The perception of hesitation in spontaneous speech¹

JAMES G. MARTIN AND WINIFRED STRANGE

CHICO STATE COLLEGE

The issue in this paper was whether attending to acoustic elements and to message elements in a speech signal were compatible operations. In four experiments Ss listened for pauses and other hesitation phenomena in spontaneous speech; in three the task was reproduction of heard speech to include hesitations; in one the task was simply the marking of heard hesitations on transcripts. Experimental variables were instructions, degree of "ungrammaticality" of hesitations in speech inputs, time interval between listening and reproduction, and task manipulations along a continuum between simple hesitation detection and hesitation detection plus simultaneous speech decoding. Results were: (1) In all experiments Ss displaced within-constituent hesitations to constituent boundaries, suggesting a grammatical organization between input and output. (2) Instructional set to reproduce hesitations increased hesitations and words but at the expense of per cent words correct, suggesting that attending to acoustic elements such as hesitations was an interfering task during speech decoding. (3) The hesitation shift persisted in the hesitation-marking task when simultaneous speech decoding was required by the nature of the task, indicating that speaking (encoding) characteristics may not completely account for the shift. (4) The distribution of hesitation marking errors toward grammatical organization seemed to require an account in terms of perceptual processes during listening.

This paper is concerned with whether certain acoustic characteristics of a spontaneous speech signal, viz., hesitations, are perceived while the speech is being decoded. There seem to be at least two reasons for considering hesitations in speech, along with words and other linguistic units. Hesitations may provide hints of psychological processes at work in the production and reception of speech. In addition, they are probably among the relatively noninformative elements of ordinary speech signals, and hence the task of consciously discriminating or detecting these acoustic effects presumably can be separated from that of decoding or understanding the message in the sound signal. Such a separation may not be possible if listeners are asked to identify or detect, say, words, which are a part of the message. Both of these kinds of tasks are called perception (cf. Garner, 1966; Gibson, 1966; Lane, 1965; Liberman, 1957).

The experiments reported below concern the perception of hesitations in speech and were based

upon two general questions. The first was about the relation between the elements as presented in an acoustic speech signal and the elements as decoded. Since this question could be answered only in terms of listener responses, it was necessary to ask also whether decoding effects could be separated from the memory and encoding characteristics of the listener's response.

To attempt an answer to these questions the general approach taken in the following experiments was to ask listeners to decode speech in one pass and attend to hesitations at the same time, in order to observe the effect of this double task on performance, or indeed, to determine if it was possible at all to impose the two tasks simultaneously. The preferred method was to measure the displacement in the listener's report of the location of hesitations in the acoustic signal, rather than to measure the adequacy of his speech decoding. The results of the experiments were considered to have some bearing on the theory that hesitations, and perhaps other acoustic features in speech as well, are not ordinarily "heard," since they are usually noninformative, and that attempts to direct a listener's attention to them will probably be resisted if he is asked also to decode the message in the signal.

The experiments to be reported have many features of experimental design and theoretical rationale in common with two earlier experiments (Martin, in press; Martin & Strange, in press). A brief outline of the theoretical model and a description of one of these experiments follows.

The model is concerned with the relation between the structure of messages and their acoustic representation. It is assumed that the syntactic-semantic structure of the speaker's intention precedes words during speaking but follows words during listening. The speaker upon intending a message is at once provided with its structure; one consequence is that he can begin talking before he has chosen all his words. Then hesitations, which are known to occur often prior to high-information words (Goldman-Eisler, 1958a; MacLay & Osgood, 1959), will mark the points of delay in matching words with intentions. These hesitations, while marking speaker uncertainty, have no utility for the listener. He must perceptually reorganize the left-to-right sound sequence in order to recover the underlying syntactic-semantic structure of the speak-

er's message. The model assumes that in doing so he biases out the noise in the sound signal; among the noisy elements are the speaker's hesitations.

In the first experiment related to the model (Martin, in press) speakers described Thematic Apperception Test (TAT) cards. Listeners yoked in pairs to speakers heard their recorded utterances and attempted to reproduce them. The speakers were called encoders, the listeners decoders. All words in encoder and decoder utterances were classified as content words (high information: noun, verb, adjective, adverb) or as function words. On the assumption that a decoder's reproductions reflect the way he decoded what he heard, it was predicted that hesitations preceding content words in encoder speech would shift relatively to precede function words in decoder speech. Such a result could be taken to indicate a displacement of hesitations from within to between grammatical constituents. The following examples illustrate what might be expected according to this point of view. The first was spoken by an encoder; slashes indicate pauses of some kind:

"An' the fella's climbing a rope/probably in a gym class/trying to/beat a certain time limit."

A decoder reported this as:

"And the fellow's tryin' to climb the rope in a gym class/and trying to beat a certain time limit."

This decoder hesitated only once, prior to a function word, and added and deleted words as well in his reproduction. These examples are "cleaner" than usual; utterances like the following are just as common:

"Somebody/looks like a girl/sitting down looks like/she's crying or something somethin' happened/sad."

It should be noted that such speech, while easily understood, nevertheless does not readily yield up an exact structural description. (Further, note the effect of the pause marks on guessing the structure of the utterance.) Given considerations like these, it is difficult to assign major, as opposed to minor, constituent boundaries, and word class. Therefore classification of words as content or function was based on data from formal written text (Miller, Newman, & Friedman, 1959).

The results of the experiment were as expected; a relatively lower proportion of decoder hesitations preceded content words.

The fact that "ungrammatical" (within-constituent) hesitations when they occur are displaced in decoder reproductions suggests that listeners normally may ignore hesitations, and perhaps other characteristics of the acoustic signal as well. In all likelihood this is for precisely the reason that while occasionally these are grammatical or informative, usually they are not. The experimental results were

thus consistent with a theory stressing the active role of speech perception mechanisms, which would assume that hesitations are filtered out during decoding.

It seems entirely possible, however, that the listeners in the experiments heard hesitations but merely omitted them during reproduction. If this is true they should be able to place them correctly if asked. A hesitation shift under these conditions could indicate the impossibility of performing both tasks simultaneously, but it could also mean that the listener was not responding to the hesitation task, since it is surely more difficult to attend to hesitations in speech than not. The problem in the experiments below was to attempt to separate these possibilities to identify the basis for the hesitation shift.

In Experiment 1, decoders were asked to include hesitations and other irregularities in their reproductions. The results were that while hesitating relatively as often as the encoders they imitated, these "exact" decoders displaced hesitations toward grammatical constituent boundaries to about the same degree as comparison decoders. In Experiment 2, decoders were given criterion-determined practice in placing hesitations to insure their understanding of the task. The encoder utterances chosen for this experiment had an unusually high proportion of hesitations preceding content words to provide a wide range of possible hesitation shifts. In this experiment too the results showed that decoders given "exact" hesitation instructions and practice trials distributed hesitations rather like ordinary decoders. It was concluded from Experiments 1 and 2 that while the hesitation shift did not necessarily implicate perceptual processes during decoding, it did seem to be a rather intractable characteristic of the decoder's response task.

Experiment 3 was used as a basis for comparing results in Experiment 4. Experiment 4 provided a decoding task with a different response measure. Listeners marked hesitations on transcripts as they listened. Two conditions required only that hesitations be detected and marked, a task approximately like that of research assistants routinely scoring speech protocols for hesitation data. A third condition required Ss to mark transcripts for hesitations and then write the words from memory. In the fourth condition Ss heard utterances first and then marked transcripts from memory. Thus in Experiment 4 the conditions did not require S to encode speech at all in the course of placing hesitations. The fourth condition seems to require the simultaneous processing of hesitations and words in a way the first three conditions do not, since it requires word recall, but only in order to perform the hesitation marking task. The results were a

progressive displacement of hesitations with increasing departure from the task of simple hesitation detection. In the last condition, Ss marked relatively higher proportions of function words which were not "there," suggesting that neither memory nor response bias processes could completely account for the shift.

EXPERIMENT 1

This experiment compared encoder and decoder hesitations as in the earlier studies. In part the purpose was to determine whether decoder hesitations were displaced because they were not heard or simply because decoders did not think it a part of their task to reproduce hesitations as well as words. The Ss were assigned into triples to include a second decoder condition in which Ss were explicitly instructed to reproduce hesitations and other irregularities in the encoder speech. The three groups were called encoders, exact decoders, and ordinary decoders. The method and results are described in detail to provide a basis for comparing later experiments.

Method

Subjects. The 129 Ss from introductory psychology classes participated for extra credit or to fulfill a class requirement. All spoke English as their native language. Assignment into triples was haphazard and determined by S's order of appearance at the laboratory. For convenience, 5-10 encoders were run successively, followed by their ordinary and exact decoders randomly assigned to one of these two conditions, then more encoders, and so on.

Procedure for encoders. The S who served as encoder in a triple was fitted with a microphone held by a wire neck loop. His instructions were, in effect, to describe in one sentence each TAT card presented to him. They read as follows, with italics indicating emphasis: "Here's what will happen in this experiment. I will hold up some pictures one at a time. Look at each picture and try to describe as accurately as you can what's going on, in just one sentence. Now, you won't have much time, so don't take too long to explain what each picture is about. Here are two for practice. (Two cards were shown. The practice criterion was any utterance in response to each of the two cards.) OK. Now we will begin."

Next the 20 experimental TAT cards were presented to S at 15-sec intervals and his utterances were taped magnetically on Ampex equipment after passing through a compressor amplifier circuit. An utterance was taken to include the total interval between onset and end of vocal response to one card. Later the encoder's utterances were re-recorded to eliminate long stretches of silent in-

terval between utterances, that is, the silent interval between the end of responding to one card and the onset of responding to the next. Typewritten transcripts were prepared.

Procedure for ordinary decoders. This S was fitted with a microphone and headphones. His instructions were: "In this experiment I will play some recorded speech to you through headphones. As soon as you hear the recorder stop you try to say what you've heard into your microphone. Do you understand? OK. Here is an example. (Two utterances were presented. The practice criterion was a response to each.) That's all there is to it. Now here's the experiment."

Then E, following along on the typewritten transcript, started the playback equipment. As he stopped it 3 sec after the end of the utterance, the decoder attempted to reproduce it. When the decoder had finished, the next utterance was presented. The decoder utterances were transcribed also.

Procedure for exact decoders. This S was treated exactly as the ordinary decoder in his triple except that his instructions added the following: "Be sure to repeat everything *exactly*. That includes mistakes, ungrammatical words, *even pauses and hesitations*."

Word classification. Miller, Newman, and Friedman's (1958) list of 363 function words was used as the basis for classification. Their list includes what are traditionally called articles, prepositions, pronouns, numbers, conjunctions, and auxiliary verbs, plus certain irregular forms, and was determined by starting with the function words listed by Fries (1952) and augmenting the list on the basis of their grammatical intuitions. The important characteristic of this classification for Miller, Newman, and Friedman (and for the present experiments as well) was that the function words were far more common than the content words in the printed text they analyzed; the 363 function word types (6.4%) accounted for 59% of the word tokens in their sample, while 5180 different content words (93.6%) accounted for the remaining 41% of the words.

Analysis of hesitations. Two scorers listened to each utterance and marked hesitations on copies of the typewritten transcripts, alternating between the different coder types in a haphazard order. They worked independently of each other and without any knowledge of the purpose of the experiment or the word classification.

Following Maclay and Osgood (1959) hesitations were classified as Repeats, False Starts, Filled Pauses, and Unfilled Pauses. Briefly, Repeats were all nonsemantic repetitions, False Starts were incomplete or self-interrupted utterances, and Filled Pauses were occurrences of the common hesitation devices in English, *ε*, *æ*, *ɪ*, *ə*, *m*. Unfilled Pauses were hesitations between words judged to be abnormal for the speaker yielding the utterance.

To familiarize the scorers with their task, one of the authors collaborated with each in marking the utterances of one speaker. Then each scorer marked the remainder, listening to each utterance at least twice.

Results

All statistical tests below unless indicated otherwise were based upon the standard A by B by S repeated measures analysis of variance with coder types (encoder, decoder, etc.), word types (content vs function), and Ss as factors, and three within-Ss error terms. The mean square for coder types was tested against the Coder Types by Ss interaction, etc. Newman-Keuls tests were used to compare pairs of means.

Words. The distribution of words given by each coder group is shown in Table 1. The pattern for encoders and ordinary decoders was quite like earlier results with spontaneous speech (Martin, in press; Martin & Strange, in press), particularly the split between content and function words (near 42 vs 58%). While these percentages cannot be compared directly with Maclay and Osgood's sample of spontaneous speech since they used the Fries (1952) word classification, they are quite like those in Miller, Newman, and Friedman's printed text, in which the investigators found 59% function words. This distribution of content and function words seems to be a fairly stable characteristic of English speech. More relevant for the present experiment was that exact decoders on the average yielded virtually the same number of words as ordinary decoders. Apparently the instructions to reproduce hesitations and other irregularities had little effect on the number or classification of words produced by exact decoders.

Hesitations. The basic data were the proportions of content and function words preceded by hesitations. Thus all hesitation measures were in terms of hesitations relative to number of words, over all Ss and for individual Ss where appropriate. Table 2 shows these proportions (times 100 = percentages) over all Ss for each kind of hesitation as well as Total Hesitations. (The percentages for the four types of hesitations do not sum to the percentages for Total Hesitations since for the latter measure only one hesitation was counted if there were two different hesitations before the same

Table 1. Mean (M) and Per cent (P) Words in Experiment 1

	Encoders		Ordinary Decoders		Exact Decoders	
	M	P	M	P	M	P
Content	129.1	42.0	117.5	43.1	115.5	42.4
Function	178.4	58.0	154.9	56.9	156.7	57.6
Total	307.5	100.0	272.4	100.0	272.2	100.0

Table 2.

Percentages of Words Preceded by Hesitations in Experiment 1

		Ordinary		Exact	All
		Encoders	Decoders	Decoders	
Total Hesitations	Content	14.58	7.81	12.46	11.71
	Function	11.90	8.77	13.02	11.24
	Both	13.02	8.36	12.78	11.45
Unfilled Pauses	Content	10.93	5.22	9.12	8.50
	Function	9.41	6.28	10.13	8.65
	Both	10.05	5.82	9.70	8.59
Filled Pauses	Content	1.53	1.37	1.87	1.59
	Function	0.96	1.24	1.31	1.16
	Both	1.20	1.29	1.55	1.34
Repeats	Content	1.44	0.51	0.89	0.96
	Function	0.52	0.54	0.67	0.57
	Both	0.91	0.53	0.76	0.74
False Starts	Content	0.74	0.87	0.69	0.77
	Function	1.12	0.95	1.05	1.04
	Both	0.96	0.91	0.90	0.93

word.) For the statistical analysis six proportions were calculated for each S triple, e.g., number of Repeats before content words by the first encoder divided by the number of content words he yielded, etc. The standard arcsine transformation was applied to each proportion for analysis of variance.

Counts of hesitations by the two scorers were averaged. Scorer agreement in this experiment as measured by percentage overlap of hesitations heard was about 81% for Total Hesitations. False Starts were lowest (70%), and Filled Pauses highest (87%). (Scorer agreement varies among pairs of scorers. Overlap by scorers for Experiment 4 was 90%. The same scorers of course were used across all within-experiment comparisons.)

First considered were the analyses of Total Hesitations and Unfilled Pauses; most hesitations typically are the latter and thus these two measures provide the most reliable data. Encoders placed relatively more Total Hesitations before content words than did either decoder group. As Table 2 shows, the differences between content and function words were: encoders, +2.68%; ordinary decoders, -0.96%; exact decoders, -0.56%. The test of the differences between these differences was the interaction between coder types and word types ($F=6.59$, $df=2/84$, $p<.005$). The corresponding differences for Unfilled Pauses, +1.52, -1.06, and -1.01%, respectively, were also reliable ($F=4.32$, $df=2/84$, $p<.025$). The results for encoders and ordinary decoders were quite similar to those in earlier experiments (Martin, in press; Martin & Strange, in press). In addition, the present experiment appears to show the predicted hesitation shift for exact decoders as well.

It was considered that the exact decoders may

not have responded to their instructions. The instructions must have had some effect, however, since their hesitation rate exceeded that for ordinary decoders and was nearly equal to that for encoders. Newman-Keuls tests showed the ordinary decoder group below the others ($F=39.86$ and 39.57 , $df=2/84$, $ps<.01$, for Total Hesitations and Unfilled Pauses, respectively). The increased exact decoder hesitation rate, since it did not seem to affect the content-function shift, was taken to be an instruction-induced motivational effect.

The remaining three types of hesitations, while typically infrequent, were potentially relevant in this experiment. Two of these at least, Filled Pauses and Repeats, are presumably rather easily detected and should show up in exact decoder reproductions. For these kinds of hesitations the following effects were significant. There were more Filled Pauses preceding content words ($F=6.49$, $df=1/42$, $p<.05$). Repeats produced differences among coder types ($F=4.59$, $df=2/84$, $p<.05$), between word types ($F=5.11$, $df=1/42$, $p<.05$), and an interaction ($F=5.36$, $df=2/84$, $p<.01$). The simple interaction comparing Repeats between ordinary and exact decoders was not reliable, indicating that the content-function difference in the encoder utterances contributed the significant variance to the main interaction. All other overall comparisons of experimental effects in the data for Repeats, False Starts, and Filled Pauses were nonsignificant. Thus there is little statistical evidence that exact decoders reproduce hesitations better than ordinary decoders. Nevertheless, the exact decoders' relatively higher proportions of Filled Pauses before content words suggest that they can after all reproduce some if not all encoder hesitations. The data were analyzed for hesitations correctly placed.

Table 3. Total Selected Encoder Hesitations Preceding Content (C), Function (F), and Total (T) Words Imitated (I) and Produced (P) by 43 Pairs of Decoders in Experiment 1

	Encoders		Ordinary Decoders			Exact Decoders		
			I	P	I+P	I	P	I+P
Filled Pauses	C	76	4	57	61	20	67	87
	F	54	4	70	74	7	72	79
	T	130	8	127	135	27	139	166
Repeats	C	65	1	20	21	7	28	35
	F	34	5	23	28	4	32	36
	T	99	6	43	49	11	60	71
False Starts	C	22	2	32	34	3	18	22
	F	43	3	45	48	4	49	53
	T	65	5	77	82	7	68	75
Combined	C	163	7	109	116	30	114	144
	F	131	12	138	150	15	153	168
	T	294	19	247	266	45	267	312
Imitations X 100			7.14%			14.42%		
Total								

Decoder Filled Pauses, Repeats, and False Starts scored as imitations or productions appear in Table 3. Decoder hesitations were counted as imitations if at least one scorer heard a hesitation in the decoder protocol and one scorer a hesitation in the encoder's protocol. A decoder pause was counted as a production only if both scorers heard it. The encoder hesitations included in the table, however, were only those on which scorers agreed, to provide the most liberal basis for comparison; because of the differences in scoring, Tables 2 and 3 are not equivalent. There were few imitations, as the table shows. Decoder groups were compared statistically by combining all hesitations (as in the bottom line of Table 3) for any pair of decoders each of whom produced at least one, but because of the small Ns there were no reliable differences. Perhaps the most reasonable conclusion from the results shown in Table 3 is that decoders will imitate a few hesitations, particularly if instructed to do so.

The results of Experiment 1 can be summarized as follows. (1) Decoder instructions to imitate hesitations increased hesitation rate without affecting the shift. (2) Few decoder hesitations were imitations.

EXPERIMENT 2

While the results for Experiment 1 are not inconsistent with the hypothesis that decoders usually filter out hesitations, it is possible that most exact decoders did not respond to their instructions to reproduce hesitations but resisted the extra task or failed to grasp its nature. Experiment 2 provided exact decoders with a practice session in placing hesitations. In addition, only one encoder was used.

Method

Encoder tape. The recorded tape for one female encoder from Experiment 1 was selected for use in this experiment; it had an unusually high content-pause rate. All four types of hesitation, and number of words, exceeded the average in Experiment 1.

Procedure. The ordinary decoder was treated exactly as in Experiment 1. The exact decoder was also, except that he was given two practice utterances, each of which was repeated in trials to a criterion of perfect reproduction by five trials or an alternate later criterion of all hesitations correct, insofar as E could judge. The point of the practice trials was to require S to demonstrate his awareness of what counted as hesitations or irregularities in the signal. The practice utterances together included all four types of hesitations. The S was told what he had missed after each reproduction attempt. The practice utterances, particularly one including 14 words and six hesitations, proved to be so difficult that none of the first six

exact decoders achieved the first criterion, and one required as many as 12 trials to get all hesitations correct. Since their attempts were correspondingly difficult for E to judge, and since a difficult practice session contributed nothing to the purpose of the experiment, the practice utterances were changed after six Ss in each group had been run to three examples with fewer hesitations and words. After the change the practice trials as well were recorded for later analysis. It seemed, however, during preliminary analysis that the earlier two groups of six decoders each might differ in potentially significant ways from the later groups, so two additional groups of six decoders each were run with the original exact decoder two-utterance practice session following 18 pairs using the three-utterance practice sessions. There were then 12 exact decoders with difficult, 18 with easier practice sessions, alternating with the 30 ordinary decoders.

Subjects. The 60 Ss were summer session psychology students, some in the introductory course, most in advanced courses. They were assigned alternately to the two conditions, ordinary and exact decoder, in order of appearance at the laboratory.

Results

The data were analyzed first with the two groups differing in practice materials as a dimension. The only difference between the two groups was overall hesitation rate, so they were combined.

Words. The encoder in Experiment 2 gave over 130 words more than the average in Experiment 1 (Table 1). The decoders, while reproducing more words absolutely on the average than Experiment 1 decoders, reproduced relatively less than the encoders imitated, presumably because of greater memory demands. Experiment 1 decoders produced about 89% as many words as their encoders, Experiment 2 decoders produced about 79%. The split between content and function words remained close to 42 vs 58% for all coder groups. (See Table 4.)

Hesitations. Total Hesitations and Unfilled Pauses only are included in Table 5; relevant results for the remaining types of hesitations are presented in Table 6. The encoder difference (Total Hesitations) between percentages of content and function words following hesitations was +14.92%. These differences became +4.74 and +5.77% in the ordinary and exact decoder protocols, a shift of over 9%. Analysis of

Table 5.
Percentages of Words Preceded by Hesitations in Experiment 2

		Encoder	Ordinary Decoders	Exact Decoders	Both Decoders
Total Hesitations	Content	29.46	18.00	24.07	21.09
	Function	14.54	13.26	18.30	15.80
	Both	20.87	15.32	20.83	18.11
Unfilled Pauses	Content	22.97	14.61	18.87	16.78
	Function	10.76	10.33	14.05	12.20
	Both	15.94	12.18	16.17	14.20

variance showed exact decoders hesitated more frequently per word ($F=25.73$, $df=1/58$, $p<.01$), and the content-word rate was higher ($F=36.63$, $df=1/58$, $p<.01$), but there was no differential shift, as shown by the nonsignificant interaction ($F<1$).

For Unfilled Pauses the same factors were reliable ($F=19.81$, $df=1/58$, $p<.01$ for coder types; $F=35.63$, $df=1/58$, $p<.01$ for word types). These results are quite like Experiment 1 in the lack of differential hesitation shift between decoder types and, in addition, seem to mean that the more disproportionate are the encoder hesitations within grammatical constituents, the greater will be the decoder shift to constituent boundaries.

Table 6 gives the decoder imitations in Experiment 2. Again, there were few matches between encoder and decoder Filled Pauses, Repeats, and False Starts. The even poorer showing of exact decoders in Experiment 2 than in Experiment 1 is probably explained by the fact that some of the Experiment 1 encoder utterances were much shorter than in Experiment 2.

The decoder protocols were analyzed further to provide a comparison of words and hesitations correct. Only content words and corresponding hesitations were counted, for reasons discussed later. The liberal scoring measure described earlier was used for hesitations. Table 7 shows the comparison. While the two decoder groups reproduced about the same mean number of correct content words, a higher proportion of the ordinary decoders' words were correct ($F=8.02$, $df=1/58$, $p<.01$), after arcsine transformation of correct content over total content words. At the same time exact decoders imitated more hesitations correctly by the measure represented by absolute mean correct ($F=8.53$, $df=1/58$, $p<.01$), but there was no difference in percentage hesitations correct ($F<1$), again after arcsine transformation of content hesitations correct over total content hesitations. Table 7 thus seems to indicate that encouraging decoders to reproduce utterances exactly induces them to hesitate more without increasing their accuracy of imitations, and while increasing their word rate the instruction does so at the expense of per cent words correct. These results are consistent with the notion that attending

Table 4. Mean (M) and Per cent (P) Words in Experiment 2

	Encoder		Ordinary Decoders		Exact Decoders	
	M	P	M	P	M	P
Content	185.0	42.4	148.4	43.3	154.3	43.9
Function	251.0	57.6	194.1	56.7	197.1	56.1
Total	436.0	100.0	342.5	100.0	351.4	100.0

Table 6. Total Selected Encoder Hesitations Preceding Content (C), Function (F), and Total (T) Words Imitated (I) and Produced (P) by 30 Pairs of Decoders in Experiment 2

		Encoder X 30	Ordinary Decoders			Exact Decoders		
			I	P	I+P	I	P	I+P
Filled Pauses	C	120	6	61	67	4	116	120
	F	120	6	74	80	6	142	148
	T	240	12	135	147	10	258	268
Repeats	C	210	1	26	27	3	42	45
	F	60	2	34	36	4	36	40
	T	270	3	60	63	7	78	85
False Starts	C	30	0	15	15	0	13	13
	F	30	0	22	22	0	34	34
	T	60	0	37	37	0	47	47
Combined	C	360	7	102	109	7	171	178
	F	210	8	130	138	10	212	222
	T	570	15	232	247	17	383	400
Imitations Total X 100			6.07%			4.25%		

to hesitations and words simultaneously is a double task with motivating effects but little if any increase in performance accuracy.

The results for Experiment 2, particularly in comparison with previous results, can be summarized as follows: (1) Decoders apparently imitate few hesitations despite understanding the requirements of the task. (2) Exact instructions stimulate decoders to produce more correct hesitations, but the percentage is not proportionate to the increase in the total. (3) The greater the disproportion in favor of encoder content-word hesitations the greater is the decoder displacement toward function words, within the range of conditions studied in Experiments 1 and 2.

EXPERIMENT 3

Experiment 3 had two purposes. One was to provide comparison data for Experiment 4 (to follow), the other was to vary time after listening in order to observe the effects of memory on hesitation shift.

Method

Composite encoder tape. All Ss listened to a tape with the following characteristics. There were 60 utterances drawn from the protocols of 27 encoders in several previous experiments. No S contributed more than four utterances. There were 26 utterances spoken by males, 34 by females. Utterances describing all 20 TAT cards were represented, no card being represented more than four times. Utterances ranged from 3-26 words with a mean of about 15. A preponderance of hesitations preceded content words in each utterance to provide maximum latitude for decoder hesitation shifts;

per cent hesitations (content minus function) in each utterance ranged from +10 to +50%, with a difference of about 22% over all 60 utterances. The tape was so constructed that obvious characteristics of the utterances, e.g., sex of talker, length, etc., were approximately evenly distributed. There were 5 sec between utterances to permit various task manipulations. A copy of the tape was made which repeated each utterance; repetitions were separated by a 2-sec interval.

Subjects and Procedure. The 21 Ss were assigned to one of the three counterbalancing conditions in sequential order of appearance at the laboratory. Each S heard the 60-utterance composite encoder tape. His instructions were the same as for ordinary decoders in Experiments 1 and 2, except that E introduced a delay interval between the end of the encoder utterance and the signal for the decoder reproduction. The delay intervals were 0, 4, or 8 sec and were signalled by E saying "now." The delay interval varied successively within Ss in the order 0, 4, and 8 for utterances N, N+1, and N+2, and one-third of the Ss started on each interval for counterbalancing, the first S on 0, the second on 4, etc. Two practice utterances also had differing delay intervals.

Results

Table 8 indicates a slightly higher than usual percentage of content words by speakers on the composite encoder tape, imitated rather closely by the decoders after each delay interval. Table 9 gives hesitation data. Only Total Hesitations are shown since other measures gave essentially the same results. Apparently the delay interval had no effect on hesitation shift. If memory is a factor in hesitation shifts it probably exerts its effect during the course of listening to the utterance itself. There were no differences among coder types and no interaction, but the difference between word types ($F=42.69$, $df=1/20$, $p<.01$) was reliable. These results, combined with Experiment 2 and other hesitation data, seem to indicate a shift proportionate to the degree of ungrammaticality of the encoder distribution of hesitations.

EXPERIMENT 4

The previous experiments are consistent with the

Table 7. Mean (M) and Per cent (P) Correct Content Words and Correct Hesitations Preceding Content Words in Experiment 2

		Ordinary Decoders	Exact Decoders
Words	M	125.1	125.1
	P	84.3	81.4
Hesitations	M	12.5	16.2
	P	41.0	39.7

Table 8. Mean (M) and Per cent (P) Words in Experiment 3

	Composite		Decoders							
	Encoder		0-sec Delay		4-sec Delay		8-sec Delay		Combined	
	Tot/ 3	P	M	P	M	P	M	P	M	P
Content	133.3	45.0	119.9	45.4	123.3	45.1	120.8	45.2	121.3	45.2
Function	163.0	55.0	144.4	54.6	150.2	54.9	146.2	54.8	146.9	54.8
Total	296.3	100.0	264.3	100.0	273.5	100.0	267.0	100.0	268.2	100.0

notion that listeners fail to notice most hesitations during decoding. But it may be that decoders hear hesitations perfectly well when asked to attend to them but find them difficult to include in their reproductions; that is, the hesitation shift could be simply an encoding bias in the decoder's reproduction task. Thus one purpose of Experiment 4 was to rule out this possibility by substituting a new decoder task. Another was to attempt to establish the difference between the task of listening for hesitations and that of listening for hesitations while decoding speech as well.

Four groups each listened to the composite encoder tape (Experiment 3) and marked hesitations on transcripts. One group listened to each utterance twice, marking it as they listened (the Mark-2 condition). A second group heard once and simultaneously marked each utterance (Mark-1). Another group heard the utterance once, marked it simultaneously, and afterwards wrote the utterance (Mark-Write). The last group listened to the utterance without a transcript and then marked the transcript from memory (Recall-Mark). These tasks seem to depart from ordinary hesitation scoring in a rough order of increasing difficulty. Mark-2 approximates most closely the research assistant's scoring task, Mark-1 somewhat less. In the Mark-Write condition the S, though able to follow along on the transcript, must attend to the utterance "as a whole," decoding speech as he goes along, if he takes his writing task seriously. The S in the Recall-Mark condition must both decode the utterance and keep track of hesitation locations. The Recall-Mark S is of particular interest since, though he need not produce speech, he must decode it in order to correctly place hesitations. Thus the first two groups need only detect hesitations, the latter two must decode and store the words as well.

Method

Subjects and Procedure. The 104 Ss serving as markers included 80 regular session and 24 summer session students, the latter divided nearly evenly (± 2) between the four experimental conditions. They were run in groups varying in size from one to 12. The Ss assigned themselves into conditions by their choice of participation time. Many had served as encoders or decoders in previous experiments; some had been in Experiments 1 or 2. In all cases prior participation was 4-12 weeks earlier.

The Ss were provided with headphones and booklets. Booklets were either 60 pages, one utterance per page, or 120 pages with each utterance preceded or followed by a blank page. (Some Ss in the Mark-2 and Mark-1 conditions had unnecessary blank pages in their booklets; they were told to ignore them.)

Mark-2 groups were told in effect (1) to listen to the utterance, marking hesitations as they listened, (2) to listen again making corrections or additions, and (3) to turn the page to the next transcribed utterance on signal. The E read the following instructions, using the blackboard in the front of the room for demonstrating the examples:

Here is what will happen in this experiment.

You will listen to speech through your earphones and read a transcript of the speech at the same time. Your task is to mark on the transcript where the speaker paused or hesitated. When I say 'ready' and then a number turn the page to the group of words with that number. While you listen, mark with a slash like this (demonstrate) any place you think the speaker paused or hesitated as he was saying the words. Then you will hear the group of words again and you can make any changes, additions, or corrections in the marks you made. When I say 'ready' turn to the next page and repeat the procedure. So if you heard 'The boy/went to/school/in the/morning,' you would mark it like this. (Demonstrate.) Be sure to mark as you hear the words. Are there any questions before we have the first example? Get ready to turn the page, listen, and mark. (The first example was marked on the blackboard by E as Ss marked in their booklets.) This is one possible way of marking the pauses. (The example was repeated and followed by two repetitions of the second example with no further comment.)

The E monitored the tape and signalled Ss to turn pages at about 3 sec after the second repetition of each utterance.

The Mark-1 condition was the same except that the tape presented each utterance once and instructions were altered accordingly.

The Mark-Write Ss were told (1) to listen to the utterances, marking hesitations as they listened, and (2) to turn the page and write the utterance on the following blank page upon signal from E. The E signalled the writing task about 3 sec after the

Table 9.

Percentages of Words Preceded by Hesitations in Experiment 3

		Decoders				
		Composite Encoder	0-sec Delay	4-sec Delay	8-sec Delay	Combined
Total	Content	34.43	13.77	14.27	14.73	14.26
Hesitations	Function	12.22	9.35	9.08	10.10	9.51
	Total	22.12	11.35	11.42	12.19	11.66

utterance. When nearly all Ss had finished writing E signalled the page turn to the next utterance which followed 2 sec later. Their instructions (omitting the section regarding examples) were as follows:

Here is what will happen in this experiment. You will listen to speech through your earphones and read a transcript of the speech at the same time. Your task is to mark on the transcript where the speaker paused or hesitated, and then write down what you heard. When I say 'ready' and then a number turn the page to the group of words with that number. While you listen, mark with a slash like this (demonstrate) any place you think the speaker paused or hesitated as he was saying the words. When I say 'turn,' turn the page and write what you heard as accurately as possible. When I say 'ready' turn to the next page and repeat the procedure...

The Recall-Mark group was required (1) to listen to an utterance once, with a blank page up in his booklet, and (2) to turn the page upon signal to the transcript of the utterance and mark hesitations as recalled. Their instructions were:

Here is what will happen in this experiment.

You will listen to speech through your earphones; then you will read a transcript of the speech. Your task is to mark on the transcript where the speaker paused or hesitated. When I say 'ready' and then a number turn to the blank page with that number on it and listen to the speech. When I say 'turn,' turn the page to the group of words you just heard and mark with a slash like this (demonstrate) any place you think the speaker paused or hesitated as he was saying the words. When I say 'ready,' turn to the next page and repeat the procedure...

The marking interval was signalled at the end of the utterance. The repeated-utterance tape was used so that with the second repetition monitored through headphones by E, but switched off of S's headphones, S had time about equal to the length of the utterance to record the hesitations he recalled. The marking interval was thus about equivalent to that for the Mark-1 and Mark-Write groups. The waiting time before marking was about like that before decoder reproduction in Experiments 1 and 2.

Results

Hesitation shift, like the content-function word split, was considered to be a relative measure and inherently proportional. In this experiment, however, a fixed number of words was presented and responded to, so content words marked for hesitations divided by total words marked for hesitations was used as the first basis for evaluating hesitation shift. For convenience, words marked for hesitations will be referred to as hesitations, that is, content hesitations and function hesitations.

The bottom line of Table 10 presents content hesitations over total hesitations for the composite encoder productions, ordinary decoder reproductions (Experiment 3) and hesitation markers. The remaining part of the table gives means for content and function words. The hesitation measure for these comparisons was the sum of Filled plus Unfilled Pauses only, for all conditions represented in the table, since these were the hesitations considered in the marking experiment. While only the four hesitation-marker groups seem to be strictly comparable statistically, the differences on the bottom line of Table 10 are about what might be expected on the notion that practiced (research assistant) listeners define one end and ordinary decoders the other of a task continuum on which can be represented the relative ease and accuracy of hesitation identification. The experimental hypothesis implies increasing shift directly with departure from the hesitation scoring task, particularly the Mark-Write and Recall-Mark conditions, which require speech decoding while marking pauses.

Analysis of variance of the differences among marker conditions represented on the bottom line of Table 9 was significant following arcsine transformation ($F=15.04$, $df=3/100$, $p<.0005$). All non-adjacent groups were reliably different ($ps<.01$) by Newman-Keuls tests. The Recall-Mark group was significantly lower than any other ($p<.01$). The results held over an analysis in which content and function hesitations were weighted in terms of opportunities, to control for the fact that the proportions of content and function words in the com-

Table 10. A Comparison of Hesitations (Filled plus Unfilled Pauses) Preceding Content and Function Words which were Marked (Experiment 4) or Produced (Experiment 3)

		Composite Encoder	Hesitation Markers (Exp. 4)				Ordinary Decoders (Exp. 3)
			Mark-2	Mark-1	Mark- Write	Recall- Mark	
M	Content	132.0	107.1	96.3	100.7	79.6	92.9
	Function	62.0	60.7	60.7	69.9	64.0	77.0
	Total	194.0	167.8	157.0	170.6	143.6	169.9
%	$\frac{\text{Content}}{\text{Total}} \times 100$	68.0	63.8	61.3	59.0	55.4	54.7

posite encoder speech were not equal. To the extent that the differences in proportions of content-word hesitations represent a hesitation shift, the experimental hypothesis appears to be supported.

It was considered, however, that a comparison of correct vs incorrect hesitations would provide a better measure, since the model implies that the real meaning of a hesitation shift is the imposition of new structure on an acoustic signal, i.e., the decoding of grammatical constituents as "units" despite disunitizing elements, viz., within-constituent hesitations in the signal. A hesitation shift would then be the consequence of not only forgotten content-word hesitations but also of increased hesitations marked at function words where there had been really no hesitation, i.e., incorrect hesitations marked at function words.

The marker conditions were compared according to the measure correct/(correct + incorrect) for both content and function words. Since there were disproportionate opportunities for guessing because of unequal content vs function proportions and unequal correct vs incorrect proportions within word types, all hesitations were weighted by dividing each term by opportunities, e.g., correct (C) hesitations marked at content words over number of encoder hesitations preceding content words, number of incorrect (I) hesitations marked at content words over number of words not preceded by encoder hesitations, etc. Then

$$\frac{C_c/132}{C_c/132 + I_c/268} = \frac{C_c}{C_c + (.4925)I_c}$$

$$\frac{C_f/62}{C_f/62 + I_f/427} = \frac{C_f}{C_f + (.1452)I_f}$$

where subscripts refer to content and function words.

The basis for correct was the judgments of the two research assistants acting as scorers for hesitations on the composite encoder tape. The most liberal scoring measure was used. A hesitation was counted correct if either scorer heard it, incorrect if neither did. Table 11 gives mean correct and incorrect scores; the means for correct plus incorrect are in Table 10.

Table 11. Mean Hesitations Correctly (C) and Incorrectly (I) Marked in Experiment 4

	Mark-2		Mark-1		Mark-Write		Recall-Mark	
	C	I	C	I	C	I	C	I
Content	102.8	4.3	91.7	4.6	93.9	6.8	70.7	8.9
Function	45.2	15.5	40.3	20.4	42.1	27.8	30.2	33.8
Total	148.0	19.8	132.0	25.0	136.0	34.6	100.9	42.7

Table 12 gives weighted proportions. Using correct vs incorrect scoring the groups again differed from the scorers in appropriate order, lending support to the validity of the comparison. Marker types, word types, and the interaction were all significant in analysis of variance following arc-sine transformation ($F=21.45$, $df=3/100$; $F=501.71$, $df=1/100$; $F=8.37$, $df=3/100$; all $ps < .0005$). Marker types was a between-Ss factor, the remaining two were within-Ss factors.

Newman-Keuls tests between means over both word types and the between-Ss error term showed the Recall-Mark condition lower than any others ($ps < .01$). Tests of content-word differences by Newman-Keuls tests and the within-Ss error term showed all conditions except Mark-2 and Mark-1 different (all $ps < .01$ except between Mark-1 and Mark-Write, $p < .05$). All differences between function words were significant ($ps < .01$). In addition, tests of simple interactions between adjacent groups showed the Recall-Mark condition different ($F=5.81$, $df=1/100$, $p < .025$).

These results provide some support for the expectation that hesitation shift varies directly with departure from the hesitation scorer's task and toward the speech decoding task. The clearest evidence seems to be located in the Recall-Mark condition, in which the simultaneous task apparently cannot be avoided. The evidence suggests also the decoder's perceptual contribution to the hesitation shift, since the effect was partly due to a disproportionate increase in incorrectly marked function-word hesitations.

Supporting evidence was found in the analysis of absolute scores represented in Table 11, although this evidence must be considered with caution, since the differences among within-cell variances were as large as 45-fold. Analyzing incorrect scores only, the increasing difference between content and function errors was reliable according to the test for interaction ($F=8.08$, $df=3/100$, $p < .0005$). No single simple interaction within adjacent conditions was reliable by conventional risk levels, but the increasing differences between means for function word errors were all reliable by Newman-Keuls test using the within-Ss error term ($p < .01$ except between Mark-2 and Mark-1, for which $p < .05$).

Summarizing Experiment 4: (1) The hesitation shift seems to persist when a response measure other than speech production is used. (2) The shift

Table 12. Proportions Correct (Correct + Incorrect) Hesitations, Weighted for Guessing, Marked in Experiment 4

	Mark-2	Mark-1	Mark-Write	Recall-Mark
Content	.9802	.9758	.9652	.9416
Function	.9522	.9316	.9125	.8600
Total	.9715	.9619	.9483	.9155

increased directly with logical dissimilarity to a simple hesitation detection task. (3) Error scores in placing hesitations seemed to indicate that memory and response characteristics together would not completely account for the organization implied in the hesitation shift.

Further results. In order to illustrate some possible ways of characterizing the distribution of words in decoder reproductions, the content words correct from Experiment 2 were classified in several ways. Each of the 185 content words in the single encoder's protocols were given a score representing the number of decoders (out of 60) who reproduced the word. Table 13 shows some of these classifications. The top of the table indicates Thorndike-Lorge (1944) frequency count. The middle refers to the number of encoders in Experiment 1 (out of 43) who used a given word in describing a given TAT card. Thus the bottom 41 words in the classification are the idiosyncratic words used by the Experiment 2 encoder (who was one of those in Experiment 1). The classification by parts of speech required some grammatical judgment and should be regarded as tentative. A number of words classified as content according to Miller, Newman, and Friedman (1958) seemed to play the role of function words in the present experimental context, not only "look" and "like," but "see," "maybe," "probably," etc. These words were classified separately. The table shows a positive relation between the index of correct reproductions and both types of frequency count. Verbs among the content words were less accurately reproduced.

DISCUSSION

In these experiments Ss were asked to attend to acoustic characteristics of speech while decoding it. The result of this presumed double task was

Table 13. Mean Correct Reproductions by Experiment 2 Decoders of 185 Encoder Content Words Classified in Several Ways

Classification		N	Mean
Thorndike-Lorge Frequency Count	AA	143	42.14
	A	18	41.44
	1-50	24	30.67
	Total	185	
Experiment 1 Frequency Count	30-43	9	51.67
	20-29	23	48.17
	10-19	41	45.76
	2-9	71	37.18
	1	41	34.61
Total		185	
Parts of Speech	Nouns	59	43.78
	Adjectives and Adverbs	27	43.15
	Verbs	39	40.59
	Pseudo-Content		
	"Look(s)," "Like"	30	43.70
	Remainder	30	19.87
Total		185	

evaluated by assessing the outcome of the hesitation task rather than the adequacy of speech decoding. The reason for this choice was that there was no simple and convincing way to evaluate decoding adequacy on the basis of reproductions. Rote recall is not a satisfactory measure since it is an easy matter to substitute words during reproduction with little change in meaning; in fact some substitutions improve upon the original.

Martin and Strange (in press) have presented empirical justification for identifying constituent boundaries on the basis of the function-content word dichotomy in printed text (Miller, Newman, & Friedman, 1958) rather than by direct inspection of speech protocols. They analyzed Miller, Newman, and Friedman's printed text and found that 82% of initial words in sentences but only 16% of final words were function words. These values would be about 58% if the distribution of function words was even throughout sentences. The implications of these results should extend to constituents of any size.

Concerning the reliability of pause judgments, it has earlier been pointed out that scorer agreement is fairly good. In addition, it should be noted that in Experiment 4 the unpracticed "one-shot" Ss in the Mark-2 group placed 88% of their pauses in agreement with the scorers.

One question for future research with spontaneous speech is that of the nature of the acoustic environment in which listeners hear hesitations. Unfilled Pauses of as much as several seconds are obvious, and spectrograph records will show the corresponding silent intervals. Others, while obvious to anyone listening for a "hesitation," yet are not seemingly of extensive duration. Early and tentative spectrographic analyses suggest that Unfilled Pauses are often not empty intervals, but rather changes of pace. These were seemingly heard (and seen) as extended syllables, followed by a "burst" of words. Pause judgments seem to depend not only upon real silent intervals but upon elongated syllables signifying changes of pace as well.

The Basis for the Hesitation Shift

It seems fair to conclude that the tendency to pause between grammatical constituents is a stable speaking (encoding) characteristic despite the tendency to pause within them under certain conditions. Decoders distributed pauses in accordance with this tendency against instructions, even when hesitations in imitated speech were overwhelmingly within constituents. When primed with content words prior to their TAT card descriptions, encoders showed the same tendency (Martin & Strange, in press). Much other work is consistent with the notion that within-constituent hesitations are a special case. Goldman-Eisler (1961a, b) found hesitation tendencies to diminish when the speaking task was

undemanding, e.g., repetition of the same utterance. Boomer (1965) reported a correspondence between pauses and "phonemic clause" boundaries; his Ss produced speech in a relaxed interview. When pauses preceding high-information words were noted, however (Goldman-Eisler, 1958a, b), typically a spontaneous construction involving some thought, e.g., a cartoon interpretation, was required. Maclay and Osgood (1959) reached a similar conclusion—though in passing it may be noted that their work has been cited both as indicating hesitations within—(e.g., Osgood, 1964) and between—(e.g., Garrett & Fodor 1967) constituents. Whatever the speaking bias toward between-constituent pausing, however, it apparently cannot account for the hesitation shift, since the shift remains in the hesitation-marking task.

Memory is also an unlikely candidate for explaining the hesitation shift, for several reasons. First, there are logical difficulties in predicting content hesitations to be forgotten more quickly than function. Second, Experiment 4 shows the shift depends upon function-word hesitations supplied by the listener. Third, Experiment 3 shows that post-listening interval is not important. Finally, an additional data analysis of Experiment 1 showed that decoders who reproduced speech from high-output encoders shifted hesitations to the same degree as decoders who had fewer words to remember and imitate.

Finally, failure of instructions to attend to hesitations does not seem to explain the shift, since it probably cannot apply to the Recall-Mark group in Experiment 4. These Ss were asked only to mark hesitations, but they could do so only by remembering the speech as well.

Conclusions and Speculations

The results of these experiments suggest that attending to the acoustic and message aspects of a speech signal are incompatible operations. Perhaps, as they alternate between pauses and words, Ss attend to different temporal segments of the signal. A more preferable view, however, might be that the alternation was between different processing levels, to cues to pauses in the acoustic signal sometimes, but more usually, to cues to the message at some other level.

The results of these experiments point also to the organization of speech between input and output. Grammatical structure in such ongoing speech surely must usually be inferred from words, since hesitations are unreliable boundary markers. It would seem an inefficient speech decoding mechanism that registers such acoustic effects at all, let alone the details of their time of arrival.

While it is fairly easy to believe that pauses and other acoustic irregularities are not "heard" in ordinary listening, what about other aspects of the

sound signal? Redundant (e.g., function) words gated out of context are often unintelligible behind noise (Lieberman, 1963); presumably speakers do not articulate them clearly. Perhaps, because they carry little information, function words, along with pauses, also are not ordinarily heard. Some data reported elsewhere (Martin, 1968) suggest that spontaneous speech is often an incompletely encoded version of structurally deeper levels of sentences. It may be that real speech, ungrammatical as it is, is understood because the speaker and hearer are communicating at these other levels.

References

- Boomer, D. S. Hesitation and grammatical encoding. *Language & Speech*, 1965, 8, 148-158.
- Fries, C. C. *The structure of English*. New York: Harcourt Brace, 1952.
- Gamer, W. To perceive is to know. *Amer. Psychologist*, 1966, 21, 11-19.
- Garrett, M., & Fodor, J. Psychological theories and linguistic constructs. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Prentice-Hall: Englewood Cliffs, 1967.
- Gibson, J. J. *The senses considered as perceptual systems*. New York: Houghton Mifflin, 1966.
- Goldman-Eisler, Frieda. Speech production and the predictability of words in context. *Quart. J. exp. Psychol.*, 1958a, 10, 96-106.
- Goldman-Eisler, Frieda. The predictability of words in context and the length of pauses in speech. *Language & Speech*, 1958b, 1, Part 3, 226-231.
- Goldman-Eisler, Frieda. Hesitation and information in speech. In C. Cherry (Ed.), *Information theory*. London: Butterworths, 1961a.
- Goldman-Eisler, Frieda. The significance of changes in the rate of articulation. *Language & Speech*, 1961b, 4, 171-174.
- Lane, H. Motor theory of speech perception: A critical review. *Psychol. Rev.*, 1965, 72, 275-309.
- Liberman, A. Some results of research on speech perception. *J. Acoust. Soc. Amer.*, 1957, 117-123. Reprinted in S. Saporta (Ed.), *Psycholinguistics: A book of readings*. New York: Wiley, 1961.
- Lieberman, P. Some effects of semantic and grammatical context on the production and perception of speech. *Language & Speech*, 1963, 6, 172-187.
- Maclay, H., & Osgood, C. E. Hesitation phenomena in spontaneous English speech. *Word*, 1959, 1, 19-44.
- Martin, J. G. Hesitations in the speaker's production and listener's reproduction of utterances. *J. verbal Learn. verbal Behav.*, in press.
- Martin, J. G. Some acoustic and grammatical features of spontaneous speech. Paper presented at the Conference on the Perception of Language, Pittsburgh, 1968.
- Martin, J. G., & Strange, Winifred. Determinants of hesitations in spontaneous speech. *J. exp. Psychol.*, in press.
- Miller, G. A., Newman, E. B., & Friedman, A. E. Length-frequency statistics for written English. *Information and Control*, 1958, 1, 370-398.
- Osgood, C. E. Psycholinguistics. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. *Investigations of man as socius: Their place in psychology and the social sciences*. New York: McGraw-Hill, 1963. Pp. 244-316.
- Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

Note

1. Supported in part by Grant MH 10400, National Institutes of Mental Health, U. S. Public Health Service. Judith Scoles and Nancy Hurlbut ran Ss and helped Ruth Lanford, Sylvia Myers, and Judy Orendorff with the data analysis.

(Accepted for publication February , 1968.)