Journal of Psycholinguistic Research, Vol. 2, No. 1, 1973

The Significance of Pauses in Spontaneous Speech

S. R. Rochester[1]

Received February 22, 1972

Studies of filled and silent pauses performed in the last two decades are reviewed in order to determine the significance of pauses for the speaker. Following a brief history, the theoretical implications of pause location are examined and the relevant studies summarized. In addition, the functional significance of pauses is considered in terms of cognitive, affective-state, and social interaction variables.

INTRODUCTION

When words are uttered spontaneously (i.e. when they are being organized at the time of utterance), they emerge haltingly, interspersed with *ums* and *ahs* and *ers*. In the last two decades, this intermittent quality of speech has attracted many investigators seeking to understand the decision-making processes underlying speech production. The question, as Lashley (1951) posed it, is the following: How is the act of language production transformed into an external temporal sequence? The answer, it has often been suggested, may be discernable through a study of where and under what conditions pauses occur. In the present paper, the literature bearing on pauses in spontaneous speech has been reviewed critically (1) to ascertain the methodological and empirical strength of claims that pauses serve as clues to speech production; and (2) to evaluate various hypotheses concerning the location and function of pauses in speech.

Following a brief history of research into pausal phenomena, three models of the speaker are examined. For each model, alternative hypotheses

[1]Clarke Institute of Psychiatry, Toronto, Canada.

regarding the location of pauses in speech are evaluated. Then, the functional significance of pauses is considered in terms of cognitive, affective-state, and social interaction variables.[2]

A BRIEF HISTORY

Historically, the study of nonsegmental systems has been neglected in language research. The reasons for this neglect are related. On the one hand, investigators were "taking their examples from the connected language of books where the sentences are often altogether wanting in the variety found in conversational speech" (Coleman, 1914). On the other hand, real possibilities for analyzing utterances were blocked until records could be made of actual speech.

Even after some form of recording was available, however, North American investigators lagged behind their European colleagues in pursuing the problem of pauses of speech (see Crystal, 1969 and Hegedüs, 1953 for reviews of the early European work). Americans studying the nonverbal characteristics of language were preoccupied with elocutionary matters throughout the first half of this century and ignored the processes underlying unrehearsed speech (e.g., Lynch, 1934; Murray and Tiffin, 1934).

The second half of the twentieth century marked the beginning of systematic investigations of pausal phenomena in English. In linguistics, Pike's (1945) thorough description of nonsegmental systems suggested that pauses and other prosodic features were not incidental but rather necessary to linguistic description. In psychology, between 1951 and 1956, three researchers presented major papers which effectively began the experimentation in spontaneous speech: Goldman-Eisler (1951, 1952, 1954, 1955, 1956a,b,c) and the anthropologist Lounsbury (1954) were concerned with brief hesitations in normal discourse, while Mahl (1956a,b) analyzed disturbances and silences in patients' speech.

Although both linguists and psychologists became interested in pausal phenomena at about the same time, they were rarely influenced by each other's perspective. Thus, psychological studies of pause location have tended either to ignore linguistic analyses or to use only weak approximations to them. At the same time, linguistic theories have not focused on speaker

[2]The focus of the present paper has been on *pauses* in *spontaneous speech*. Consequently, it has been necessary to omit certain interesting but not directly relevant studies. Two omissions are noteworthy: (1) Studies focused on interview characteristics (e.g., Matarazzo *et al.*, 1965; Saslow and Matarazzo, 1959) have been omitted if latency has been the only silence measure used. (2) Studies investigating the distribution of pause time in reading (e.g., Brown and Miron, 1971) are likewise omitted.

performance and consequently have not provided models of production (cf. comments on this problem by Bever, 1970; Fillenbaum, 1971; Průcha, 1970; Rommetveit, 1968.)

The few theoretical efforts that have been relevant to investigations of pauses are best described by beginning with Lounsbury's (1954) description of the role of "sequential mechanisms" in behavior. This analysis appears to be the first coherent psycholinguistic explanation of pauses.

An Early Model of the Speaker

In the interdisciplinary spirit of the early 1950's, Lounsbury proposed a model of the speaker (and the listener) which linked Shannon's (1951) information measures to the immediate constituent analyses of linguistics and to behaviorist proposals about language. Following Osgood (cf. Osgood and Jenkins, 1953), Lounsbury argued that complexes of internal stimuli mediate between linguistic events on various levels. On the basis of the frequency and/or contiguity of occurrence of those events in the speaker's experience, hierarchies of habits of various strengths would be developed and these would be reflected in sets of transition probabilities linking those events to each other.

Pauses were fit into this model as an ingenious device for testing its hypotheses. Since habit strength is inversely correlated with the latency between stimulus and response, it was argued, pauses should serve as clues to the strength of associations between two linguistic events, where the antecedent event represents the stimulus situation and the subsequent event represents the response. The stronger the transitional habits, the shorter the pausal durations should be. Thus, within words and phrases that are familiar, the durations of pauses should be minimal whereas pauses found between weakly associated words and phrases should be rather long.

In order to use the pause as a measure of latency between associations, it was necessary to distinguish between pauses which might reflect encoding or speaker units and pauses which might be more pertinent to the speaker's allowance for listener processes. Pauses which are very brief (100 msec or less) and which fall between the boundaries of high-order constituents were designated as *juncture pauses* and hypothesized to serve as an aid to the listener, "to help put across the structure of a sentence. . ." (p. 99). The longer (to 3 sec) *hesitation pauses* were supposed to occur at points of lowest transition probability, reflecting the presence of very weak associations between linguistic events, and thereby marking the beginning (or end) of speaker units.

At first glance, it may appear that Lounsbury was postulating two mutually exclusive sorts of events which differed from each other in location, duration, and function. In fact, it seems more likely that the events were meant to be independent or possibly complementary. For example, while pauses at syntactic junctures were supposed to serve the listener, Lounsbury did not rule out the possibility of a speaker function. The critical factor in determining function would seem to have been duration: presumably, a pause occurring at a major syntactic boundary could serve the speaker if it were long enough. This issue was one of several to be encountered by later investigators of pauses who sought to discover the encoding (and decoding)[3] units in speech.

Testing the Model

The question of encoding units falls within the more general issue of the relationship between the location and the function of pauses in speech. Prior to investigation, it is not evident whether pauses function in terms of words, phrases, intonation units, major grammatical constituents in the surface or deep structure, or some other aspects of utterances not described by linguistic categories. Consequently, it is not obvious which locations are most relevant to an analysis of pauses. The initial explorations of an encoding function were focused, somewhat arbitrarily,[4] on the word as the fundamental speaker unit.

Goldman-Eisler (1958a,b) argued that once a speaker has uttered the first word of a sentence, only lexical decision-making remains. Thus, all pauses after the first word of a sentence should reflect word selection processes. In two experiments designed to test this hypothesis, subjects were provided with the preceding context of a recorded conversation and asked to guess the first and each successive word in a given sentence. In one case, subjects read the sentences silently (Goldman-Eisler, 1958a), and in the other case they read aloud (Goldman-Eisler, 1958b). The transition probabilities between words were estimated using the Shannon guessing technique. Apparently,[5] this was accomplished by determining the ratios of correct guesses to the total guesses made for the given word. Words which were difficult to guess had low transition probabilities and therefore represented an increase in information when uttered; words which were easy to guess had high transition probabilities and these reduced uncertainty only minimally.

In general, the results suggested that speech fluency is associated with high predictability of words while pauses are associated with low predicta-

[3]The significance of pauses in decoding processes is discussed in Rochester (1971).

[4]The decision was not entirely arbitrary in view of the prevailing linguistic view at the time which regarded words as grammatical units (cf. Hockett, 1968, p. 26).

[5]Goldman-Eisler is inconsistent in her account of the procedure for estimating transition probabilities, as Boomer (1970, p. 159) points out.

bility. The findings were as follows: (1) Words following a silent pause were less predictable, and (2) took significantly longer to replace, than words uttered in fluent contexts. In addition, (3) words preceding a silent pause tended to be even more predictable than words uttered in other fluent contexts. However, (4) the only words which were reliably related to the incidence of silent pauses were those which subjects found difficult to guess from both directions–approached both from the beginning and from the end of the sentence. These results and others bearing on the location of silent pauses are summarized in Table I.

Table I. Locating the Silent Pause

Author and date	Speech sample: Size	Speech sample: Speakers	Speech sample: Type	Independent or correlated variables	Definition of silent pause
Goldman-Eisler, 1958a	12 sentences (348 words)	4[a]	Debate Interview Dictated letter	High *vs.* low information words (Shannon technique)	Silence >250 msec
Goldman-Eisler, 1958b	4 sentences	–	–	High *vs.* low information words (Shannon technique)	Silence >100 msec
Tannenbaum *et al.*, 1965	Expt 1:- Expt 2: 440 words	1M 1M/1F	Monologue	High *vs.* low information words ("Cloze")	"Silence of unusual length" determined by judges[b]
Boomer, 1965	3 min/speaker	16M	Monologue	Grammatical location (Phonemic analysis)	Silence >200 msec
Henderson *et al.*, 1965	350-430 words/speaker 220 words	5[a] 5[a]	Interview	Read *vs.* spontaneous speech	Audible breath inspirations at silences >100 msec
Henderson *et al.*, 1966	Same as above speech sample		Spontaneous	Syntactic location (Parsing)	Silence >100 msec
Taylor, 1969	120 sentences	20[a]	Monologue	Task difficulty; Encoding operations (Johnson's Phrase-Structure Analysis)	Latency >1 sec "Hesitations"

[a]Sex not given.
[b]Sex and number not given.

The findings from these two experiments are intriguing because they suggest two forms for a model of sentence production: where findings 1 and 2 support a lexical decision-making model in which pauses are predictable on a probabilistic word-by-word basis, findings 3 and 4 call this simple model into question and imply a structural plan of some sort guiding the choice of words in sequences. However, there are serious constraints placed on these interpretations. First, in the nearly 15 years since these widely cited experiments were published, there has been no adequate replication of at least two of the major findings. Second, there are profound methodological difficulties with the original studies.

The single attempt at replication seems to be that of Tannenbaum *et al.* (1965) who investigated the relationship between transition probabilities and hesitations. They defined hesitations not only as silent pauses but also as filled pauses, phoneme and word repetitions, and false starts. Their results partially support Goldman-Eisler's findings: words following a hesitation were less predictable than words found in other contexts. However, words preceding a hesitation were not easy to guess. In fact, it was always the words uttered just before and just after hesitations that were the most difficult to predict.

Unfortunately, the Tannenbaum *et al.* study diverges rather drastically from Goldman-Eisler's work. Some departures represent attempts to improve methodology: judges in the later study were presented with a single "message" from the monologue of one male and one female speaker instead of sentences taken from a variety of situations and speakers[6]; the later study used the "Cloze" technique (Taylor, 1953) in which reverse transition probabilities are calculated somewhat differently than with the Shannon guessing technique.[7] Other departures seem almost deliberate constraints on replicability: since Tannenbaum *et al.* present only the data for "hesitations," it is not possible to assess the extent of direct replication of Goldman-Eisler's findings with silent pauses; moreover, even if subdivision of the data were possible, the reader would be faced with the task of comparing events termed "silences of unusual length" with Goldman-Eisler's silences exceeding 250 msec. At the very least, the later investigators might have reported the mean and standard

[6]It is not clear that the use of connected discourse represents an improvement in this case. Goldman-Eisler examined pauses in single sentences in an attempt to eliminate pauses due to decisions about content or sentence structure. While the adequacy of this procedure may be questioned, its rationale should not be ignored as it has been by Tannenbaum *et al.*

[7]The Cloze technique permits simultaneous estimation of antecedent and subsequent transition probabilities while the Shannon technique requires an averaging of antecedent and subsequent context scores. The difference would affect only conclusions concerning transition probabilities in the fully combined sentence context (i.e., Goldman-Eisler's findings 3 and 4).

deviation of the judged silences, and might have provided some information about the consistency of those judgments.

In addition to the lack of an adequate replication, there are problems of sampling, design, and analysis in the original studies. Because these studies have served as the basis for several later experiments (e.g., Bernstein, 1962; Goldman-Eisler, 1961a,b; Mishler and Waxler, 1970) and because they have been cited uncritically in numerous reviews (e.g., Crystal, 1969; Fillenbaum, 1971; Hörmann, 1971; Ervin-Trip and Slobin, 1966), they are worth scrutinizing in more detail.

First, as Boomer (1965, 1970) has complained, the sample used in the two experiments was not only small but highly selective. The selection took two forms: The utterances were taken from somewhat restricted situations (three from a debate, five from an interview with a neurotic patient, and four from a letter dictated "especially for the experiment") and all were required to be grammatically correct and well-constructed sentences. The latter constraint is the more serious. Not only are such samples likely to be unrepresentative of spontaneous language, but they are likely to be biased in a certain direction. By selecting sentences as the basic utterance unit, and by eliminating all of those containing any "repetition or midway changing of grammatical construction," Goldman-Eisler was virtually excluding utterances in which syntax was problematic for the speaker.

Why was this exclusion necessary? Goldman-Eisler (personal communication) has explained that it was unavoidable, given the nature of the "guessing game" involved in her modification of the Shannon technique. If subjects had been asked to supply words for illogical and/or ungrammatical sentences, she submits, the Shannon technique would have yielded transition probabilities close to zero. The technique, in other words, determined the set of possible observations. Because the technique was biased against syntactically relevant pauses, few such pauses were seen. Thus, the technique chosen to investigate a lexical decision-making hypothesis was biased in favor of the hypothesis and against alternative (or complementary) hypotheses. This strategy is most disturbing because the technical biases were not mentioned (either in 1958 or in Goldman-Eisler, 1968) in interpretation of the data.

The question of bias is raised in a different connection by Boomer (1970). He points out that the words counted as following a pause, "p" words, were not always the first word following a pause. Rather, they were the first *content* words (nouns, verbs, adjectives, or adverbs) following a pause. One-third (11/34) of all "p" words lagged one or two or three words behind the pause because the pause was in fact followed by one or more function words (prepositions, conjunctions, auxiliary verbs, pronouns). Why were content words counted and function words ignored? The only justification seems to be the following: "Introspection indicated that the utterance of such grammatical expressions is often delayed until the choice of the next content word is made" (Goldman-Eisler, 1958a, p. 100). Again, the lexical choice hypothesis seems to have determined an experimental decision.

A third problem concerns the statistical analyses used in Experiment 2 of 1958a. This is the study in which findings based on reverse transition probabilities were made. Again, briefly, these findings were (3) that words preceding silent pauses are more predictable than words uttered in other fluent contexts, and (4) that words unpredictable from either forward or reverse transition probabilities are preceded by pauses. There is very little support offered for 3. No numerical results are given and the only evidence seems to be in Fig. 3 (1958a): of the 26 words which precede pauses in this experiment, there appear to be no more than 16 or 17 represented in the figure. The data are neither self-evident nor are they explained. Finding 4 has been questioned by Boomer (1970). He draws attention to the scatter plot on which this finding is based (Fig. 8, 1958a). Each word is plotted in terms of the number of forward guesses against the number of reverse guesses generated. An apparently arbitrary "least squares space" is drawn to enclose those words which had stimulated more than ten guesses in each direction. Boomer points out that this space excludes about one-quarter of the relevant data, and, when those data are included, the positive finding disappears.

To summarize, Goldman-Eisler found that the preponderance of words immediately following pauses are difficult to guess. This has been supported by the work of Tannenbaum *et al.* In addition, Goldman-Eisler reported that words preceding pauses are particularly easy to guess and that, when both forward and reverse probabilities are estimated, all increments in information are preceded by pauses. These additional findings have not been replicated and moreover, are based upon questionable analyses. In light of the foregoing criticisms, the transition probability studies seem to support only a linear model in which pauses are dependent on moment-to-moment lexical choices.

A Later Model of the Speaker

At this point, it is useful to consider a later model of the speaker in which decisions are presumed to occur in terms of larger units of encoding based on the surface or phrase structure of the language. Models of this sort resemble Lounsbury's in their reliance on an essentially linear decision-making process based on behavioristic views about language. They depart from Lounsbury's proposals in an emphasis on major constituents as encoding units, and in a relegation of lexical choice to a low level of decision-making. There are a number of hierarchical models of sentence transmission which postulate decisions at major constituent boundaries. Among these (e.g., Johnson, 1965, 1968; Martin and Roberts, 1967; Martin *et al.,* 1968; Osgood, 1963; Yngve, 1960), Johnson's is most readily extended to make predictions for the speaker. Following Yngve, Johnson argues that listeners "decode" sentences in a number of operations corresponding to the numbers of phrase or constituent units. Presumably, speakers would engage in analogous operations and the operations would tend to be associated with time lags in speech. In particular, pauses should tend to precede grammatical constituents and the probability of a pause should decrease from the beginning to the end of a constituent. This in essence was Boomer's (1965) observation.

Boomer determined the location of each silence exceeding 200 msec in brief monologues and analyzed the data in terms of a phonemic analysis which, it is claimed,[8]

> makes . . . the recognition of immediate constituents into solidly established objective procedures, removing once and for all the necessity of defending one's subjective judgment as to what goes with what (Trager and Smith, 1951, p. 77).

The results showed that 40% of all nonjuncture (hesitation) pauses occurred after the first word of a phonemic clause. Boomer argued that this nonrandom occurrence of hestiation pauses supported the notion of grammatical encoding in speech and refuted the predictions of a word-transitional model.

In a thoughtful examination of Boomer's data, Barik (1968) has revealed an even stronger data base for a grammatical encoding hypothesis. In Boomer's original analysis, all pauses occurring before the first word of a phonemic clause were defined as juncture pauses, i.e., as conventional pauses made for the sake of the listener. Consequently, such pauses were ignored in tabulations of hesitation pauses, i.e., pauses made for the sake of the speaker. However, this definition seems unduly stringent in the case of long pauses (700 msec or longer) at syntactic junctures. Such pauses, Barik suggests, may be *combinations* of juncture and hesitation pauses, with part of the pause (the first 500 msec) serving a conventional, listener-oriented function and part (the latter 200 msec or more) reflecting decision-making operations of the speaker. Pursuing this suggestion, it appears that a grammatical encoding hypothesis can account for about half (717/1398 or 51%) of all hesitation pauses in Boomer's data. This revised estimate conforms closely to the 54% predictability of pauses at grammatical locations found by Henderson *et al.* (1966).

Another test of the hierarchical model has been made by Taylor (1969). She argues that if the structural complexity of the sentences is related to the

[8]See the evidence of Lieberman (1965) disputing the objectivity but not the relevance to immediate constituents of Trager and Smith's phonemic analysis.

complexity of speaker processing, then the length and/or frequency of pauses should bear a functional relationship of the number of inferred encoding operations. To test this prediction, she asked subjects to produce sentences using topic words of varying difficulty. It was expected that reaction time prior to sentence production would reflect preprocessing and therefore should increase as a function of number of inferred encoding operations. The results showed that subjects took the same time to begin sentences assumed to have from two to eight encoding operations. In contrast, reaction time did vary with topic difficulty, and "hesitations" (defined as pauses greater than 1 sec, false starts, repeats, and topic recitation) were more frequent within the sentences having more difficult topics. Taylor concludes that the main variable involved in formulating sentences is content and that structure plays a relatively insignificant role in such production.

Boomer's and Taylor's findings appear to contradict each other, the former supporting an hierarchical model and the latter refuting it. Resolution of the findings seems to depend on the particular version of decision-making dictated by the model. Among the possibilities, two contrasting alternatives seem tenable. The speaker formulates a sentence by making decisions from the general to the particular, and either (a) decision times decrease as the speaker proceeds from the most general decision, *sentence,* to the lexical choices; or (b) decision times decrease only as there is a reduction in the number of immediately succeeding encoding operations.

For example, in the sentence "The person who sat near you is famous," speakers proceeding according to an hierarchical model should first decide to produce *sentence,* then *subject* and *predicate* (two encoding operations), then *noun phrase* and *relative clause* from *subject* (two encoding operations), then *article* and *noun* from *noun phrase* (two encoding operations), and finally the lexical choices. In the case of (a), a primary decision would be made coincident with the *sentence* decision, and would be reflected in the speaker's time to begin sentence production. Decision times would decrease as the number of encoding decisions to be made decreased, so that the pause before *subject* (in the example above) would be longer than the pause before *relative clause.* Two predictions can be made here: (1) From sentence to sentence, the latency of sentence production will vary as a function of the total number of endocing operations to be made. (2) Within a sentence, pause times will decrease as the total number of encoding operations to be made decreases.

In the case of (b), relatively limited structural decisions are envisaged. It is as if, at any given time, the speaker can discern only the immediate encoding alternatives and is unaffected by decisions yet to be made at lower levels. Again, there are two predictions: (1) Across sentences, latencies will be unrelated to the total number of encoding decisions to be made. (2) Within a sentence, pause times will be evenly distributed at points corresponding to equal numbers of immediately succeeding encoding operations. In the sentence example, the pause preceding *noun phrase* should equal the pause preceding *relative clause* should equal the pause preceding *subject.*

Taylor has found no support for (a). Boomer's evidence of pauses preceding phonemic clauses may provide support for (b) or for some modification of (a) in which decisions are made below the highest node, thereby eliminating the primary importance of reaction times. Unfortunately, Taylor did not report within-sentence hesitations as a function of number of encoding operations, so it is not clear whether within-sentence events tend to support or dispute reaction time data.[9]

At present, the significance of surface structure in the speaker's formulation of utterances cannot be evaluated fully. However, some comments can be offered as follows. First, a serious criticism of top-to-bottom compositional models has been that the lexicon cannot be selected until after all structural decisions are made (cf. Chomsky, 1965; Jacobs and Rosenbaum, 1970; Perfetti, 1969). It may be, however, that the criticism of hierarchical models can be mitigated by imagining a less stringently linear process. That is, in actual communication the semantic and syntactic aspects of the utterance may be combined, with lexical decisions proceeding simultaneously with structural decisions. To the extent that this alternative view is valid, an hierarchical model may have value in its portrayal of segment-to-segment decision-making in sentence production.

Second, the relatively few studies discussed thus far suggest (1) that structural decisions may be relatively insignificant (or no more significant than lexical choices) in the formulation of utterances, and (2) that the most critical decisions are related to what is vaguely designated as "content." These are, of course, only suggestions and cannot be made to constitute an argument for the primacy of semantic factors in sentence production. However, such an argument might have been possible if the extensive experiments which focus on the function of pauses (to be discussed presently) had incorporated location in the syntax as a major dependent variable. These studies have tended to overlook or ignore the question of location, perhaps because the experimenters believe that structural decisions are less important than "content" decisions in deciding the time course of language production.

[9]The additional data would be particularly helpful since it is difficult to feel confident that latencies preceding one-sentence word definitions correspond to latencies preceding sentence formulation in ordinary discourse.

Molar Units of Speech Production

In models of the speaker examined thus far, the incidence of a pause has been supposed to reflect cognitive activity at that precise point in time. This is true regardless of the size of the encoding unit contemplated. Pauses are presumed to just (or nearly just) precede problematic words when the speaker is portrayed as a lexical decision maker. Similarly, when the speaker is thought to make decisions in larger, structural units, the relevant pauses are sought at the beginnings of those units.

The relationship between pauses and decision units in these models is what Boomer (1970) terms a *proximal* relationship, i.e., the pause refers to local decision-making. An alternative view has been presented in the later work of Goldman-Eisler and her colleagues (Goldman-Eisler, 1968; Henderson *et al.,* 1965, 1966) in which the pause is supposed to signal not only local but also relatively distant decision making. In Boomer's terms, the pause bears a temporally *distal* relationship to the choices to be made.

The theory of a distal function for pauses has been posited to account for consecutive time periods of speech and silence in individual speakers. Henderson *et al.* have plotted cumulative graphs of these periods and fit the data by eye with straight lines. The resulting figures suggest that periods of long pauses and short speech bursts alternate with periods of little pausing and sustained speech. These figures are said to reflect a cycle of acts of planning and production in speech. In particular, the hesitant steep slopes are thought to be devoted to planning while the fluent shallow slopes are supposed to be the outcome of this planning. The two periods are conceived as constituting one psycholinguistic unit, during which the speech from high fluency periods is the consequence of generative activity occurring during periods of hesitancy.

The pause, according to this model, reflects not only lexical choices made for local utterances but also semantic choices made for periods of high fluency, some 5 to 15 sec distant from the actual pause. This portrayal of the role of pauses and the model of the speaker implied is attractive. On the one hand, it suggests units of speech production which are functional rather than structural in nature, i.e., units based on cognitive activity rather than on linguistic analyses which may or may not be relevant to encoding processes. On the other hand, it suggests a possible series of hierarchies of decisions which might go into the composition of these large, functional units. Where the units themselves may reflect semantic decisions, they might include several smaller units of structural and lexical choices.

It is therefore discomfiting to find that such speculations are premature, and that the notion of distal pause effects has been seriously challenged on both methodological and empirical grounds. The most cogent objections derive from the exceedingly careful investigations of speech-silence patterns by Jaffe, Feldstein and their colleagues.

The first objection is methodological. Schwartz and Jaffe (1968) point out that the procedure of fitting by eye is questionable at best. They argue that in any sequence of random events "runs" will occur such that, arbitrarily, one can always find subsequences which give an apparent structure to the result. In support of this, they demonstrate that random sequences of sound-silence durations generated by computer show the same stepwise pattern claimed by Henderson *et al.* (1966). In a later study (Jaffe *et al.*, 1970), the simulation was extended to test the conclusion put forward by Henderson *et al.* (1966) that the longer the steep "planning" slope, the longer the succeeding shallow "fluent" slope. Again, the result was shown to be characteristic of the random graphs generated by the computer. Jaffe and his colleagues conclude that the correlation between successive phases found by Henderson *et al.* is not a psycholinguistic effect but rather a function of the human perceiver when judging slopes and "eyeing in" apparent patterns.

The second objection is empirical. Jaffe and Feldstein (1970) have studied the sound-silence patterns of monologues and dialogues by feeding audio recordings into an electronic signal detection system which transforms each speaker's production into a string of 1's if he is making noises above some cutoff and a string of 0's when he is silent. The strings are classified by computer into categories of vocalizations and pauses. Using a process of successive approximation to fit these data with stochastic models, Jaffe and Feldstein have been able to model the sound-silence sequences in their data as a two-state stochastic process with a 300 msec constraint between transitions. This means that over the range the model fits (98% of the pause and vocalization events), the durations of adjacent pauses and vocalizations must be statistically dependent.

The short-range constraints (300 msec) of on-off patterns found by Jaffe and Feldstein are in direct contradiction to the long-range constraints (5-15 sec) found by Henderson *et al.* In light of this data and in view of the dubious validity of the Henderson *et al.* technique, the notion of cognitive rhythm units seems untenable. In particular, there has been no adequate evidence for the view that pauses reflect decision-making about relatively distant events.

To summarize this section briefly, three models of the speaker have been examined which formulate the significance of pauses in the speaker's production. The earliest model was one in which pauses were supposed to reflect the strength or weakness of verbal habits; the most recent model portrayed pauses as signaling cognitive decisions about both immediate and later speech. With the exception of the most recent model, pauses have been assumed to stand in a temporally proximal relationship to the choices to be made. If this assumption is granted, for it seems tenable at least, then there is support for at least two particular pausal functions: (a) pauses signal some word choices, and (b) may reflect decisions at major constituent boundaries.

A third function for pauses has been discussed only briefly but appears as a theme in several studies. This is the semantic or "content" decision-making that Taylor has emphasized in her work and that Goldman-Eisler and her colleagues considered in their molar model of the speaker. The matter of content is treated somewhat more fully in the following section where the function of pauses for the speaker is examined. In these studies, formal models of the speech act have not been offered although various general relationships between pauses and production have been proposed. The studies are reviewed in detail, following a brief diversion to incorporate the often-neglected data on filled pauses.

Locating the Filled Pause

When Lounsbury identified the hesitation pause, he stipulated that it might occasionally be filled "with hemming and hawing." The possibility of a voiced pause corresponding in location and function to the silent pause was considered by Maclay and Osgood (1959) in a descriptive study of utterances at a conference. They distinguished between silent pauses (SPs) and filled (FPs), with the latter voiced typically as "ah" [a], and less often as "er" [r], "uh" [ϵ], and "mm" [m]. Though both sorts of pauses were more likely to occur before content words than before function words (e.g., before nouns and verbs rather than before prepositions and connectives), there tended to be some distributional differences. Filled pauses occurred relatively more often before function words and at phrase boundaries while silent pauses were more likely before content words and within phrases. However, Maclay and Osgood took pains to caution that the two types of pause were not clearly distinct. Not only were there individual speaker differences in the proportion of filled to silent pauses uttered, but pauses of "either type can occur in any position where the other occurs and [do] so frequently" (1959, p. 39).

To account for the distributional differences observed between filled and silent pauses, Maclay and Osgood hypothesized that FPs occur when two conditions exist simultaneously: (1) the speaker has paused a rather long time ("long enough to receive the cue of his own silence"), and (2) wishes to keep control of the "conversational ball." In order to prevent the listener from speaking during a long silent pause, the speaker presumably produces "some kind of signal . . . which says, in effect, "I'm still in control–don't interrupt me!" (p. 41).

Maclay and Osgood state that FPs should occur at points of higher uncertainty than SPs since FPs presumably follow particularly long silences. However, there have been two challenges to this view. First, Boomer (1965) found that FPs rarely follow long SPs (only 17% of FPs were followed by longer-than-average pauses). Second, Cook (1969b) offers findings suggesting that FPs and SPs occur in the same locations. He reports that words immediately following FPs are more difficult to guess than words in other contexts. Unfortunately, the speech sample used to estimate transition probabilities is not described, precluding adequate comparisons with similar studies of SPs. To sum up, it is possible that SPs and FPs occur in the same syntactic and/or functional locations in speech, but this is merely a possibility at the present time.

THE FUNCTION OF PAUSES FOR THE SPEAKER

In part, the consideration of where pauses occur in speech reflects a concern with their function for the speaker. However, the question of function has been undertaken more systematically in studies of the conditions under which pauses vary. These experiments are summarized in Table II.

Cognitive Variables

Three experiments measured both FPs and SPs under conditions of increasing task difficulty. Goldman-Eisler (1961a) asked subjects to describe the content of brief cartoon stories and to formulate the essential point of each. While the duration of SPs increased during the formulation stage of the task, the proportion of pause time occupied by FPs remained constant (Goldman-Eisler, 1961b).

In an attempt to replicate Goldman-Eisler's work, Lay and Paivio (1970) presented subjects with three tasks: self-description (e.g., age, name, sex), cartoon descriptions, and evaluation of pairs of proverbs. The results showed that both FPs and SPs increased in frequency with increases in the presumed level of abstractness required. Unfortunately, the abstractness of the task seems to be confounded with type of stimuli (verbal *vs.* nonverbal) and with type of speech required (well-practiced *vs.* spontaneous). A more adequate test of the effects of increasing abstractness was performed in a

Table II. Functional Investigations of Silent Pauses and Filled Pauses

Author and date	Speech sample: Size[a]	Speech sample: Speakers	Speech sample: Type	Main independent or correlated variables	Silent pause	Outcome	Filled pause	Outcome
				Cognitive variables				
Goldman-Eisler, 1961a	2 min	9[b]	Monologue	Task difficulty (Description *vs.* explanation)	(>250 msec) Duration	Increase		
Goldman-Eisler, 1961b	2 min	9[b]	Monologue	Task difficulty (Description *vs.* explanation)	Reaction time	Increase	$\frac{\text{FP duration}}{\text{SP duration}}$	No difference
Levin *et al.*, 1967	180 words	12M/12F (5-12 yrs)	Interview	Task difficulty (Description *vs.* explanation)	(>80 msec) Duration Frequency	 Increase Increase	$\frac{\text{FP} + \text{SD}^c}{\text{Words}}$	Increase
Reynolds and Paivio, 1968	5 min (400 words)	12M/12F	Modified monologue	Task difficulty (Definitions)	(>1.5 sec) Frequency	Increase	Ah ratio	Increase
Taylor, 1969	120 sentences	20[b]	Modified monologue	Task difficulty (Sentence duction)	Reaction time	Increase		
Lay and Paivio, 1970	6 min (600 words)	24M	Modified monologue	Task difficulty (Descriptions *vs.* explanation)	(>1 sec) Frequency/t Duration/t	Increase Increase	Ah ratio	Increase
				Affective-state variables				
Mahl, 1956a	6 50 min	1F patient	Interview	P[d]/Anxiety (Judged by therapist)	Patient silence quotient (PSQ = SP duration/t)	Increase		

Mahl, 1956b	60 min	11M/20F outpatients	Interview	P/Anxiety (Rating scale)			Ah ratio	Decrease
Panek and Martin, 1959	60 min	3M/1F outpatients	Interview	Correlated GSR			Ah ratio	No significant correlation
Krause and Pilisuk, 1961	–	13M/F	Interview	T[e]/Anxiety (Topic-induced)			"Procrastinations" (FPs, words, clichés)	No difference
Boomer, 1963	39 1.5 min excerpts	1[b] patient	Interview	Correlated body movement			Ah + repetitions	$r = +.41$
Feldstein, 1964	–	15M/F actors	Reading	Simulations of affect	(>300 msec) Duration/t	Differences		
Kasl and Mahl, 1965	60 min	45M	Interview	T/Anxiety (Topic-induced)			Ah ratio	No difference
Siegman and Pope, 1965a	–	50F	Interview	T/Anxiety (Topic-induced)	PSQ (SP>2 sec)	No difference	Ah ratio	No difference
Siegman and Pope, 1965b	60 min	50F	Interview	P/Anxiety (Inventory)	PSQ (>2 sec)	No significant correlation	Ah ratio	Decrease
				P/Anxiety (Inventory)	(>300 msec) Switching pause	No difference decrease		
Cassotta *et al.*,	4 16 min	50F	Interview	T/Anxiety (Topic-induced)	(>300 msec)	Increase		
Cook, 1969a	30 min	5M/5F	Interview	P/Anxiety (Inventories			Ah ratio	No difference
				T/Anxiety (Topic)				No difference

(Continued on next page)

Table II. Continued

Author and date	Speech sample: Size[a]	Speech sample: Speakers	Speech sample: Type	Main independent or correlated variables	Silent pause	Outcome	Filled pause	Outcome
Pope *et al.*, 1970	320 min (10 min/day)	2M/4F inpatients	Monologues	T/Anxiety (Rating scales)	PSQ	Decrease	Ah ratio	No difference
				T/Depression (Rating scales)		Increase		Decrease
				Social interaction variables				
Levin and Silverman, 1965	4 500 words	48M/F children (10 yrs.)	Monologues	Audience (*vs.* Alone)	Frequency	Increase	Ah ratio	No difference
Feldstein *et al.*, 1967	3 40 min	32M/32F	Dialogue	Screen (*vs.* Vis-a-vis)	(>300 msec) Duration	Decrease		
Lalljee and Cook, 1967	10	14 M	Modified Dialogue	Pressure on speaker (Interruptions)			Ah ratio	No difference

Preston and Gardner, 1968	5 min	50F/45M	Monologues	Factor analysis	(>1.5 sec) Frequency	Related to vocabulary size		
					Duration	Related to social approval	Ah ratio	Related to association-al skills
Ramsay, 1968	–	56F/28M	Interview monologue	Extroversion	(>10 msec) Duartion Frequency	Increase no differ-ence		
Reynolds and Paivio, 1968	(See entry under Cognitive variables)			Audience sensitivity	(>1.5 sec) Frequency	Increase	Ah ratio	Increase
Lay and Paivio, 1970	(See entry under Cognitive variables)			Audience sensitivity	(>1 sec Frequency/t Duration/t	No differ-erence Increase	Ah ratio	No differ-ence

Note: t = total speaker utterance time.
[a]Size per speaker.
[b]Sex not given.
[c]SD = Speech disturbances, which include stutters, word repetitions, and omissions and other paralinguistic effects described by Mahl (1956a).
[d]P = Predispositional anxiety with assessment procedure in brackets.
[e]T = Transient anxiety with experimental procedure in brackets.

related experiment by Reynolds and Paivio (1968). Filled and silent pauses were more frequent when subjects defined abstract rather than concrete nouns.

Silent pauses have been shown to increase with the difficulty or abstractness of the speech task in two experiments: Levin *et al.* (1967) found that SPs were longer and more frequent when children explained rather than described a series of physical demonstrations; Taylor (1969) reported that subjects took longer to begin sentence production as topic words increased in "difficulty" (defined in terms of frequency of occurrence and abstractness).

Two studies by Siegman and Pope seem pertinent to the effects of abstractness or difficulty on FPs: in one they varied the specificity of interviewer remarks (1965a) and in the other they asked subjects to discuss TAT cards ranging from high to low in ambiguity (1966). In the former study, self-rating scales indicated that subjects were generally relaxed throughout; in the latter, effects of anxiety were assessed by rating scales and held constant by a covariate analysis. In each study, the ratio of FPs to words spoken increased with situational uncertainty.

These findings show that both FPs and SPs increase as a function of increases in the presumed difficulty of the tasks. In view of this generality, it is worth asking why Goldman-Eisler's FP data diverge from those of the other studies. There are two possibilities: (1) Two of Goldman-Eisler's three FP indices depend largely or entirely on SP duration. Since SP duration increases with task difficulty, FPs would have to increase simply to yield a result of no difference. A third measure, FP duration/total words uttered, increases with task difficulty for two-thirds of the subjects. This suggests that if frequency measures had been taken, they would have increased. (2) Goldman-Eisler's speech sample was extremely small, less than half the duration required for a stable estimate of one person's speech (cf. Jaffe and Breskin, 1970). Therefore, it may be that the small proportion of FPs to SPs she found is unrepresentative of longer monologues. The high proportion (aobut one-to-one) of FPs to SPs in the data of Reynolds and Paivio supports this presumption.

A general caveat should be mentioned with regard to the vaguely defined "task difficulty" manipulation. The commonality and/or validity of "difficulty" operations has been established only for the most laboratorylike situations. Where the independent variable is defined as abstractness and the task materials are single words, it is a relatively simple matter to select a set of abstract and a set of concrete words and to have this selection verified by a few judges (cf. Reynolds and Paivio, 1968; Taylor, 1969). Where the independent variable is task difficulty, however, or where the task is closer to ordinary conversation, appropriate validity measures are more elusive.

An example of problems encountered in validating task difficulty manipulations is seen in the case of postexperimental self-reports. Although such reports (e.g., Lay and Paivio, 1970) are convenient means of confirming validity where the difficulty manipulation has been effective, they are less useful where that manipulation may have failed. Does a lack of difference in self-reports indicate a failure of the experimental operation, or is the operation judged ineffective only if no differences in pause incidence are found? A validity check should be potent regardless of changes in the dependent variable. It is not clear that self-report has this merit.

Within the methodological constraints outlined above, these studies demonstrate repeatedly that pauses are relevant to cognitive processing. In contrast to the data concerning the location of pauses, however, the present findings cannot be examined in the context of a language model. The problem is the lack of a model which would predict the language functioning of adults at various levels of cognitive difficulty. This lack of psychologically relevant language models becomes more apparent in the next two sections where pauses are related to affective-state and social-interaction variables.

Affective-State Variables

One of the first interpretations of silences and FPs was offered by Mahl (1956a,b) who assumed that these events, along with a variety of vocal clues or signs, constituted disruptions in the behavior of the speaker. He reasoned that since anxiety tends to disrupt complicated behavior, disfluencies in speech might be caused by anxiety. In an exploratory study, he found that silences and FPs tended to be more frequent during phases of interviews judged high rather than low in anxiety.

To make the measures independent of the speaker's verbal output, Mahl defined two ratios. The Ah ratio consists of the number of FPs (at first simply "ah" sounds but later expanded to include the sounds listed on p. 64) divided by the number of words uttered by the speaker within brief interview segments. The "word" count was defined to include all incompleted and completed words, sounds caused by stuttering and incoherency, and FPs. The second measure, the Patient-Silence Quotient (PSQ), was defined as the seconds of silence divided by the seconds available to the patient to talk.

Subsequent experiments have revealed a relationship between pauses and anxiety that is decidedly more complex than Mahl's original proposal. To examine these studies, it is helpful to classify them according to the two primary constructs investigated, predispositional anxiety and situational anxiety, and the two dependent variables of interest here, SPs and FPs.

Predispositional Anxiety

Cassotta *et al.* (1967) divided female college students into high and low scorers on the basis of an anxiety questionnaire, and had them participate in four interviews. No significant correlations were found between anxiety and SP duration. However, conversational switching pauses were significantly different for the high and low anxious subjects. A switching pause was defined as the silence between the speech of participants, and it was assigned to the participant who terminated the silence. High anxious subjects had consistently shorter switching pauses than low anxious subjects across the four interviews. Siegman and Pope (1965b) found insignificant correlations with Manifest Anxiety Scale scores and PSQs, when nursing students spoke on two assigned topics. Pope *et al.* (1970) obtained 10-min daily monologues for a 3-month period from inpatients diagnosed as psychosomatic. On the basis of nurses' ratings on a series of behavior scales, 8 days were selected as high-anxious and 8 days were selected as low-anxious for each subject. Silences (PSQs) decreased on high anxiety days.

These three investigations cover a wide range of SPs from silences longer than 300 msec noted by Cassotta *et al.* to the PSQs used by Siegman and Pope and their colleagues in which only silences of 3 sec or longer were included. Where significant results were obtained, predispositional anxiety was negatively correlated with pausing.

The incidence of FPs was either the same (Cook, 1969a) or lower (Mahl, 1956b; Siegman and Pope, 1965b) in high anxiety groups as it was in low anxiety groups. The observation that high anxious subjects tend to utter fewer FPs than low anxious subjects was dismissed by the investigators who were testing directional hypotheses (i.e., that FP incidence increases with increasing anxiety). However, it is interesting that these negative correlations parallel the negative correlations observed for anxiety scale scores and SPs in some experiments. It appears that when differences are found between high and low anxious subjects, high anxious subjects have fewer switching pauses, fewer SPs and fewer FPs than low anxious subjects. This result is the reverse of Mahl's prediction.

Situational Anxiety

Four within-group designs have been used to test the effects of experimentally induced anxiety on pauses. Krause and Pilisuk (1961) asked subjects to describe their feelings about a wide range of topics. Responses were sorted into high and low anxiety categories by a questionable procedure which eliminated about 80% of the data and responses from one-third of the subjects. Results showed that "Procrastinations" (FPs plus such cliché phrases as "as a matter of fact") were no more predictable than chance on the basis of the anxiety sorting. Kasl and Mahl (1965) found that anxiety induced by selection of interview topics and confirmed by palmar sweat readings did not lead to differences in Ah ratio. Siegman and Pope (1965a), in a similar study, repeated the findings.

In contrast to the generally negative outcomes with FPs, there is evidence that SPs increase with increases in situational anxiety. Siegman and Pope (1965a) found that both reaction times and frequency of SPs increased when "high anxiety" topics were introduced. This was so although, in postexperimental reports, subjects claimed to feel relaxed throughout the interviews. Some confirmation of the effects of topic variation is seen in data (Cassotta *et al.*, 1967) showing that both SPs and switching pauses are longer when subjects are interviewed about topics intended to be stressful rather than neutral.

Thus, situational anxiety seems to be positively related to SPs though possibly not to FPs. The trend for SPs here is the reverse of that seen with predispositional anxiety: when "high anxiety" topics are introduced, subjects tend to pause longer and more often than when topics are designated as "low anxiety." This result conforms to Mahl's early predictions.

To summarize, the data suggest (1) that there is a relation between anxiety and SPs, and (2) that there may be an analogous relation between anxiety and FPs, although this latter point is quite speculative at present. It may be objected that this ignores the central contrast in the data, viz. when "anxiety" refers to the stress tolerance or vulnerability of subjects, it is negatively related to pausing; when "anxiety" refers to the stress level of the situation, it is positively related to pausing. However, it is possible to resolve this apparent contradiction by postulating a U-shaped functional relationship between anxiety and pausing. This U-curve hypothesis has been specifically applied to verbal productivity by Murray (1971). Briefly, Murray argues that situational and dispositional anxiety factors are additive. On the one hand, as situational stress is increased, pausing drops to some minimum and then increases. In the "high anxious" conditions of experiments with humans (which probably involve only moderate stress levels), pausing will appear to be a monotonic function of situational anxiety. However, the performance of highly anxious persons placed in a mildly to moderately anxiety-producing situation should fall beyond the minimum of the U, somewhere on the ascending right shoulder of the curve. Consequently, it should often be the case that pausing is negatively related to predispositional anxiety. This interpretation of the overall verbal productivity data seems to account nicely for the present SP data, and may also be relevant to investigations of FPs. It has the advantage of including the notion of anxiety as a disruptor of speech within the general view of a necessary level of activation for any (coherent) speech production to occur.

Unfortunately, there remain a number of methodological problems which plague attempts to relate pauses to affective states in general and to anxiety in particular. One of these problems is the discovery of an adequate validation procedure for the independent variable. Solutions have ranged from Kramer's (1963) argument that the opinions of heterogeneous judges constitute a series of validity checks on the stress levels of a situation, to the view (Krause and Pilisuk, 1961) that self-reports are sufficient evidence of vulnerability to stress. The value of self-reports has been discussed above; the usefulness of judgments from a heterogeneous sample is dubious, given the customary lack of agreement in homogeneous groups of judges (cf. Boomer and Gooderich, 1961).

Two further solutions to validity assessment have been attempted. Feldstein (1964) used actors' simulations of affect on the assumption that stereotypes might yield clearer relationships among the vocal dimensions and affects than could be expected from "naturalistic emotional expression." Although the data obtained from this approach are interesting, their relevance to spontaneous speech is problematic. In an alternate approach, a few correlational studies have been performed in which speech disfluencies were related to other putative covariants of anxiety. Results of these studies have been generally discouraging, yielding very modest (Boomer, 1963) or negligible (Panek and Martin, 1959) correlations.

In addition to general difficulties associated with validation, there are knotty problems peculiar to the between-group designs used to study predispositional anxiety. First, as Preston (1967) points out, patient and nonpatient groups may differ not only in affective state but also in the attention devoted to encoding speech. For example, they may differ in the sampling and/or evaluation of alternative responses. Second, groups may differ in characteristic pause duration and speed of utterance. Since there is no way to account for the differences in tempo and rhythmicality with the present measurement techniques, it is preferable to use within-group designs where each speaker can serve as his own control.

Social Interaction Variables

Thus far, the speaker has been seen simply as a language generator which pauses either in the course of normal decision-making operations or because of disruptions in those operations. From another perspective, however, the speaker must be seen as part of a larger unit, as a participant in the social act of speech. According to this view, pauses and other phenomena of spontaneous speech should be functionally related to changes in the interpersonal situation and/or to changes in the responsiveness of the speaker, given a constant interpersonal situation.

There is some evidence that SPs increase as the emphasis on interaction increases. For example, 10-year-old children pause more frequently when telling stories before an audience of adults than when they are alone speaking into a microphone (Levin and Silverman, 1965). Moreover, differential sensitivity to others seems to affect SP incidence. Subjects scoring high in an audience sensitivity test paused more frequently when addressing an audience than did low scorers (Raynolds and Paivio, 1968) but these differences were not found in the absence of an audience (Lay and Paivio, 1970). Pause frequency remained constant but duration increased when utterances of subjects scoring high in concern for approval (Preston and Gardner, 1968) and extroversion (Ramsay, 1968) were compared with the vocalizations of low-scoring subjects.

In an important sense, the above results are ambiguous. While they may be due entirely to interpersonal variables, they may also result from the mediating effects of these variables in either extending the normal decisions involved in speech production or increasing the disruptions due to anxiety. This interpretation is particularly tenable since either cognitive or affective-state hypotheses could be used to predict increases in the duration and frequency of SPs. All of the studies cited above conform in part to these predictions. Therefore, none of them can be considered as evidence for a peculiarly interactional hypothesis.

In contrast, there are a few experimental results which seem predictable only from a social psychological point of view. These predictions are derived by extending Maclay and Osgood's (1959) interpretation of FP utterances to include SPs. If long pauses typically result in a loss of control of the "conversational ball," it follows that speakers should utter shorter pauses when they wish to maintain such control. Moreover, to the extent that any silence is an opportunity for a transfer of control, the speaker who wishes to continue speaking should utter fewer SPs of all durations. There are several predictions which derive from this hypothesis. If the speaker's desire to maintain control of a conversation is signaled by an increase in FPs and a decrease in both duration and frequency of SPs, then (1) these events should be more likely in dialogues than in monologues, (2) they should be less likely where the subject wishes to break off speaking, (3) they should be unchanged where the number of potential speakers remains constant, and (4) they should be more likely where the speaker lacks visual means of controlling the conversation.

There is some support for each prediction. (1) Speakers from the same subject population produce higher Ah ratios in interviews (Siegman and Pope, 1965a) than in monologues (Siegman and Pope, 1966). (2) Patients have longer SPs but fewer FPs when they are depressed (and presumably uninterested in conversing) than when they are anxious (Pope *et al.,* 1970). (3) Subjects presenting monologues do not differ significantly in rate of FP utterance regardless of whether an audience is present or absent, or whether they have scored high or low in an audience-sensitivity test (Lay and Paivio, 1970; Reynolds and Paivio, 1968).[10] (4) When an opaque screen is placed between conversationalists, they utter more FPs (Kasl and Mahl, 1965) but shorter SPs (Cassotta *et al.,* 1967) than when they are face-to-face.

To date, there has been only one published attempt to test Maclay and Osgood's "control" hypothesis directly. Lalljee and Cook (1969) had a confederate either remain silent or interrupt subjects in an artificial conversation. Although the manipulation served to increase the subject's verbal productivity and interruptions, it had no noticeable effect on FPs (SPs were not reported). The authors conclude that Maclay and Osgood's theory is not correct, at least for dialogues (p. 27).

It is possible, however, that a "high pressure" manipulation of three interruptions in 2 min represents an ordinary rate of verbal interchange, unlikely to prompt the subject to "hold the floor" under any circumstances. In general, it seems unwise to assume that some arbitrary number of interruptions will be adequate to describe a "high" rather than a "moderate" or "low" pressure condition. Indeed, it is likely that a conversation with no interruptions is more extraordinary than one with several. In the absence of any systematic investigation of conditions under which interruptions occur, the Lalljee and Cook study cannot be considered an adequate test of the "control" hypothesis.

To summarize, it appears that two sorts of social interaction variables influence pausing in spontaneous speech: (1) mediating variables, such as changes in the audience situation and predispositional responsiveness to listeners, and (2) control variables, such as number of potential speakers and individual desire to speak. The effects of these manipulations on pauses seem to be antithetical. Where mediating variables increase the frequency and duration of SPs and generally have no effect on FPs, control variables appear to reduce the frequency and duration of SPs and to increase the rate of FPs.

[10]Although failure to reject the null hypothesis does not constitute support for an alternate hypothesis, it is nevertheless noteworthy that a "control hypothesis" is the only one among the three suggested that clearly does *not* predict an increase in FPs when audience-sensitive subjects speak before a large group of listeners.

It is plausible that mediating variables actually disrupt or make more complex the processes of speech encoding.

CONCLUDING REMARKS

What is the next step? Given that pauses are relevant to cognitive processing and to social and affective-state variables, and that pause location indicates at least lexical and structural levels of decision-making, what remains to be established?

From a methodological viewpoint, there is clearly a need for further investigations of the interrelationships between location and function of pauses. For example, when pauses increase in frequency as a function of cognitive variables, where do the increases occur? When the same question is posed with regard to social variables, are the locations the same or different?

Certainly, there is insufficient support at the moment for Goldman-Eisler's data favoring an overall scheme or compositional plan and for Taylor's results demonstrating the significance of "content" as opposed to structural decisions. The former findings should be replicated and the latter extended to reveal the relationships between latency and within-sentence hesitations.

Further methodological problems include the following: (1) little is known about the distributions of FPs relative to SPs or relative to possible grammatical encoding units, (2) SPs and FPs are accepted as distinct entities and possibilities for their cooccurrence have been ignored, and (3) number of participants in the speech act and type of participation have not been explored either as an experimental variable or an aspect crucial to replication.

The more significant requirements, however, are in the realm of theory. How may the speaker be modeled to conform to the characteristics of pauses described thus far? Three directions seem clear. First, it seems necessary to expand the narrow view of a single-leveled decision-making process occurring at any given moment to include possibilities of multileveled models in which content or theme decisions are made initially while later structural and lexical decisions proceed simultaneously. Speakers hesitate at the boundaries of major constituents and prior to low probability words, and these behaviors are too close in time to permit acceptance of a totally hierarchical model in which the lexicon is a minor concern, left till last in the speaker's composition.

Second, there is surely a need to consider new levels of utterance analysis. On the one hand, the potential of sophisticated linguistic analyses such as those of transformational grammar has not been tapped. The recent work of Brown and Miron (1971) in the study of reading, however, suggests that such an approach can be fruitful. On the other hand, nonstructural analyses should also be considered. This direction is suggested by the findings of two investigations which are difficult to incorporate within present-day linguistic endeavors: (1) In both active and passive sentences, Wilkes and Kennedy (1970) found only the subject-predicate break marked by pausing. (2) Significant increases in the frequency and duration of SPs were found (O'Connell *et al.,* 1970) when German subjects recounted stories containing a semantically anomalous sentence. These studies suggest that pauses may be related to semantic or propositional factors in the utterance, supporting the observations of Taylor discussed earlier.

Finally, it may be that the search for *units* of encoding has itself been the most serious stumbling block to a view of multileveled decision-making by the speaker. Derived from probabilistic models of verbal behavior, the notion of units seems to require a stringently sequential process in which segregated groups of elements have stronger or weaker associations to each other. If the speaker makes a simultaneous lexical-structural decision, what is the "unit?" Perhaps it is wise to leave the search for units to engage in a pursuit of processes or operations which can and often do cooccur.

ACKNOWLEDGMENTS

For helpful comments on an earlier draft of this paper, I am indebted to Peter Reich, Oliver Mott, and Michael Dobrovolsky of the University of Toronto, Peter Dean of DCIEM at Toronto, Wayne Sailor of the University of Kansas, and Insup Taylor of Guelph University.

REFERENCES

Barik, H. C. (1968). On defining juncture pauses: A note on Boomer's "Hesitation and grammatical encoding." *Language and Speech* **11**: 156-159.

Bernstein, B. (1962). Linguistic codes, hesitation phenomena and intelligence. *Language and Speech* **5**: 31-46.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In Hayes, J. R. (ed.), *Cognition and the Development of Language.* Wiley, New York, pp. 279-352.

Boomer, D. S. (1936). Speech disturbance and body movement in interviews. *J. Nerv. Ment. Dis.* **136**: 263-266.

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

Boomer, D. S. (1970). Review of F. Goldman-Eisler, Psycholinguistics: experiments in spontaneous speech. *Lingua* **25**: 152-164.

Boomer, D. S., and Goodrich, D. W. (1961). Speech disturbance and judged anxiety. *J. Consult. Psych.* **25**: 160-164.

Brown, E., and Miron, M. S. (1971). Lexical and syntactic predictors of the distribution of pause time in reading. *J. Verb. Learn. Verb. Behav.* **10**: 658-667.

Cassotta, L., Feldstein, S., and Jaffe, J. (1967). The stability and modifiability of individual vocal characteristics in stress and nonstress interviews. Research Bulletin No. 2. New York: William Alanson White Institute.

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

Coleman, H. O. (1914). Intonation and emphasis. *Misc. Phonetica* 1: 6-26.

Cook, M. (1969a). Anxiety, speech disturbance and speech rate. *Brit. J. Soc. Clin. Psych.* 8: 13-21.

Cook, M. (1969b). Transition probabilities and the incidence of filled pauses. *Psychonom. Sci.* **16**(4): 191-192.

Crystal, D. (1969). *Prosodic Systems and Intonation in English.* Cambridge University Press, Cambridge, Great Britain.

Ervin-Tripp, S. M., and Slobin, D. I. (1966). Psycholinguistics. *Ann. Rev. Psych.* 17: 435-474.

Feldstein, S. (1964). Vocal patterning of emotional expression. In Masserman, J. H. (ed.), *Development of Research.* Grune & Stratton, New York, pp. 193-210.

Fillenbaum, S. (1971). Psycholinguistics. *Ann. Rev. Psych.* 22: 251-308.

Goldman-Eisler, F. (1951). The measurement of time sequences in conversational behavior. *Brit. J. Psych.* 42: 355-362.

Goldman-Eisler, F. (1952). Individual differences between interviewers and their effect on interviewees' conversational behavior. *J. Ment. Sci.* 98: 660-671.

Goldman-Eisler, F. (1954). On the variability of the speed of talking and on its relation to the length of utterances in conversations. *Brit. J. Psych.* 45: 94.

Goldman-Eisler, F. (1955). Speech-breathing activity–a measure of tension and affect during interviews. *Brit. J. Psych.* 25: 53-63.

Goldman-Eisler, F. (1956a). Speech-breathing activity and context in psychiatric interviews. *Brit. J. Psych.* 29: 35-48.

Goldman-Eisler, F. (1956b). A contribution to the objective measurement of the cathartic process. *J. Ment. Sci.* **102**: 78-95.

Goldman-Eisler, F. (1956c). The determinants of the rate of speech and their mutual relations. *J. Psych. Res.* 2: 137-143.

Goldman-Eisler, F. (1958a). Speech production and the predictability of words in context. *Quart. J. Exper. Psych.* **10**: 96-106.

Goldman-Eisler, F. (1958b). The predictability of words in context and the length of pauses in speech. *Language and Speech* 1: 226-231.

Goldman-Eisler, F. (1961a). Hesitation and information in speech. In Cherry, C. (ed.), *Information Theory.* Butterworth, London, pp. 162-174.

Goldman-Eisler, F. (1961b). A comparative study of two hesitation phenomena. *Language and Speech* 4: 18-26.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech.* Academic Press, New York.

Hegedüs, L. (1953). On the problem of the pauses of speech. *Acta Ling. Acad. Sci. Hung* 3: 1-34.

Henderson, A., Goldman-Eisler, F., and Skarbeck, A. (1965). Temporal patterns of cognitive activity and breath control in speech. *Language and Speech* 8: 236-242.

Henderson, A., Goldman-Eisler, F., and Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech* 9: 207-216.

Hockett, C. F. (1968). *The State of the Art.* Mouton, The Hague.

Hörmann, H. (1971). *Psycholinguistics.* Springer-Verlag, New York.

Jacobs, R. A., and Rosenbaum, P. S. (1968). *English Transformational Grammar.* Blaisdell, Toronto.

Jaffe, J., and Breskin, S. (1970). Temporal patterns of speech and sample size. *J. Speech Hear. Res.* **13**: 667-668.

Jaffe, J., Breskin, S., and Gerstman, L. J. (1970). On the range of sequential constraint in monologue rhythms. *Psychonom. Sci.* 19: 233-234.

Jaffe, J., and Feldstein, S. (1970). *Rhythms of Dialogue.* Academic Press, New York.

Johnson, N. F. (1965). The psychological reality of phrase-structure rules. *J. Verb. Learn. Verb. Behav.* **4**: 469-475.

Johnson, N. F. (1968). Sequential verbal behavior. In Dixon, T. R. and Horton, D. L. (eds.), *Verbal Behavior and General Behavior Theory.* Prentice-Hall, Englewood Cliffs, N.J., pp. 421-450.

Kasl, S. V., and Mahl, G. F. (1965). The relationship of disturbances and hesitations in spontaneous speech to anxiety. *J. Personal. Soc. Psych.* **1**: 425-433.

Kramer, E. (1963). Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psych. Bull.* **60**: 408-420.

Krause, M. S., and Pilisuk, M. (1961). Anxiety in verbal behavior: A validation study. *J. Consult. Psych.* **25**: 414-419.

Lalljee, M. G., and Cook, M. (1969). An experimental investigation of the function of filled pauses in speech. *Language and Speech* **12**: 24-28.

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

Lay, C. H., and Paivio, A. (1969). The effects of task difficulty and anxiety on hesitations in speech. *Can. J. Behav. Sci.* **1**: 25-37.

Levin, H., and Silverman, I. (1965). Hesitation phenomena in children's speech. *Language and Speech* **8**: 67-85.

Levin, H., Silverman, I., and Ford, B. L. (1967). Hesitations in children's speech during explanation and description. *J. Verb. Learn. Verb. Behav.* **6**: 560-564.

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists. *Word* **21**: 50-53.

Lounsbury, F. G. (1954). Transitional probability, linguistic structure, and systems of habit-family hierarchies. In Osgood, C. E., and Sebeok, T. A. (eds.), *Psycholinguistics: A Survey of Theory and Research Problems.* Waverly Press, Baltimore, pp. 93-101.

Lynch, G. (1934). A phonophotographic study of trained and untrained voices reading factual and dramatic material. *Arch. Speech* **1**: 9-25.

Maclay, H., and Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word* **15**: 19-44.

Mahl, G. F. (1956a) Disturbances and silences in the patient's speech in psychotherapy. *J. Abnorm. Soc. Psychol.* **53**: 1-15.

Mahl, G. F. (1956b). Disturbances in the patient's speech as a function of anxiety. Paper delivered before the Eastern Psychological Association, Atlantic City, N.J. Reprinted in Pool, I. (ed.), *Trends in Content Analysis.* Univeristy of Illinois Press, Urbana, 1959, pp. 89-130.

Martin, E., and Roberts, K. H. (1967). Sentence length and sentence retention in the free-learning situation. *Psychonom. Sci.* **8**: 535-536.

Martin, E., Roberts, K. H., and Collins, A. M. (1968). Short-term memory for sentences. *J. Verb. Learn. Verb. Behav.* **7**: 560-566.

Matarazzo, J. D., Wiens, A. N., and Saslow, G. (1965). Studies of interview speech behavior. In Krasner, L. and Ullmann, L. P. (eds.), *Research in Behavior Modification: New Developments and Their Clinical Implications.* Holt, Rinehart & Winston, New York.

Mishler, E. G., and Waxler, N. E. (1970). Functions of hestiations in the speech of normal families and families of schizophrenic patients. *Language and Speech* **13**: 102-117.

Murray, D. C. (1971). Talk, silence, and anxiety. *Psych. Bull.* **75**: 244-260.

Murray, E., and Tiffin, J. (1934). An analysis of some basic aspects of effective speech. *Arch. Speech* **1**: 61-83.

O'Connell, D. C., Kowal, S., and Hörmann, H. (1970). Semantic determinants of pauses. In d'Arcais, G. B. F., and Levelt, W. J. M. (eds.), *Advances in Psycholinguistics.* American Elsevier, New York, pp. 218-223.

Osgood, C. E. (1963). On understanding and creating sentences. *Amer. Psych.* **18**: 735-751.

Osgood, C. E., and Jenkins, J. J. (1954). A psycholinguistic analysis of decoding and encoding. In Osgood, C. E. and Sebeok, T. A. (eds.), *Psycholinguistics: A Survey of Theory and Research Problems.* Waverly Press, Baltimore, pp. 126-153.

Panek, D. M., and Martin, B. (1959). The relationship between GSR and speech disturbances in psychotherapy. *J. Abnorm. Soc. Psych.* **58**: 402-405.

Perfetti, C. A. (1969). Lexical density and phrase structure depth as variables in sentence retention. *J. Verb. Learn. Verb. Behav.* **8**: 719-724.

Pike, K. L. (1945). *The Intonation of American English.* University of Michigan Press, Ann Arbor.

Pope, B., Blass, T., Siegman, A. W., and Raher, J. (1970). Anxiety and depression in speech. *J. Consult. Clin. Psych.* **35**: 128-133.

Preston, J. M. (1967). Encoding and decoding in oral communications involving pleasant and unpleasant content. Unpublished doctoral dissertation, University of Western Ontario.

Preston, J. M., and Gardner, R. C. (1967). Dimensions of oral and written language fluency. *J. Verb. Learn. Verb. Behav.* **6**: 936-945.

Průcha, J. (1970). Problems of generative model in psycholinguistics. Paper presented at the Symposium of Algebraic Linguistics, Smolenice, Czechoslovakia, February, 1970.

Ramsay, R. W. (1968). Speech patterns and personality. *Language and Speech* **11**: 54-63.

Reynolds, A., and Paivio, A. (1968). Cognitive and emotional determinants of speech. *Can. J. Psych.* **22**: 164-175,

Rochester, S. R. (1971). The significance of pauses in decoding utterances. Unpublished manuscript.

Rommetveit, R. (1968). *Words, Meanings, and Messages.* Academic Press, New York.

Saslow, G., and Matarazzo, J. D. (1959). A technique for studying changes in interview behavior. In Rubinstein, E. A. and Parloff, M. B. (eds.), *Research in Psychotherapy.* American Psychological Association, Washington, D.C., pp. 125-159.

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

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Tech. J.* **30**: 50-64.

Siegman, A. W., and Pope, B. (1965a). Effects of question specificity and anxiety producing messages on verbal fluency in the initial interview. *J. Personal. Soc. Psych.* **2**: 522-530.

Siegman, A. W., and Pope, B. (1965b). Personality variables associated with productivity and verbal fluency in an initial interview. In *Proc. 73rd Ann. Convention of the American Psychological Association.* American Psychological Association, Washington, D.C., pp. 273-274.

Siegman, A. W., and Pope, B. (1966). Ambiguity and verbal fluency in the TAT. *J. Consult. Psych.* **30**: 239-245.

Tannenbaum, P. H., Williams, F., and Hillier, C. S. (1965). Word predictability in the environment of hestiations. *J. Verb. Learn. Verb. Behav.* **4**: 134-140.

Taylor, I. (1969). Content and structure in sentence production. *J. Verb. Learn. Verb. Behav.* **8**: 170-175.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readibility. *J. Quart.* **30**: 415-433.

Trager, G. L., and Smith, H. L., Jr. (1951). *An outline of English structure,* Studies in Linguistics, Occasional papers, **3**. (American Council of Learned Societies, Washington; Battenburg Press, Norman, Okla.

Wilkes, A. L., and Kennedy, R. A. (1970). Response retrieval in active and passive sentences. *Quart. J. Exper. Psych.* **22**: 1-8.

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proc. Amer. Phil. Soc.* **104**: 444-466.