented, it is helpful to consider several possible alternative functions for the pause.

First, and perhaps least interesting is the possibility that hesitation pauses reflect breath pauses in speech. Certainly breathing and speaking are highly interrelated processes and it would be surprising if none of the variability in pause patterns could be accounted for by this factor.

Second, pausing may be symptomatic of the need for time to "compute" on the part of the speaker. Decisions as to the syntactic form of the utterance or the lexical items to be chosen may be reflected in periods of hesitation on the part of the speaker.

Third, pausing may serve a listener function. The periods of silence may provide time to decode and comprehend the speaker's communication, or anticipate the next word.

There is also a relatively uninteresting suggestion in the literature that pausing plays no real functional role (except perhaps for the first suggestion above) but merely reflects a random generative process in the production of language. Given the paucity of research in the area to the contrary, this hypothesis might seem somewhat tenable. However, we will review the available evidence and it can be said at this point that this trivial possibility will prove to be rather unlikely.

Background history[1]

Rochester (1973) points out that only during the last several decades has the study of spontaneous speech behavior become technologically feasible due to a previous lack of recording and analyzing equipment. During the 1930's however, the "interaction chronograph" (Chapple, 1939; Matarazzo et al., 1956) was developed. This instrument was basically an electronic stop watch. An observer would depress a switch when the speaker would begin, and release it at the end. With an improvement in technology, the "interaction chronograph" was replaced by the "automatic speech analyzer" (Verzeanno and Finesinger, 1949). Basically, the device used sound activated switches in place of the human observer. This technology has generated several interesting findings in the area of dyadic interactions, which will only be briefly reviewed here since the major emphasis of this literature review section is to focus upon the function of pauses in the single speaker. A significant finding in dyadic interactions however, is that speech unit durations can be described as an exponential decay process. In particular, Hargreaves (1960) notes that "...a fairly accurate model for speech unit duration may be developed by considering speech as fitting a simple random decay process." Hargreaves studied a married couple continuously for a period of two weeks and found a mathematical fit to frequency of speech unit durations to be obtainable by using an exponential function. This type of "random" generation of speech unit durations is also typified in the work of Starkweather (1959) who also examined dyadic interactions.

What is very important about the preceding two studies is their relevance for the production of hesitation pauses. Although the authors were concerned with speech duration their findings tell us a great deal about the production of pauses, at least in dyadic relationships. Since speech durations (i.e., the talking time between pauses) are randomly distributed in time so must the hesitation pauses that demarcate them. This is important because it bears on a hypothesis concerning pauses, namely that they represent a random production process. In fact, Schwartz and Jaffe (1968) have elaborated this into a mathematical model to predict the succession of periods of silence and speech in spontaneous speech. This model is basically Markovian in that "...the duration of an event (speech or pause) is independent of any preceding event." (p. 28). Using the model, they generate results consistent with the findings of Henderson et al. (1966) who examined the sequences of speech and silence in spontaneous speech. Henderson et al. found that the overall duration of the subject's activity could be broken down into periods where short pauses alternate with long speech segments and periods where long pauses alternate with short speech segments. Schwartz and Jaffe's (1968) work suggests that these periods are the result of a random generative process where the durations of pauses and speech segments are random and therefore the alternating high and low pause periods found in speech reflect a random process. However, in any case where a mathematical system is being used to model data we must always be wary of the fact that even if the fit of the model to the data is good, the underlying process responsible for the data in the organism may or may not be the same process captured by the mathematical model. It is not necessary for the model to generate data the same way that organisms do. Two entirely spearate processes may yield similar results. In the present context it is important to note whether or not any other relevant variables are found to be co-extensive with the natural phenomena. In particular, are there distinct differences in the type of speech outputed during periods of long pauses vis à vis periods of short duration pauses. If no distinct differences exist then there is no reason to reject Schwartz and Jaffe's hypothesis.

Goldman-Eisler (1972) has presented evidence which bears on this issue. Pauses occurring in spontaneous speech were found to correlate with the type of syntactic juncture (i.e., subordinate clauses, coordinate clauses, and sentences) occurring after the pause. As Goldman-Eisler states, "...the grammatical description of each of the clause types has its quantitative reflection in the length or absence of pauses which precede it, in natural, unprepared and spontaneous speech." (p. 107). If pause durations are generated in a random fashion there seems no a priori reason to expect syntactic structure to correlate with it, unless we are willing to suppose syntax is governed by the same random process. Furthermore, Goldman-Eisler (1968) offers convincing evidence that the occurrence of pauses correlate with the degree of planning which the utterance reflects. In one experiment subjects were shown cartoon stories and were required to describe the content of the story and also to offer an interpretation of the underlying general meaning or moral of the story (the latter presumably requiring more planning and spontaneous generation of ideas). It was found that the proportion of pause to speech duration was significantly greater for the interpretation task. Periods of greater hesitancy in speech, it was concluded, represent periods of generation of information in the speaker. It seems unlikely that a random process should correlate with an intellectual one. Overall, then there is no real support for the hypothesis that pausing is the result of a random production process.

An important phenomenon that occurs during speech (or non-speech for that matter) is the act of breathing. What is its relevance for the speech process? Fónagy and Magdics (1960) studied the relationship between the speed of an utterance and the length of that utterance and asked whether speech rate would be indicative of the respiratory sequence. Speed of utterance was defined as "...the number of sounds uttered in a second." Although they don't explicitly define the term "sound," they denote the size of their corpus in number of syllables, and they indicate that six sounds are about equivalent to four syllables whereas ten sounds equate to approximately seven syllables (p. 188). However, they also refer to sounds as being syllables elsewhere (p. 191). A phrase is characterized "...by the rising pitch in the stressed syllable," (usually the first syllable of the phrase), "...and the falling intonation of the rest of the phrase. The phrase cannot be interrupted by any pause. If the speaker breaks off the sentence at an unusual place, within a section that would naturally be regarded as a phrase-unit, we consider the section as two different phrases. In some exceptional cases the phrase may be shorter than a word." (p. 180). The preceding difficulties in defining a "sound," notwithstanding, their results bear on the hypothesis that pauses represent a stochastic process. Specifically, they found that mean duration (i.e., the inverse of speed) of a sound decreased as the length of the utterance increased indicating that when a speaker has something lengthy to say he speeds up his rate of saying it. They then examined the relationship between this "speeding-up" process and breath pauses and found it not to be viable primarily because the rate of speech might be expected to be greater at the end of a phrase just before inspiration (assuming that "air" demands are high at that point), than immediately following it. In fact there is only a one msec. mean difference, in favor of this hypothesis but hardly demanding it.

The above study, however, suggests that perhaps hesitation pauses reflect breath pauses in speech. This hypothesis has been examined by Goldman-Eisler (1968) who examined individual speakers. The criterion for a pause was a period of silence of at least 250 msec. in duration. Pauses in speech were segmented into two types: breath and non-breath pauses. It was found that the type of speech is an important variable. When a subject is reading a prepared text the breath pause predominates (77% vs. 23% for non-breath pauses). However, in spontaneous speech the opposite findings were found with most pauses being of the non-breath variety (66%). Thus, the hypothesis that breathing accounts for most pauses in spontaneous speech receives little support.

Another line of inquiry on the functional significance of pausing in spontaneous speech stems from a set of hypotheses by Lounsbury (1954). Basically, they assume that the points of uncertainty in spontaneous speech are reflected in pauses occurring at these decision points during speech production. Specifically they are:

"Hypothesis 1: Hesitation pauses correspond to the points of highest statistical uncertainty in the sequencing of units of any given order....

Hypothesis 2: Hesitation pauses and points of high statistical uncertainty correspond to the beginning of units of encoding....

Hypothesis 3: Hesitation pauses and points of high statistical uncertainty frequently do not fall at the points where immediate-constituent analysis would establish boundaries between

higher-order linguistic units or where syntactic junctures or 'facultative pauses' would occur....
Hypothesis 4: The units given by immediate-constituent analysis, and especially those bounded by facultative pause points, do correspond to units of decoding, however....
Hypothesis 5: Units of encoding for easy oft-repeated combinations approach coincidence with those of decoding."
(pp. 99-100)

Evidence for hypothesis one was provided by Goldman-Eisler (1968) in a series of experiments using a modification of the Shannon (1951) guessing technique. In Goldman-Eisler's procedure (involving all written materials) subjects had to guess succeeding words in a sentence (generated by someone else) given the context of the preceding words. Then the transition probabilities of each guessed word was estimated by the ratio of the frequency of correct guesses to the total number of guesses. Thus a word with a low transition probability was one less likely to be guessed. In one experiment using this procedure, it was found that where the original speaker had spoken without pauses (defined as at least 250 msec. of silence) the transition probability of the next word was high (i.e., contained little information) whereas where there had been a pause preceding the word, the guessing subject had made many more errors, yielding a low transition probability (thus the word was of high information content). Therefore, pauses do correlate with statistical uncertainty, in support of Lounsbury's first hypothesis. However, Boomer (1970) has criticized the sample of spontaneous speech chosen on the grounds that it was biased and not representative, "...the experimental design dictated that only grammatical well constructed sentences be chosen." (p. 149). The study is really only suggestive and hardly definitive with regard to Lounsbury's hypothesis.

Maclay and Osgood (1959) also provide evidence for Lounsbury's first hypothesis. They analyzed a sample of approximately 50,000 words and found that both filled pauses (i.e., sounds occurring within the break, such as "m") and unfilled ones occurred more frequently before lexical than before function words (according to the Fries (1952) classificatory scheme). Since the class of lexical words is far larger than that of function words, a choice among lexical words by the speaker most probably (given Goldman-Eisler's data) involves more uncertainty. The findings may be viewed as supporting Lounsbury's hypothesis. On the other hand, Blankenship and Kay (1964), looking at filled pauses found a greater percentage of them to occur before function than before lexical words. Although the authors provide information on the percentage of lexical and function words their data are difficult to interpret since a statistical analysis was not performed. The authors do suggest that individual speaker differences may have been responsible for this outcome. These contrasting results suggest that whether or not a choice among words (lexical or function) will be difficult is determined by the constraints present due to what has and what will be said. Goldman-Eisler's (1968) choice of the Shannon guessing technique prevented these problems.[2]

Boomer (1965) segmented a corpus of spontaneous speech into phonemic clauses using the Trager and Smith (1951) procedure. He then examined frequency of occurrence of hesitations preceding each word within each clause and found a modal value occurring before the second word regardless of clause length (in clauses of two to ten words) and a decreasing frequency occurring before each succeeding word. This finding, he argued, goes against Lounsbury's first hypothesis (see above for the hypothesis). His argument is that primary stress and word frequency have been found to be negatively related; the less frequent a word, the more likely it will receive phrase stress. In addition, phrase stress is typically found toward the end of a phonemic clause. Therefore, infrequent words occur towards the end of phonemic clauses and pauses should be expected to precede them given Lounsbury's first hypothesis. Instead, Boomer found pauses to be sparsely distributed near the end of the clause, thereby arguing against a word-unit model. On the other hand, Lounsbury's second hypothesis indicates the possibility of units larger than the word and it also predicts that pauses should predominate at the beginning of these units. If we are willing to accept the phonemic clause as the basic unit of encoding, as Boomer suggests, the predominance of pauses toward the beginning of these units is in support of Lounsbury's second hypothesis.[3] There is a major problem here however. A phonemic clause analysis may not really be valid. As Lieberman (1968) says in summary of his own work, "...the results of the experiment suggest that the phonemic pitch levels and terminal symbols of the Trager-Smith system often have no distinct physical basis." We are left in a somewhat shaky position regarding the legitimacy of the phonemic clause.

Henderson et al. (1966) have provided evidence for Lounsbury's third hypothesis (see above). Using recorded interviews they found that 55% of all pauses occurred at grammatical junctures and 45% at non-grammatical ones in speech production. This 45% figure agrees closely with Maclay and Osgood's (1959) finding that 47% of all pauses occur within phrases. They make the assumption that syntactical junctures do not occur within a phrase and thus the 47% finding is in support of Lounsbury's prediction. It should be noted, however, that Maclay and Osgood do not adequately describe the criteria they used when parsing the corpus into phrases, and thus it is unclear if syntactical junctures existed within the sequences they chose as phrases.

In reference to the last of Lounsbury's hypothesis (number five, see above) Lounsbury (1954, p. 100-101) added that in repeated material pauses will be greatly reduced in number with those remaining being located at major syntactic junctures. In the Henderson et al. (1966) paper it is noted that pauses in prose readings were exclusively at grammatical junctures. In addition to this Goldman-Eisler (1968, ch. 1) found the number of words in a phrase, defined as a sequence uttered without breaks (i.e., breaks of at least 250 msec.) to be dependent upon whether the speech sample was rehearsed or spontaneous. In particular, for spontaneous speech she found, "...50% of speech is broken up into phrases of less than three words, 75% into phrases of less than five words, 80% into less than six words, 90% less than ten words, and that phrases of more than ten words uttered with fluency constitute only 10% of speech...". When the same situation (i.e., describing or interpreting cartoons) is repeated six times, "...only 35% of speech is broken into phrases of less than three words, while 50% are less than five words in length, 65% less than six words, 85% less than ten words and 90% less than twelve words. However, even in speech as well-learned as this, phrases with more than ten words are uttered only in 15% of cases." Spontaneous speech then is characterized by fewer words between pauses and there are fewer pauses in repeated material. Taken together with Henderson et al.'s (1966) finding, the pauses that remain in repeated speech are probably grammatical junctures in support of Lounsbury's prediction.

The work so far reviewed has focused upon pausing from the viewpoint of its utility to the speaker. There is a very attractive hypothesis that pausing in speech may be of value for both the listener as well as the speaker. Indeed, the listener may be viewed as taking an active role in the communicative process, utilizing the speaker's pauses in decoding. Goldman-Eisler (1968), as previously discussed, used a "guessing-game" procedure in determining the relationship between a pause and the transition probability of the succeeding word. In one of these experiments, the subject's delays in guessing the missing words were compared to the original speaker's delays (who had generated the sentence in the first place). It was found that in those cases where the subject's pauses were proportional to those of the original speaker there was a much higher likelihood of the "guesser" (i.e., the subject) being correct. In other words, where the pausing patterns of speaker and listener matched, greater comprehensibility seemed to have occurred. Of course this is somewhat speculative since no test for comprehension was administered.

Another study also provides evidence that speaker pause time may be utilized by the listener. Goldman-Eisler and Cohen (1974) had subjects listen to a tape-recorded speech (unfortunately no indication is given of its spontaneity) while performing a mental arithmetic task which was of varying degrees of complexity. It was found that complex tasks tended to be performed when the speaker was pausing only if the pause occurred at the end of a sentence. On the other hand, if the pause was within a sentence the task tended to be inhibited with it being performed primarily during fluent passages of the speech. The authors relate these results to the previous work of Goldman-Eisler (1968) which suggested that speaker pause time represents a period of high information transfer, since the words uttered after pauses tend to have a low statistical probability of occurrence given the preceding context. Within a sentence, speaker pause time represents complex cognitive processing on the part of both the speaker (who is generating these high-information words) and the listener (who has difficulty in anticipating what will be said because of its low probability of occurrence). Fortunately the present experiment did provide a test of comprehensibility in order to be certain that the subjects were paying attention to the speech. In fact the results reported here are for those subjects performing well on the test.

Thus the hypothesis that speaker pause time is utilized by the listener has received some support in the literature, albeit more of a suggestive rather than a definitive nature. It is a fascinating hypothesis and one that deserves further research, and indeed as will shortly become apparent, receives just such attention here.

There are definite problems with the preceding work on spontaneous speech. Firstly, those studies that looked at the distributional constraints found in the language typically stop short of saying anything further. For example, although Maclay and Osgood (1959) provide information on the distribution of pauses (filled and unfilled) they stop short of applying any experimental manipulations in order to ascertain the underlying processes responsible for these distributions. Secondly, much emphasis has been placed on the verification of Lounsbury's (1954) hypotheses which may not have been the most fruitful in studying unrehearsed speech. In discussing his first hypothesis, Lounsbury (1954), suggests that the word as a "unit" be studied. This choice of a unit is a direct outgrowth of information theory and the supposition that speech can be treated as a Markov process, a proposition which is well known to be of insufficient power to describe the grammar of language (Chomsky, 1957). Although the Lounsbury hypotheses also suggest units larger than the word be studied, the research has not been very productive. As we have seen the work of Boomer (1965), which deals with pause-unit contingencies where the unit is larger than the word (i.e., a phonemic clause), has problems, as previously mentioned, since it utilizes a questionable linguistic analysis.

The previous research can also be rather broadly broken down into two distinctive categories: those dealing with the pause patterns of the speaker and those concerned with its utilization by the listener in normal communication. The literature has neglected the latter category thereby leaving out an important component of the total communicative process. The purpose of the present set of studies was to provide some inroads into this largely unexplored region of spontaneous speech.

FOOTNOTES

1. It should be noted that a good deal of research has been done in a clinical setting, typically in attempting to determine the relationship between pausing and anxiety (cf. Mahl, 1956), and will not be discussed here.

2. The probability that a given word occurs at a given location is probably determined by many factors. Among them might be expected to be the number of members of the words class, which for example, is the factor Maclay and Osgood (1959) used. Another factor may be, as suggested in the present dissertation, the constraints present due to the context of what is being said, since certain words would be more likely to occur given a particular speech "environment." Maclay and Osgood only used one of these factors to predict the word probability (i.e., size of the words class) and, hence, may have neglected more overpowering determinants. However, Goldman-Eisler avoided the problem of determining what these factors are by having subjects give an estimate of their combined effect (i.e., the transition probability of a given word at a given location), and so did not overlook any factors which were involved.

3. Boomer (1965) argues against the assumption in Lounsbury's second hypothesis that hestiations co-occur with places of statistical unpredictability since he found pauses to be most frequent at the beginning of phonemic clauses but he argued that infrequent (i.e. unpredictable) words occur toward the end. If this assumption, given in the Lounsbury hypothesis, were deleted, Boomer would (and in fact does) note that the predominance of hesitation pauses at the beginning of phonemic clauses argue for this clause as a unit of encoding, in agreement with the rest of Lounsbury's second hypothesis.

STUDY ONE

Previous research has focused upon the word or the phonemic clause as the basic unit of analysis. Each of these units was chosen with a particular theoretical rationale, and in each case the rationale was essentially linguistic not psychological. In the case of the word (cf. Lounsbury, 1954) the motivation stemmed from the use of a Markov type information theory as a model of language. As is well known this type of approach has been shown to be inadequate in describing the language production process. It therefore would seem to be of little value in analyzing spontaneous speech and it will not be pursued here.

As for the phonemic clause (cf. Boomer, 1965), its motivation stemmed from the belief that phonemic clauses would be unambiguously determinable and would provide insights into the underlying generative process. As we have seen these ambitions were not fulfilled and a phonemic clause analysis will not be pursued, either.

In addition, these units have principly been proposed as encoding units for the speaker. It may very well be the case that units of encoding and those of decoding are different. Indeed, Lounsbury (1954) has distinguished between the two. Since we have rejected the analytical units previously proposed there is an obligation to provide a more functionally and theoretically appropriate one. Rather than commit the same error that we feel previous researchers have committed, we felt it necessary to approach this problem from as purely "psychological" a perspective as possible. To this end we assume initially a totally neutral stance with regard to all psycholinguistic considerations save the essential one: the communication of meaning. We wish to identify a "unit" of analysis and rather then let it be dictated by existing theory, we decided to let our subjects tell us what it may be. The procedures then almost suggest themselves. Provide normal fluent adults with a sample of spontaneous speech spoken by a normal fluent adult and request them to isolate "units" of meaning.

METHOD

Speech sample

A mature, fluent speaker of the language was recorded in a situation where the intent was to convey information to an audience, and the listeners were present with the intention of comprehending what was being said. The speaker was primarily engaged in a monologue, however, there were occassional interruptions which were excluded from the data analyses. The total monologue was over an hour in duration. Approximately twenty minutes were used for the present study of which less than one minute represented scattered listener interruptions, so that the speech corpus was essentially totally composed of the single speaker.

Pause patterns

The pause pattern of the corpus of spontaneous speech was obtained by playing the tape through a voice-operated relay (Model E7300 A-1; Grason-Stadler, West Concord, Mass.). The relay output was channeled through the remote input of a single marker-EEG and polygraph data recording system (Models SMT7BC and 79C, Grass Instrument Co.; Quincy, Mass.), which fed an ink writing oscillograph recording at 10mm/sec. (Model 7WC8PA, Grass Instrument Co.; Quincy, Mass.). In this way a visual record was made of the time course of periods of speech and silence.[1] The actual placement of each segment of speech in relation to this visual record was made by going over the tape and carefully noting the correspondences.

Transcript

A verbatim transcript of the speech corpus was prepared. It was in completely lower case form. Successive verbalizations were placed two spaces apart with triple spacing between lines. All forms of punctuation were omitted. Basically, it contained all of the sounds uttered by the speaker including non-lexical ones (e.g. um, ah, repeated segments of words, stutterings). In effect all of the sounds that were produced by the speaker were included in order to provide a complete record.

General procedure

Subjects. Five sophisticated native English speakers agreed to act as observer-raters in this study. One was a member of the doctoral faculty, three were doctoral students, and the fifth an advanced undergraduate. They were knowledgeable in the area of psycholingusitics and cognitive psychology and were treated as quasi-experts for this task.

Identification of units. The observers were each provided with a copy of the transcript and the tape which they listened to using a Revox tape deck (type A77, Germany) in conjunction with a set of Sharpe headphones (Model HA-10A, New York) modified to play the single channel recording through both earpieces. Their only specific instruction was to consider a unit to be any group of words or non-lexical sounds which went together more coherently than they went with the words or sounds preceeding or following them. That is, they were asked to isolate those groups of words or sounds which together conveyed a single, cohesive thought or idea. The transcript was used to indicate the boundaries of these units by the placement of slash marks at the particular locations. They were permitted to re-play the tape (or portions of it) as necessary. As a conservative estimate, a unit boundary location was defined as any location where at least 80% (four out of five) of the observers had so indicated.

RESULTS

In order to establish that there was a non-random, high level of agreement among our listeners as to the units of meaning, the following analysis was performed. A random model was used to predict the expected frequencies of different degrees of agreement among the observer-raters (ranging from none to complete agreement).[2] A chi-square analysis was performed comparing these expected frequencies with the actual data (Table 1). Since two of the expected frequencies were below the typically used minimum, the calculation of chi-square was modified according to the procedure outlined by Nass (1959) and reviewed by Katti and Sastry (1965). The results were strongly in favor of the conclusion that the observers agreed very highly in choosing unit boundaries (χ^2=9,283.76, df=3.310, *p* is vanishingly small). These results are quite evident from the table where it can be seen that the criterion of defining a unit boundary as occurring where at least four out of the five listeners had chosen one was well justified (note the extreme departure from random expectation for four and five subjects in agreement).

Thus we can be very confident that the units of meaning are quite real and we can now proceed with reviewing their properties. These can be divided into "intra-unit" and "unit boundary" categories. The former will be presented first.

In order to aid in the presentation it will be helpful to define the following terms:

a) When referring to the distribution of pauses in relation to the units of meaning, pauses occurring at unit boundaries will be distinguished from those occurring within units.

b) A period of silence in the speech stream of at least 150 msec.

TABLE 1

Observed and expected frequencies of different degrees of subject agreement that a given location is a unit boundary

No. of inter-observer agreements	Expected Frequency	Observed Frequency
0	561.311	823
1	341.672	45
2	83.150	31
3	10.169	16
4	0.598	24
5	0.015	58

will be referred to as a silent pause.

c) A non-lexical sound (e.g. um, ah, repeated segments of words, stutterings, etc.), or a series of repeated sounds and/or words which do not provide any new linguistic information will be referred to as pause fillers.[3]

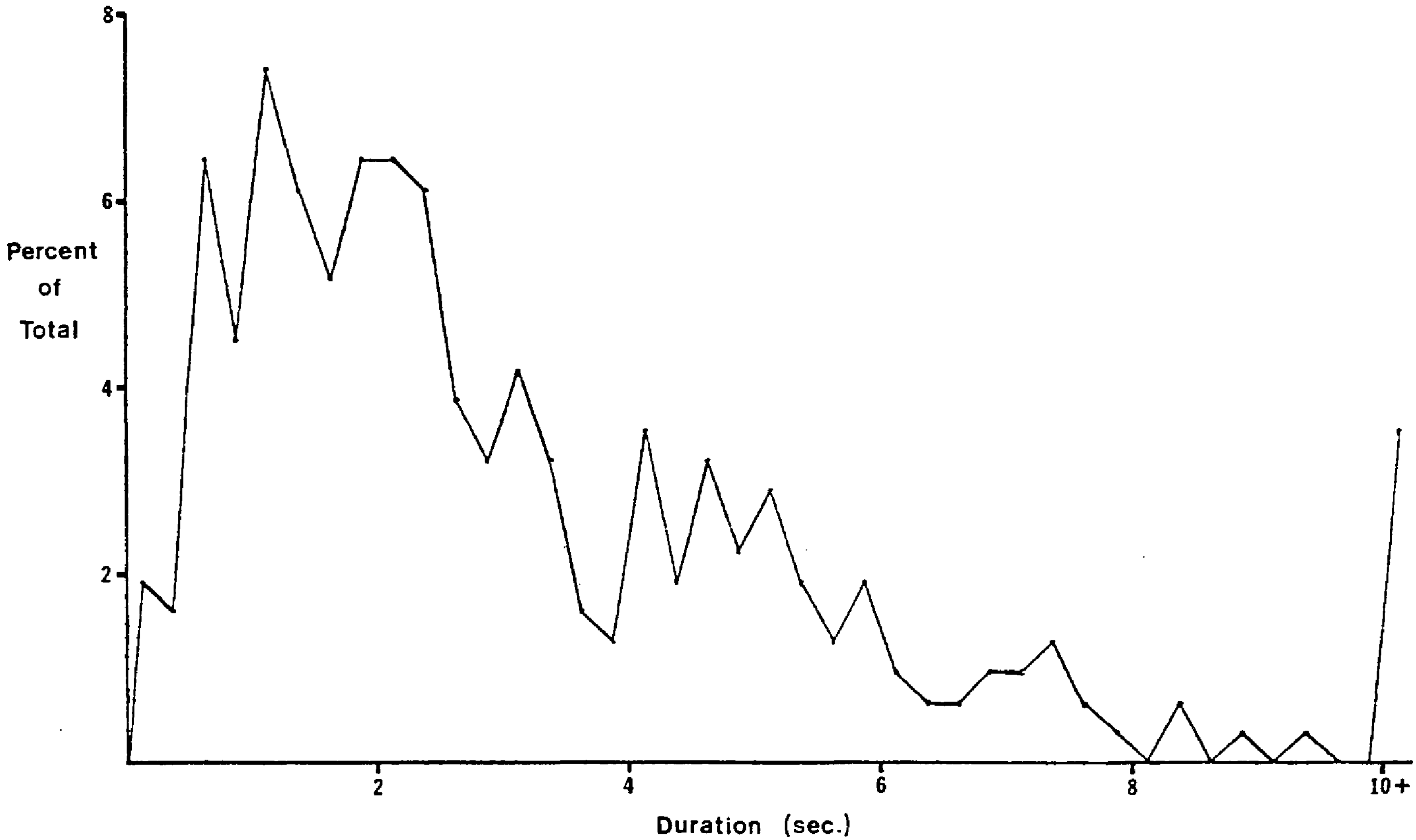
I. Intra-unit properties

This section is primarily concerned with the presence and distribution of pause fillers.[4] It is hoped that these analyses will provide some information concerning their function. Further research will be necessary to fully explicate the matter, since it fell beyond the scope of the present set of studies which were concerned with the communicative role of silent pauses.

A. Percentage distribution of unit durations.

There appears to be a suggestion that there may be an optimal duration for a unit. This can be seen in Figure 1 where the percentage of units falling at different durations is plotted in 250 msec. increments. As can be seen in the Figure there is a preponderance of relatively short duration units. It is unclear if this is due to a restriction on the amount of cognitive complexity that longer units presumably involve, and if so, is it a speaker and/or listener restriction. At this juncture it is unclear what to attribute this result to. It is likely that it may result from some form of limit on cognitive processing. The issue is undoubtedly extremely complex since the limitation could be either speaker dependent or listener dependent (or perhaps both). For example, it is known that parents adjust their utterance lengths when speaking to children and make continuous adjustments that reveal that they are sensitive to listener limitations (Snow, 1972; Newport,

Figure 1. Percentage distribution of total unit durations.


Percent of Total
8
6
4
2
2
4
6
8
10+
Duration (sec.)

1976). Just how pervasive this adjustment process with adults is is not yet known but is assuredly an important issue.

B. Percentage distribution of unit lengths.

If we shift perspectives to the individual items making up a unit an interesting picture emerges concerning the relationship of pause fillers to the units themselves. The percentage of different length units when pause fillers are included can be compared with the corresponding distribution when they are excluded from the length computation (see Figure 2). Interestingly there is the same indication as in Figure 1 that units containing fewer items (in the present case) are preferred. The exclusion of pause fillers would be expected to result in a greater divergence between the two graphs as length increases since there presumbably would have been more pause filler's represented. This is somewhat difficult to see in the Figure, although there is this suggestion.

In order to gain a clearer understanding the corresponding cumulative distributions are presented in Figure 3. As can be seen the two curves do diverge somewhat more toward larger unit sizes. The exact amount of divergence has been computed and is presented in Table 2. The mean differences do increase rapidly toward the beginning and there is a slow drop-off toward the end which suggests that more pause fillers were present in larger units. However, their type of generation (i.e. random or otherwise) is left open.

C. Proportion of pause fillers to total size of unit.

The question of whether or not pause fillers are generated randomly can be approached by examining if the ratio of pause fillers to the entire unit (including pause fillers) changes as a function of length. If it does it suggests that the generation is not random. This comparison

Figure 2. Percentage distribution of unit lengths when pause fillers are either included (●—●) or excluded (●- -●).

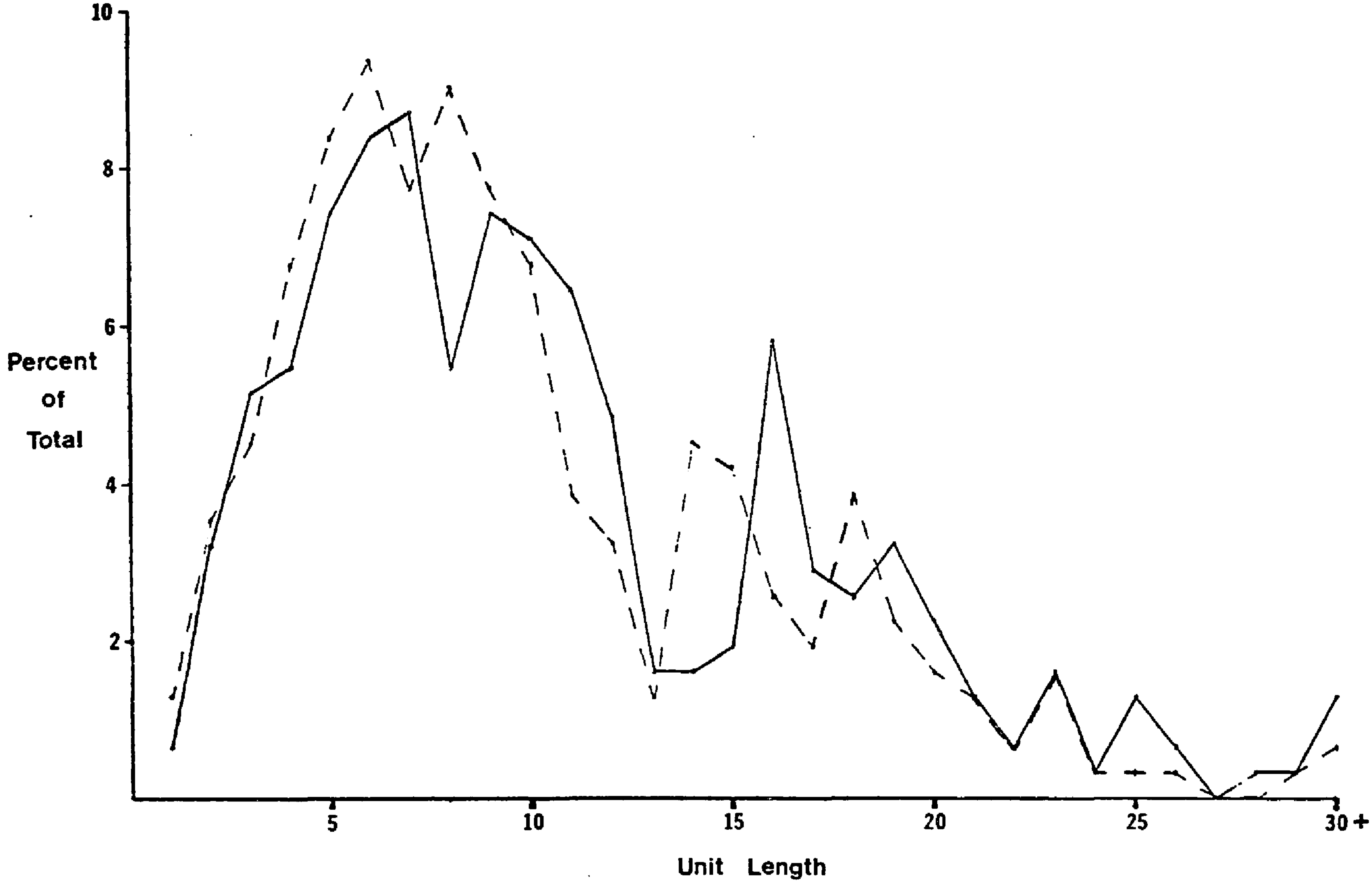

10
8
6
4
2
Percent of Total
5
10
15
20
25
30+
Unit Length

Figure 3. Cumulative distribution of unit lengths when pause fillers are either included (●——●) or excluded (● -●).

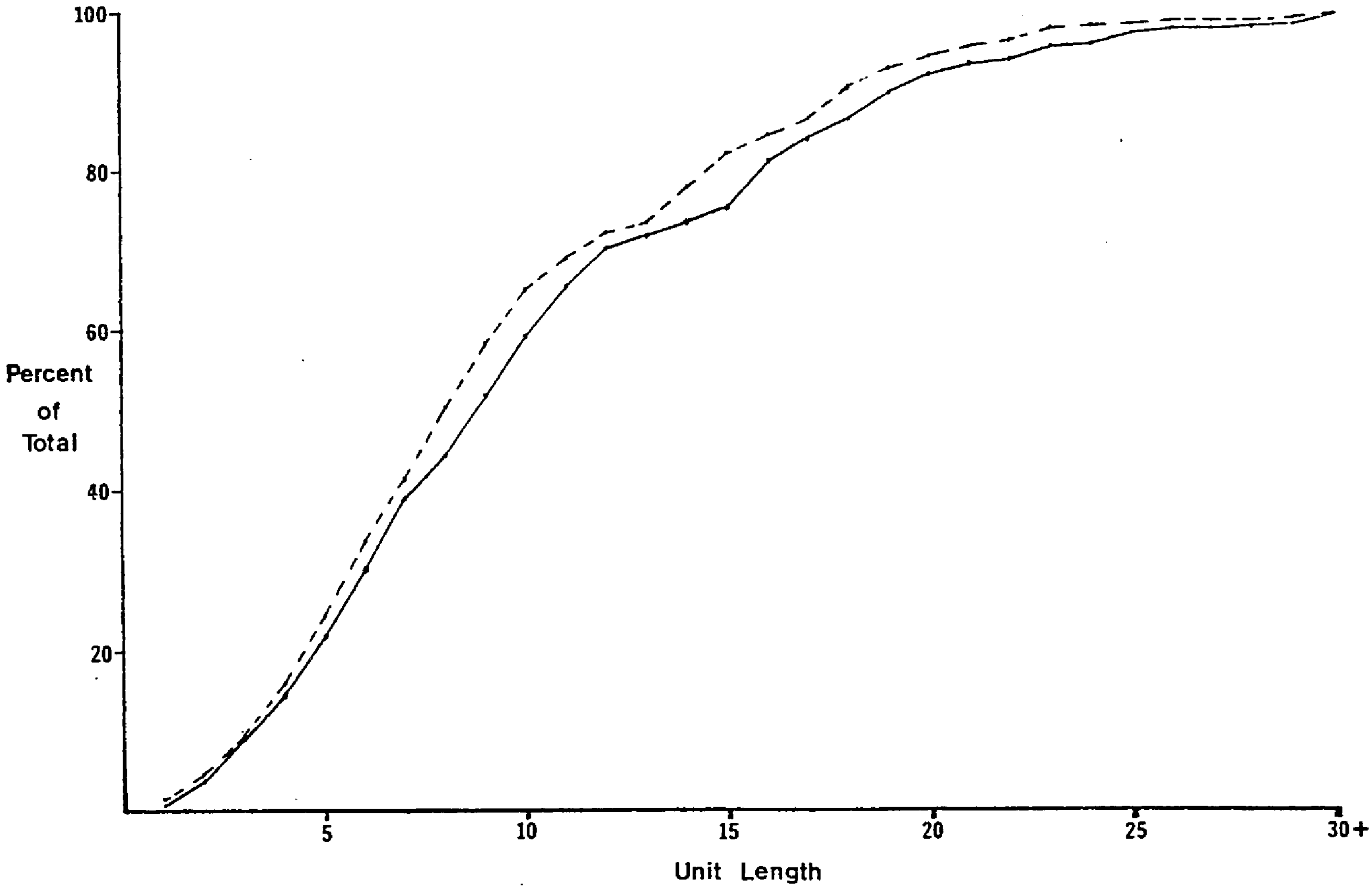

100
80
60
40
20
Percent of Total
5
10
15
20
25
30+
Unit Length

TABLE 2

Differences between the mean cumulative percentage distributions of unit lengths (including or excluding pause fillers, respectively) in blocks of five

Length of Unit	Mean Differences
1-5	1.22
6-10	4.97
11-15	3.68
16-20	3.03
21-25	2.06
26-30	0.71
31+	0.00

was performed and as can be seen in Table 3 the ratio appears to increase somewhat toward the middle of the Table with further increases coming with the very large units. This is not what would be expected with a random model and obviously the suggestion is that the pause filler is a more frequent phenomenon at large unit sizes. If we assume that the pause filler represents computing time on the part of the speaker then the increments with larger units suggests that cognitive demands are greater there due to the speaker having to keep track of more words. Of course this interpretation also assumes that the units which are listener defined are also salient to the speaker. That is that perhaps decoding and encoding units are the same, at least at this level of analysis. Further research in the area of pause fillers is clearly indicated.

II. Inter-unit properties

The present section forms the backbone of this set of studies. The units were here analyzed in terms of the distribution of silent pauses in order to gain a clearer understanding of their role for the listener. It should be noted that only those durations of silence lasting at least 150 msec. were counted as silent pauses.

A. Percentage distribution of silent pause durations at unit boundaries.

What is the nature of the naturally occurring silent pauses that were located at unit boundaries? This question is interesting because it has been suggested that decoding units may be bounded by pauses (Lounsbury, 1954). Figure 4 displays the percentage distribution of silent pause durations that fell at unit boundaries. The distribution is exponential-like, and is characterized by many contributions of pauses of up to a second or longer.

TABLE 3

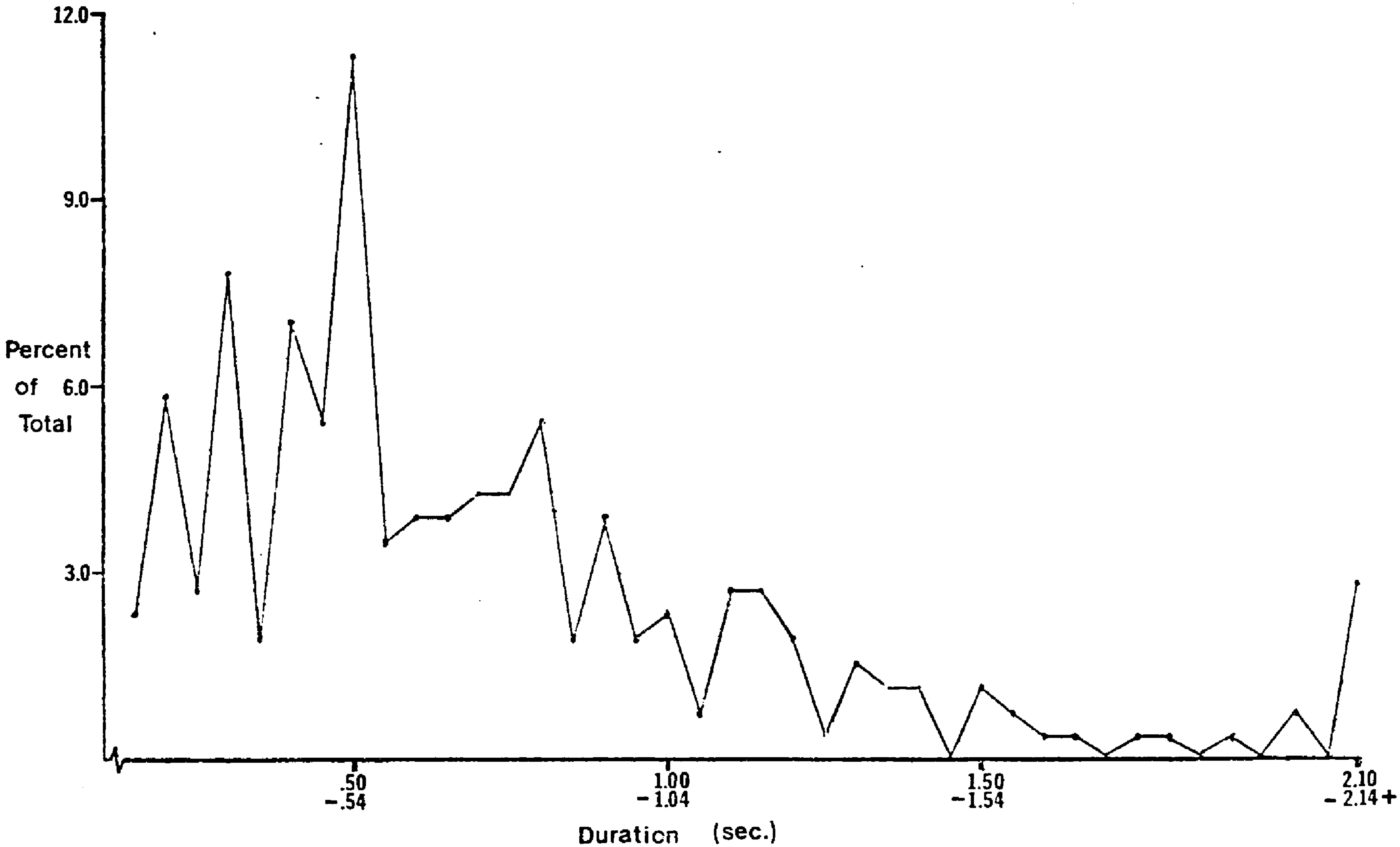
Mean proportions of pause fillers to total number of items in a unit, in block intervals

Length of Unit	Mean Proportion
2-5[a]	.049
6-9	.048
10-13	.094
14-17	.072
18-21	.067
22-25	.084
26-29	.164[b]
30-32	.130[b]
44-46	.156

[a] Unit lengths of one were not included.

[b] Since there were no instances of unit lengths of 27 or 31, these computations were based on the remaining unit lengths.

Figure 4. Distribution of pauses that occurred at unit boundaries.


12.0
9.0
Percent of Total
6.0
3.0
.50 -.54
1.00 - 1.04
1.50 -1.54
2.10 - 2.14+
Duration (sec.)

B. Percentage distribution of silent pause durations at other locations.

If we examine the corresponding distribution of pauses that fell at other than unit boundary locations (i.e. within a unit) it is evident that the two distributions are quite different (see Figure 5). Although also exponential-like, the shorter pauses predominate in the second distribution. In fact, approximately 73% of these non-unit-boundary pauses were less than 300 msec. whereas approximately 72% of the unit-boundary pauses were less than 900 msec. Thus we see that long pauses tended to be more likely to occur at unit boundaries (a point which will be elaborated upon shortly).

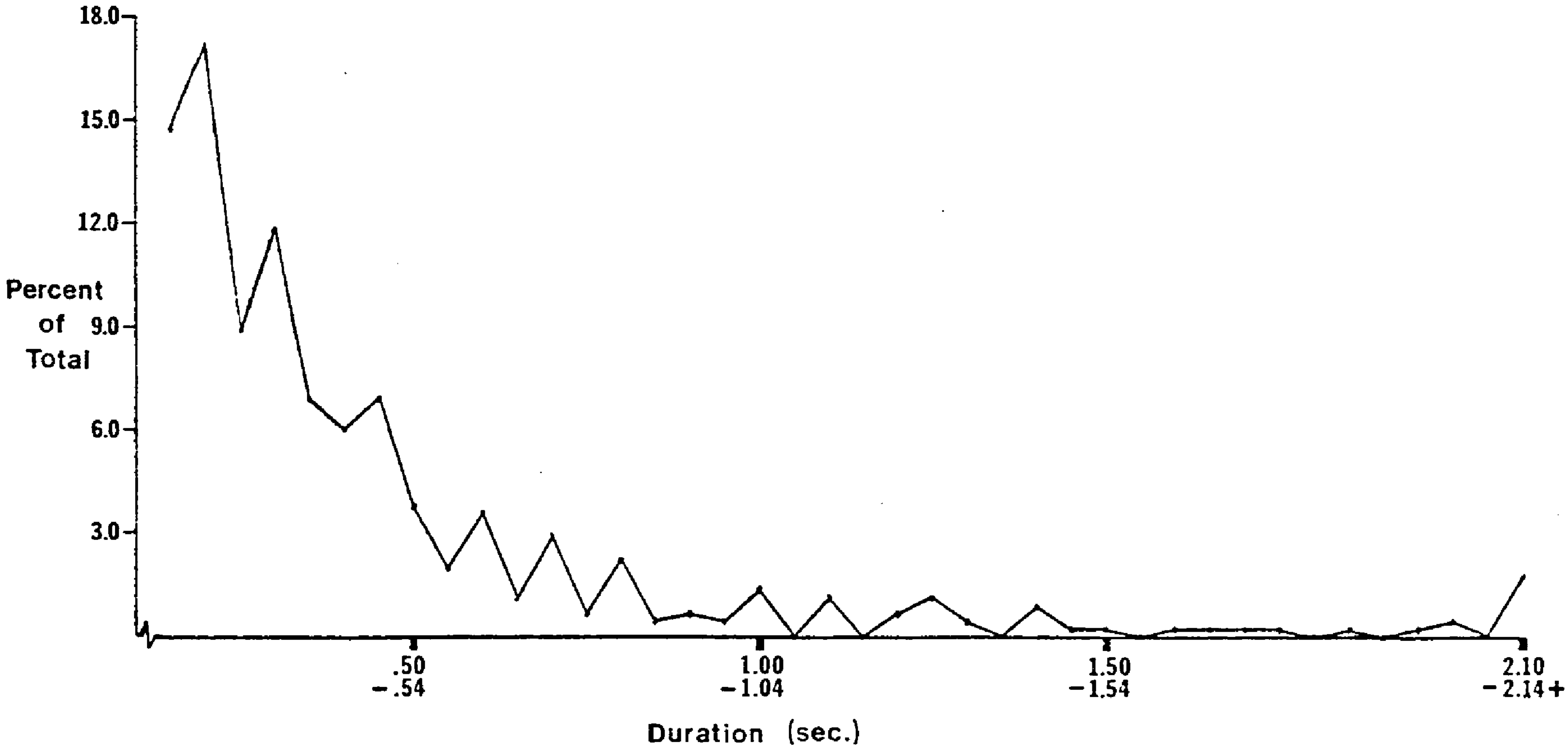
C. Relationship between pausing, syntactic breaks, and unit boundaries.

The task of our observer-raters was to parse the speech corpus into meaningful units. Therefore, we might reasonably expect a relationship between the syntactic properties of the speech and the choice of units by the observer-raters. In addition, Lounsbury (1954) has suggested that a "unit of decoding" might be expected to have pauses at its boundaries (as well as syntactic breaks). We therefore examined the relationship between the pausing and syntax variables, and the choice of units by our subjects.

A two-way chi-square analysis was performed on the frequencies of all possible combinations of unit and non-unit boundaries with major syntactic breaks (or their absence). The result was quite significant (χ^2=1,883.910, df=1, _p_ vanishingly small). Unit boundaries were very likely to have been chosen at syntactic breaks.

Another two-way chi-square was performed on the frequencies of all

Figure 5. Distribution of pauses that occurred at non-unit boundaries.


18.0
15.0
12.0
9.0
6.0
3.0
Percent of Total
.50 -.54
1.00 -1.04
1.50 -1.54
2.10 -2.14+
Duration (sec.)

possible combinations of unit and non-unit boundaries with pauses (or their absence). The pause variable was divided into three components: none (less than 150 msec.), short (150-690 msec.), and long (700 msec. or longer).[5] The data are presented in Table 4 and again the result was quite significant (χ^2=957.317, df=2, p <.05), indicating an important relationship between pausing on the part of the speaker and the choice of unit boundaries by the listener. In order to clarify this relationship specific comparisons were performed using Ryan's procedure (cf. Linton and Gallo, Jr., 1975). The results were that locations that had either long or short pauses were significantly chosen as unit boundaries in comparison to locations where no pauses (i.e. less than 150 msec.) had occurred (χ^2=1,057.545, df=1, p <.05 and χ^2=491.252, df=1, p <.05, respectively). Further, long pause locations had been chosen more often than short pause ones as unit boundaries (χ^2=59.276, df=1, p <.05).

This picture is complicated by the fact that the pause and syntactic variables were not independent of each other, as is evident from the results of a two-way chi-square which was performed on the frequencies of all possible combinations of major syntactic breaks (and other locations) with pauses (using the previous pause partition). The data are presented in Table 5 and here, too, the result was strongly significant (χ^2=663.236, df=2, p <.05), indicating that both variables had been associated. Specific comparisons revealed that locations that had either long or short pauses were significantly likely to also have had major syntactic breaks as compared with locations where no pause (i.e. <150 msec.) had occurred (χ^2=748.744, df=1, p <.05 and χ^2=332.125, df=1, p <.05, respectively). Further, locations having long pauses also

TABLE 4

Observed f(O) and expected f(E) frequencies for the different combinations of the pause and unit variables

Pauses	Unit Boundary Locations f(O)	Unit Boundary Locations f(E)	Other Locations f(O)	Other Locations f(E)
None ($<$150 msec.)	54	245.409	2,579	2,387.591
Short (150 msec. $\leq$ Pause $\leq$690 msec.)	143	47.162	363	458.838
Long ($\geq$ 700 msec.)	113	17.429	74	169.571

TABLE 5

Observed f(O) and expected f(E) frequencies for the different combinations of the pause and syntactic variables

Pauses	Syntactic Boundary Locations f(O)	Syntactic Boundary Locations f(E)	Other Locations f(O)	Other Locations f(E)
None (< 150 msec.)	36	169.411	2,597	2,463.589
Short (150 msec. ≤ Pause ≤ 690 msec.)	98	32.557	408	473.443
Long (≥ 700 msec.)	80	12.032	107	174.968

were more likely to have a major syntactic break as compared to short pause locations (χ^2=37.995, df=1, p <.05).

In sum, the long pause was associated with a significant portion of the listener's unit boundary markings in comparison with no pause or short ones. Short pause locations were also significantly chosen as unit boundaries in comparison with no pause. However, the relationship between pausing and unit boundaries is unclear since major syntactic breaks went with unit boundaries and also were highly associated with pauses (especially long ones). Therefore we cannot determine if the pause variable acted alone or with the syntactic one (or even if pauses acted at all) in establishing unit boundaries. The problem is that pauses and major syntactic breaks naturally co-occurred.

In order to clarify this problem, locations where there were no major syntactic boundaries were examined. If it would be found that pauses were still associated with unit boundaries, it would indicate that the syntactic variable would not have been wholly responsible for the choice of unit boundaries. This analysis was performed by doing a two-way chi-square on the frequencies of all possible combinations of unit and non-unit boundaries with pauses (using the previous three-fold pause distinction). The data are presented in Table 6 where it can be seen that there were obvious differences between the observed and expected frequencies (χ^2=403.451, df=2, p <.05). Specific comparisons revealed that both long and short pauses were significantly more likely to co-occur with unit boundaries in comparison to locations where there were no pauses (χ^2=454.332, df=1, p <.05 and χ^2=176.613, df=1, p <.05, respectively).[6] In addition, long pause locations were also chosen more often as unit boundaries than locations where short pauses were

TABLE 6

Observed f(O) and expected f(E) frequencies for the different combinations of the pause and unit variables, at locations where no major syntactic boundaries had occurred

	Unit Boundary Locations		Other Locations	
Pauses	f(O)	f(E)	f(O)	f(E)
None ($<$ 150 msec.)	26	92.631	2,571	2,504.369
Short (150 msec. $\leq$ Pause $\leq$ 690 msec.)	50	14.553	358	393.447
Long ($\geq$ 700 msec.)	35	3.817	72	103.183

(χ^2=24.275, df=1, p <.05). Thus, the same results were obtained at locations with no major syntactic breaks as had been obtained when they were included. The obvious conclusion follows that the syntax variable was not the sole factor in the listener's choice of unit boundaries.

Given the relationship between pausing on the part of the speaker and unit boundary selection by the listener, the question naturally arises if indeed the syntax variable played any part in this determination, since there was a high association between pauses (especially long ones) and major syntactic breaks? In order to answer this question a two-way chi-square analysis was performed on the frequencies of all possible combinations of unit and non-unit boundaries with major syntactic breaks (or their absence), where only those locations that had no pause (i.e.< 150 msec.) were included. The results were significant indicating that the no-pause locations that contained a major syntactic break were more likely to be chosen as a unit boundary than were non-syntactic locations (χ^2=1,004.060, df=1, p <.05).[7]

Overall, the pause and syntactic variables, separately, exhibited an important relationship with the listerner's determination of unit boundaries. In the case of pauses, this effect was especially pronounced for long durations of silence. These results were obtained by excluding from the total data either those pauses that had co-occurred with major syntactic breaks and then examining the effect of the pause variable only where there had been no such linguistic cues, or by excluding those major syntactic breaks which had co-occurred with pauses and then examining the effect of the syntactic variable using the remaining syntactic boundary locations. The original analyses, which had included all of the data, had indicated that both the syntax and pause variables had been important determinants of the listener's unit boundary selections. The second set of analyses (which had had the aforementioned data excluded) also had indicated that both variables, separately, had been important. The remaining question is, was the data which was used in the second set of analyses solely responsible for the significant findings observed when all of the data has been pooled? That is, what is the relationship between the syntax and pause variables that is found at only those locations where both co-occur? In order to answer this remaining question, two two-way chi-square analyses were performed. One examined the relationship between the frequencies of all possible combinations of unit and non-unit boundaries with major syntactic breaks (or their absence), at only those locations that had a pause. The results were still significant (χ^2=356.058, df=1, $p < .05$). The other analysis looked at all possible combinations of unit and non-unit boundaries with pauses (using the previous three-fold pause distinction) at only those locations where the syntactic variable had co-occurred. The data are presented in Table 7 and the resulting chi-square was significant (χ^2=15.825, df=2, $p < .05$). Rather than using Ryan's procedure for specific comparisons (as was done in the previous analyses), Fisher's exact probability test was used since otherwise each comparison would have violated the restriction on minimum expected frequencies to be used with chi-sqaure. The results were that both short and long pause locations had been significantly chosen as unit boundaries ($p < .05$ in both cases) in comparison to locations where no pause (i.e. less than 150 msec.) had occurred. However, there was no significant difference between long and short pauses in terms of their co-occurrence with unit boundaries ($p > .05$). The reason for this last result appears to be due to a "ceiling effect." That is, the

TABLE 7

Observed f(O) and expected f(E) frequencies for the different combinations of the pause and unit variables, at locations where only major syntactic boundaries had occurred

	Unit Boundary Locations		Other Locations	
Pauses	f(O)	f(E)	f(O)	f(E)
None ($<$150 msec.)	28	33.477	8	2.523
Short (150 msec. $\leq$ Pause $\leq$690 msec.)	93	91.131	5	6.869
Long ($\geq$ 700 msec.)	78	74.393	2	5.607

presence of even a short pause at a location which also contained a major syntactic boundary practically assured its being chosen as a unit boundary (approximately 94.9% of such locations were chosen). Therefore there was very little room for improvement, although it did increase slightly with long pauses to 97.5%.

In sum, both syntax and pauses played important roles in the listener's choice of unit boundaries. This was the case when either variable was isolated from the other as well as when they co-occurred at a given location.

DISCUSSION

The silent pause naturally co-occurred with many major syntactic boundaries and both were found to be good predictors of unit boundary locations. This was found to be the case when either variable was in isolation and when in combination. Lounsbury (1954) has hypothesized that syntactic boundaries (especially those which co-occur with pauses) will be found at decoding boundaries. Obviously the syntactic boundary would be expected to end a group of words which convey a related "chunk" of information. The finding that the units of meaning were highly related to syntactic boundaries suggests that they were taken as decoding units by the observer-raters. We can view the listener as having to decide whether or not a preceding group of words is to be understood as conveying a single thought or idea. These choices will occur at those points in the speech stream where an affirmative decision is probable. Such locations might be expected to be found when the speaker pauses, which could be taken as an indication that a new thought is about to begin. Should the location be at a major syntactic boundary, the listener is more likely to be correct if he decides affirmatively. This conclusion is strongly indicated by the data which showed that pauses were likely candidates for unit boundary choices, especially if a syntactic boundary was also present.

The further course of this investigation requires that the silent pause be unconfounded from its naturally co-occurring linguistic counterpart, the syntactic boundary. Some information in this regard was obtained by partitioning the data into portions which selectively excluded those locations that had both a pause and a syntactic boundary. The resulting analyses examined the pause and the syntactic boundary as separate "attracters" of unit boundaries, with the affirmative result that both are quite effective. Thus, there is definite evidence that pausal information was being used by the observer-raters in their decisions. This is an entirely reasonable result given the high level of natural concurrence between pauses and syntactic boundaries, which would allow usage of the pause as a reliable indicator. However, to fully explicate the underlying processes inherent in natural communication, this natural confounding between the pause and the syntactic boundary has to be eliminated. Nevertheless, it is important to retain the natural properties that define spontaneous speech or else the very process of tightly constraining the stimulus parameters will have destroyed the very essence of the phenomenon that we are concerned with. The philosophical approach here followed represents a melding of a traditional structural linguistic investigation with its emphasis on the micro-detailed analysis of a well defined stimulus source with that of the more empirically oriented philosophy of experimental psychology.[8] This amalgam also crosses the borders of other more traditionally sociological areas. However, given the nature of spontaneous speech, its role in communicative and social processes, this approach quite naturally follows. The area under present investigation, then, is the monologue situation in which the speaker is a highly informed individual in the topic area. Dyadic interactions are, therefore, excluded from consideration. The concern of the present research is directed at the listener's utilization of the speaker's pauses.

The speech corpus chosen for investigation is very representative of speech obtained in natural settings, thus strongly favoring its choice for further study. Schwartz and Jaffe (1968) provide useful information in this regard. They have noted that durations of speech as well as durations of silence are of the same general mathematical form which they suggest is the exponential. Hargreaves (1960) has noted that a better approximation to speech durations is obtained by sampling from different exponential distributions rather than a single one. Taken together with the findings by Schwartz and Jaffe (1968) that speech unit durations match silence durations, the distribution of periods of silence obtained in the present speech corpus is found to be of the same exponential-like form, as was pointed out in the results section. This type of distribution indicates that the long pause is less frequent than the short one and, as was noted, is also more informative of a syntactic boundary than the short pause is.

One of the distinctive forms found in spontaneous speech is the filled pause (as contrasted with the pause filler - see footnote 4, page 51). As noted previously in the results section, it has important implications for the processes involved in speech production. In particular, the filled pause may be a useful indicator of "computing" time on the part of the speaker, a position which also receives some support from the present study. The fact that there is normative data available on filled pauses in spontaneous speech provides further clarification on the representativeness of the present corpus. Maclay and Osgood (1959), as previously discussed, examined the production of filled pauses in a naturalistic setting. They found that the mean number (averaged across different speakers) emitted per 100 words was 3.87. The same analysis was performed on the present speech corpus. The results were essentially a complete replication of the Maclay and Osgood finding with the present number of filled pauses (per 100 words) averaging 3.81. The present speech corpus, then, exhibits physical and verbal properties that have been shown to be reliable indicators of spontaneous speech across a variety of sources.

The next step in this investigation is to experimentally manipulate some of the natural co-occurrences of silent pauses with syntactic boundaries. In so doing the underlying "essence" of spontaneous speech must also be preserved. An obvious method almost suggests itself. The pauses can be manipulated artifically by physically removing them from the tape via splicing and then re-distributing them to designated portions of the same tape. In this way the only difference that will occur will be the distribution of pausing with all other stimulus properties remaining unchanged. This procedure will effectively eliminate the natural confounding which has been observed and will allow the silent pause to be investigated without the ambiguity inherent in the natural environment. At the same time the procedure is conservative in that the essence of spontaneous speech will be preserved.

FOOTNOTES

1. A pause was defined as occurring if at least 150 msec. of silence had elapsed.

2. Basically, we assumed each observer responded randomly with an equal probability across observers of marking a given location. We pooled the observers' responses (using approximately every third page of the transcript) to estimate this probability and then used the binomial distribution to generate the probabilities of from none to all observers choosing the same location. From these probabilities the expected frequencies were generated.

3. In defining repetitions of words (or parts) certain instances required arbitrary decisions. Take for example a not atypical segment "...it became it first of all..." The second "it" was not counted as a repetition since it might well have been part of introducing another idea instead of being the start of "...it became it became."

4. It should be noted that the term "filled pauses" has typically been used to include only non-lexical sounds such as "um" or "ah." Maclay and Osgood (1959) have suggested that they may function as "floor holders" in dialogue situations when the speaker is in need of time to decide how to proceed. Filled pauses however are well practiced and their presence in monologue situations would not be very surprising, although this "floor holding" function would be unnecessary.

5. Pauses which occurred within a word were excluded from this and the remaining analyses in this study since they would not have been chosen as unit boundaries.

6. The former chi-square analysis was performed where one of the expected frequencies was less than five which is the traditional lower limit for use with chi-square with one degree of freedom. However, Cochran (1954) has recommended the use of chi-square in cases where the total N is greater than 40. Since we easily met this criterion we proceeded. Further, as shall next be presented, another chi-square analysis (which met traditional expected frequency standards) demonstrated that long pauses were more likely to be associated with unit boundaries than short ones. Therefore, given the fact that short pauses were more likely to co-occur with unit boundaries then no pauses were, it can be safely concluded that long pause locations were more often chosen as unit boundaries in comparison with locations that did not have a pause.

7. These findings must also be qualified since there was a failure to meet traditional minimum expected frequency standards. Again, Cochran's (1954) recommendations were followed.

8. For a well illustrated example of one type of extensive analysis of a stimulus source see Pittenger et al. (1960).

STUDY TWO

One of the major findings of Study 1 was a relationship between the occurrence of long pauses in natural speech and the demarcation of units of meaning by the listener. However, in the natural environment of the spontaneous speaker, many variables may be confounded. Specifically, other variables such as surface syntactic structure, semantic subtleties and listener decoding operations may also be operative. To take an example, it is simply unclear at this point whether or not the listener even utilizes the speaker's pauses in decoding and ultimately comprehending the message. It would not be unreasonable to take as a starting point the assumption that a listener naturally assumes that when a speaker pauses it's because he has just finished outputing a given "idea" and therefore that the preceding few seconds of speech (since perhaps the last pause) are to be taken as a "unit" and understood as such. Such an arrangement would not be unexpected since in the natural stream of speech the listener would always have to "wait and see" at those points in the speech stream where the preceding context indicates that what has just been said may be complete, unless there were some other way to determine this. If pausing is this "other way" then it would indicate that the functional utility of pausing for the listener is to act as a guide in decoding natural speech. Study 1 did demonstrate a relationship between pausing on the part of the speaker and its utilization by the listener in demarcating units of meaning. However, there was no attempt made to unconfound the pause variable from any other that may have been present that also co-occurred with units. As we saw, syntactic boundaries were also prime candidates for unit choices. Therefore the present study was designed to artifically manipulate the pause variable independent of the syntax.

METHOD

Speech sample

A three minute and forty-seven second portion of the corpus from Study 1 was used here. The sample had several important factors that led to its choice. First, it had a wide distribution of pausing represented so that both relatively long and short pauses could be altered. Second, it contained a large amount of factual information of a kind which the typical undergraduate in an introductory psychology course would not know, so that it is appropriate for use in a test of comprehension. See Appendix A for a full transcript of the corpus.

Two forms of the sample were used in this study. The first was an essentially unaltered version, referred to here as the Control tape[1], the second was "doctored" by the deletion of some pauses and the addition of others, referred to here as the Experimental tape.

Experimental tape. The entire tape segment contained 149 pauses. For the sake of any possible warm-up effects, the last 116 pauses were chosen to be within the critical area of pausal manipulations. Within these 116 pauses, 13 were selected to be reduced to 200 msec. or less[2], and five to be lengthened by splicing in "silence." In addition, four word boundaries that were uttered without pausing in the original corpus had pauses longer than 250 msec. spliced in. These alterations were carried out by a careful splicing of the tape and the use of a dolby noise reduction unit (model 101, Advent Corp., Cambridge, Mass.) to reduce extraneous noises. A check run with several observers showed that they could not detect that the tape had been doctored. The total playing time of this tape was three minutes and forty-seven seconds which matched the control version. The exact pause manipulations were made with the following considerations in mind:

First, the decision was made to leave at least one unit boundary (as indicated by our listeners in Study 1) untouched between any two deletions.[3] This criterion was used with additions, also, to minimize the possible interactive contaminations between our manipulations and the intrinsic structure of the speech corpus. Note however that this spacing criterion is a within-class restriction. It did not prevent us from closely juxtaposing a deletion and an addition. Indeed of the nine, six additions were within four words of a deletion. Of interest here is the possibility that the closer an addition is to a deletion the more likely it will be to attract unit marker responses.[4]

Second, the exact deletions and additions made were selected with an eye to evaluating the various functions that pauses may play. Let us take the deletions first.

a) Deletions. All twelve deletions were placed at locations which had been chosen as units by the observers in Study 1.

Seven of the deletions were located at boundaries that were both "unit" and "syntactic," as those terms were used in Study 1. They ranged in duration from 750 msec. to 1,100 msec. in the original speech corpus and were reduced to 200 msec. or less (see footnore 2). If it is important for the listener to have a period of silence coincide with the boundary point of a major syntactic and semantic unit then these deletions should adversely affect the listener's comprehension.

Three of the deletions were pauses (1,170, 1,200, and 1,210 msec. in the original corpus) that co-occurred with points that were "unit" boundaries in Study 1 but were not major syntactic breaks. Here we are interested in trying to separate out the contributions of syntax and the intuitively meaningful units our subjects indentified in Study 1.

The two remaining deletions removed a filled pause and a sequence of a filled pause followed by a silent pause. The filled pause is an extraneous verbal gesture (um, ah) which is so common in spontaneous speech. Each of these deleted filled pauses was located immediately adjacent to a unit boundary (see Appendix B and footnote 3).

b) Additions. Four of the nine additions were made at locations where there had been no pause time in the original corpus. The remaining five had pause durations ranging from 300 to 810 msec. in the original corpus. None of the nine additions was made at either a unit or a syntactic boundary because of the following considerations:

First, we were interested in determining the role of pauses in listener comprehension. The analyses presented in Study 1 indicated that an alteration of the normal sequence of pauses might affect the placement of marks by the individual subjects. Since certain locations (identified as unit boundaries in Study 1) had already been demonstrated to co-occur with pauses, the addition of extra pause time at these locations would seem superfluous.

Second, Study 1 had indicated that given that a syntactic boundary had occurred, the probability was high that a unit had co-occurred at that location. In addition, pauses were also found to co-vary strongly with the syntactic variable. Therefore, these syntactic-pause-unit locations also suffered from a ceiling effect in terms of "attracting" the marks of the listener and adding extra pause time there would not be expected to provide useful information.

By placing the nine additions at non-unit and non-syntactic locations, the importance of the pause alone could be assessed. (See Appendix A for the exact locations of all pause manipulations.)

Practice tape. A third tape was prepared, of another speaker reading from the transcript prepared for Study 1 of a segment appearing elsewhere in the tape of Study 1. It was given to all Experimental and Control subjects as a "warm-up" aid and ran for one minute and fifty seconds.

Transcripts

Complete transcripts of the corpus including all sounds uttered by the speaker were prepared according to the same procedure used in Study 1. All subjects, of course, used the same transcripts regardless of whether they listened to the Control or Experimental tape.

Comprehension test

A sixteen item true-false comprehension test was prepared from the material covered in the speech sample. In addition to indicating whether a statement was correct or incorrect, each item required the subject to rate his degree of confidence on a five point scale, from "completely guessing" to "absolutely sure." (See Appendix G for test items.)

Procedure

Subjects. Eighty-nine native, male and female, English speakers took part in the experiment. They were all students in introductory psychology courses, at Brooklyn College.

General procedure. The subjects were split into the following three groups: Control (N=39), Experimental (N=28), and a Transcript Control (N=22) which was run using only transcripts, no tape was used. The reason why this latter group was needed is outlined below.

The subjects were group tested in quite classrooms.[5] The procedure for the Control and Experimental groups was the same in all respects (except the tape, of course). The subjects were told that they would hear several tape recordings while they had a transcript of the speech in front of them. Their task was to indicate on the transcript which group of words together expressed a single thought or idea, by their placement of slash marks which would act as demarcators of these "units." They were also told to let the tape recordings be their guide and the transcripts only a means for indicating their choices. They were cautioned that once they placed a slash mark down at a particular location they should not change it (this task requires continuous attention and speed and any attempt at changes would have left them hopelessly behind the tape). The experimenter then answered any questions that they had (for a complete description of the prepared instructions read to the subjects, see Appendix C).

Then a short warm-up tape (1 min, 50 sec.) was played and the corresponding transcripts marked by the subjects. Afterwards, the warm-up transcripts were collected and the other one handed out. The second tape was played three times, with subjects indicating their choices each time using a different colored pencil so as to allow us to determine which units had been marked on each playing. They were told to feel perfectly free to indicate their choices without regard to their previous decision(s).[6] After the final playing, the transcripts were collected and the comprehension test handed out (it should be noted that the subjects were unaware that they would be given a test of comprehension).

The Transcript Control group was included in order to rule out the possibility that the subjects were ignoring the tape and basing their responses solely upon the written material. If this indeed turned out to be the case we would expect no differences among the three groups. The procedure for the Transcript Control group was essentially the same with the exception that no tape was played. The subjects were told to mark their choices on the transcripts, however, without the rapid pace of the tapes to guide them, these subjects took longer at the task. They spent approximately six-seven minutes on the practice transcript and approximately eight minutes on the other transcript the first time through it. The second and third times through it, however, they spent approximately the same time as the other two groups. As will be seen in the results section, this extra time did not really place them above the other two groups in terms of performance on the comprehension test.

RESULTS

Part I - Transcript data

One of the difficulties in working with naturalistic materials such as spontaneous speech is the lack of suitable statistical analysis techniques. In the present case, each of the deletion and addition locations had to be analyzed separately since there was no legitimate way to combine them given that the same subject was free to contribute to each. Each of the three groups was compared with the other two in tests of the significance of the difference between two independent proportions. Out of the twelve deletions, only ten are included in the following analyses since the other two involved filled pause manipulations and therefore are not applicable in explicating the role of silent pauses. It should be noted that given the sample sizes that were readily available, the power of the tests differed depending upon which groups were being compared. Power also varied depending upon whether or not a continuity correction factor was required in a particular test. The tests were powerful enough to detect a minimum difference in proportions of 0.24 between the Experimental and Control groups, 0.28 for the Transcript Control-Experimental comparison, and 0.26 for the Control-Transcript Control comparisons. These values changed, if a continuity correction factor was used to 0.27, 0.32, 0.30, respectively.

First, let us examine the Control-Experimental comparisons. Table 1 shows the proportion of subjects who marked each of the nineteen critical locations as representing a unit boundary. Those classified as "deletions" were positions that were all marked as unit boundaries in Study 1. Of these ten, the first tabled seven were also positions which corresponded to syntactic boundaries. Those classified as "ad-

TABLE 1

Proportion of subjects in the Control (C) and Experimental (E) groups marking unit boundaries at each of the critical locations; the first seven deletions were at syntactic boundary locations

Deletion				Addition			
Location Number	Group C	Group E	Z	Location Number	Group C	Group E	Z
1	.46	.25	1.77	1	.15	.00	b
2	.46	.18	2.15*a	2	.15	.46	2.51*a
3	.62	.46	1.23	3	.41	.50	-0.73
4	.74	.46	2.33*	4	.21	.18	-0.04a
5	.46	.46	-0.02c	5	.00	.00	0.00
6	.85	.79	0.31a	6	.00	.00	0.00
7	.82	.75	0.40a	7	.00	.11	b
8	.62	.32	2.37*	8	.08	.14	b
9	.41	.32	0.74	9	.18	.46	2.24*a
10	.59	.43	1.30				

* $p < .05$

a Correction for continuity was necessary on these comparisons; here sign of Z is not necessarily indicative of the difference in proportions.

b Fisher's exact probability test was run on these comparisons because of failure to fulfill requirements of Z. None were significant.

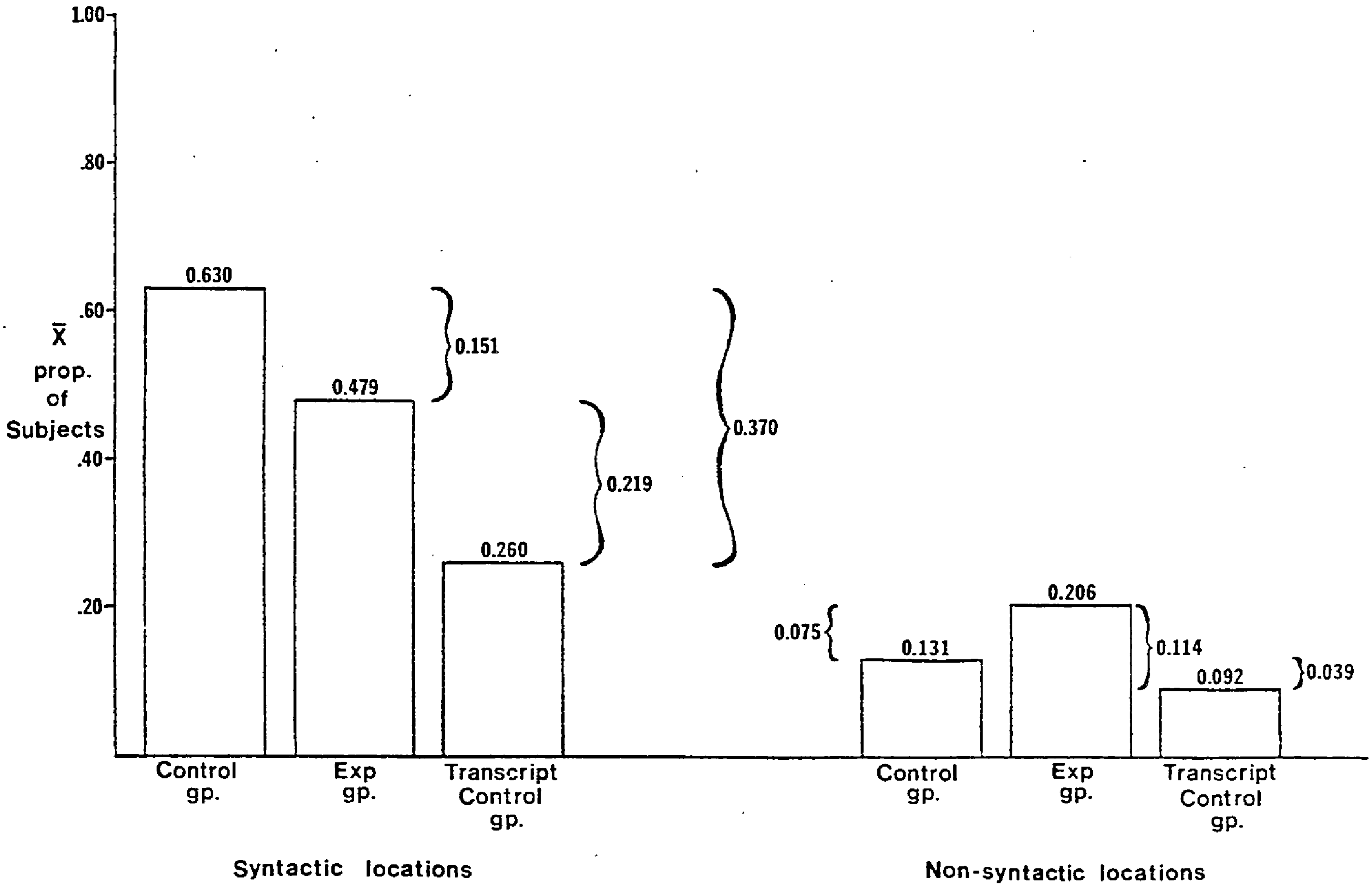
c Slight differences between proportions were responsible for $Z \neq 0$.

ditions" were all locations which were not marked as units in Study 1.[7]

Initially note that certain patterns emerge. Those locations from which pauses were deleted still attract a relatively large proportion of unit markings, whereas those where pause time had been added are still relatively weak in this regard. However, there are important differences between the two tape versions. Note that in the case of the deletions the proportions are lower with the doctored tape than with the control one, whereas in the case of the additions there are more differences in the opposite direction, as expected (although this latter effect is weaker). As will be recalled, our deletions did not typically result in the removal of the complete pause (see footnote 2) but rather in their reduction in duration. Thus the finding of these experimental effects attests to the importance of relatively long pauses.

In order to gain a clearer understanding of the differential role of the pausing and syntactic variables in guiding the subject's choices a comparison has been plotted of the mean proportion of subjects responding at syntactic and non-syntactic locations, in each group (Figure 1). Since most of the deletions were from major syntactic boundaries, the syntactic means were computed using the appropriate deletion proportions; the non-syntactic means were computed using the additions data. Thus if the pause variable were exerting an influence on the subject's responding we would expect to find a higher proportion of Control subject's (compared to Experimental ones) responding at syntactic boundary locations but a lower mean proportion of Control subject's (compared to Experimental ones) responding at non-syntactic boundary locations. As can be seen in the figure, this is precisely what occurred. A more fine grained analysis reveals that there is evidence of an interactive

Figure 1. The mean proportion of subjects responding at syntactic and non-syntactic locations, for each group.


1.00
.80
.60
.40
.20
X̄ prop. of Subjects
0.630
0.479
0.260
0.151
0.219
0.370
0.075
0.131
0.206
0.114
0.092
0.039
Control gp.
Exp gp.
Transcript Control gp.
Control gp.
Exp gp.
Transcript Control gp.
Syntactic locations
Non-syntactic locations

effect of the syntactic variable. That is, the difference in mean proportions between the Experimental and Control groups was greater at syntactic as compared to non-syntactic locations. We also note that the mean proportions of both Control and Experimental subjects were greater at syntactic locations. Overall, then, the findings indicate that, not surprisingly, both the pause and syntactic variables were important in determining the subject's responses.

When we examine the Transcript Control group we see (Table 2) there were differences between this and the other groups. Overall, these subject's responded rather infrequently at locations which were used for deletions and additions. This is also clearly evident in Figure 1 where it can be seen that there was an additional effect of the syntax variable in the same direction as found in the other groups (i.e., there was a lower mean proportion of subject's responding at non-syntactic locations). The subjects in this group silently read the transcript and it is unclear if they implicitly imposed their own pausing pattern while reading. Therefore, we cannot analyze what variable(s) were responsible for this effect. However, clearly this group differed from the other two, indicating that the presence of the tape was an important factor in determining the subject's responses.

In order to gain a clearer understanding of what variables were effective in guiding the subject's responses, the mean proportion of subjects, in each group, responding at either purely syntactic locations, purely unit ones (as defined by the subjects of Study 1), their combination, or other locations are presented in Table 3.[8] Pausing was effective in attracting the subject's responses independent of syntax or unit structure in the Control and Experimental groups, with all pause mean

TABLE 2

Proportion of subjects in the Control (C), Transcript Control (TC), and Experimental (E) groups marking unit boundaries at each of the critical locations, where the first seven deletions were at syntactic boundary locations

Deletion

Location Number	Group C	Group TC	Group E	Z(C,TC)	Z(E,TC)
1	.46	.09	.25	2.68^{*a}	b
2	.46	.18	.18	1.91^{a}	b
3	.62	.41	.46	1.55	0.10^{a}
4	.74	.55	.46	1.30^{a}	-0.57
5	$.46^{c}$	.00	$.46^{c}$	3.50^{*a}	3.39^{*a}
6	.85	.23	.79	4.51^{*a}	3.93^{*}
7	.82	.36	.75	3.33^{*a}	2.46^{*a}
8	.62	.14	.32	3.35^{*a}	1.19^{a}
9	.41	.18	.32	1.54^{a}	0.79^{a}
10	.59	.23	.43	2.73^{*}	1.19^{a}

Addition

Location Number	Group C	Group TC	Group E	Z(C,TC)	Z(E,TC)
1	.15	.00	.00	b	0.00
2	.15	.05	.46	b	2.96^{*a}
3	.41	.59	.50	-1.36	-0.64
4	.21	.00	.18	b^{*}	b
5	.00	.00	.00	0.00	0.00
6	.00	.14	.00	b	b
7	.00	.00	.11	0.00	b
8	.08	.00	.14	b	b
9	.18	.05	.46	b	2.96^{*a}

FOOTNOTES TO TABLE 2

* $p < .05$

[a] Correction for continuity was necessary on these comparisons; here sign of Z is not necessarily indicative of the difference in proportions.

[b] Fisher's exact probability test was run on these comparisons because of failure to fulfill requirements of Z. One was significant.

[c] Slight differences between proportions were responsible for $Z(C, TC) \neq Z(E, TC)$.

TABLE 3

Proportion of times that each given location in the transcript was marked by the subjects in the three groups. P = pause location, NP = non-pause one

	Location Type							
	Unit		Syntactic		Both		Neither	
Group	P	NP	P	NP	P	NP	P	NP
Control	.471	.333	.474	----	.632	.551	.082	.010
Transcript Control	.165	.182	.227	----	.351	.364	.032	.006
Experimental	.335	.036	.446	----	.524	.304	.067	.004

proportions being higher than the corresponding no-pause entries. This finding is further supported by the lack of such an effect in the Transcript Control group, which never listened to the tapes. The number of locations is small in some cases (e.g., there were no cases of syntactic boundaries without a pause and only two instances of a non-pause location being a major syntactic and unit boundary) so general trends are difficult to spot with regard to units vs. syntax. However, note that overall the "Both" condition is uniformly higher than either taken separately so that the general semantic aspects seem to combine with the syntactic frames in controlling subject's responses. These results must be qualified by the fact that there were large variances present and statistical analyses would not be appropriate since the subjects had the opportunity to contribute to the categories one or more times, depending upon the particular subject. Nevertheless, the results are quite suggestive and indeed support the findings of Study 1. In that as well as in the present study, both pausing and syntax were found to play important roles in the subject's responding. Further, in the present study as well as the previous one, those locations which neither had a pause nor major syntactic break received few responses. This further buttresses the conclusion that subjects in both studies were performing the task we had required of them (namely to demarcate those boundaries that set off meaningful "chunks").

<u>Part II - Comprehension data</u>

Each subject had to answer true or false to each comprehension question and had to indicate their degree of confidence on a five-point scale (from 1-"completely guessing" to 5-"absolutely sure"). Each question was analyzed separately in pairwise comparisons among the three groups in two different ways: a test of the significance of the difference between two independent proportions (i.e., the proportion of subjects responding correctly), and a t-test of the significance of the difference between the mean confidence ratings. This separate-question analysis was mandated due to the fact that the same subject contributed to most, if not all of the questions.

The power of the tests differed depending upon which question and groups were being compared due to different N's in each group, loss of occasional data points because of illegible responses, and whether or not a correction for continuity was required. Also, tests of independent proportions are generally less powerful than t-tests which could not be done on these data. The mean and range of the smallest difference in proportions that could be detected were 0.26 (range: 0.24-0.28) for the Control-Experimental comparisons, 0.33 (range: 0.32-0.33) for the Experimental-Transcript Control comparisons, and 0.30 (range: 0.27-0.31) for the Control-Transcript Control comparisons. Power analyses were not performed on the t-tests since they require the variances of the actual responses.[9] This is not true with the tests of independent proportions where we know the maximum possible variances for any sample sizes.

Table 4 shows the proportion of Control and Experimental subjects who correctly answered each of the comprehension test items. As can be seen in the table, the Experimental group performed somewhat more poorly than the Control group (they were less correct on 12 of the 16 items), however, none of the differences were significant. This is not surprising since, as we shall later discuss, the subjects had access to several types of information in answering each question and would not necessarily

TABLE 4

Proportion of subjects in the Experimental (E) and Control groups (C) answering correctly to each of the comprehension questions

Question Number	Group E	Group C	Z
1	.33	.53	-1.54
2	.89	.89	b
3	.50	.57	-0.54
4	.74	.78	0.04[a]
5	.74	.79	0.16[a]
6	.79	.86	b
7	.82	.89	b
8	.32	.30	-0.06[a]
9	.37	.53	-1.24
10	.86	.95	b
11	.82	.86	b
12	.36	.42	-0.53
13	.85	.79	b
14	.86	.76	0.69[a]
15	.75	.66	0.53[a]
16	.86	.87	b

* $p < .05$

[a] Correction for continuity was necessary on these comparisons; here sign of Z is not necessarily indicative of the difference in proportions.

[b] Fisher's exact probability test was run on these comparisons because of failure to fulfill requirements of Z. None were significant.

do poorly. The pause manipulation, however, becomes apparent when we look at the results of the t-tests that were performed on the confidence ratings (see Appendix D for t-test values). Here we find that the Experimental group was less confident of their answers than the Control group on 25 out of the 32 comparisons. A more fine grained analysis reveals that the Experimental group was less confident regardless of whether they had been correct or incorrect in their response. (in fact for 12 out of the 16 correct choices and 13 out of the 16 incorrect ones, they had a lower mean confidence rating). The pause manipulation had resulted in a lowering of the degree of confidence that the subjects had of their answers. These effects came through even though our manipulations had been extremely conservative. As will be recalled, we typically did not fully remove a pause at a deletion location, but rather only decreased its duration (see footnore 2). In addition, pause deletions and additions were made at only a very small proportion of the actual word boundaries which were located within the critical region of the tape (i.e., approximately five percent of all word-boundary locations were manipulated). Therefore, the pause variable exerts a strong influence in the listener's natural comprehension of spontaneous speech.

The Transcript Control group can be compared to the other two groups using the same two types of analyses. Table 5 shows the proportion of subjects in all groups who correctly answered each comprehension test item. As can be seen, there were few differences between the three groups. When we move on to t-tests, again, certain patterns emerge. The Experimental subjects were less confident of their answers (23 out of the 32 tests). In addition, overall, it did not appear to matter whether they had been correct or incorrect in their answers; either way they were

TABLE 5

Proportion of subjects in the Control (C), Transcript Control (TC), and Experimental (E) groups answering correctly to each of the comprehension questions

Question Number	Group C	Group TC	Group E	Z(C, TC)[a]	Z(E, TC)[a]
1	.53	.43	.33	0.72	0.38
2	.89	.76	.89	b	b
3	.57	.50	.50	0.21	-0.29
4	.78	.81	.74	b	b
5	.79	.76	.74	b	-0.17
6	.86	.90	.79	b	b
7	.89	.86	.82	b	b
8	.30	.33	.32	-0.01	-0.22
9	.53	.52	.37	-0.25	0.77
10	.95	.86	.86	b	b
11	.86	.85	.82	b	b
12	.42	.48	.36	0.13	0.55
13	.79	.81	.85	b	b
14	.76	.48	.86	1.87	2.55*
15	.66	.67	.75	-0.22	0.32
16	.87	.95	.86	b	b

* $p < .05$

[a] Correction for continuity was necessary on all Z-test comparisons; here sign of Z is not necessarily indicative of the difference in proportions.

[b] Fisher's exact probability test was run on these comparisons because of failure to meet requirements of Z. None were significant.

less confident. They had a lower mean confidence rating on 11 out of the 16 correct choices and 12 out of the 16 incorrect ones (there was also one case of an incorrect answer where the mean confidence ratings were equal in both groups).[10] The Control subjects, on the other hand, did not differ in essentially any respects from the Transcript Control ones. On almost half of the 32 tests (15 to be exact) the Control mean confidence ratings were lower, with the remaining tests going in the opposite direction. Also, whether or not the subject was answering correctly did not make too much of a difference, with the Controls having lower mean confidence ratings on nine out of the 16 correct choices and six out of the 16 incorrect ones. Overall, then the Controls and Transcript Controls did not differ in any real respects (see Appendix D for t-test values).

DISCUSSION

The results of Study 1 indicate that there is a relationship between the presence of pauses in natural speech and the demarcation of units of meaning by the listener. However, since there was also a confounding of pauses and syntactic boundaries it was unclear just what the functional significance of pauses was. This study replicates these general findings and at the very least shores up the reliability of our procedures. The manipulation of pausal patterns, however, allows us to probe further into their role for the listener. Consider the likely interpretation that pauses are used by the listener as an aid in deciding which group of words should be decoded as a unit. When one reads a book there are agreed upon symbols which demarcate which words go together in conveying information. In natural speech the pause (especially long ones) may be a form of "oral-punctuation," setting off the preceding word as a unit for decoding. Indeed, Lounsbury (1954) specifically hypothesized that units derived from an immediate-constituent analysis would correspond to these units of decoding, in particular those where pauses also co-occurred. Thus, the implication is that syntactic boundaries (especially those with pauses) are the most salient units for decoding by the listener. The present experiment, by manipulating the pause variable, permits its unconfounding from syntactic and other naturally co-occurring variables.

The results indicated that indeed the listener uses the speaker's pauses as a cue for decoding. The mean proportions of subjects choosing a deletion or addition location was smaller or larger, respectively, for the Experimental subjects compared to the Controls. The role of the syntactic variable was also found to be important, with the mean proportions of subjects within each group being greater at syntactic locations. In fact, there appears to be an interactive relationship between the presence of pauses and syntax in the listener's unit demarcations, which indicates that both variables are important ones. This redundancy of cues if interferred with, can result in detrimental effects, as evidenced by the decreased confidence ratings in the Experimental group.

Since a pause manipulation will affect the demarcation of units by the subject it suggests that memorial organization of the material by the subject will be affected as well. Indeed, Lounsbury (1954) specifically suggested that an experiment on pause manipulations may be found to affect comprehension. The present study looked at this variable and found that it did not directly affect comprehension but rather confidence in what was comprehended. Why did it exert an effect upon the latter and not the former? In order to answer this question it may be helpful to look at recent models of semantic memory. There is a general class of such models referred to as network models (cf. Rips et al., 1973). The subject is viewed as performing a "parsing" operation upon the inputted verbal string in extracting out the propositions which underly its surface structure. The propositions are the basis of the memory structure for the communication with related propositions being interconnected. Since for a given communication there are typically many such interconnections, the resulting memory base is a highly complex structure.

When the subjects listened to the tape the assumption is that they decoded it into propositions which were interrelated to each other in memory. These links within and among propositions formed an interconnected structure. In natural speech pauses act as cues as to where to decode and typically these will be expected to come at the end of a set of one or more propositions. Normally this arrangement works quite effectively. The subject uses pauses as cues to decode the surface string into its constituent propositions and builds up an interconnected structure of propositions related to the text of the speech. Thus the listener has, quite probably on an implicit level, learned that pausing is an indication of a new idea about to be generated and in fact part of this knowledge is due to the listener's own experience in the role of a speaker. By adopting a state of "pausal empathy" with the speaker, the listener maximizes this communicative process. This hypothesis has received some support from an experiment by Goldman-Eisler (1968) which used a variation of the Shannon (1951) guessing game procedure. Goldman-Eisler had subjects read aloud sentences that had been generated by someone else. The sentences had certain words missing and the subject's task was to guess these words. It was found that a subject was more likely to be correct if the durations of his pausing before the critical words were proportional to those of the original generator of the sentence. Although Goldman-Eisler was primarily concerned here with word-to-word predicability, the implications of the experiment are clearly relevant to the present model of the speaker-listener interrelationship.

If we artifically insert pauses or delete them we will interfere with this memory structure because the listener will decode at inappropriate times and will in certain cases be breaking up underlying propositions. Therefore, the memory structure will have links which are freely "dangling" instead of being connected as they usually would be.

In answering the comprehension test items a subject must search the memory structure he has set up for the relevant information. However, in answering any particular item, it is not necessary that all of the propositional pathways be intact since if an indirect path can be found or if certain alternative answers can be ruled out the item can be responded to correctly. The subject's accuracy greatly depends upon his strategy in searching the memory structure and the degree to which it matches the input information.

It will be recalled that the present study made pause manipulations at only a small proportion of the available pause locations (i.e. there were over 100 pauses). The resultant disruption of the subject's memory structure can be inferred by examining the confidence rating data. Let us propose that a subject will answer a comprehension test statement "true" or "false" by the process previously mentioned. That is by either establishing some propositional route (direct or indirect) to the relevant information or by ruling out alternative possibilities (which may also involve a search through the structure). Now when he is asked to rate the confidence of his answer, he can do so by keeping track of how many pathways exist that lead to the same conclusion. In other words, the more sources of redundancy the more certain he can be of his answer. If we interfere with the organization of the memory structure by cueing the decoding of the speech input at inappropriate locations, there will be fewer pathways leading to a given conclusion and thus the subject exhibits less confidence in his answer. The Experimental subjects diminished confidence thus can be interpreted as indicating that their memory structure had not been fully interconnected due to the erroneous cueing information supplied to them by the pause manipulations.

We are evolving here a model of the listener as actively abstracting information from the speech signal and interrelating the information into a memory framework. The listener makes use of the speaker's pauses in decoding since pauses are a good indicator that a group of words should be treated as a meaningful unit. When we disrupt this natural cue by artifically manipulating pauses we interfere with the memory framework being set up because the listener follows the new pause structure in deciding which group of words is to be treated as a unit. Thus, pauses form a very important part of the communicative chain.

The research in propositional memory indicates that immediately after hearing verbal material, listeners have available two forms of information stores: the surface structure of what was actually said and the underlying propositional structure representing the meaning of the communication (Anderson, 1974; Begg and Wicklegren, 1974; Sachs, 1974). The present study can be viewed as putting the subject into a set to respond based upon this latter form of memory. That is, the insistence that the subject pay careful attention to the underlying meaning of the tape and the procedure's three-fold repetition would seem to have made the comprehension test a task of retrieval from propositional memory; an interpretation which the data supports. However, under suitable task conditions the subject may not be expected to rely upon this information store and it would be interesting to note the results of the confidence ratings given by using this variation.

FOOTNOTES

1. See Appendix B for a precise specification of this tape.

2. Efforts to reduce this beyond 200 msec. tended to produce detectable changes in the sound of the neighboring words (although in a couple of instances essentially complete pause removals were made). Thus, our manipulations here are actually extremely conservative. What we will be referring to as a "deleted" pause is, in most cases, actually a "reduced" pause. The fact that these manipulations produced experimental effects strengthens the arguments we shall make below about the role of pausing in normal speech.

3. For purposes of clarification, a pause manipulation will be defined as having been performed at a location if there is a difference there in the duration of silence between the control and experimental versions of the tape. If the pause time is less in the experimental version, at that location, then it is a deletion, if greater it is an addition. However, there were two manipulations involving a filled pause removal in the experimental tape but not in the control tape. Specifically, at both of these locations the filled pause was followed by a silent pause which was removed with the filled pause. The control tape at these two locations had at one of the locations the silent pause removed, whereas nothing was altered at the other location on the control tape. Both of these locations are considered deletions. With this in mind, of the thirteen pauses which were reduced in the experimental tape, twelve of them are classified as deletions. The one exception is a pause reduction that was necessitated in both the Control and the Experimental tape (see page 103 line 19, of Appendix B).

4. This manipulation turned out not to produce any unambiguous effects and thus is not included in analyses discussed later.

5. Due to problems in securing large numbers of subjects at the same time, where necessary (in the experiments reported in this dissertation) a treatment group (for example, Experimental) was run in smaller groups and the data combined for the analyses. It is also possible that several of the subjects were tested in very small groups as, for example, a group size of two (or even tested alone) and the test environment may or may not have been a classroom but rather the laboratory.

6. In the data analyses only the first run through the tape was used.

7. The Fisher exact probability tests were not included in the power analyses in Studies 2 and 3.

8. The pause variable was obviously irrelevant for the Transcript Control group. Therefore the tabled entries for pause and no pause refer to the corresponding locations for the other groups. Further, in order to allow comparability between the groups, the locations immediately affected by the two deletions which had involved filled

pauses as well as the four addition locations where no pause time had occurred in the original tape, were eliminated from this analysis.

9. The power of the t-test is lower with a restricted range of responses (i.e., subjects responded on a five-point scale), and in fact the rejection of the null hypothesis requires that subject's responses be clustered over a portion of the scale which can result in problems concerning the applicability of the t-test. However, despite these considerations the t-test was the most appropriate form of analysis and was performed in this and in Study 3.

10. However, we will not pursue the conclusion that accuracy did not interact with the direction of confidence ratings since the strongest cases of the Experimental subjects being less confident were all for correct answers.

STUDY THREE

Study 2 found that the subjects who had listened to the Experimental tape were more uncertain about the material imparted by the speaker. The possibility therefore exists that this result can be attributed to stimulus novelty effects. In order to determine if stimulus novelty was responsible, the following study was carried out. The primary modification in the present study was to present the tape stimulus only once, rather than three times as had been the case in Study 2.

METHOD

Speech sample

The tapes used in Study two (i.e. the Experimental and Control tapes) were used in the present study. The warm-up tape of Study 2 was not included.

Comprehension test

The items were the same as in Study 2, and as in that study each item required the listener to indicate whether an item was correct or incorrect, and to indicate their degree of confidence, using a five point scale.

Procedure

Subjects. Eighty native, male and female, English speakers took part in the experiment. They were all students in introductory psychology courses, at Brooklyn College.

General procedure. The subjects were split into the following two groups: Control (N=41) and Experimental (N=39).

They were group tested in quiet classrooms with each group listening only once to their respective tape. They were told to pay careful attention since they would be asked questions afterwards (see Appendix E for the exact prepared instructions read to the subjects). After the playing, they answered the comprehension questions.

RESULTS

The same analyses used in Study 2 for the comprehension test data were also used here. That is each question was analyzed separately in two different ways: a test of the significance of the difference between two independent proportions (i.e. the proportion of subjects responding correctly), and a t-test of the significance of the difference between the mean confidence ratings.[1]

As in Study 2 only the power of the proportions tests will be reported. Due to the fact that the present study employed only two groups, a sufficient number of subjects was available so that we could exclude any subjects who did not fill out the comprehension tests properly (rather than excluding individual questions as in Study 2). The smallest difference in proportions that could be detected was 0.22 (0.24 if a correction for continuity was required).

Table 1 shows the proportion of subjects who correctly answered each of the comprehension test items. As can be seen in the Table there is absolutely no overall difference between the groups. In Study 2 there was an indication that the Experimental subjects had not performed as well as the Control ones; in the present study we find just as many proportions differences going in one direction as in the other.

When we turn to the t-test data we get the same kind of results (see Appendix F for t-test values). Both groups were essentially matched in terms of confidence (specifically, the Experimental subjects were less confident on 15 of the 32 tests with the remaining tests going in the opposite direction). Also it didn't appear to really matter whether or not the response was correct, in both cases both groups are similar.

TABLE 1

Proportion of subjects answering each comprehension item correctly

Question Number	Group		Z
	Experimental	Control	
1	.59	.76	-1.59
2	.82	.80	-0.11[a]
3	.64	.59	0.51
4	.79	.76	0.15[a]
5	.62	.76	-1.36
6	.74	.88	1.25[a]
7	.87	.73	1.29[a]
8	.41	.41	-0.04[c]
9	.49	.54	-0.44
10	.95	.85	b
11	.85	.78	0.47[a]
12	.49	.27	2.02*
13	.72	.80	0.65[a]
14	.74	.68	0.60
15	.36	.66	-2.68*
16	.85	.98	b

* $p < .05$

[a] Correction for continuity was necessary on these comparisons; here sign of Z is not necessarily indicative of the difference in proportions.

[b] Fisher's exact probability test was run on these comparisons because of failure to fulfill requirements of Z. Neither was significant.

[c] Slight differences between proportions were responsible for $Z \neq 0$.

That is, for six of the 16 correct choices and nine of the 16 incorrect ones the Experimental subjects were less confident of their responses.

DISCUSSION

The subjects of the present study did not exhibit a differential degree of confidence as a function of the group which they had been in, whereas in the second study the Experimental subjects had been more uncertain of the material than the Control subjects. There doesn't seem to be any reason to expect stimulus novelty effects only in the second and not in the third study. Indeed, with only one playing of the tape in the present study, stimulus novelty effects would be expected to be more effective than in a situation where the subjects have listened three times and have therefore become more familiar with the stimulus as had been the case in the second study.

It should be noted that these results cannot be due to different levels of overall comprehension between the subjects of the two studies since the mean proportions of correct answers were almost exactly the same for the Control and Experimental groups of both studies. Specifically the mean proportions of subjects responding correctly (obtained from Table 1) for the Experimental and Control groups of the present study were 0.683 and 0.703, respectively. The corresponding mean proportions from Study 2 were 0.678 and 0.715.

On the other hand the propositional model is seen to offer a very convincing explanation for the difference observed between the second and third studies. Research in the area of propositional memory has found that immediately after hearing verbal material, listeners have two types of information stores available: what actually was said (i.e. surface structure memory) and the meaning for what has been said (i.e. deep structure or propositional memory) (Anderson, 1974; Begg and Wickelgren, 1974; Sachs, 1974). The former rapidly decays while the latter is more durable and can influence subject's decisions at later intervals. Study 2 had several interesting features in this regard.

First, the subjects were told to concentrate on the meaning of what was being said (the transcript marking task). Second, they marked the transcripts three times and each time had to concentrate on the meaning of the material. Third, they were uninformed as to the fact that they would be later tested on details that were covered in the material.

These factors lead us to believe that the subjects were placed in a propositional "set." That is, they were primarily concerned with the abstract meaning of the material rather than the particular words being uttered and therefore would be expected to develop a rather extensive propositional network, which they would rely upon when unexpectedly handed a comprehension test. Indeed, the second study's findings support the hypothesis that the pauses exerted their debilitating influence via problems which developed at this level of memory. If this is the case then a change in instructions and procedure might be expected to alter our results. If we set up conditions where the subjects do not pay as careful attention to the meaning, but concentrate upon holding as much information as possible, we might expect them to perform differently. In the present study we simply presented the tape only once without a transcript task and informed our subjects that they would be tested on the material. According to our model of communication, given these conditions the subjects would not be expected to demonstrate the differential degree of loss of confidence that was the case between the Experimental and Control subjects in the previous study.

FOOTNOTES

1. For the statistical and methodological considerations which led to our choice of these analyses see Study 2 - comprehension data.

GENERAL DISCUSSION

The communicative function of language has only recently begun to be studied in its most natural setting - everyday spontaneous speech. One of the most influential and insightful proponents in the area has been Lounsbury whose hypotheses concerning the functional significance of pausing have guided the research in spontaneous speech for many years (Lounsbury, 1954). However, the bulk of these studies have been concerned primarily with the role that pauses play in the speaker's encoding of an utterance. It seems that this one-sided emphasis has been the result of the theoretical issues raised in transformational generative grammar which have also been predominatly speaker oriented. Lounsbury (1954), however, was concerned with the total communicative chain and the role that the speaker's pauses play in the listener's comprehension of the speech. Indeed he specifically suggested that the listener utilizes some of the speaker's pauses in deciding upon the way in which to decode the speech. The resulting decoding units, he suggested, would be an appropriate reflection of the speaker's intentions. The boundaries of these units could be ascertained by noting the points in the speech stream which an immediate-constituent analysis would designate as boundaries, particularly if the speaker had chosen to pause at these points as well. Therefore, the role of pausing is not strictly a speaker oriented one but rather also includes the listener as well. After all the listener can simply be viewed as a speaker in another mode, and therefore can be considered to have all of the requisite knowledge, probably of an implicit nature, concerning the placement of pauses between separate ideas. It would seem rather peculiar that this should be otherwise given that the function of language is principally to communicate.

Human communication should not be expected to be less efficient than that of the rest of the animal kingdom where the participants generally are found to be quite naturally well organized in this respect.

The present set of studies focused upon the listener side of the communicative chain. Rather than examine this process by using analytical units which were primarily designed with the encoding problem in mind, a new approach was undertaken which originates from as purely a psychological perspective as possible. Several sophisticated observer-raters listened to a rather lengthy recording of a mature, fluent speaker of American English in a situation where communication was the primary goal. The observers provided units whose sole criterion of selection was that they represent single, meaningful thoughts or ideas. As was seen there was an exceptionally high degree of reliability among the raters in the selection of these "units" of meaning. Indeed, as Lounsbury (1954) suggested, the units were found to be highly related to the overall syntactic structure of the speech corpus. In addition, the silent pause was found to be an excellent predictor of the unit's boundaries especially if the pause co-occurred with a major syntactic break.

The speech corpus used in these studies was examined in terms of its properties in relation to the literature on spontaneous speech. It was found that the corpus was highly representative of similar corpuses obtained from a wide variety of speakers, both in terms of physical as well as linguistic properties. Therefore, results obtained with this speech corpus are important and reliable which accounts for our continued use of it in the other two studies.

The pause was found to be a reliable indicator of unit boundaries. Of course, in a completely naturalistic setting it is difficult to unambiguously conclude that pauses and not some other correlated variable are responsible for these results. Therefore, the second study experimentally manipulated the pause variable independently of the syntactic one in order to unconfound it from this naturally co-occurring variable. Both variables were found to be important determinants so that the results obtained in this new situation paralleled the findings of the first study in a very precise manner, although the present study had used ordinary subjects who were not expected to have any pre-conceptions of what would constitute a unit.

Of course, the very interesting question concerns how communication is affected by the speaker's pause placements. In a normal situation the speaker's pauses which fall at these unit boundaries would be expected to be a cue that the listener uses in decoding and comprehending the communication. Recent memory research indicates that a fruitful way of approaching the problem of comprehension is from a propositional point of view. That is, the listener is viewed as actively abstracting the propositions that underly the message and interrelating them into a memory framework which can be viewed as a network of propositions. When faced with a question concerning the communication, the memory structure is searched and the subject responds accurately if a "route" can be established to the relevant information. The particular paths taken will depend upon the strategies the subject uses in searching. It is not necessary that all pathways be completely intact since if an indirect one can be found or if alternative answers can be ruled out, the subject will respond accurately. The insertion of pauses at inappropriate locations will interfere with the decoding of the communication and its subsequent memorial organization. Since direct routes are not necessary for accuracy, the subject can still answer questions concerning the message, in an accurate manner. However, the memorial structure has not been fully interconnected and this results in the subject being less certain of their answers since there are fewer pathways leading to the same conclusions. The results of the second study are viewable in the light of this analysis. The Experimental subjects listened to a tape where pause time had been manipulated and the subjects responded by marking their unit boundaries accordingly. Further, their performance on the comprehension test demonstrated that their memory structure had been detrimentally affected which was reflected in terms of the lower degree of confidence they had of the information inherent in the communication.

A change in procedure which resulted in a subject not relying upon their propositional memory for underlying meaning would be expected to result in no such lowered confidence if the preceding analysis of this situation is correct (i.e. if it involves a problem with the organization of propositional memory). A change in instructions which resulted in a reliance upon memory for the surface structure of the material rather than with its underlying meaning, as had been the case in Study 2, would provide important information in this regard. The third study fulfilled these requirements and indeed it was found that the confidence ratings were not differentially affected between the Experimental and Control groups. Further, this was not a consequence of different levels of overall comprehension between the subjects of the two studies since the mean proportions of subjects answering correctly were almost exactly the same for the Control and Experimental groups of both studies. In addition, the Experimental tape used in Studies 2 and 3 was not perceived as sounding "unnatural" and as such the lowered confidence ratings given by the Experimental subjects of the second study cannot be attributed to stimulus novelty effects. This is further supported by the lack of any differences in confidence which was found to be the case in the third study. There doesn't seem to be any reason to expect novelty effects only in the second and not the third study. On the other hand, these results are explainable in terms of the model, and in fact the propositional analysis is strengthened by them.

These conclusions must be tempered by the realization that even in a situation where the stimulus materials are spontaneous, the task imposed on the subjects is not typically encountered in the "real" world. Thus, although listeners do extract information from speakers, the actual way in which it is done as well as the format in which they remember it is probably not independent of the task the listener must perform. It is hoped that some insights into this process were illuminated via this dissertation. Nevertheless, the distinction between the "real" world and that of experimental research must be borne in mind. The proposition has been found to be of utility in explicating results obtained under the latter condition. Hopefully, this approach has meaning in the "real" world as well.

Overall, then the listener can be viewed as making use of the natural pause cues which the speaker provides. Since spontaneous speech is the most typical form in which everyday communication occurs, it is not surprising that such a relationship should exist. How the listener acquires this knowledge is beyond the scope of the present dissertation but it is probably the result of a long-term implicit learning experience. Further research is necessary to provide the information necessary to answer this question.

APPENDIX A

The additions are marked as "A". Seven of the deletions were at syntactic boundaries and are marked as "D", the two that involved filled pauses are marked as "Df", and the remaining three as "d". The tape begins as indicated (arrow) and the critical areas where manipulations occurred are indicated by vertical lines. It should be noted that the last items on the tape are labeled "(student comment)" and consisted of the speaker telling a student to ask a question and a portion of the student's remarks.[1] We left these few extra words in so that the last items covered in the critical area would be set off as complete. Transcript is as follows:

uh the whole thing ah could not have been possible without without the Darwinian theory and the approach ah ah that it reflects in terms of the study of individual differences and the adaptive value of behavior the adaptive value of consciousness ah and so forth um ya (student comment) no no what was I supposed to look for I forgot (student comment) being used as social Darwinism um no forgot all about it I'll try again um did I uh you ask me also one time what what wha would William James's family did (student) did I ever tell you that I found out the answer (student) no (student) no his grandfather his grandfather was was a millionaire uh his grandfather came over as a as a nothing immigrant poor immigrant immigrant stepped over the backs of thousands of peasants who worked his ↑ way up to being a multimillionaire that's where the money came from ok ah let's look at the Chicago school ok first and um let's start by saying a few things about uh John Dewey who's dates love these dates love them let me make sure I get them right because they're a little staggering um 1859 to 1952 heh um a long and rich life heh 93 years John Dewey is best known as an educator and as a matter of fact he is I think beyond a shadow of a doubt the most progressive innovative uh and profound theoretician in in pedagogic theory um ever perhaps maybe |D ah Dewey goes to the University of Vermont a nice New England boy D ah after graduating teaches high school for a while and then goes off to Johns Hopkins ah for his doctorate degree A in philosophy d which he takes in 1884 um however remember what's happening um at this time in philosophy it's really becoming very very psychological and indeed Dewey gets caught up in in this in this flurry of psychological psychological research A that's going on and he accepts a teaching post over the next several years at Michigan and then at Minnesota and during this period of time he he begins to orient to himself very clearly toward toward psychology A coming from the philosophical end of it and uh uh a scant two years after getting his degree he publishes a text book which is called uh not surprisingly Psychology uh and it is a uh very very popular text for about four years D um if you count those four years you will notice that's the date in which William James publishes the Principles of Psychology which very quickly lowers the popularity of every other popular text that happened to be in use at at at that time uh Df four years after that uh that's that's in 1894 he goes to Chicago he goes to Chicago um Df where he stays for ten years now during that ten year period he is sort of a philosopher dash psychologist and after he leaves Chicago in 1904 [A] to go to Columbia [d] he never really uh has anything to say about psychology again he is primarily a philosopher a representative of the pragmatic school and of course uh a pedagogic theorist and and an educator of uh great renown [D] um and the extraordinary thing about him is that in his entire life he publishes exactly one paper in psychology that's all [D] just one [A] paper that could ever be really clearly called psychology that he himself would call psychology and it really sets the spark to a very volatile mixture that exists at the time [D] the paper is published [A] in 1896 two years after he arrives at Chicago [d] and [A] it provides the focus for the development of the functionalist school in Chicago and it's called um the reflex arc concept [A] in psychology [D] and it literally starts the new movement it is this it is the approach that is that is laid out in this paper that formu d d that forms the the uh the nexus th that Titchener is attacking when he talks about these functionalists [A] in his in his 1898 paper | (student comment) that's interesting I was try I was trying to recover that myself um um what what I was saying you know you know one track of the mind was was trying to put together things in time I'm not sure I'm really not sure um how much he actually had to deal with the creation of TC in particular teacher's college itself I just don't know I just don't know does anybody does anybody have any idea when TC started or alright that's another thing I want to try to (garbeled) to check out both of those things how Darwin felt about um um

Spence's development of social darwinism and whether John Dewey really originated TC um I'll try to remember to do those things okay lets look at this paper though it's the important thing

FOOTNOTE

1. It should be noted that due to a slight mishap in the preparation of the tapes used in Studies 2 and 3, there was a small difference between the Experimental and Control tapes as to where they ended during the student's comment. The speech utterances which occurred at the end of the tape (labeled "(student comment)" on the transcript) were the following:

   Speaker: you have a question

   Student: yeah I was just curious when you mentioned Columbia and that when he....

   The Control tape ended while the student was saying "Columbia" whereas the Experimental tape ended immediately after the word "he." This minor difference was considered to be quite unimportant since it fell outside of the critical area.

APPENDIX B

Spontaneous speech has certain interesting properties. Among them is the sequences of periods of hesitation observed. Henderson et al. (1965) have noted that these are periods where pauses take up a substantial proportion of total speaking time (i.e., speech and pause times) and periods where few pauses are noted (referred to as hesitant and fluent periods, respectively). In addition, the pauses are differentially distributed as a function of the type of period. Pauses during hesitant periods occur predominantly at non-grammatical junctures whereas those during fluent periods are predominantly located at grammatical junctures. Thus, spontaneous speech is a series of alternating high and low frequencies of pausing.

Within the context of Studies 2 and 3, the Control tape's pauses were modified at four places out of the over 400 possible word boundaries, thereby resulting in a little less than a one percent modification of the natural speech (an essentially trivial departure from complete spontaneity). There was one instance of a 280 msec. pause following a filled pause which was removed in both the Control and Experimental versions; the Experimental also had the filled pause removed.[1] We were here interested in the function of filled pauses by themselves. There was also a 2.24 sec. pause which was reduced to 200 msec. in both the Control and Experimental tapes. The particular location had been chosen as a unit boundary (and was a syntactic one as well) by the listeners of Study 1 and while it was located too close to other manipulations to be deleted its reduction was expected to give us some further information as to the interrelationship between pausing and the environment in which it is found. The information expected from these two modifications was tangential to the present studies and are not included here. The third one was a naturally occurring 0.50 sec. pause which was reduced to 0.30 sec. The corresponding location in the Experimental tape was an addition cite, which due to this modification resulted in a larger net difference between both tapes in hopes of providing a sufficient pause differential to detect experimentally. Finally, there was an extraneous short burst of sound (possibly someone snapping their fingers) which occurred during one of the pauses. The corresponding location had been chosen for deletion in the Experimental tape. It was felt that leaving the sound in the deleted area would sound unnatural (since it was not "surrounded" by silence). Therefore it was eliminated in both tape versions.

As will be noted at the beginning of the Appendix, spontaneous speech is characterized by several properties, among them is the duration of pauses (and their frequencies). Another is the type of speech encountered (e.g., filled pauses). Our Control modifications resulted in no changes in the latter and only minor ones in the former, since the reductions in pause time left pauses which were more probable than the original ones. That is, shorter pauses are more likely than longer ones (and in fact the most likely category is "no pause"). Given this information and the fact that it was less than a one percent modification, we felt very confident that spontaneity had been essentially preserved.

FOOTNOTE

1. Both the Experimental and Control groups received the same transcripts. This resulted in an elimination of any possible confounding that could have resulted if the filled pause had been eliminated in the Experimental transcript (both for this location and one other where the filled pause was eliminated in the Experimental tape but no mofifications had been made in the Control tape at that location). This minor departure from a veridical transcript was not mentioned by any of the subjects.

APPENDIX C

The following are the instructions that were read to the Experimental and Control subjects in Study 2. The Transcript Control subjects did not listen to any tape and their instructions were modified appropriately:

> "You will listen to some tape recordings of someone speaking and you will have in front of you a transcript of the actual speech that you will hear. When we normally hear someone speak, we tend to process some of the words together because together they express a single thought or idea. They form a meaningful and cohesive unit. I want you to mark off on the transcripts these units. Whenever you feel a meaningful unit has just ended, place a slash mark after the last item of the unit, just before the beginning of the next unit. When deciding where a unit is to go let the tape recording be your guide and the transcript only a means for indicating your choice. Also, once you have put a slash mark on the transcript don't change it. Are there any questions?"

After the warm-up tape was played, the following instructions were used:

> "Now you will hear another tape, this time of someone else speaking. I want you to do exactly the same task you did with the other tape. That is mark with a slash where a meaningful unit has just ended by placing the slash, as you did before, after the last item of the meaningful unit, just before the first item of the next unit. The tape begins where you see the red arrow on the transcript. Again once you have put a slash

mark down, don't change it. And again let the tape recording be your guide for deciding where a unit is to go. Are there any questions?"

It was explained to them that if they could not continue (ex., pencil breaks) they were to just sit quietly.

For the second and third playings of the major tape, the following instructions were used:

> "I will play the tape again and I want you to indicate your choices on the same transcript. If you feel the same way about where a particular unit is just place the slash mark alongside the other one. Feel perfectly free to place it anywhere you feel it's correct. It need not necessarily be in the same position it was before, it's all up to you. Again let the tape recording be your guide and after you have placed a slash mark down on the transcript don't change it. Are there any questions?"

It was also explained to them to sit quietly if they could not continue.

APPENDIX D

Mean confidence ratings for subjects in the Experimental (E), Control (C), and Transcript Control (TC) groups answering correctly or incorrectly to each comprehension question in Study 2

EXPERIMENTAL - CONTROL

Question Number	Correct				Incorrect			
	E	C	t	df	E	C	t	df
1	1.67	2.85	-2.53*	27	3.17	3.33	-0.41	34
2	3.20	3.38	-0.55	57	2.00	3.75	-1.43	5
3	2.29	3.24	-2.18*	33	2.36	2.88	-1.17	28
4	3.15	3.79	-1.61	46	2.57	2.88	-0.49	13
5	3.05	2.97	0.24	48	2.00	3.63	-2.32*	13
6	2.59	3.09	-1.50	52	2.50	3.20	-0.99	9
7	2.17	3.12	-2.32*	55	1.60	1.75	-0.30	7
8	1.78	4.00	-4.52*	18	3.05	3.15	-0.27	43
9	2.40	2.30	0.17	28	2.53	2.89	-0.85	33
10	3.29	3.42	-0.40	58	2.25	2.50	-0.25	4
11	3.39	3.06	1.07	53	1.80	2.60	-1.26	8
12	2.30	2.25	0.09	24	1.89	2.55	-2.15*	38
13	2.96	3.13	-0.47	51	2.25	1.88	0.52	10
14	3.46	3.54	-0.21	50	3.00	2.33	0.72	11
15	2.81	2.92	-0.28	44	1.29	1.69	-0.82	18
16	3.00	3.18	-0.52	55	2.50	2.00	0.88	7

EXPERIMENTAL - TRANSCRIPT CONTROL

Question Number	Correct				Incorrect			
	E	TC	t	df	E	TC	t	df
1	1.67	3.22	-2.72*	16	3.17	2.83	0.75	28
2	3.20	3.06	0.41	39	2.00	2.80	-0.75	6
3	2.29	2.70	-0.84	22	2.36	2.50	-0.30	22
4	3.15	3.94	-2.10*	35	2.57	2.75	-0.21	9
5	3.05	3.13	-0.20	34	2.00	2.60	-1.07	10
6	2.59	3.37	-2.11*	39	2.50	2.50	0.00	6
7	2.17	3.44	-2.86*	39	1.60	1.33	0.45	6
8	1.78	3.86	-5.32*	14	3.05	3.21	-0.36	31
9	2.40	3.09	-0.96	19	2.53	3.40	-1.57	25
10	3.29	3.56	-0.60	40	2.25	2.00	0.28	5
11	3.39	3.00	1.01	38	1.80	2.00	-0.26	6
12	2.30	2.40	-0.17	18	1.89	2.00	-0.36	27
13	2.96	2.76	0.45	38	2.25	2.50	-0.25	6
14	3.46	3.00	0.81	32	3.00	3.55	-0.63	13
15	2.81	2.57	0.47	33	1.29	1.86	-1.29	12
16	3.00	3.25	-0.58	42	2.50	4.00	-1.34	3

CONTROL - TRANSCRIPT CONTROL

Question Number	Correct				Incorrect			
	C	TC	t	df	C	TC	t	df
1	2.85	3.22	-0.67	27	3.33	2.83	1.30	28
2	3.38	3.06	0.86	48	3.75	2.80	1.02	7
3	3.24	2.70	1.03	29	2.88	2.50	0.93	24
4	3.79	3.94	-0.43	43	2.88	2.75	0.15	10
5	2.97	3.13	-0.43	44	3.63	2.60	1.19	11
6	3.09	3.37	-0.79	49	3.20	2.50	0.81	5

CONTROL - TRANSCRIPT CONTROL
(continued)

Question Number	Correct				Incorrect			
	C	TC	t	df	C	TC	t	df
7	3.12	3.44	-0.75	50	1.75	1.33	1.02	5
8	4.00	3.86	0.27	16	3.15	3.21	-0.15	38
9	2.30	3.09	-1.56	29	2.89	3.40	-0.91	26
10	3.42	3.56	-0.40	52	2.50	2.00	0.60	3
11	3.06	3.00	0.19	47	2.60	2.00	0.88	6
12	2.25	2.40	-0.28	24	2.55	2.00	1.34	31
13	3.13	2.76	0.89	45	1.88	2.50	-0.94	10
14	3.54	3.00	1.27	36	2.33	3.55	-1.94	18
15	2.92	2.57	0.76	37	1.69	1.86	-0.29	18
16	3.18	3.25	-0.19	51	2.00	4.00	-2.58	4

* $p < .05$

APPENDIX E

The following are the instructions that were read to the Experimental and Control subjects of Study 3 before they listened to their respective tape:

> "This is a study of language communication and conversations. You will hear a tape of someone speaking in a lecture. The tape begins in the middle of the lecture. Pay careful attention to the tape because I will ask you some questions later. Are there any questions? Please don't interrupt once the tape is playing."

APPENDIX F

Mean confidence ratings for subjects in the Experimental (E) and Control (C) groups answering correctly or incorrectly to each comprehension question in Study 3

Question Number	Correct				Incorrect			
	E	C	t	df	E	C	t	df
1	3.57	3.26	0.89	52	3.06	2.70	0.85	24
2	3.59	4.21	-2.45*	63	2.71	2.75	-0.06	13
3	3.88	3.88	0.01[a]	47	2.64	3.47	-2.07*	29
4	4.10	4.45	-1.42	60	3.50	2.70	2.12*	16
5	3.42	3.00	1.09	53	2.87	3.30	-0.74	23
6	3.79	3.61	0.61	63	2.70	3.60	-1.28	13
7	3.38	3.17	0.53	62	2.00	1.00	2.83*	14
8	4.38	3.71	1.41	31	3.39	3.42	-0.07	45
9	2.89	3.45	-1.10	39	2.55	2.32	0.61	37
10	3.95	3.57	1.50	70	2.50	3.67	-1.25	6
11	3.52	3.50	0.06	63	2.67	2.67	0.00	13
12	3.79	4.27	-1.14	28	3.25	3.27	-0.04	48
13	3.96	3.73	0.72	59	3.45	2.63	1.55	17
14	3.69	3.89	-0.53	55	2.40	2.77	-0.70	21
15	1.57	2.26	-1.73	39	2.52	1.71	1.67	37
16	3.64	3.63	0.04	71	3.33	5.00	-1.13	5

* $p < .05$

[a] Slight differences between the means were responsible for $t \neq 0$.

APPENDIX G

The following are the comprehension test items used in Studies 2 and 3:

1. Dewey stayed at Columbia for ten years.
2. Dewey received a doctorate degree in philosophy.
3. Dewey received his doctorate degree at the University of Vermont.
4. Dewey spends essentially all of his career working as a psychologist.
5. Dewey is considered to have started the functionalist movement.
6. Philosophy at the time of Dewey was becoming psychological.
7. Dewey refused to teach at Minnesota.
8. Four years after getting his degree, Dewey published a textbook.
9. William James' textbook was extremely popular up until the publication of Dewey's text.
10. Dewey was a "philosopher dash psychologist" during his stay at Chicago.
11. At Columbia, Dewey is best known as a philosopher and an educator.
12. Dewey published an important paper in 1890.
13. Dewey's paper did not cause a reaction in the world.
14. Dewey's paper was called "The reflex arc concept in psychology".
15. Titchener attacked the functionalist school of thought.
16. Dewey approached psychology with a philosophical viewpoint.

REFERENCES

Anderson, J. R. Verbatim and propositional representation of sentences in immediate and long-term memory. Journal of Verbal Learning and Verbal Behavior, 1974, 13 (2), 149-162.

Begg, I., & Wickelgren, W. A. Retention functions for syntactic and lexical vs. semantic information in sentence recognition memory. Memory and Cognition, 1974, 2 (2), 353-359.

Blankenship, J., & Kay, C. Hesitation phenomena in English speech: A study in distribution. Word, 1964, 20 (3), 360-372.

Boomer, D. S. Hesitation and grammatical encoding. Language and Speech, 1965, 8 (3), 148-158.

Boomer, D. S. Frieda Goldman-Eisler Psycholinguistics; experiments in spontaneous speech. Lingua, 1970, 25 (2), 152-164.

Chapple, E. D. Quantitative analysis of the interaction of individuals. Proceedings of the National Academy of Sciences of the United States of America, 1939, 25, 58-67.

Chomsky, N. Syntactic Structures. The Hague: Mouton and Co., 1957.

Cochran, W. G. Some methods for strengthening the common χ^2 tests. Biometrics, 1954, 10 (4), 417-451.

Fónagy, I., & Magdics, K. Speed of utterance in phrases of different lengths. Language and Speech, 1960, 3 (4), 179-192.

Fries, C. C. The Structure of English. New York: Harcourt, Brace and Co., 1952.

Goldman-Eisler, F. Psycholinguistics: Experiments in spontaneous speech. New York: Academic Press, 1968.

Goldman-Eisler, F. Pauses, clauses, sentences. Language and Speech, 1972, 15 (2), 103-113.

Goldman-Eisler, F., & Cohen, M. An experimental study of interference between receptive and productive processes relating to simultaneous translation. _Language and Speech_, 1974, _17_ (1), 1-10.

Hargreaves, W. A. A model for speech unit duration. _Language and Speech_, 1960, _3_ (3), 164-173.

Henderson, A., Goldman-Eisler, F., & Skarbeck, A. Sequential temporal patterns in spontaneous speech. _Language and Speech_, 1966, _9_ (4), 207-216.

Katti, S. K., & Sastry, A. N. Biological examples of small expected frequencies and the chi-square test. _Biometrics_, 1965, _21_ (1), 49-54.

Lieberman, P. On the acoustic basis of the perception of intonation by linguists. _Word_, 1965, _21_, 40-54. (As found in Oldfield, R. C., & Marshall, J. C. (Eds.), _Language_. Baltimore: Penguin, 1968.)

Linton, M., & Gallo, P. S., Jr. _The Practical Statistician: Simplified handbook of statistics_. Belmont, California: Wadsworth, 1975.

Lounsbury, F. G. Pausal, juncture and hesitation phenomena. In, Osgood, C. E., & Sebeok, T. A. (Eds.), _Psycholinguistics: A Survey of Theory and Research Problems_. Bloomington, Indiana: Indiana University Press, 1965. (Original publication, 1954.)

Maclay, H., & Osgood, C. E. Hesitation phenomena in spontaneous English speech. _Word_, 1959, _15_ (1), 19-44.

Mahl, G. F. Distrubances and silences in the patient's speech in psychotherapy. _Journal of Abnormal and Social Psychology_, 1956, _53_ (1), 1-15.

Matarazzo, J. D., Saslow, G., & Matarazzo, R. G. The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. _Journal of Psychology_, 1956, _41_, 347-367.

Nass, C. A. G. The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism. Biometrika, 1959, 46, 365-385.

Newport, E. L. Motherese: The speech of mothers to young children. In Catellan, N., Pisoni, D., & Potts, G. (Eds.), Cognitive Theory. V. II, Hillsdale, New Jersey: LEA Press, 1976.

Pittenger, R. E., Hockett, C. F., & Danehy, J. J. The First Five Minutes. New York: Martineau, 1960.

Rips, L. J., Snoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 1973, 12 (1), 1-20.

Rochester, S. R. The significance of pauses in spontaneous speech. Journal of Psycholinguistic Research, 1973, 2 (1), 51-81.

Sachs, J. S. Memory in reading and listening to discourse. Memory and Cognition, 1974, 2 (1A), 95-100.

Schwartz, J., & Jaffe, J. Markovian prediction of sequential temporal patterns in spontaneous speech. Language and Speech, 1968, 11 (1), 27-30.

Shannon, C. E. Prediction and entropy of printed English. The Bell System Technical Journal, 1951, 30 (1), 50-64.

Snow, C. E. Mothers' speech to children learning language. Child Development, 1972, 43 (2), 549-565.

Starkweather, J. A. Vocal behavior: The duration of speech units. Language and Speech, 1959, 2, 146-153.

Trager, G. L., & Smith, H. L. An outline of English structure. Norman, Oklahoma: Battenburg Press, 1951.

Verzeano, M., & Finesinger, J. E. An automatic analyzer for the study of speech in interaction and in free association. Science, 1949, 110 (2845), 45-46.