[redacted]

JOURNAL OF SPEECH AND HEARING DISORDERS

MONOGRAPH SUPPLEMENT 7
JUNE 1961

A Journal of the
AMERICAN SPEECH AND HEARING ASSOCIATION

Studies of Speech Disfluency and Rate of Stutterers and Nonstutterers

Wendell Johnson

and

James N. Neelley
Eric K. Sander

Harris Winitz
Martin A. Young

Special Editor
Wendell Johnson

THE JOURNAL OF SPEECH AND HEARING DISORDERS
MONOGRAPH SUPPLEMENT 7 JUNE, 1961

The Journal of Speech and Hearing Disorders Monograph Supplement

7

Studies of Speech Disfluency and Rate of Stutterers and Nonstutterers

Andrea

Editor's Note

The planning and preparation of this Monograph substantially antedate the term of service of the present Editor of Monographs. Wendell Johnson conceived and supervised the research reported herein, and exercised editorial responsibility for its publication in his capacity as Editor of the Association. The attention of the reader is to be called to Dr. Johnson's designation as Special Editor of Monograph Supplement 7.

Frederic L. Darley
Editor of Monographs
American Speech and Hearing Association

Introduction

WENDELL JOHNSON

The research program represented by the five studies reported in this Monograph is concerned with two of the fundamental dimensions of speech behavior, those of rate and fluency, or disfluency, and with the problem generally referred to as stuttering. The program is motivated basically by the view that normative and experimental data on speaking rate and disfluency are important in their own right, and for the light they throw upon the biological uniqueness of man by virtue of the distinctively human character of the process of symbolization which they represent. So far as the process of symbolization, mediated by the nervous system, interacts with other bodily functions, its relation to health and disease and to behavioral effectiveness and inadequacy clearly warrants intensive investigation. Essential to such investigation are methods of measurement and the standards of reference that are provided by relevant normative data. Methods of sampling speech behavior and of quantifying its disfluency and rate characteristics have been developed in certain ways in earlier studies in this program of research. These studies, reported in *Stuttering in Children and Adults* (University of Minnesota Press, 1955, pp. 155-196) and *The Onset of Stuttering* (University of Minnesota Press, 1959, pp. 196-220), have also yielded representative disfluency and rate data for male and female speakers, classified as stutterers and nonstutterers, respectively, primarily within the age range of about 2 to 8 years. The first study presented in this Monograph was designed to provide data of essentially the same kinds for male and female stutterers and nonstutterers at the college, or young adult, age level. The findings yielded by this study have also been sought with a view to facilitating consideration of the question of the relation of disfluency, per se, to the problem called stuttering.

The second and third studies, by Sander and Young, respectively, were intended to provide evaluation of the reliability, and the relationship to other variables, of the data derived from disfluency analysis and measurements of rate. Using essentially the method described in the first report, Sander compared two samplings of the speech and oral reading of a group of adult stutterers; the samplings were separated in each case by an interval of 24 hours. He also performed a correlational analysis of his data and on the basis of the results suggested a modification of the procedure for measuring disfluency.

Young related measures of disfluency to ratings of 'severity of stuttering.' He also refined and simplified the method of disfluency analysis employed in the first study referred to above. Young's multiple correlations proved to be effectively sensitive in distinguishing the specific kinds of disfluency that listeners do and the kinds they do not associate semantically with 'severity of stuttering.'

His findings suggest, therefore, a kind of operational and an extensional definition of 'stuttering' or, more precisely, 'severity of stuttering' that represents a clarification of the distinction between stuttering and disfluency, per se, that is of considerable theoretical and clinical significance. It is also to be noted that the structural correspondence between Young's report and his investigative operations, as shown in Figure 1 on page 33, represents in certain respects an innovation in scientific writing that reflects provocatively a consideration of the order of events in the process of abstracting.

The fourth report, by Winitz, and the fifth, by Neelley, include additional normative data and also serve to focus attention particularly on the problem of clarifying the nature and extent of the relationship between disfluency and the problem called stuttering. Winitz, in an analysis of repetitions in samples of the vocal productions of children during the first two years of life, found that the repetitive form of disfluency is a relatively prominent feature of the vocalization and early speech of presumably representative 'nonclinical' children. Neelley compared the speech of stutterers and nonstuttering normal speakers under the conditions of normal and delayed auditory feedback. He found that the two groups of subjects reacted in substantially similar ways to delayed auditory feedback, and that listeners readily discriminated the speech of both stutterers and nonstutterers as affected by delayed auditory feedback from that of the stutterers under normal feedback conditions. The stutterers reported, moreover, that the experience of speaking with delayed auditory feedback was qualitatively different from that of stuttering. The report of this study is concluded with these words: "The hypothesis that stuttering may be somehow related to a delay in auditory feedback, on the ground that speech produced under conditions of delayed auditory feedback appears to behave like stuttering, to sound like stuttering, and to be an experience like stuttering, is not supported by the findings of this experiment."

The findings presented in this Monograph serve to make clear that the terms 'stuttering' and 'disfluent speech' are not extensionally equivalent. This means that an adequate theory of stuttering must be designed to do more than account for certain kinds and amounts of disfluency. Indeed, the findings here reported indicate that consideration of some of the forms and contexts of disfluency is irrelevant to a theory of stuttering. Basically, the comprehensive problem called stuttering is appropriately to be distinguished from the speech phenomena classifiable as disfluency, per se.

The detailed description of speech disfluency in its several forms, the identification of the variables functionally related to it, and the determination of the relevant patterns of interaction are major aspects of a program of relatively basic research. Another fundamental area of investigation is that of the measurement of the rate of utterance in oral reading and in speaking in their various modes and circumstances, together with the identification of the variables which interact with rate and the determination of the patterns of interaction involved. The scientific investigation of the problem called stuttering is another matter. The rate and disfluency of utterance of the speaker are of interest in the study of the stuttering problem, but primarily in their distinctive relationships to the perceptual and evaluative reactions of the listener—and of the speaker functioning as his own listener—and in their interactions in given circumstances with the other relevant behaviors and conditions of the speaker.

The data obtained in the studies here reported provide a basis not only for further experimental research but also for the development of practical procedures of measuring disfluency and rate for clinical purposes. Young suggests the elimination of those measures of disfluency which are least predictive of ratings of severity of stuttering, and the combining of the remaining measures. In like vein, Sander suggests a single measure representing the number of 'disfluent words.' These and other possibilities of simplified methods of measurement are under review. Meanwhile, the measurements of rate and disfluency here reported may be taken, with due qualifications, as normative for the sampled populations of speakers.

The assistance of many persons who have shared the work and responsibility involved in the research here presented is to be acknowledged with appreciation. The specific acknowledgments made in the text and in footnotes are necessarily far from exhaustive. The research program which these studies represent has been carried on at the University of Iowa for over 35 years, and today's findings and interpretations are, in varying fashions and extents, the fruits of yesterday's insights and wonderings—and of yesterday's mistakes. It is impossible in these circumstances to know precisely what has been contributed by each member of the large and growing company of collaborators, or to be sure indeed that all of them are known. No one deserving of recognition for labors performed in the planning and execution of the research reported in this Monograph has been disregarded except unwittingly. Particular appreciation is to be expressed for the contributions that have been made to this research, in all of the obvious and subtle ways that defy enumeration, by James Curtis, Frederic Darley, D. C. Spriestersbach, Dorothy Sherman, and Dean Williams. Special acknowledgment is to be made of the counsel provided by Professor Merle Tate, particularly in connection with the analysis and evaluation of the data in the studies by Johnson and Neelley. In addition to discharging his responsibilities as a research associate, Walter Cullinan, under the gracious tutelage of Dorothy Moeller, performed the duties of editorial assistant in the publication of this Monograph. The secretarial services involved were performed by Ruth Farstrup. To those specifically mentioned, and to all the others who know that they too should have been mentioned, my warm thanks.

The basic institutional groundwork for the research reported in this Monograph has been provided by the University of Iowa and the Louis W. and Maud Hill Family Foundation and, through Grant RD 319, its specific operational aspects have been supported by the Office of Vocational Rehabilitation of the Department of Health, Education, and Welfare.

Measurements of Oral Reading and Speaking Rate And Disfluency of Adult Male and Female Stutterers and Nonstutterers

WENDELL JOHNSON

There were three main purposes of this study. The first was to develop procedures for obtaining samples of speech and of oral reading and for analyzing them with respect to discernible varieties of disfluency. The study was designed to provide also for measures of rate of utterance. The second purpose was to obtain normative and comparative data respecting rate and disfluency in the speech and oral reading of adult male and female stutterers and nonstutterers. The third purpose was to explore the implications of the assembled data, in relation to certain findings from other studies, with a view to the further clarification of the fundamental nature of the problem called stuttering. The design of the study serves to place in focus particularly the question of the referential equivalence of the terms 'stuttering' and 'disfluency.' It also brings under scrutiny, therefore, the related, and more basic, question of whether the stuttering problem is definable more appropriately by reference to the speaker and his disfluency or by reference to the interaction between speaker and listener, or, more precisely, between the processes of vocal utterance and those of the perceptual and evaluative reactions to it, or associated with it.

The methodology of the present study constitutes an extension and refinement of that developed in previous investigations (*3*, *8*, *10*, *15*), and the normative data assembled in this study are an addition to the normative data obtained in those investigations. The theoretical explorations stimulated by the findings of the present study constitute an elaboration of interpretative discussions of data obtained in earlier phases of the research program of which the studies reported in this Monograph are representative (*8*, *9*, *10*).

Wendell Johnson is Professor of Speech Pathology and Psychology, University of Iowa. The study reported here is the culmination of work extending over several years and many persons have been associated with the various phases of the relevant research. Those who have contributed directly to the investigation as here reported are Richard Boehmler, F. Lee Brissey, Walter Cullinan, Robert Duffy, James Frick, Claire Hanley, Joseph Kools, James Neelley, Elizabeth Prather, Earl Schubert, Merle W. Tate, William Trotter, Dean Williams, and Martin Young. The research done by Dr. Robert J. Duffy in the preparation of his M.A. thesis at the University of Iowa is represented by the data for the female stutterers included in this report. Drs. Richard Boehmler, William Trotter and Martin Young contributed substantially to the compilation, analysis, and description of the data. The extensive assistance of Professor Merle W. Tate as statistical analyst and consultant in preparing the final report is to be acknowledged with particular appreciation. This study was supported, in part, by the Louis W. and Maud Hill Family Foundation and by Grant RD 319 of the Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

Procedure

In this study measurements were made of rate and disfluency in samples of the oral reading and speaking of 100 male and 100 female adult speakers, of whom 50 in each group were classified by themselves and others as stutterers and 50 as nonstutterers.

Experimental Operations. With the subject seated in full view of the recording equipment, a tape-recorded speech sample was obtained. First, the tape recorder was turned on and the subject was asked for identifying information such as name, age, level of education, and marital status. He was then asked why he had come to college and what previous experience he had had in having his speech recorded. The main purpose of this interview was to accustom the subject to the experimental situation. After two or three minutes of conversation the recorder was turned off and instructions were given for the first speaking performance, the Job task. The subject was instructed to perform this task by talking for three minutes or so about his future job or vocation. It was suggested that he tell about the vocation, why he chose it, and anything else about it that he wished to discuss. If the subject had not yet chosen a vocation he was asked to tell about jobs he had held in the past. He was allowed one minute to think about what to say. The recorder was then turned on and the subject was asked to begin speaking. If he stopped before the end of two minutes he was encouraged by leading questions to continue. An effort was made to avoid formal structuring of the speaking performance.

Upon completion of the first task the recorder was turned off and instructions were given for the second speaking performance, the TAT task. The subject was presented with Thematic Apperception Test (TAT) card number 10 (*11*) and asked to develop a dramatic story based on the picture. He was asked to be prepared to speak for about five minutes about what was happening at the moment in the pictured situation, what events had preceded those shown in the picture, and what the outcome of the story was to be. Up to one minute was allowed the subject to prepare the story. The recorder was then turned on and the subject was asked to begin; if he stopped talking before the end of three minutes he was asked leading questions to stimulate him to continue speaking.

At the conclusion of the second task the recorder was turned off. The subject was handed a 300-word reading passage and instructed to read it aloud as he ordinarily would. The recorder was then turned on again and the subject was asked to begin reading. This performance was called the Reading task. The reading passage was constructed by Darley (*3*) and may be found in Fairbanks (*5*).

All subjects performed the three tasks in the same order, as indicated above: Job, TAT, and Reading. The experimenter was the only observer.

Subjects. One hundred subjects, all of whom were considered by themselves and by their speech clinicians to be stutterers, constituted one of the two groups in this study. Fifty of these were male college students drawn from seven Midwestern colleges and universities; all were of college age and 45 of the 50 for whom exact age data were obtained ranged in age from 16 to 24 years, with a mean of 19.6 years. The remaining subjects in the clinical group consisted of 50 female speakers drawn from seven Midwestern and two Eastern colleges and universities and from private speech clinics in various parts of the country. The age range of this group was 17 to 41 years, the middle 80% ranging from 18 to 28 years, with a mean age of 21.4 years. Thirty of the 50 subjects in the female clinical group were attending college at the time the speech samples were obtained; the remaining 20 were not attending or had not attended a college or university. The mean reading rates for the college and noncollege female subjects were, respectively, 110 and 111 words per minute, and the corresponding mean total numbers of disfluencies per 100 words were 21 and 16. The distributions were such that the indicated differences in means did not contraindicate pooling of the data for those attending and those not attending college.

The second group was composed of 100 subjects, 50 males and 50 females, attending the University of Iowa. All were considered by themselves and by the investigators to be normal speakers. They ranged in age from 17 to 24 years; the mean age of the male group was 19.2 years and that of the female group was 19.3 years. These subjects were drawn from beginning courses in psychology and speech and from students who had recently enrolled in the University of Iowa for the first time as transfers from other institutions. They were majoring in a variety of fields including premedicine, nursing, commerce, music, psychology, education, art, and speech.

Analysis of Disfluency. The features of speech emphasized in evaluation of the samples were the ones thought to be representative of, or most closely related to, disfluency. One obvious omission was pause time. This measure was omitted because of the relatively unsystematic judgment involved in deciding whether a given pause is or is not part of the meaningfully fluent production of speech. The following types of speech behavior were classified as disfluencies:

1. Interjections of Sounds, Syllables, Words, or Phrases. This category includes extraneous sounds such as 'uh,' 'er,' and 'hmmm' and extraneous words such as 'well,' which are distinct from sounds and words associated with the fluent text or with phenomena included in other categories. An instance of interjection may include one or more units of repetition of the interjected material; for example, 'uh' and 'uh uh uh' are each counted as one instance of interjection. The number of times the interjection is repeated (units of repetition) within each instance is also noted; 'uh uh' is an example of an interjection repeated once and 'uh uh uh' is an example of an interjection repeated twice.

2. Part-word Repetitions. Repetitions of parts of words—that is, syllables and sounds—are placed in this category. Within each instance of repetition the number of times the sound or syllable is repeated is counted; 'buh-boy' involves one unit of repetition and 'guh-guh-girl' involves two units. No attempt is made to draw a distinction between sound and syllable repetitions.

'Ruh-ruh-run,' 'cuh-come,' 'ba-ba-baby,' and 'a-bou-bout' are examples of part-word repetitions.

3. Word Repetitions. Repetitions of whole words, including words of one syllable, are counted in this category. Both the number of instances and the number of repetition units within each instance are counted. 'I-I-I,' 'was-was,' and 'going-going' are samples of instances of word repetition; the first involves two units of repetition and each of the other two involves one unit. A word repeated for emphasis, as in 'very, very clean,' is not counted as a disfluency. A part-word repetition, or an interjection, does not nullify a word repetition; for example, 'going uh going' or 'guh-going going' is classified as a word repetition. In any such case, the interjected or associated disfluency is also tabulated in the appropriate category.

4. Phrase Repetitions. Repetitions of two or more words are included in this category. 'I was I was going' is an example of this type of disfluency.

5. Revisions. Instances in which the content of a phrase is modified, or in which there is grammatical modification, are counted as instances of revision. Change in the pronunciation of a word is also counted as a revision. 'I was—I am going' is an example of this category.

6. Incomplete Phrases. An incomplete phrase is one in which the thought or content is not completed and which is not an instance of phrase repetition. 'She was—and after she got there he came' contains an example of an incomplete phrase.

7. Broken Words. This category is typified by words which are not completely pronounced and which are not associated with any other category, or in which the normal rhythm of the word is broken in a way that definitely interferes with the smooth flow of speech. 'I was g— (pause) —oing home' is an example of a broken word.

8. Prolonged Sounds. This category includes sounds judged to be unduly prolonged. If a sound is prolonged twice, it is counted both as a prolonged sound and a part-word repetition.

Disfluencies are identified in each case from a verbatim transcript while listening to a play-back of the recording. As much relistening as necessary is allowed until the operator is satisfied that an accurate identification of disfluencies has been achieved.

Treatment of Data. Verbal Output. The subject's verbal output in each task is defined as the number of words spoken. In computing this measure each word repeated singly or in a phrase is counted only once, and interjected sounds or words not regarded as integral parts of the meaningful context are not counted. In any instance of revision only the words in the final forms are counted. The verbal output for the reading passage is always taken as 300 words even though some subjects may have omitted or added words. It should be noted that verbal output in the two speaking tasks is under arbitrary control of the investigator to the degree that he encourages some of the subjects to continue speaking when they stop short of what he regards as an adequate sample.

Speaking and Oral Reading Rates. Speaking and oral reading rates are computed in terms of words per minute, as determined by calculating the ratio of verbal output, as defined above, to reading or speaking time as measured by a stop watch.

Disfluency Category Index. A computation is made of the number of instances of each type of disfluency per 100 words for each of the three tasks. The following formula is used to compute this index: Disfluency Category Index = (ND/NW)100, in which ND represents the total number of instances of disfluency of the designated type in the speech sample of the subject, and NW represents the number of words, or the verbal output, as defined above, of the subject for the sample.

Task Disfluency Index. A disfluency index for each task is computed for each subject by obtaining the sum of the subject's category indexes for the task.

Average Number of Units per Instance of Repetition. For each of the first three categories (interjections, part-word repetitions, and word repetitions) the number of units of repetition in each instance of disfluency is computed. For example, as has been indicated, 'buh-boy' and 'guh-guh-girl' represent one and two units of repetition, respectively. An average is computed by dividing the total number of units of repetition by the total number of instances of repetition for each category.

Scorer Reliability of the Disfluency Analysis. Scorer reliability is indicated by information from three separate investigations in the research program which this Monograph represents. In one aspect of the present study Duffy (*4*) obtained data from two observers who listened to 12 recorded speech samples of adult female stutterers and marked individual moments of disfluency, employing the disfluency analysis described above. Duffy correlated the total frequencies registered by the two observers in identifying disfluences in each disfluency category, and obtained Pearson coefficients of correlation which ranged from .90 to .99. As part of the study by Young (*17*), reported in this Monograph, 10 tape-recorded 200-word speech samples of adult male stutterers were analyzed twice, with intervening periods of from three weeks to two months, with reference to five types of disfluency. An index of agreement was computed by means of the formula $C/\sqrt{xy}$, in which C represents the number of disfluencies identified in relation to the same words in both analyses, and x and y the numbers of disfluencies identified in the two separate analyses. This is an index of the self-agreement of the observer in identifying types as well as specific loci of occurrences of disfluency. Indexes ranged from .91 to 1.00, or perfect agreement, for the 10 samples, with an index of agreement of .97 for all samples combined into one large sample of 2,000 words. As reported in this Monograph, Sander (*12*), using the eight disfluency categories employed in the present study, made two analyses of the tape recordings of 12 Job tasks and 12 Reading tasks performed by adult male stutterers. In each case the two analyses were separated by at least one month. In order to estimate the scorer's self-agreement, Sander used the formula Agreement Index $= a/(a + d)$, in which a represents agreements and d disagreements (the discrepancy in each case between the first and second scor-

TABLE 1. Summary of measures of verbal output, defined as number of words spoken, for the Job and TAT tasks for 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN).

Task	*Range*	*Mean*	*Standard Deviation*	*10th %-ile*	*Median*	*90th %-ile*
Job						
MS	55-570	280.4	108.5	136	276	430
FS	48-553	248.1	118.3	94	228	400
MN	60-689	359.4	107.7	247	358	498
FN	174-687	392.6	108.7	258	398	550
TAT						
MS	83-766	378.6	175.1	160	381	700
FS	54-960	293.5	179.7	118	264	594
MN	158-1064	518.2	161.0	339	524	703
FN	129-996	565.7	186.7	376	578	828

ings in number of disfluencies noted). He computed indexes of 96% for both the Job and Reading tasks.

Results

A summary of the obtained measures concerned with verbal output in the Job and TAT tasks for the male and female stutterers and nonstutterers is presented in Table 1. The Reading task, as has been stated, was assumed to involve 300 words for all subjects. In Table 2 there is a summary of the measures of time spent on each task by the male

TABLE 2. Summary of measures of amount of time, in minutes and seconds, spent on each task by 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN).

Task	*Range*	*Mean*	*Standard Deviation*	*10th %-ile*	*Median*	*90th %-ile*
Job						
MS	1:32-5:43	3:02	0:39	2:24	2:59	3:55
FS	0:55-4:45	2:36	0:42	1:58	2:25	3:43
MN	1:25-4:15	2:41	0:36	1:48	2:42	3:31
FN	1:26-3:54	2:40	0:33	1:57	2:38	3:17
TAT						
MS	2:09-7:57	4:48	1:08	3:24	4:53	6:16
FS	1:29-6:00	3:42	1:03	2:15	3:36	5:25
MN	2:11-6:17	4:16	1:00	2:55	4:34	5:23
FN	1:59-6:23	4:14	1:03	2:44	4:22	5:33
Reading						
MS	1:30-9:24	3:01	1:42	1:39	2:26	5:25
FS	1:30-14:45	3:20	2:16	1:38	2:40	5:36
MN	1:23-3:00	1:47	0:17	1:34	1:43	2:05
FN	1:22-2:13	1:43	0:10	1:32	1:43	1:58

and female subjects in each of the two groups. The ranges and deciles of the distributions of oral reading and speaking rates for the three tasks for all subgroups of subjects are presented in Table 3. Measures of disfluency are summarized in Tables 4-9. Table 10 contains coefficients of correlation among the three tasks for measures of each of the disfluency variables and of rate.

Measures of Rate and Disfluency of Stutterers and Nonstutterers. The differences between stutterers and nonstutterers in measures of rate of oral reading and of speaking were highly significant, the nonstutterers, of course, showing the higher rates. In general, the nonstutterers were considerably less disfluent than the stutterers. The extent of overlap of the distributions and the frequencies of zero values in the majority of the disfluency distributions of both groups, however, are to be considered as well as the averages.

The degrees of overlap for the various distributions are to be seen particularly in Tables 4-9. In Tables 4-8 there are represented 60 distributions of disfluency measures for the nonstutterers, 30 for the males and 30 for the females, and 60 corresponding distributions for the stutterers. In the great majority of the 60 pairs of distributions there is substantial overlapping. Considering the data in Table 4 for the Job task, the overlap of the stutterers' and nonstutterers' distributions is virtually complete for the category of revisions and it is considerably extensive for interjections, incomplete phrases, and for both word and phrase repetitions. The proportions of both major groups presenting no broken words or prolonged sounds were sufficiently large to warrant the statement

TABLE 3. Ranges and deciles of distributions of values for speaking and reading rates in words per minute, for each task for 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN).

		Decile*								
Task	*Range*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Job										
MS	24.7-184.4	39.3	67.1	81.4	92.5	102.0	105.5	121.0	133.0	139.4
FS	12.9-183.3	44.1	64.8	70.6	81.0	98.9	103.1	120.0	148.3	170.2
MN	42.3-201.2	105.4	112.6	120.3	129.7	136.2	141.5	146.6	158.1	160.0
FN	94.7-198.4	121.8	131.1	135.7	140.9	147.0	150.0	154.8	164.7	185.1
TAT										
MS	18.3-148.6	29.4	48.9	68.2	78.6	86.1	91.8	102.0	111.9	135.7
FS	9.9-177.2	31.7	44.7	56.6	70.4	78.6	84.0	104.7	113.2	141.4
MN	72.5-197.8	99.6	101.6	112.3	114.7	119.2	127.2	130.9	138.0	148.6
FN	58.6-202.7	108.8	117.1	119.9	122.9	130.5	138.2	144.3	151.4	162.4
Reading										
MS	31.9-200.0	55.4	76.5	102.4	116.7	123.5	131.6	142.9	162.2	181.8
FS	20.3-200.0	53.6	67.7	84.1	92.3	109.8	128.6	146.5	155.2	181.8
MN	104.9-217.4	151.5	160.4	164.8	171.4	176.5	179.6	181.8	187.5	202.7
FN	135.1-219.0	155.4	163.9	171.4	173.4	176.5	181.6	184.1	187.5	197.4

that approximately half of the stutterers were indistinguishable from most of the nonstutterers with respect to these types of disfluency. With regard to word repetitions, roughly 20% of the stutterers were more fluent than from 30% to 40% of the nonstutterers. The most disfluent 30% of the nonstutterers performed more phrase repetitions than did the most fluent 30% of the stutterers. Half of the male nonstutterers were no more fluent than 30% of the male stutterers, and 40% of the female nonstutterers were no more fluent than 20% of the female stutterers, as far as interjections were concerned. There is only about a 10% overlap of the two distributions, however, for part-word repetitions. Essentially the same statements are to be made concerning the data in Table 5 for the TAT task. The main differences between the distributions for the speaking tasks and the Reading task are to be seen in the large proportions of zeros, each zero indicating that no disfluencies of the designated type were performed, and the relatively reduced decile values for both groups for oral reading.

In each of 36 of 48 comparisons of the highest values for the sex subgroups of stutterers and nonstutterers represented in Tables 4-6, the most disfluent nonstutterer was less fluent than 50% or more of the stutterers. For example, the male nonstutterer with the most interjections in the Job task had more interjections than 70% of the male stutterers; and the female nonstutterer with the most prolonged sounds in the Reading task had more of them than 50% of the female stutterers. In all three tasks, for both males and females, the nonstutterer with the highest number of revisions and of incomplete phrases, respectively, had more of them than 80% of the corresponding subgroups of stutterers in two of the 12 comparisons and more than 90% in the remaining 10 comparisons.

Differences Among Tasks. Tables 4, 5, 7, 8, and 9 indicate that the distributions of disfluencies per 100 words were very similar for the two speaking tasks. According to the sign test,[1] there was only one significant difference between the Job and TAT tasks. The female stutterers performed significantly more disfluencies, all types combined, (Table 7), on the TAT task than on the Job task ($P < .01$ from the sign test); repeated words and phrases and incomplete phrases contributed particularly to this difference, as can be seen in Tables 4 and 5.

Both male and female stutterers showed significantly fewer total disfluencies in the Reading task than in the speaking tasks, with $P < .01$ in all comparisons according to the sign test. The relevent data are summarized in Table 7. Differences in interjections and word repetitions, as represented in Tables 4 and 5, contributed most to these differences in total disfluencies. As expected, of course, the nonstutterers were decidely more fluent in read-

[1]As indicated by inspection of Tables 4-9, all distributions of disfluencies in all subgroups are severely skewed to the right and in nearly all there is a relatively high frequency of zero scores. Hence, it was necessary to use distribution-free or nonparametric methods in testing for significant differences. The sign test and the Kolmogorov-Smirnov test, referred to below, are such methods. The sign test is discussed by Siegel (*13*) and Tate and Clelland (*14*) and the Kolmogorov-Smirnov test is described by Goodman (7).

TABLE 4. Ranges and deciles of distributions of values for number of disfluencies of each type per 100 words for the Job task for 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN).

		Decile*								
	Range	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Interjections										
MS	1.0-80.0	2.2	3.5	4.4	4.9	7.3	10.3	12.1	15.5	23.5
FS	0-52.1	2.4	3.1	4.9	5.9	7.5	9.3	11.1	15.8	22.9
MN	0-13.7	1.2	2.3	3.2	3.8	4.4	5.6	6.7	8.2	9.3
FN	0.4-6.6	0.7	1.6	1.8	2.2	2.8	3.1	3.6	4.5	5.6
Part-Word Repetitions										
MS	0-44.6	1.0	1.4	2.3	3.4	3.7	6.0	7.3	12.8	34.5
FS	0-59.5	0.5	0.8	1.6	2.4	4.8	6.5	8.9	11.8	19.9
MN	0-2.2	0	0	0	0	0	0.2	0.3	0.5	1.0
FN	0-1.0	0	0	0	0	0.2	0.2	0.3	0.4	0.6
Word Repetitions										
MS	0-15.5	0.4	0.8	1.5	2.0	2.7	3.4	4.0	5.4	6.6
FS	0-14.9	0.3	0.6	1.3	1.9	2.4	3.1	4.3	5.5	8.4
MN	0-2.9	0	0	0.3	0.5	0.7	0.9	1.1	1.3	1.5
FN	0-2.4	0	0	0.2	0.3	0.4	0.5	0.7	0.9	1.6
Phrase Repetitions										
MS	0-8.5	0	0	0.3	0.5	0.6	1.0	1.5	2.0	2.9
FS	0-3.8	0	0	0.3	0.4	0.7	0.9	1.1	1.3	2.2
MN	0-2.2	0	0	0	0	0	0.2	0.3	0.4	0.6
FN	0-0.6	0	0	0	0	0	0.2	0.3	0.3	0.5
Revisions										
MS	0-5.5	0	0.5	0.7	0.8	0.9	1.2	1.5	2.1	2.8
FS	0-1.4	0	0	0	0	0.3	0.5	0.6	0.9	1.1
MN	0-4.3	0.5	0.6	0.8	0.9	1.1	1.3	1.7	2.0	2.2
FN	0-2.7	0	0.2	0.4	0.5	0.6	1.0	1.2	1.5	1.8
Incomplete Phrases										
MS	0-5.9	0	0	0	0	0	0	0	0	0.5
FS	0-2.3	0	0	0	0	0.2	0.4	0.6	1.0	1.7
MN	0-0.8	0	0	0	0	0	0	0	0	0.3
FN	0-1.3	0	0	0	0	0	0	0	0.3	0.5
Broken Words										
MS	0-28.2	0	0	0	0	0	0.2	0.6	0.9	2.1
FS	0-5.7	0	0	0	0	0.3	0.4	1.1	1.7	3.2
MN	0-0.5	0	0	0	0	0	0	0	0	0
FN	0-0.3	0	0	0	0	0	0	0	0	0.2
Prolonged Sounds										
MS	0-21.8	0	0	0	0.5	0.7	1.4	2.7	4.4	9.8
FS	0-25.7	0	0	0	0	0.5	1.2	2.2	4.5	10.6
MN	0-1.7	0	0	0	0	0	0	0	0	0.3
FN	0-0.4	0	0	0	0	0	0	0	0	0.2

*Computed from ungrouped data.

TABLE 5. Ranges and deciles of distributions of values for number of disfluencies of each type per 100 words for the TAT task for 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN).

		Decile*								
	Range	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Interjections										
MS	0-71.6	1.0	2.7	3.3	5.8	7.2	8.9	11.8	18.2	27.7
FS	0.2-79.6	1.6	5.4	6.5	7.5	8.8	10.8	13.6	16.6	24.8
MN	0-15.3	0.8	1.1	1.7	2.2	2.6	3.7	6.4	7.1	9.4
FN	0-11.6	0.3	0.6	1.0	1.4	2.0	3.1	4.3	5.0	7.5
Part-Word Repetitions										
MS	0-52.2	1.0	1.6	2.2	3.5	4.2	5.0	8.2	10.7	33.6
FS	0-57.3	0.4	0.8	1.0	3.1	5.2	6.9	12.3	14.1	25.4
MN	0-1.2	0	0	0	0	0.2	0.3	0.4	0.4	0.6
FN	0-1.2	0	0	0	0.1	0.2	0.2	0.3	0.5	0.6
Word Repetitions										
MS	0-13.5	0.6	1.1	1.4	1.7	3.1	3.8	4.6	5.9	9.9
FS	0-17.7	0.5	1.2	1.6	2.0	2.9	3.7	5.0	6.7	8.6
MN	0-2.5	0	0.2	0.4	0.6	0.6	0.7	1.0	1.3	1.9
FN	0-4.5	0	0.2	0.3	0.3	0.4	0.5	0.7	1.0	1.5
Phrase Repetitions										
MS	0-5.5	0	0.2	0.4	0.6	0.9	1.3	1.5	2.2	3.1
FS	0-2.9	0	0	0.3	0.6	0.8	1.2	1.4	1.7	2.3
MN	0-1.3	0	0	0	0	0.2	0.2	0.4	0.7	1.1
FN	0-1.5	0	0	0	0	0.1	0.2	0.3	0.5	0.6
Revisions										
MS	0-5.4	0.3	0.8	1.0	1.1	1.2	1.4	1.7	2.4	2.6
FS	0-2.2	0	0	0	0	0.3	0.4	0.8	1.2	1.7
MN	0.3-4.3	0.6	0.7	0.8	1.0	1.2	1.4	1.7	2.4	2.8
FN	0-3.1	0.2	0.4	0.6	1.0	1.1	1.2	1.5	1.6	1.8
Incomplete Phrases										
MS	0-26.6	0	0	0	0	0	0	0	0.2	0.6
FS	0-5.4	0	0	0	0.2	0.3	0.5	0.8	1.0	1.5
MN	0-1.3	0	0	0	0	0	0	0	0.2	0.3
FN	0-1.2	0	0	0	0	0	0.1	0.2	0.3	0.6
Broken Words										
MS	0-3.5	0	0	0	0	0	0	0.3	1.1	1.9
FS	0-7.0	0	0	0	0	0.3	0.7	1.0	1.2	2.1
MN	0-0.6	0	0	0	0	0	0	0	0	0.2
FN	0-0.9	0	0	0	0	0	0	0	0	0.2
Prolonged Sounds										
MS	0-17.6	0	0	0.3	0.7	1.1	2.0	2.9	4.4	11.0
FS	0-23.9	0	0	0	0.5	0.6	0.9	2.4	5.7	11.7
MN	0-0.6	0	0	0	0	0	0	0	0.2	0.2
FN	0-0.3	0	0	0	0	0	0	0	0.1	0.2

*Computed from ungrouped data.

ing than in speaking. The great majority of the nonstuttering subjects showed little or no disfluency in reading. The most disfluent nonstutterer had a total of only 12 disfluencies in the Reading task.

While it has been generally noted previously that stutterers perform better as a rule in oral reading than in speaking, the present data are somewhat more substantial than most comparable sets of data previously reported. In view, however, of the fact that there are individual exceptions to the rule that most stutterers are more fluent in oral reading than in speaking (16% of the males and 30% of the females in the present samples were more fluent in speaking), further study would be expected to provide a desirably more substantial basis for systematic interpretation.

Correlations Among Tasks. The correlation coefficients presented in Table 10 were computed separately for the stutterers and nonstutterers, with sex subgroups combined. Although the differences between sexes in both groups were significant for certain disfluency variables in both speaking tasks, as noted below, none of the differences was great enough to inflate seriously the coefficient of correlation when sexes were combined. Of greater concern than the small inflation of the coefficients were the departure from normality and the relatively high frequencies of zero scores in the distributions. In such circumstances the results of tests of significance of the product-moment correlation coefficient must be interpreted cautiously.

The coefficients of correlation were generally higher for the stutterers, with 23 out of their 30 coefficients significant at the 1% level. Of the coefficients for the measures in the Job task and the TAT task, four, all involving the stutterers, were .90 or higher. The only nonsignificant coefficient for stutterers' measures in the two speaking tasks was that of .17 for incomplete phrases. For this disfluency variable there was a significant coefficient of .31 for the Job task and the TAT task for the nonstutterers; the difference between the values of .17 and .31 was not statistically significant, however. The other coefficients involving incomplete phrases for the other pairs of tasks were not significant for either group of speakers. Seven out of ten coefficients for the nonstutterers' measures in the Job and TAT tasks were statistically significant at the 1% level; those that were not significant were for the part-word repetitions, broken words, and prolonged sounds, types of disfluency performed with relatively low frequency by this group.

For both groups of speakers the measures for the two speaking tasks were more highly related to each other than were the measures for either of these tasks to those for the Reading task. In general, the coefficients of correlation between tasks were higher for the stutterers than for the nonstutterers, except those for phrase repetitions, incomplete phrases, and revisions, which were relatively low and in most instances not statistically significant at the 1% level. The higher coefficients of correlation for the stutterers for most of the disfluency categories and for rate are to be evaluated with reference to the generally greater spread of the scores of the stutterers. The only two coefficients of even moderate magnitude for the nonstut-

TABLE 6. Ranges and deciles of distributions of values for number of disfluencies of each type per 100 words for the Reading task for 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN).

	Range	*Decile** *1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Interjections										
MS	0-85.5	0	0	0	0.3	0.3	0.7	1.0	2.0	9.7
FS	0-84.6	0	0	0	0.3	0.7	1.3	2.0	2.7	12.8
MN	0-1.0	0	0	0	0	0	0	0	0	0
FN	0-1.3	0	0	0	0	0	0	0	0	0.3
Part-Word Repetitions										
MS	0-47.8	0.7	1.0	1.3	2.3	4.0	5.3	7.0	12.3	16.7
FS	0-88.0	0.7	1.0	1.3	2.0	3.0	5.7	12.3	19.3	25.3
MN	0-2.0	0	0	0	0.3	0.3	0.3	0.7	1.0	1.0
FN	0-1.3	0	0	0	0	0	0.3	0.3	0.3	0.7
Word Repetitions										
MS	0-6.7	0	0	0	0.3	0.3	0.7	1.0	2.0	2.7
FS	0-13.9	0	0	0	0.3	0.3	1.0	1.3	2.0	2.7
MN	0-1.0	0	0	0	0	0	0	0.3	0.3	0.7
FN	0-0.7	0	0	0	0	0	0	0	0.3	0.3
Phrase Repetitions										
MS	0-7.7	0	0	0	0.3	0.3	0.7	0.7	1.0	1.7
FS	0-6.7	0	0	0	0.3	0.3	0.3	0.7	1.3	1.7
MN	0-0.7	0	0	0	0	0	0	0	0.3	0.3
FN	0-0.3	0	0	0	0	0	0	0	0	0
Revisions										
MS	0-1.3	0	0	0	0.3	0.3	0.7	0.7	0.7	1.0
FS	0-2.3	0	0	0	0	0.3	0.3	0.3	0.7	1.0
MN	0-2.0	0	0	0.3	0.3	0.7	0.7	0.7	1.0	1.3
FN	0-3.0	0	0	0.3	0.3	0.3	0.3	0.7	0.7	1.0
Incomplete Phrases										
MS	0-0	0	0	0	0	0	0	0	0	0
FS	0-0.7	0	0	0	0	0	0	0	0	0
MN	0-0	0	0	0	0	0	0	0	0	0
FN	0-0	0	0	0	0	0	0	0	0	0
Broken Words										
MS	0-9.7	0	0	0	0	0	0	0.3	0.3	2.7
FS	0-19.3	0	0	0	0	0	0.7	0.7	1.3	3.1
MN	0-0.7	0	0	0	0	0	0	0	0	0
FN	0-0.7	0	0	0	0	0	0	0	0	0
Prolonged Sounds										
MS	0-27.7	0	0	0	0.3	0.7	1.0	2.7	4.7	8.7
FS	0-14.3	0	0	0	0.3	0.3	0.7	1.3	4.3	10.7
MN	0-0.3	0	0	0	0	0	0	0	0	0.3
FN	0-0.3	0	0	0	0	0	0	0	0	0

*Computed from ungrouped data.

terers involving the Reading task and the other two tasks were those for phrase repetitions (Job task vs. Reading task, .51) and total task index (TAT task vs. Reading task, .45). For the nonstutterers the only other coefficient involving reading that was significant at the 1% level was that for revisions (Job task vs. Reading task, .28). For the stutterers 14 out of 20 coefficients of correlation between measures for the Reading task and the two speaking tasks were significant.

Sex Differences. Differences between male and female stutterers and between male and female nonstutterers in verbal output, rate of oral reading and of speaking, and numbers of the various kinds of disfluency per 100 words were tested for significance by use of the Kolmogorov-Smirnov test.

The verbal output of the male stutterers was greater in both the Job and the TAT tasks than that of the female stutterers and significantly so in the latter ($P < .005$). The relevant data are shown in Table 1. For the nonstutterers, the verbal output was greater for females in both tasks, but the differences were not significant. If it turns out in further research that the verbal output of nonstuttering females in such speaking tasks is equal to or greater than that of male nonstutterers and that female stutterers are significantly lower in verbal output than male stutterers, at least in the TAT sort of task, explanation would be in order.

There were no significant differences in rate of oral reading and of speaking, measured in words per minute, between the males and females in either group. The rates were markedly faster, as would be expected, for the nonstutterers than the stutterers in all tasks.

Male stutterers showed significantly more revisions and significantly fewer incomplete phrases than female stutterers in both speaking tasks ($P < .02$ in all four comparisons). The relevant data are shown in Tables 4 and 5. Male nonstutterers showed more revisions on both tasks than female nonstutterers, the difference being significant for the job task ($P < .005$). Since revisions and incomplete phrases are complementary to some extent, the bulk of evidence suggests that female speakers in general make fewer revisions and more incomplete phrases than male speakers in the Job and TAT kinds of task. To the extent that this is so, it is important to ask why these sex differences in speech behavior are more pronounced in the stuttering group.

Male nonstutterers showed more interjections in both tasks than did female nonstutterers, the difference for the Job task being significant ($P < .02$). In total number of disfluencies, as summarized in Table 7, male and female stutterers showed small and insignificant differences in all three tasks. Male nonstutterers showed consistantly higher total disfluency values in all three tasks, but the difference was significant only for the Job task ($P < .01$).

There were no significant differences between the sex subgroups in extent of repetition, expressed as the number of units per instance of repetition. The relevant data are summarized in Table 9.

Although caution is indicated in generalizing from a few significant differences when numerous comparisons are made, it seems probable that there are in fact a number of kinds of speech behavior differences between male and female speakers in the respective populations represented by the samples in hand.

Discussion

The discussion that follows is designed to point up some of the more important aspects of the data obtained in this study, and to explore some of their more significant implications with respect to the fundamental question of the appropriate formulation of the problem called stuttering.

Fundamentally, the findings imply that the problem is not to be adequately described or accounted for by reference to only the disfluency and other presumably relevant characteristics of the speaker. The problem would seem to involve distinctive patterns of interaction of speaker and listener, of the observer and the observed, of the processes of perception and evaluation and vocal utterance, of speaking and reactions to it. In order to describe, or construct, the indicated patterns of interaction, it would seem necessary to give consideration to all of the processes involved, not only to the specific matter of speech disfluency, per se.

The issue is pointed up particularly in the degree of overlapping of the distributions for the two groups of subjects for the various types of disfluency. The obvious meaning of the overlapping is that some speakers who are classified as stutterers are more fluent, at least in certain respects and at certain times, than are some speakers who are regarded as normal speakers. This observation serves to bring into focus the question of why specified disfluencies are sometimes classified as stuttering and not at other times. Data obtained in exploration of the origins

TABLE 7. Ranges and deciles of distributions of values for total numbers of disfluencies per 100 words for each of the three tasks for 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN).

		Decile*								
Task	*Range*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Job										
MS	2.7-127.3	8.2	12.8	13.9	15.1	19.0	25.7	29.6	44.9	57.7
FS	0.8-101.3	8.3	12.4	13.8	16.6	17.9	22.7	32.3	42.7	62.5
MN	1.6-20.1	4.2	5.0	5.8	6.2	7.0	8.1	9.0	10.8	12.3
FN	0.4-9.1	2.3	2.7	3.1	3.8	4.9	5.9	6.4	7.4	7.9
TAT										
MS	4.6-135.8	8.3	10.5	14.7	18.4	22.6	26.6	33.3	47.3	66.1
FS	1.8-109.3	8.5	13.8	16.2	19.2	22.3	27.7	38.5	49.0	73.7
MN	0.7-19.9	2.2	2.9	3.9	5.3	6.6	8.0	8.8	9.7	12.6
FN	0.5-17.1	1.4	2.0	3.0	3.7	4.6	5.7	7.1	8.4	10.4
Reading										
MS	0-141.5	1.7	2.3	4.0	6.3	8.3	15.7	17.3	26.3	38.0
FS	0-112.8	1.3	2.7	3.7	6.3	9.7	18.7	21.3	35.5	57.0
MN	0-4.0	0.3	0.7	0.7	1.0	1.0	1.3	1.7	2.3	3.0
FN	0-4.0	0	0.3	0.3	0.7	0.7	1.0	1.3	1.7	2.3

*Computed from ungrouped data.

of the stuttering problem (*8*) and other data reported by Williams and Kent (*16*), Giolas and Williams (*6*), Boehmler (*2*), and Young (*17*) indicate that of all the types of disfluency, part-word repetitions are most likely to be classified by listeners, at least in our general culture, as 'stuttering' and that certain other kinds of disfluency, most notably perhaps interjections, revisions, and phrase repetitions, are more commonly considered as 'normal' disfluencies. In addition, listener agreement on whether or not a particular disfluency 'is stuttering' appears to be highest when the disfluency is identified as a part-word repetition (*2*).

Among the possible interpretations of the data pertaining to speaker performance and listener evaluation are these: (a) a listener is more likely to classify a given disfluency as stuttering if he is set to evaluate some disfluencies as stuttering; (b) some disfluencies, particularly those associated with apparent struggle reactions, and those that are part-word repetitions even when relatively simple and effortless, are more likely than other disfluencies to be classified as stuttering by the listener; (c) the more disfluencies the speaker displays the more likely the listener is to regard him as a stutterer; (d) the listener is more likely to classify the speaker's disfluencies as stuttering if he regards the speaker as a stutterer.

Williams and Kent (*16*) have discussed this problem in relation to the clinical counseling of parents who are inclined to regard their children's speech disfluencies as stuttering. They report that when parents who feel that their children are beginning to stutter are advised to give attention to the 'normal' repetitions in their children's speech and to note which are 'normal' and which are 'stutterings,'

TABLE 8. Ranges and deciles of distributions of values for number of part-word, word, and phrase repetitions, combined, per 100 words for all three tasks for 50 male stutterers (MS), 50 female stutterers (FS), 50 male nonstutterers (MN), and 50 female nonstutterers (FN).

		Decile*								
Task	*Range*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Job										
MS	0.3-62.0	2.4	3.9	5.1	6.3	7.9	10.1	13.5	21.6	38.4
FS	0.3-60.7	2.6	3.6	4.5	7.4	8.3	10.4	13.5	16.8	28.3
MN	0-3.7	0	0.4	0.7	0.8	1.1	1.4	1.8	2.3	2.9
FN	0-2.9	0.2	0.3	0.5	0.6	0.8	0.9	1.2	1.4	2.2
TAT										
MS	1.7-65.2	2.7	3.3	5.2	7.1	9.4	10.6	14.1	18.3	42.7
FS	0.8-68.4	1.8	2.8	4.5	8.4	12.0	13.1	16.7	23.4	29.9
MN	0-3.6	0	0.6	0.8	1.0	1.2	1.5	1.9	2.3	2.7
FN	0-5.3	0.4	0.4	0.6	0.8	0.9	1.0	1.5	1.9	2.4
Reading										
MS	0-51.0	1.0	1.6	2.7	4.0	5.6	7.0	9.9	14.4	23.7
FS	0-88.6	1.0	1.7	2.3	2.7	4.7	9.3	15.3	22.7	31.0
MN	0-2.7	0	0.3	0.4	0.5	0.6	0.7	0.8	1.2	2.0
FN	0-2.0	0	0	0	0	0.4	0.5	0.6	0.7	0.9

*Computed from ungrouped data.

TABLE 9. Ranges and deciles of the distributions of numbers of units of repetition per instance of repetition for all tasks and for all subgroups of subjects.

Task	*Range*	*Decile* 1	2	3	4	5	6	7	8	9	N*
Job											
MS											
Interjection	1.00-4.73	1.01	1.03	1.05	1.07	1.08	1.12	1.18	1.34	1.43	50
Part-Wd Rep.	1.00-4.13	1.05	1.11	1.16	1.27	1.42	1.54	1.72	2.26	3.04	48
Word Repet.	1.00-2.00	1.01	1.03	1.04	1.06	1.07	1.09	1.18	1.41	1.64	48
FS											
Interjection	1.00-3.18	1.02	1.03	1.05	1.07	1.09	1.17	1.28	1.56	2.24	49
Part-Wd Rep.	1.00-3.85	1.07	1.15	1.25	1.41	1.55	1.82	2.09	2.34	3.04	49
Word Repet.	1.00-3.00	1.01	1.03	1.05	1.07	1.08	1.14	1.27	1.46	1.58	45
MN											
Interjection	1.00-1.08	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.02	49
Part-Wd Rep.	1.00-1.50	1.00	1.01	1.01	1.02	1.02	1.03	1.04	1.04	1.34	21
Word Repet.	1.00-1.50	1.00	1.01	1.01	1.02	1.02	1.03	1.04	1.04	1.25	37
FN											
Interjection	1.00-1.09	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	50
Part-Wd Rep.	1.00-1.67	1.00	1.01	1.01	1.02	1.02	1.03	1.04	1.04	1.33	26
Word Repet.	1.00-1.20	1.00	1.01	1.01	1.02	1.02	1.03	1.03	1.04	1.04	36
TAT											
MS											
Interjection	1.00-3.71	1.02	1.04	1.06	1.08	1.11	1.16	1.24	1.36	1.74	50
Part-Wd Rep.	1.00-5.50	1.08	1.17	1.26	1.34	1.46	1.69	2.02	2.42	2.82	49
Word Repet.	1.00-1.88	1.02	1.04	1.06	1.08	1.11	1.18	1.28	1.40	1.55	49
FS											
Interjection	1.00-4.02	1.02	1.04	1.06	1.08	1.11	1.16	1.24	1.32	1.60	50
Part-Wd Rep.	1.00-3.46	1.08	1.17	1.35	1.50	1.64	1.81	2.07	2.21	2.75	47
Word Repet.	1.00-2.00	1.01	1.03	1.05	1.07	1.08	1.13	1.23	1.33	1.40	48
MN											
Interjection	1.00-1.16	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.03	48
Part-Wd Rep.	1.00-2.00	1.02	1.04	1.06	1.09	1.11	1.13	1.16	1.18	1.70	29
Word Repet.	1.00-2.00	1.00	1.00	1.00	1.01	1.01	1.01	1.02	1.02	1.11	43
FN											
Interjection	1.00-1.29	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	47
Part-Wd Rep.	1.00-2.00	1.02	1.04	1.07	1.09	1.11	1.14	1.16	1.18	1.24	32
Word Repet.	1.00-1.25	1.00	1.00	1.00	1.01	1.01	1.01	1.02	1.02	1.02	42
Reading											
MS											
Interjection	1.00-2.49	1.01	1.03	1.05	1.06	1.08	1.12	1.28	1.50	1.96	34
Part-Wd Rep.	1.00-3.32	1.08	1.17	1.28	1.42	1.58	1.76	1.92	2.06	2.18	47
Word Repet.	1.00-2.50	1.01	1.03	1.04	1.06	1.08	1.09	1.20	1.34	1.56	33
FS											
Interjection	1.00-4.01	1.01	1.03	1.04	1.06	1.08	1.09	1.21	1.36	1.88	32
Part-Wd Rep.	1.00-3.98	1.05	1.11	1.17	1.36	1.51	1.64	1.77	2.18	2.82	46
Word Repet.	1.00-4.00	1.01	1.02	1.04	1.05	1.07	1.08	1.10	1.21	1.36	33

(*Continued on next page*)

TABLE 9 (*Continued*)

Task	Range	Decile 1	2	3	4	5	6	7	8	9	N*
MN											
Interjection	1.00-2.00										4
Part-Wd Rep.	1.00-3.00	1.01	1.01	1.03	1.05	1.06	1.07	1.08	1.14	1.17	31
Word Repet.	1.00-1.00										14
FN											
Interjection	1.00-1.00										8
Part-Wd Rep.	1.00-2.00	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09	23
Word Repet.	1.00-1.00										12

*For each distribution the N is necessarily restricted to those subjects who performed the indicated type of disfluency in the specified task. For the four distributions with the smallest N's meaningful decile values could not be computed.

they tend increasingly to evaluate the perceived repetitions as 'normal interruptions' rather than as 'stutterings.' Johnson (*8*, pp. 255-257) has approached the problem from the speaker's point of view in presenting a possible interpretation of motivational differences between speakers classified as stutterers and those classified as nonstutterers in relation to differences in relative frequency of part-word as distinguished from whole-word and phrase repetitions. The part-word repetitions may differ from the other two types in being more representative of conflict between the drive to speak and the felt need to avoid disapproval. If so, it could be reasonably assumed that with continuing disapproval by the listener of the speaker's repetitions of whatever kind, the speaker will become more and more strongly motivated to perform the repetitions in the relatively avoidant part-word form rather than in the relatively conflict-free whole word and phrase forms. The shift in the direction of relatively higher frequencies of part-word, rather than whole word and phrase, repetitions as the problem called stuttering develops would appear, then, to be a function of a learning process subsequent to evaluation of repetition as undesirable.

Additional relevant tentative generalizations are suggested by an examination of findings by Johnson and his co-workers (*9*, pp. 206-207). They found that 43 speech pathology students were unable to discriminate between 12 tape-recorded samples of oral reading obtained from speakers considered to be stutterers and 12 samples obtained from speakers considered to be nonstutterers, the two groups of recorded samples having been equated on the basis of the total numbers of disfluencies they contained. This was not a study of any one type of disfluency. As part of a larger study, Boehmler (*2*) obtained 66 five-second segments of recorded speech, each segment containing one instance of a part-word repetition. Twenty-three of these recorded segments of speech were from nonstutterers and 43 were from stutterers. The two groups of speakers from which these samples had been obtained had been equated previously on the basis of rated severity of disfluency for all types of disfluency combined. Using as his base measure the number of judges ($N = 30$) who labeled each segment of recorded speech as an example of stuttering (there were 400 segments in all representing many types of disfluency), Boehmler computed two medians, one for the samples of part-word repetition obtained from the nonstutterers and the other for the samples of part-word repetition obtained from the stutterers. He found that 17 of the 23 samples from the nonstutterers and 28 of the 43 from the stutterers exceeded the respective median value computed for all samples for each of the groups. The difference between these two proportions, 17:32 and 28:43, was not significant at the 5% level of confidence. In other words, part-word repetitions were classified by the judges as 'stuttering' with about the same relative frequency when they were performed by nonstutterers as when they were performed by stutterers. The samples of repetition had been removed from their original contexts, and the listeners were not aware of the classification, as stutterer or nonstutterer, of the speaker in each case. When set to evaluate some disfluencies as stutterings, the listeners were about as likely so to evaluate the part-word repetitions of the speakers generally reputed to be nonstutterers as those of the speakers reputed to be stutterers. In fact, Boehmler found that his listeners classified as stutterings 42.2% of the disfluencies (of all types) of the subgroup of nonstutterers whose disfluencies were given an overall severity rating of 'severe,' and they classified as stutterings only 34.6% of the disfluencies (of all types) of the subgroup of stutterers whose disfluencies were given an overall severity rating of 'mild.' Of all of the nonstutterers' 300 five-second speech sample segments containing disfluencies, 27.5% were classified as stutterings by Boehmler's 30 listeners; of the stutterers' 300 samples of disfluencies, 68.6% were classified as stutterings.

Generalizations from the findings of Johnson (*9*) are to be drawn with due emphasis on certain limitations. First, the speech samples used were matched only on the basis of total number of disfluencies, not on a rating of severity of disfluency or an equating of the frequencies of particular types of disfluency. In Boehmler's study (*2*), the matching of the samples from which the five-second segments were drawn was made on the basis of ratings of severity of disfluency for all types of disfluency combined, not for part-word repetitions alone.

Further comments are to be made with reference to studies by Tuthill (*15*), Boehmler (*2*), and Bloodstein, Jaeger, and Tureen (*1*), as well as Johnson's investigations of the origins of the problem called stuttering (*8*). From these studies the generalization may be drawn that the listener's response to disfluency is in considerable part determined by factors which are associated with the listener himself, and which are therefore relatively independent of the speaker's behavior. Does the listener consider himself to be a stutterer or a nonstutterer? Is the listener a speech clinician or a lay observer? If the listener is a speech clinician, at what institution did he receive his professional training? Is the listener a parent who is inclined to consider his child to be a stutterer or one who is disposed to regard his child as a nonstutterer? Factors such as these questions represent appear to influence the observer's perceptual and evaluative reactions to the speaker's disfluency.[2] Further investigation of the interaction between discernible features of the speaker's vocal utterance and the listener's perceptual and evaluative sets or tendencies may be expected to yield additional and more extended generalizations.

With respect to the matter of methodology, the detailed disfluency analysis used in this study, while possessing advantages for purposes of exploratory investigation, has certain limitations so far as various specific research purposes and general clinical use are concerned. Much time is needed, often several hours per subject, to obtain accurate transcriptions and identifications of the disfluencies in individual speech samples. The program of research represented by the studies reported in this Monograph has been developed in consideration of, among other things, the need for investigation of intra- and inter-observer agreement and the problem of adapting disfluency analysis procedures to given purposes. This research program is also concerned with the patterns of relation among measures of disfluency, and between measures of rate and disfluency, as well as the relation between measures of disfluency and rate, on the one hand, and, on the other, judged severity of stuttering. The disfluency analysis developed in the course of the present study has been evaluated and refined in the investigations reported in this Monograph by Sander (*12*) and Young (*17*).

[2]The interaction between the speaker and the listener as it may be considered to relate to the onset and development of the problem of stuttering has been more fully considered elsewhere by Johnson (*8*, pp. 236-264).

TABLE 10. Coefficients of correlation among Job, TAT, and Reading tasks for each of nine measures of disfluency and of rate in words per minute for 50 male and 50 female stutterers (S) and 50 male and 50 female nonstutterers (NS).

	Tasks Correlated					
	Job vs. TAT		*Job vs. Reading*		*TAT vs. Reading*	
Variables	*S*	*NS*	*S*	*NS*	*S*	*NS*
Rate	.86*	.26	.66	.03	.64	.17
Interjections	.90	.42	.76	.10	.80	.15
Part-word Repetitions	.94	.18	.82	.01	.82	.19
Word Repetitions	.83	.47	.39	.08	.45	.01
Phrase Repetitions	.62	.31	.25	.51	.21	.25
Incomplete Phrases	.17	.31	-.05	.00	-.03	.00
Revisions	.29	.44	.13	.28	.15	.23
Broken Words	.48	-.01	.58	-.07	.72	.32
Prolonged Sounds	.91	.25	.66	.05	.67	.15
Task Index	.93	.56	.67	.13	.75	.45

*For 98 *df*, $r = .26$ is needed for significance at 1% level, but the extreme skewness of most of the distributions involved suggests caution in interpreting the results.

Summary

In the present study normative data concerning rate and disfluency were obtained through analysis of tape-recorded samples of speaking and oral reading secured from 100 male and 100 female adult speakers, half of whom were regarded by themselves and by the investigators as stutterers and half as normally speaking nonstutterers. Measures of eight types of disfluency and of speaking and reading rate were presented for each group of subjects, together with relevant comparisons between main groups and sex subgroups, and between types of performance or task. The varying degrees of overlapping of the distributions of disfluency measures for the subjects classified as stutterers and those not so classified imply that the problem called stuttering is not to be adequately identified or defined solely by reference to speech disfluency, as such. It is suggested that variables associated with the perceptual and evaluative reactions of the listener and of the speaker, as well as those associated with the frequencies and forms of disfluency of the speaker, are to be included in an adequately comprehensive and systematic consideration of the problem called stuttering.

References

1. Bloodstein, O., Jaeger, W., and Tureen, J., A study of the diagnosis of stuttering by parents of stutterers and nonstutterers. *J. Speech Hearing Dis.*, 17, 1952, 308-315.
2. Boehmler, R. M., Listener responses to nonfluencies. *J. Speech Hearing Res.*, 1, 1958, 132-141.
3. Darley, F. L., A normative study of oral reading rate. M.A. Thesis, University of Iowa, 1939.
4. Duffy, R. J., A quantitative study of the speech nonfluencies of fifty adult female stutterers. M.A. Thesis, University of Iowa, 1957.
5. Fairbanks, G., *Voice and Articulation Drillbook*, first edition. New York: Harper & Brothers, 1940, 144.
6. Giolas, T. G., and Williams, D. E., Children's reactions to nonfluencies in adult speech. *J. Speech Hearing Res.*, 1, 1958, 86-93.
7. Goodman, L. A., Kolmogorov-Smirnov tests for psychological research. *Psychol. Bull.*, 51, 1954, 160-168.
8. Johnson, W., and associates, *The Onset of Stuttering: Research Findings and Implications*. Minneapolis: University of Minnesota Press, 1959.
9. Johnson, W., Brown, S. F., Curtis, J. F., Edney, E. W., and Keaster, Jacqueline, *Speech Handicapped School Children*, second edition. New York: Harper & Brothers, 1956.
10. Johnson, W., (Ed.), *Stuttering in Children and Adults*. Minneapolis: University of Minnesota Press, 1955.
11. Murray, H. A., *Manual for the Thematic Apperception Test*. Cambridge: Harvard University Press, 1943.
12. Sander, E. K., Reliability of the Iowa Speech Disfluency Test. *J. Speech Hearing Dis.*, Monograph Supplement 7, 1961, 21-30.
13. Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill, 1956.
14. Tate, M. W., and Clelland, R. C., *Nonparametric and Shortcut Statistics*. Danville, Illinois: Interstate Printers and Publishers, 1957.
15. Tuthill, C., A quantitative study of extensional meaning with special reference to stuttering. *Speech Monogr.*, 13, No. 1, 1946, 81-98.
16. Williams, D. E., and Kent, Louise R., Listener evaluations of speech interruptions. *J. Speech Hearing Res.*, 1, 1958, 124-131.
17. Young, M., Predicting severity of stuttering. *J. Speech Hearing Dis.*, Monograph Supplement 7, 1961, 31-54.

Reliability of the Iowa Speech Disfluency Test

ERIC K. SANDER

Stuttering behavior is observed clinically, and is reported by stutterers themselves, to be variable from day to day and from week to week.

Several authors have commented on the variability or intermittency of stuttering behavior. West (*11*) has stated that stuttering is characterized not only by 'the occurrence of sudden breaks in the automatic processes of articulate speech or of phonation for speech' but by 'more or less extended periods of freedom from these breaks, and obversely, clustering of these breaks in various social situations.' Johnson (*5*) has remarked:

> It is, of course, obvious that 'stuttering' in almost any meaning the term could be given, refers to a series of discrete events, or to a process or phenomenon that is for all practical purposes intermittent and definitely variable. The average stutterer has difficulty on something like 10 percent of his words, and half or more of his stutterings last one second or less; he has 'good' and 'bad' days; he stutters more in some situations than he does in others.

Berry and Eisenson (*2*) have stated:

> The amount and severity of stuttering varies considerably for the individual from time to time and according to the situation. . . . Almost all stutterers are relatively free from difficulty in some situations, of which they may or may not be aware. Even the most severe stutterers have many moments of fluency, and some may be fluent for days on end.

Naylor (*8*), studying 24 stutterers, observed that the individual 'stutterer's estimate of the severity of his stuttering while reading a short passage for recording did not consistently reflect his estimate of the severity of his stuttering for the preceding several months.' He also reported no statistically significant relationships between the stutterers' estimates of the severity of their stuttering for the preceding several months and judges' ratings of the severity of the subjects' stuttering from recordings. Perhaps findings such as Naylor's led Milisen (7) to conclude that observations of the speech behavior of stutterers 'need not involve highly accurate measurements of overt symptoms or attitudes, because the conditions change so markedly from one period to another and from one situation to another.'

The fact that stuttering behavior varies is of particular importance in research attempts now underway to quantify the results of stuttering therapy. Progress is seldom linear during the course of therapy. We select a particular cross-section of time, for example at the beginning or close of a semester, at which to administer a cer-

---

Eric K. Sander (M. A., University of Iowa, 1959) is Research Associate in Speech Pathology, Cleveland Hearing and Speech Center, Western Reserve University. This article is based on the author's M. A. thesis completed under the direction of Professor Wendell Johnson. The research was supported, in part, by the Louis W. and Maud Hill Family Foundation and by Grant RD 319 of the Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

tain test or a battery of tests. Unresolved is the degree of variability that might be observed in the stuttering behavior of an individual were it possible to observe him periodically, such as at hourly, weekly, or monthly intervals, in a particular situation, to say nothing of behavior changes that would undoubtedly become evident in different situations.

Various methods have been used to evaluate the speech improvement made by stutterers in the therapy program at the University of Iowa Speech Clinic: ratings by acquaintances, clinicians' ratings, and self-ratings of severity of stuttering; speech situation rating sheets; speaking time records; adaptation and consistency tests; etc. There is a need for more descriptive procedures that will bypass the relative subjectivity of rating scales. Trotter and Kools (*10*) speculated that much of the improvement reported by clinicians was a result of change in the listener rather than any change in the stutterer's speech. They reported a systematic listener adaptation effect to stuttering with repeated stimulation, and concluded that if estimated improvement were 'based on the clinician's memory of how severe the case was at the beginning of the therapy period . . . probably every speech case would show some improvement.'

A procedure that has recently come to be known as the Iowa Speech Disfluency Test (*4*) is adaptable to the problem of evaluating the clinical progress of stutterers. The test involves the eliciting of a recorded sample of reading and spontaneous speech. The recording is then replayed as often as necessary to determine and classify the speaker's disfluencies. The final test result is usually expressed in terms of the number of disfluencies per one hundred words spoken or read by the speaker.

Johnson (*4*) has determined norms for the Iowa Speech Disfluency Test for a group of stutterers and nonstutterers. Thus it is possible to express a speaker's performance in terms of his decile rank, that is, the particular decile at or below which he lies, for a population of stutterers or for a population of nonstutterers.

Problem

The present study is concerned with establishing the test-retest reliability for both the reading and speaking tasks of the Iowa Speech Disfluency Test administered to a group of stutterers. The question asked is: How consistent are the disfluencies of stutterers on these two tasks over an interval of 24 hours with the speaking situation held relatively constant?

A measure of temporal stability indicates, according to Anastasi (*1*), 'the degree to which scores on a given test are affected by the random daily fluctuations in the conditions of the subject or of the testing environment.' For example, a subject at the onset of therapy might rank in the eighth decile in comparison with other stutterers with respect to the speaking task of the Iowa Speech Disfluency Test. When the task is repeated at the close of the semester the individual's performance may have shifted to the sixth decile. Unless the reliability of such a task is established we have no means of gauging the significance of such a change. With what confidence, in other words, can we assume that discrepancies of any given magnitude between two speaking performances on a given task are not attributable to chance? According to Anastasi, 'short-range, random fluctuations which occur during intervals ranging from a few hours to a few months are generally included under the error variance of the test score. Thus in checking this type of test reliability, an effort is made to keep the interval as short as feasible.' Changes over long periods of time, that is, over six months, 'are apt to be cumulative and progressive rather than entirely random.'

Experimental Procedure

Subjects. The subjects were 40 stutterers, all but one participating in therapy at the University of Iowa Speech Clinic. Thirty-four of the stutterers were males, six were females. They ranged in age from 17 to 37 years with a mean age of 22.6 years and a median age of 21.5 years. All but two of the subjects were university students at the time the recordings were made.

General Procedure. There was a 24-hour interval between the initial and subsequent administration of the reading and speaking tasks. Such an interval was considered to be sufficiently long to eliminate any decrease in stuttering due to familiarity with the reading material, yet short enough to obviate any genuine improvement or changes due to therapy. It was thought that the formulation of the initial spontaneous speaking task could conceivably result in a reduction in stuttering with reference to the material in the second session. Therefore, for purposes of this experiment the speaking task was subdivided into separate units for each session in order to assure spontaneity: (a) description of *past* jobs held by the stutterer and (b) description of a *future* job for which the stutterer is preparing or an ideal job he would like to hold. The reading passage used was the *Test Passage for Measurement of Reading Rate* (*3*). The same reading passage was used for both sessions. The speaking situation for both days was held constant. The tape recorder was placed in full view of the subject. The experimenter was the only other person present in the room during the recording. The order of the speaking task and the reading task as well as the suborder of the two speaking tasks was counterbalanced.

Disfluency Analysis. Disfluencies were identified from a verbatim transcript made for each speaking task. Exactly 250 words of the speaking task, excluding word and phrase repetitions and revisions, and the complete 300-word reading passage were scored for disfluencies. Each recording was replayed at least three times, usually more often, to determine and classify the disfluencies. A total of 10,150 disfluencies were classified. After the disfluencies for a particular task had been marked, but not counted, the reading passage or transcript was not referred to again until the complementary task from the other session had been completed. A minimum interval of 24 hours was established between the analysis of a reading or speaking task for one session and that of another in order to minimize any possible bias due to memory of the speaker's previous performance. The order in which the recordings of a particular task were analyzed, for example session one first, then session two, was alternated. For all subjects the reading tasks were analyzed first.

The following disfluency categories were employed: (a) interjection of syllables, sounds, words, or phrases; (b) sound and syllable repetitions; (c) repetition of words; (d) repetition of phrases; (e) revisions; (f) incomplete phrases; (g) broken words; and (h) prolonged sounds.[1] In addition to these categories of disfluency, for each of which the total number of disfluencies was tabulated, two other measures, number of disfluent words and rate of utterance, were used in this experiment.

Number of Disfluent Words. A word was considered to be disfluent if it involved prolonged sounds, was classified as a 'broken word,' was involved in a sound, syllable, or word repetition, or was interrupted by an interjection. Words preceded by interjections or involved in phrase repetitions were not counted as disfluent words.

Rate of Utterance. The total time taken to read the 300-word passage or to speak 250 fluent words in the speaking task was determined. In a few instances some prompting was necessary during the speaking task. When such prompting occurred the timing was stopped with the last words uttered prior to the prompting and resumed with the first words uttered following the prompting. The experimenter interjected promptings whenever pauses of longer than 10 to 15 seconds occurred.

Self Scoring Reliability. Twelve reading and 12 speaking tasks, selected at random, were rescored for total number of disfluencies to determine the reliability of the experimenter's analysis. In each case a minimum of one month elapsed between the original analysis and the rescoring. Only one of the two sessions of a particular task was rescored for any individual; that is, if one of the reading or speaking tasks was rescored, the second was arbitrarily excluded from consideration. For two subjects both a reading and a speaking task were rescored. Thus a reliability check was obtained on at least one task for 22 of the 40 subjects in this experiment.

Results and Discussion

Self-Agreement in Scoring. The formula used to establish scoring reliability for total disfluencies was Agreement Index $= a/(a + d)$ in which $a =$ agreements and $d =$ disagreements (the discrepancy in total disfluencies between the original and rescored task). For the 12 reading tasks there was a total of 557 agreements and 23 disagreements; for the speaking tasks, 767 agreements and 35 disagreements. In both instances the coefficient of agreement was .96. These results are for total number of disfluencies only and do not indicate the extent of agreement for the individual disfluency categories nor for the occurrences of particular disfluencies.

Group Differences Between Test and Retest Sessions. The means, standard deviations, and ranges for total disfluencies, disfluent words, and time are presented in Table 1. The range in total number of disfluencies for both sessions was from 8 to 255 on the 250-word speaking task and from 0 to 270 on the 300-word reading task.

Test-Retest Correlations. Test-retest Pearson product-moment correlation coefficients were computed for the reading and speaking tasks repeated after 24 hours by 40 subjects. The test-

[1] See Johnson (*4*) for a more complete description of these disfluency categories.

TABLE 1. Means, standard deviations, and ranges of total disfluencies, disfluent words, and time for Sessions I and II of the Reading and Speaking Tasks (N = 40).

	Speaking			Reading		
	Mean	*SD*	*Range*	*Mean*	*SD*	*Range*
Total Disfluencies						
Session I	78.1	55.4	8- 230	48.2	64.3	0- 270
Session II	81.2	56.6	15- 255	46.0	54.7	1- 210
Disfluent Words						
Session I	38.2	33.4	1- 134	29.0	38.7	0- 184
Session II	40.0	32.0	3- 122	29.1	35.4	0- 160
Time in Seconds						
Session I	242.4	207.3	86-1235	218.4	206.9	89-1028
Session II	237.7	167.4	92- 853	198.7	141.0	85- 782

retest coefficients of correlation for the reading and speaking tasks, respectively, were .94 and .91 for total number of disfluencies and .97 and .94 both for number of disfluent words and for time.[2]

In reporting these rather high correlations attention should be directed not only to the short interval (24-hours) between test and retest, but to the wide range and variation in disfluent behavior displayed by the subjects. On the reading tasks, for example, the standard deviations of disfluency exceeded the mean values (Table 1).

Individual Changes after 24 Hours. Johnson (*4*) has reported decile values of measures of disfluency and rate of utterance for 50 male and 50 female stutterers of college age. Each subject in the present experiment was placed in a 'decile group' with respect to the norms established by Johnson. For example, male subjects whose disfluencies fell at or below the first decile, according to Johnson's norms for male stutterers, were placed in the first decile group; subjects whose disfluencies fell at or below the second decile but above the first were placed in the second decile group, etc. Tables 2 and 3 indicate the number of subjects in each decile group for the speaking and reading tasks. The numbers of subjects within each decile group showing decile changes for the second session and the averages (means) of the decile changes for each group are also listed.

For both reading and speaking tasks approximately half of the stutterers (47.5% for speaking, 52.5% for reading) showed no changes in decile placement for tasks repeated after 24 hours. Extremely mild or severe stutterers seemed to show the least changes. On the reading task, 15 of the 19 subjects belonging to decile groups three through seven, inclusive, shifted in decile placement, but only four of the remaining 21 subjects, the mild or severe stutterers, showed any shifts. For the speaking task, 13 of the 18 subjects from the third to the seventh decile groups, inclusive, showed shifts in their decile placement while only eight of the remaining 23 subjects, the mild and severe stutterers, showed changes.

[2]For *df* = 38, *r* = .40 is significant at the 1% level of confidence.

TABLE 2. The number of subjects at or below a particular decile for Speaking Session I according to the disfluency norms established by Johnson (*4*), the number of subjects within each group showing decile shifts for Speaking Session II, and the average (mean) decile shift for each group for the Speaking Task.

Decile	*Number of Subjects*	*Number of Subjects Showing Shifts*	*Average Decile Shift (each group)*
1	3	0	0.0
2	4	2	0.8
3	1	1	1.0
4	2	1	1.0
5	6	5	1.8
6	6	3	1.0
7	3	3	1.0
8	5	2	0.4
9	5	3	1.0
10	5	1	0.2

TABLE 3. The number of subjects at or below a particular decile for Reading Session I according to the disfluency norms established by Johnson (*4*), the number of subjects within each group showing decile shifts for Reading Session II, and the average (mean) decile shift for each group for the Reading Task.

Decile	*Number of Subjects*	*Number of Subjects Showing Shifts*	*Average Decile Shift (each group)*
1	7	0	0.0
2	3	2	1.0
3	6	5	1.8
4	6	4	1.2
5	1	1	1.0
6	4	3	1.3
7	2	2	1.0
8	4	1	0.3
9	1	0	0.0
10	6	1	0.2

The conclusion may be drawn that very mild or severe stutterers are more consistent in their stuttering behavior; however, the apparent consistency of the severe and mild groups was due in part to the relatively greater numbers of disfluencies separating adjacent higher and lower deciles within the relevant segments of the distribution.

Approximately one-third of the subjects (32.5% for speaking, 30% for reading) changed one decile in their grouping from the first to the second session. Thus a combined total of 80% of all subjects fell within one decile of their original grouping when reading and speaking tasks were repeated after 24 hours. The greatest shift, recorded for one subject (2.5%), was four deciles on the reading task. Although no shifts greater than three deciles were evident on the speaking task, five subjects (12.5%) showed shifts of three deciles. The percentages of stutterers falling in other categories of decile shift are: for reading, 12.5% in the two and 2.5% in the three decile shift groups; for speaking, 7.5% in the two decile shift group.

The extent and direction of change in decile grouping for both tasks can be seen from an examination of Figures 1 and 2. Both of these frequency polygons appear to be roughly symmetrical, indicating that approximately the same number of stutterers moved up in decile ranking as moved down, and that the magnitude of the changes in both directions was also about the same.

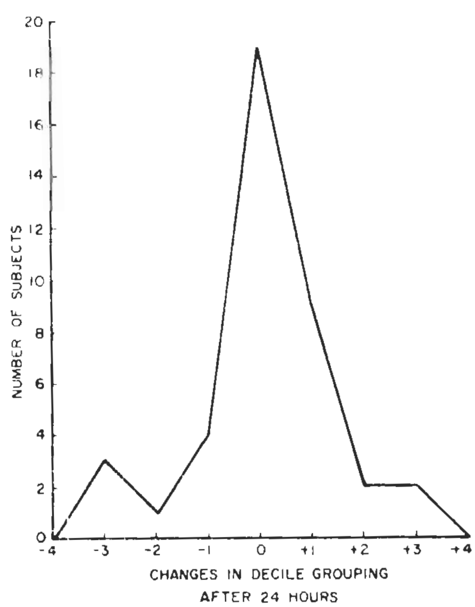

FIGURE 1. The extent and direction of change in decile placement (according to Johnson's (*4*) norms) for test-retest (24-hour interval) performance of 40 subjects on the Speaking Task of the Iowa Speech Disfluency Test. (Negative and positive changes indicate lower and higher decile placements, respectively, on the second test than on the first.)

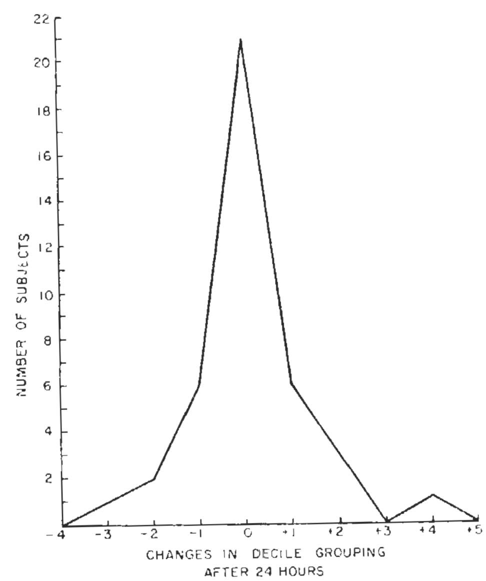

FIGURE 2. The extent and direction of change in decile placement (according to Johnson's (*4*) norms) for test-retest (24-hour interval) performance of 40 subjects on the Reading Task of the Iowa Speech Disfluency Test. (Negative and positive changes indicate lower and higher decile placements, respectively, on the second test than on the first.)

The sample of stutterers in the present experiment did not differ substantially from the normative group studied by Johnson; they tended to be slightly less fluent on the speaking task, but more fluent on the reading task. It was possible to arrive at independent decile values from the present sample and to compute decile shifts accordingly; however, the final results for changes in total disfluency were virtually the same as those already reported: half of the subjects showed no shifts (45% for speaking, 52.5% for reading) and over four-fifths of the subjects (82.5% for speaking, 85% for reading) showed changes not greater than one decile. Using values from the present sample, 32 of 40 subjects changed not more than one decile in their reading rate, and 29 of 40 subjects remained within one decile of their speaking rate when these tasks were repeated 24 hours later.

Intercorrelations Among Measures of Stuttering Severity and Between Reading and Speaking. For the reading task (Session I) the total number of disfluencies correlated .86 with disfluent words and .86 with rate. For the speaking task (Session I) total disfluencies correlated .87 with disfluent words and .81 with rate. The correlation coefficients (reading vs speaking) for total disfluencies and disfluent words were .72 and .70, respectively. A correlation coefficient of .90 was found between reading rate and speaking rate.[3]

The significance of these coefficients is to be viewed in the light of the need for streamlining the Iowa Speech Disfluency Test as a practical clinical tool. The present method of determining and classifying the total disfluencies of a subject is tedious and time-consuming. A simple counting measure, such as number of disfluent words together with a time measure, seems to be an adequate substitute, as an index of improvement, for total disfluencies. A significant fact emerging from the data already reported is the unusual stability of rate as a measure of stuttering severity. Test-retest correlations for rate of .97 for reading and .94 for speaking were found. Rate was shown to correlate .86 with total disfluencies for reading and .81 for speaking. In addition, nearly all subjects in this study who underwent substantial speaking or reading changes from Session I to Session II showed corresponding shifts in rate. Measures of rate thus appear to be of value insofar as they reflect changes which correspond to individual speech improvement or relapse.[4]

Changes in the reading fluency of subjects after 24 hours were not always accompanied by concomitant changes in speaking fluency or vice versa. For example, one subject showed a decrease in total speaking disfluencies from 120 to 55, yet his reading performance remained stable. Another subject increased his speaking disfluencies from 59 to 131 while his reading disfluencies changed only slightly. For reading, one subject showed a drop after 24 hours from 270 to 167 disfluencies but an increase in speaking disfluencies from 230 to 255. Another subject showed a dramatic rise from 9 to 52 in his reading disfluencies while his speaking performance remained stable.

These findings in part confirm the observation that those stutterers in this study who showed the greatest decile shifts in reading or speaking from one day to the next were apparently not influenced by gross situational factors or cyclic changes in their emotional mood.[5] For want of a better term, we

---

[3]For $df = 38$, $r = .40$ is significant at the 1% level of confidence.

[4]What we label 'improvement' constitutes a a clinical judgment. Kent and Williams (*6*) rightly stress that '"improvement" for one individual may involve changes that are opposite in direction to changes which may constitute "improvement" for another individual.' For severe stutterers, at least, speaking rate may reflect a more significant clinical dimension of improvement than a count of total disfluencies.

[5]Quarrington (*9*) notes the frequent occurrence, particularly among younger stutterers, of cyclical variations in stuttering frequency superimposed upon situational fluctuations. He states: 'The length of these periods is reported by stutterers as being regular, although varying from individual to individual from several weeks to about two months.'

might label as 'random' the observed changes in their stuttering behavior. Information regarding the extent and prevalence of cyclic variations in stuttering would require an extended series of observations uncomplicated by the effects of therapy.

Summary

The primary purpose of this study was to investigate the temporal reliability of the Iowa Speech Disfluency Test and its general usefulness as a tool for evaluating the speech improvement of stutterers following therapy.

The Iowa Speech Disfluency Test was administered to a group of 40 college-age stutterers and repeated 24 hours later. Self-agreement of the experimenter in scoring total disfluencies was found to be adequate. Test-retest Pearson product-moment correlation coefficients of .91 and .94 were found for total disfluencies on the speaking and reading tasks, respectively. These results seem to indicate rather high temporal reliability, although the extreme variation in disfluency exhibited by the representative sample of stutterers used in the present study undoubtedly served to account in part for the magnitude of the obtained correlation coefficients. More detailed attention might be given to the test performance of homogeneous subgroups of stutterers. In addition the question of cyclical and situational sources of stuttering variability is deserving of further study.

Subjects in the present experiment were classified in decile groups according to disfluency norms established by Johnson (*4*). Approximately one-half of the stutterers showed no changes in decile grouping for reading and speaking tasks repeated after 24 hours. Shifts no greater than one decile were reported for four-fifths of the subjects. Extremely mild or severe stutterers seemed to show the least changes in decile placement after 24 hours.

Intercorrelations between reading and speaking tasks were .72 for total disfluencies, .70 for disfluent words, and .90 for rate. Coefficients of correlation between total disfluencies and disfluent words, as defined, were .87 for speaking and .86 for reading. Total disfluencies and rate correlated .81 for the speaking task and .86 for the reading task. These results suggest the feasibility of employing, for certain purposes, a relatively simplified method of disfluency analysis consisting of the counting of disfluent words, as defined in this study, and the measurement of rate of utterance.

References

1. Anastasi, Anne, *Psychological Testing*. New York: The MacMillan Co., 1954.
2. Berry, Mildred and Eisenson, J., *Speech Disorders: Principles and Practices of Therapy*. New York: Appleton-Century-Crofts, Inc., 1956.
3. Fairbanks, G., *Voice and Articulation Drillbook*. New York: Harper & Brothers, 1940.
4. Johnson, W., Measurements of oral reading and speaking rate and disfluency of college-age male and female stutterers and nonstutterers. *J. Speech Hearing Dis.*, Monograph Supplement 7, 1961, 1-20.
5. Johnson, W., Perceptual and evaluation factors in stuttering. Chapter 28 in Travis, L. E., *Handbook of Speech Pathology*, New York: Appleton-Century-Crofts, Inc., 1957.
6. Kent, Louise R., and Williams, D. E., Use of meprobamate as an adjunct to stuttering therapy. *J. Speech Hearing Dis.*, 24, 1959, 64-69.
7. Milisen, R., Methods of evaluation and

diagnosis of speech disorders. Chapter 8 in Travis, L. E., *Handbook of Speech Pathology*. New York: Appleton-Century-Crofts, Inc., 1957.

8. Naylor, R., A comparative study of methods of estimating the severity of stuttering, *J. Speech Hearing Dis.*, 18, 1953, 30-37.
9. Quarrington, B., Cyclical variation in stuttering frequency and some related forms of variation. *Canadian Journal of Psychology*, 1956, 10, 179-184.
10. Trotter, W., and Kools, J., Listener adaptation to the severity of stuttering, *J. Speech Hearing Dis.*, 20, 1955, 385-387.
11. West, R., Ansberry, M., and Carr, Anna, *The Rehabilitation of Speech*, New York: Harper & Brothers, 1957.

Predicting Ratings of Severity of Stuttering

MARTIN A. YOUNG

Severity of stuttering is a useful and important construct when considered as an interaction between a speaker and a listener. The research procedures to be described were oriented to this hypothesis. The tentative goal was to estimate the accuracy with which a rating of severity of stuttering could be predicted from an analysis of the fluency aspects of tape-recorded samples of the spontaneous speech of persons with the problem of stuttering. A secondary aim was to develop a procedure for estimating the severity of stuttering that would have research and clinical utility.

Relevant research was reviewed; and the following summary is provided to call attention to certain findings.

(a) Speech disfluency categories have been modified and refined since the early studies of fluency were performed. The categories selected for examination in the present investigation, based on categories employed in previous research, were considered to characterize best those aspects of speech disfluency to which the listener might respond (*4*, *5*, *6*, *8*, *11*, *12*, *13*, *15*, *21*).

(b) A listener's rating of a speaker's disfluency with respect to severity of stuttering represents, in large part, an evaluation and not a description of that behavior, although certain variations in type and degree of disfluency do appear to be associated with concomitant evaluatory changes. (*1*, *3*, *10*, *16*, *17*, *18*, *19*, *20*).

(c) Measures of observer reliability in classifying and counting disfluencies of various types and speaker consistency in performing the speaking task to be employed in the present study have been demonstrated to be satisfactory enough to allow the use of similar experimental methods of measurement in the present analysis (7, *16*,).

Problem

The present study was designed to suggest answers to the following questions:

(a) With what precision can ratings of severity of stuttering be predicted from specified measures of rate and disfluency?

(b) What is the relative importance of frequencies of occurrence of different types of disfluency in predicting ratings of severity of stuttering?

(c) To what degree may the meas-

Martin A. Young (Ph.D., University of Iowa, 1960) is Coordinator of Research, Cleveland Hearing and Speech Center, and Assistant Clinical Professor of Speech Pathology and Audiology, Western Reserve University. This article is based on a doctoral dissertation completed under the direction of Professor Wendell Johnson. The research was supported, in part, by the Louis W. and Maud Hill Family Foundation and by Grant RD 319 of the Office of Vocational Rehabilitation, Department of Health, Education, and Welfare.

urement of disfluency be simplified through reduction in the number of categories of disfluency while retaining enough information to allow for reliable predictions of ratings of severity of stuttering?

(d) To what extent do listeners agree in ratings of severity of stuttering?

Experimental Operations and Results: Part 1

Figure 1 represents the model employed in the present research for estimating accuracy in predicting ratings of severity of stuttering. Each italicized item is meant to indicate 'a fact on record' or an event, while the items in parentheses represent the experimental operations performed in moving from one 'level' of event to another. The direction of movement from level to level is shown by the arrows. No two levels are the same, of course, and the experimental operations tend to focus on and select from certain aspects of any one level, ignoring all the rest of the details. The event on each level results from the experimental operations performed on the event shown directly above it on the model. For this reason the order of presentation of the present report will follow the outline shown in Figure 1, and the experimental operations and results will be presented as a single unit. Events 17 and operation 18, as shown in Figure 1, lead to the event labeled 1′, the modifications of the original assumptions. The 'Etc.' at the bottom of Figure 1 is intended to indicate that the process described in the model is potentially continuous.

1. *Assumptions.* In the present research the following working assumptions were employed: (a) the estimate or rating of severity of stuttering involves an interaction between a speaker and a listener, and (b) a listener's evaluation of severity of stuttering is associated to a considerable degree with certain measurable dimensions of a speaker's fluency.

2. *Speakers.* The predicted ratings of severity of stuttering made in this study are to be considered with reference to a hypothetical population of speakers composed of males (a) attending college, (b) considered by themselves and their speech clinicians to have the problem of stuttering, and (c) participating in speech therapy. This limited hypothetical population of speakers excludes a substantial number of persons with the problem of stuttering, such as female speakers, children and adolescents, persons not attending college speech clinics, etc. Generalizations drawn from the data obtained are to be interpreted accordingly. This particular population of speakers represented the largest homogeneous group of subjects available to the experimenter.

3. *Selection of Speakers.* Thirty-seven of the speakers used in Part 1 of the present study were originally used as subjects in a fluency study of college-age speakers (*11*). These subjects were selected for use in the present experiment on the basis of (a) quality of the tape recordings available, (b) length of the speech samples recorded as measured by the number of words spoken, and (c) the desire of the experimenter to secure a representative range of stuttering severity. From speakers who were locally available,

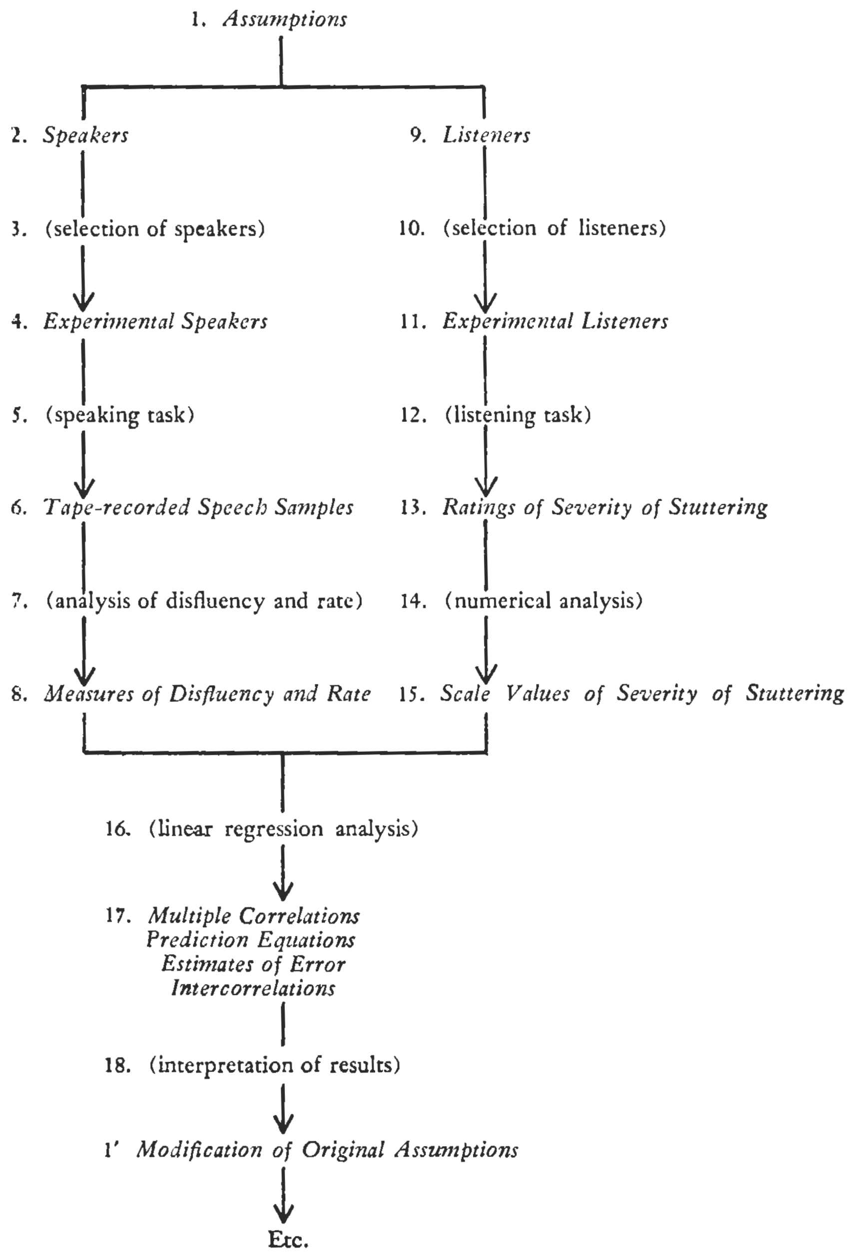

FIGURE 1. Model employed in the present research for estimating accuracy in predicting ratings of severity of stuttering.

an additional 13 subjects were selected who, on the basis of the experimenter's judgment, exhibited relatively more severe stuttering than was exhibited by most of the 37 subjects mentioned above. All 50 speakers performed the same experimental task under essentially identical conditions.

4. *Experimental Speakers.* The 50 experimental speakers were considered by themselves and by their speech clinicians to have the problem of stuttering. They were all college-age males attending Midwestern universities and participating in speech therapy when the tape recordings of samples of their speech were made. The 13 speakers selected by the investigator were generally similar in age and academic attainment to the other 37 speakers who were used in the fluency study reported by Johnson (*11*).[1]

5. *Speaking Task.* The Job task, as described by Johnson (*11*), was employed to elicit the tape-recorded samples of speech. This task was performed as follows:

Each subject was seated in full view of the recording equipment, the tape recorder was turned on, and an informal interview was begun. This subject was asked for appropriate identifying information such as name, age, education, and marital status. Included in the interview were questions pertaining to previous experience with tape recorders and present course of study in college. The purpose of the interview was to familiarize the subject with the experimental situation. The recorder was turned off and the subject was told to talk about his future job or vocation or about jobs he had held in the past. It was suggested that he talk for at least three minutes. The subject was allowed one minute to think about what he might say. The recorder was then turned on and the subject was asked to begin speaking. Leading questions were asked whenever the subject appeared unable to think of anything further to say. When the experimenter was satisfied that a speech sample of sufficient length had been recorded, the subject was asked to discontinue speaking, the recorder was turned off, and the subject was dismissed.

6. *Tape-recorded Speech Samples.* The 50 tape-recorded samples of speech were edited to remove the experimenter's voice and any unduly long pauses which appeared to be due to the subject's inability to think of anything to say. Each sample as prepared for analysis was 200 words long, this length being fixed by taking the first 200 words that would have been spoken had the speaker performed no disfluencies counted in the present research.

The 50 tapes were then randomized and combined into a single test tape. The instructions for the listening task as well as 15-second pauses between adjacent samples were added at this time. The 50 samples of speech, together with the instructions and pauses, were approximately two hours and 15 minutes in duration.

Low recording level, distortions, and some background noise, were apparent when listening to several of the 50 tape recordings. It was believed, however, that adverse effects due to using these

[1]For 45 of Johnson's 50 male speakers with the problem of stuttering the age range was 16 to 24 years and the mean age was 19.6 years. Exact age data for five of the male stutterers employed in this study were not available.

particular tape recordings were negligible in view of the use to which they were to be put. Measurements of agreement among listeners evaluating these tapes substantiated this opinion.

7. *Analysis of Disfluency and Measurement of Rate*. Time. This measure was the duration, in seconds, of each tape-recorded speech sample.

The categories employed by the experimenter to characterize the disfluencies he observed while listening to the 50 tape-recorded samples of speech were as follows:[2]

Interjections. This category included interjected sounds, syllables, words, or phrases which were clearly distinct from the context.

Part-word Repetitions. Repetitions of sounds or syllables or any other parts of words were placed in this category.

Word-phrase Repetitions. Repetitions of whole words, including words of one syllable, or of phrases (two or more words) were counted in this category.

Prolongations. Sounds or parts of words that were prolonged, broken words, and words spoken with unusual stress were classified together as prolongations.

Revisions. This category represented disfluencies in which either the content or the grammatical construction of a phrase or sentence was modified, or in which formulation of the statement or remark that had been started was not completed.

If more than one type of disfluency were performed during the production of a single sound or syllable, an event which usually occurred as a combination of a prolongation and a part-word repetition, only one moment of disfluency was tallied, and categorization was made in accordance with the listener's judgment as to which disfluency type appeared to be more predominant.

✓ The decision to modify Johnson's disfluency classification scheme, using five rather than eight categories, was made in view of the rarity of occurrence of phrase repetitions, incomplete phrases, broken words, and prolonged sounds, and the difficulty occasionally experienced in discriminating between broken words and prolonged sounds, and between revisions and incomplete phrases. Accordingly, the categories designated as revisions and as prolongations in the present study represented a combination of revisions and incomplete phrases and of prolonged sounds and broken words, respectively. In addition, it was felt that a measure of rate was necessary which would take into account pause time and other aspects of fluency to which the listener might respond in rating severity of stuttering, but which are represented quite indirectly or not at all in measurements of the frequency of disfluencies. This variable was labeled 'time.'

The procedures used in making the fluency analyses of the 50 tape-recorded speech samples were essentially identical. First a verbatim transcript was made and then the disfluencies were identified and classified. No limit was set on the amount of replaying of each tape-recorded sample; the decision as to when sufficient accuracy had been obtained rested on the judgment of the investigator. To estimate the experimenter's self-agreement in making the fluency analysis, 10 tapes were selected at random and a second independent

[2] See Johnson (*11*) for more elaborate descriptions of disfluency categories as well as representative examples.

fluency analysis of each of these was performed. A minimum of two weeks and a maximum of three months separated the two analyses. Since the total sample of 50 tapes contained well over 10,000 words, with 2615 disfluencies identified during the first analysis, it seemed reasonable to assume that memory did not play a significant role in re-analyzing the 10 randomly selected tapes.

Indexes representing the self-agreement of the investigator in identifying and classifying disfluencies in the 10 tapes ranged from .91 to 1.0 (perfect self-agreement) as computed for individual samples, while for all 10 samples combined the indexes were .97 for observing occurrences of disfluency, .99 for the categorizing of types of disfluency (where there was agreement that a disfluency had occurred), and .97 for observing both types and occurrences of disfluency at the same time.[3]

8. *Measures of Disfluency and of Rate.* For each of the 50 speakers there were five measures of disfluency, a frequency count of the observed occurrences of each of five types of disfluency, and a measure of speaking time. A summary of these measures, together with a measure of the total frequency of all types of disfluency, in terms of mean frequency, range, and standard deviation, is to be found in Table 1.

The average speaker performed 52.3 disfluencies of all types per 200 words with interjections accounting for the largest proportion, 36.7%, of that total. Revisions were performed least often, accounting for only 5.1 % of the total number of disfluencies. Of some interest is the range of disfluent behavior of the speakers. The most fluent speaker performed only four and the least fluent speaker 223 disfluencies.

9. *Listeners.* No precise attempt was made to define the hypothetical population of listeners to which the results of the present study were to be generalized. Data obtained by Tuthill (*19*), Boehmler (*3*), and Bloodstein, Jaeger, and Tureen (*2*), would seem to indicate that an evaluation of severity of stuttering depends upon certain characteristics and attitudes associated with the listener. For this reason it was considered desirable to secure groups of listeners whose evaluations of severity of stuttering might differ. The decision was made, therefore, to select speakers with the problem of stuttering, speech clinicians, and laymen generally regarded as normal speakers to represent the heterogeneous population of listeners. The experimental operations described below that were employed in selecting the listeners serve to define the hypothetical population from which these listeners were drawn.

10. *Selection of Listeners.* Listeners were obtained from three subgroups of subjects; listeners in Group 1 were to be persons with the problem of stuttering, those in Group 2 were to be speech clinicians, and those in Group 3 were to be laymen generally reputed to have normal speech. The experimenter made a request for volunteers from the students enrolled in the course entitled Introduction to Speech Pathology and Audiology at the Uni-

[3]Reliability of the investigator in making the fluency analysis was estimated by means of the formula $C / \sqrt{xy}$, in which C = number of items showing agreement for both analyses and xy = product of the number of items identified independently in analysis 1 and analysis 2.

versity of Iowa and from adults with the problem of stuttering attending the University of Iowa Speech Clinic. The student clinicians enrolled in the Stuttering Practicum were also asked to participate. Scheduled times were announced for the listening sessions, one such session for each subgroup, and the subjects who presented themselves at these listening sessions became the experimental listeners.

11. *Experimental Listeners* (L1). Forty-eight subjects, divided into three subgroups, served as the experimental listeners in Part 1. Group 1 (G1) consisted of 16 persons with the problem of stuttering who were attending the Speech Clinic in the University of Iowa; Group 2 (G2) consisted of 13 student clinicians; and Group 3 (G3) was made up of nonstuttering students not majoring in speech pathology who were selected from the introductory course in speech pathology and audiology at the University of Iowa. The 19 students in Group 3 constituted the 'lay' subgroup of listeners in the experiment. All listeners in G1 and G2 were engaged in speech therapy, as clients and clinicians, respectively, when the listening sessions took place. Both G1 and G2 had had some experience with equal-appearing intervals scales prior to the experimental listening sessions, while G3 was considered to be essentially naive with respect to scaling procedures of this type.

12. *Listening Task.* Three separate listening sessions were held, one for each of the three listener groups. The listeners in each session attended to the 50 tape-recorded speech samples, taking four short rest periods while the experimenter changed the test tapes, and rated the severity of stuttering of each sample on the basis of a nine-point equal-appearing intervals scale.[4] Two short samples of speech were chosen to represent the extremes on this scale and were incorporated as part of the instructions to the raters. These segments were chosen from the tape-recorded speech samples of the two speakers who exhibited the highest and lowest total numbers of disfluencies.

13. *Ratings of Severity of Stuttering.* Twenty-four hundred ratings of severity of stuttering were obtained from the 48 experimental listeners. The mean rating of severity of stuttering for all 48 listeners was 3.84, with the means for G1 (stutterers), G2 (clinicians), and G3 (laymen) being 4.01, 3.88, and 3.68, respectively. The group differences in rating severity of stuttering were evaluated by means of analysis

---

[4]*Instructions for Raters:* You are about to hear some samples of speech from speakers who consider themselves to be stutterers. Your task will be to make an over-all rating of the severity of stuttering in each sample. You will do this on a nine-point scale on which a rating of 1 means no stuttering. (You may or may not judge any given sample to contain stuttering.) A rating of 2 means very mild stuttering and a rating of 9 means very severe stuttering. A rating of 5, in the middle of this scale, indicates an average severity of stuttering. The other values on the scale, 3, 4, and 6, 7, 8, represent equal intervals between these scale points. Before you start, however, you will hear a few segments of speech arbitrarily chosen to represent extremes on this nine-point scale. Now we will begin the judging procedure. Remember, a rating of 1 means no stuttering and a rating of 9 indicates very severe stuttering, while a rating of 5 is to be considered a middle point between these two extremes. *Please be sure to rate every sample, giving only one rating.* Each sample is numbered, and the number of each sample will be announced just before the sample is presented. Write your rating in the space provided in each case. You will be allowed a short pause between each sample to record your rating. Are there any questions?

TABLE 1. Means, ranges, and standard deviations of measures of disfluency and rate based on 200 words spoken by 50 adult males with the problem of stuttering.

Disfluency Category	*Mean Number of Disfluencies*	*Percentage of Total Disfluencies*	*Range of Number of Disfluencies*	*SD*
Interjections	19.2	36.7	2- 95	16.8
Part-word Repetitions	14.4	27.5	0- 79	16.1
Word-phrase Repetitions	5.5	10.6	0- 28	5.1
Prolongations	10.5	20.0	0- 70	15.5
Revisions	2.7	5.1	0- 8	1.8
Total*	52.3	100.0	4-223	40.8
	Mean Number of Seconds	*Range of Number of Seconds*	*SD*	
Time	134.0	61-595	82.0	

*Sum of measures for all disfluency categories.

of variance, and the results of this analysis are to be presented later.

14. *Numerical Analysis.* Listener agreement in evaluating severity of stuttering with respect to a nine-point equal-appearing intervals scale was estimated by means of intraclass correlations.[5] The coefficients for G1, G2, G3, and for all 48 listeners (L1) considered as a single group were, respectively, .79, .83, .87, and .83.

15. *Scale Values of Severity of Stuttering.* Two hundred scale values of severity of stuttering were computed, four for each of the 50 tape-recorded speech samples, each scale value being the median rating by one of the three separate listener groups and the fourth scale value being the median of the ratings of all 48 listeners. The mean of these median ratings, or scale values of severity of stuttering, for G1, G2, G3, and L1 were 3.95, 3.88, 3.59, and 3.81, respectively. The scale values obtained from the ratings of L1 were used as the dependent variable in the linear regression analysis to be described in sections 16 and 17 following.

16. *Linear Regression Analysis.* In the linear regression analysis that was performed the scale values of severity of stuttering served as the dependent variable and the measures for the five disfluency categories, together with that for time, were the independent variables. Certain aspects of linear regression procedures and subsequent interpretation require the investigator to make the assumptions of homoscedasticity and linearity in order for the interpretation to be meaningful beyond that resulting from a mere manipulation of numbers.[6]

[5]This particular coefficient of agreement may be interpreted as the ratio of the variance of the true ratings to the variance of the obtained ratings for this population of raters (*14*). It may also be considered as an average intercorrelation of ratings of N speakers from all possible pairs of k listeners (*10*). Using the symbols of Lindquist (*14*), the formula employed for computing this coefficient was $(ms_A - ms_{AS}) / [ms_A + (k\text{-}1)ms_{AS}]$ in which the between-listeners variance is not considered.

[6]The assumption of homoscedasticity is analogous to requiring homogeneity of variance in analysis of variance situations, while linearity refers to the type of relationship required for all zero order correlations between the dependent and independent variables.

TABLE 2. Summary of values derived in Part 1 by means of multiple correlation procedures.

		Original Analysis		Deletion of Variable 4		Deletion of Variable 2		Deletion of Variable 6	
	r_{oi}	β_i	*t value** of β_i	β_i	*t value* of β_i	β_i	*t value* of β_i	β_i	*t value* of β_i
Time	.6810	.3796	3.02	.3602	2.91	.2871	3.58	.2655	3.31†
Interjections	.4531	−.1193	−1.05	−.0826	−.78				
Part-word Repetitions	.8258	.4371	4.53	.4692	5.21	.4778	5.35	.5128	5.82†
Word-phrase Repetitions	.3689	.0720	.92						
Prolongations	.7462	.2810	3.12	.2696	3.02	.2774	3.12	.2691	3.00†
Revisions	.1771	.1228	1.74	.1323	1.90	.1149	1.74		
		R**	*F value*†† of R	R	*F value* of R	R	*F value* of R	R	*F value* of R
		.91	33.41***	.91	40.06	.90	50.36	.90	63.34

*$t\beta_i = \beta_i/\sigma\beta_i$; $df = N - m$
†Significant at 1% level.
**$R^2 = \Sigma\beta_i r_{oi}$
††$F = \frac{R^2}{1 - R^2}(N - m/m - 1)$; $df = m - 1, N - m$
***All *R*s significant at 1% level.

17. *Multiple Correlations, Prediction Equations, Estimates of Errors, and Intercorrelations.* Table 2 contains a summary of the results of the multiple correlational procedure. It includes the values for the multiple correlation coefficient (R) and the regression weights (βs) for the original analysis and for each deletion, along with the corresponding values of t and F involved in tests of significance. The values for the regression coefficients, that is, the Beta values, can be used to compare the relative contribution of each variable with that of any other variable. The first column of Table 2 which includes the values for r_{oi}, the zero order correlations between the dependent variable and each independent variable, provides a more complete picture of the relationships involved.

Considering, first, the two columns of Table 2 labeled *Original Analysis*, it can be seen that variable 4, word-phrase repetitions, contributed least to the multiple correlation; that is, it had the smallest Beta value, and the smallest value of t.[7] Accordingly, variable 4 was dropped from the analysis and a new R and new Beta weights were computed.[8] The results of the new analysis are shown in the first of the two columns labeled *Deletion of Variable 4.* Again, the second of these two columns shows the corresponding values of F and t for the variables involved. Since

[7]There is some question as to whether variables are more appropriately deleted on the basis of the size of the Beta value or the size of t. The latter procedure was followed in this study.

[8]For t.01 (df = 42), the value of t necessary for significance was 2.70. It can be seen that variables 1, 3, and 5 were significant at this level, the variables 2, 4, and 6 were not.

variable 2 now had the smallest t value, it was dropped from the next analysis. Following the same procedures for the remainder of the table, variable 6 was dropped.

In the two columns labeled *Deletion of Variable 6*, the final analysis is presented. Here it can be seen that with variables 1, 3, and 5, an R of .90 was obtained, differing from the multiple correlation coefficient in the original analysis by only .01. The differences between β_1 - β_3, β_1 - β_5, and β_3 - β_5 were tested, and no difference was found to be significant at the 5% level of confidence.

The regression or prediction equation in raw score form was: $Y_1 = .0071\, X_1 + .0697\, X_3 + .0381\, X_5 + 1.4396$ in which Y_1 = predicted rating of severity of stuttering, X_1 = time, in seconds, to speak 200 words, X_3 = frequency of part-word repetitions, and X_5 = frequency of prolongations. The standard error of estimate was .97, indicating that, for these data, a predicted rating of severity of stuttering would lie within .97 scale points of the obtained rating about two-thirds of the time.

A matrix of intercorrelations between the dependent and independent variables is presented in Table 3. This table also shows estimates of the strength of relationship between T (total number of disfluencies) and the obtained median ratings of severity of stuttering, and between time, in seconds, and total number of disfluencies. The highest correlation coefficient obtained was .85, between total number of disfluencies and rated severity of stuttering. The highest correlation coefficient between rated severity of stuttering and the frequency of any one type of disfluency was .83, for the variable labeled part-word repetitions. Among the disfluency categories alone, excluding time, part-word repetitions and prolongations were most highly related with a correlation coefficient of .65. The disfluency variable which appeared to be least related to other variables was frequency of revisions. Referring to the top row of correlation coefficients in Table 3, it can be seen that frequency of revisions was the only disfluency variable not significantly correlated with rated severity of stuttering.

TABLE 3. Intercorrelation matrix for dependent variable *0* and independent variables *1* through *6*.*

	1	*2*	*3*	*4*	*5*	*6*	*T*†
0	.68**	.45	.83	.37	.76	.18	.85
1		.76	.52	.26	.55	−.06	.76
2			.33	.42	.33	.13	
3				.39	.65	.16	
4					.15	.29	
5						.02	

**0*: rated severity of stuttering; *1*: speaking time; *2*: interjections; *3*: part-word repetitions; *4*: word-phrase repetitions; *5*: prolongations; *6*: revisions.

†Sum of disfluency categories *2* through *6* (total number of disfluencies).

**For $df = 48$, rs of .279 (5%) and .361 (1%) are significantly different from zero.

Disfluencies classified as part-word repetitions and prolongations were then combined into one disfluency category for the following reasons: (a) the contribution of frequency of prolongations to the prediction of a rating of severity of stuttering, although significantly different from zero, was smaller than the predictive potential of frequency of part-word repetitions, and (b) combining these two disfluency classifications would eliminate the often troublesome discrimination problem of placing disfluencies in one or the other category. The resultant combined category was designated as repetitions. The linear regression analysis was computed again using rated severity of stuttering as the dependent variable and repetitions and time as the two independent fluency variables with the following results. The multiple correlation coefficient was .89, and the regression weights in deviation form were .71 and .26 for repetitions and time, respectively. Both of these regression weights were significantly different from zero.[9] The prediction equation in raw score form was: $Y_{\text{severity}} = .0542\ X_{\text{repetitions}} + .0069\ X_{\text{time}} + 1.5149$ and the standard error of estimate was .99. The negligible change in multiple correlation coefficient and standard error of estimate obtained by using two independent variables rather than three, that is, by combining repetitions and prolongations into one disfluency category, is to be noted.

18. *Interpretation of Results.* Indexes of investigator agreement in identifying and classifying disfluencies and of listener agreement in rating severity of stuttering were sufficiently high for purposes represented by the experimental design of this study. An examination of the intraclass correlation coefficients would seem to indicate that listener agreement was more than adequate for purposes of this experiment, especially in view of the type of task the listeners were asked to perform, that of evaluating a long, complex sample of speech and making a rating of severity of stuttering. On a priori grounds one would have little reason to expect such relatively good agreement among the listeners in this experiment.

The size of the multiple correlation coefficient (.89) and the standard error of the estimate (.99) obtained by employing frequency of repetitions (part-word repetitions plus prolongations) and time as the predictors of rated severity of stuttering indicated that the model for predicting severity of stuttering presented in Figure 1 may tentatively be considered reasonably valid. Some limitations, however, are to be placed on generalizations derived from data obtained by means of multiple correlational procedures. First, the assumptions of homoscedasticity and linearity were not subjected to close examination. Second, and more important, the multiple correlation coefficients represent the maximum values that could be obtained from the data. That is, the regression weights were determined so as to make the correlation coefficients maximal. It is to be expected, therefore, that a smaller multiple correlation coefficient would be obtained if the same regression weights were to be applied to data based on responses of a new sample of speakers and listeners. With this in mind, then, the next task was to repeat the entire study using a new sample of speakers and listeners. This was carried out and will be reported in Part 2.

Experimental Operations and Results: Part 2

✓ Figure 1 was also employed in Part 2 as the model for estimating the accuracy with which a rating of severity of stuttering might be predicted. Differences in the operations and results

[9] For $\beta_{\text{repetitions}}$, $t = 8.80$; for β_{time}, $t = 3.20$.

between Part 1 and Part 2 are to be noted.

1'. *Modifications of Original Assumptions.* On the basis of the results and interpretation of the results presented in Part 1 the assumptions underlying the goal of estimating the accuracy of predicting a rating of severity of stuttering were retained or modified as follows: (a) the rating of severity of stuttering involves an interaction between speaker and listener; and (b) listeners' ratings of severity of stuttering may be predicted from measurements of the two speech variables designated as 'repetitions' and 'time.'

2'. *Speakers.* The population of hypothetical speakers from which the experimental speakers were selected was considered to be the same as described in Part 1; that is, college males with the problem of stuttering who were participating in speech therapy.

3'. *Selection of Speakers.* Speakers were selected from persons with the problem of stuttering who had attended or were currently attending the University of Iowa. As in Part 1, the quality of the tape recording available, length of speech sample recorded, and the investigator's evaluation of the severity of stuttering represented by the tape-recorded speech sample were the primary criteria of appropriateness of a tape recording in Part 2 of the present study. Many more tape recordings were available in Part 2 than in Part 1, and the experimenter was able to be relatively selective in applying the criteria just enumerated.

4'. *Experimental Speakers.* Fifty experimental speakers were employed in Part 2, none of whom had served as an experimental speaker in Part 1. All the speakers were considered by themselves and by their speech clinicians to have the problem of stuttering and were enrolled for therapy at the Iowa Speech Clinic when the tape recordings were made. They were considered to be similar to the experimental speakers employed in Part 1 with respect to age and academic attainment.[10]

5'. *Speaking Task.* The same speaking task used in Part 1, the Job task, was employed to secure tape-recorded samples of speech. Speech samples of longer duration were secured in Part 2, although only the first 200 words were used as before.

6'. *Tape-recorded Speech Samples.* The tape recordings were edited and randomly combined as described in Part 1, except that the instructions and samples representing extremes of severity of stuttering were not part of this second set of 50 tape-recorded speech samples. The duration of the second test tape, consisting of 50 200-word samples and 15-second pauses between adjacent samples, was approximately two hours and 55 minutes. It seemed to the experimenter that the tape recordings used in Part 2 were of better quality than those employed in Part 1.

7'. *Analysis of Disfluency and Rate.* The results of Part 1 indicated that only two disfluency variables were necessary for the adequate prediction of median ratings of severity of stuttering, and for that reason the 50 tape-recorded samples of speech were analyzed in terms of these two variables. The measures used were: (a) the number of words in relation to which a part-word repetition, a sound prolongation, a broken utterance, or unusual

[10]See Part 1, section 4 and footnote 1.

stress was observed, a word being counted only once no matter how many or how often these types of behaviors were observed during its production; and (b) the time, measured in number of seconds, used by the speaker to say 200 words. These two categories of disfluency were designated as 'repetitions' and 'time,' respectively. It is to be noted that the disfluency category labeled 'repetitions' in Part 2 included the categories called 'part-word repetitions' and 'prolongations' employed in Part 1.

TABLE 4. Summary of measures of disfluency, in terms of number of repetitions and time, in seconds, obtained for 50 speakers in Part 1 and 50 speakers in Part 2, and *t* tests between means.

	Speakers, Part 1		*Speakers, Part 2*		
	Mean	*SD*	*Mean*	*SD*	*t*
Repetitions	24.9	28.7	28.0	24.9	.57*
Time	134.0	82.0	189.1	123.7	2.60†

*Not significant.
†Significant at the 1% level.

8′. *Measures of Disfluency and of Rate.* The mean frequency of repetitions and the mean time, together with the standard deviations for each of these distributions, are to be found in Table 4. Similar data for the sample of speakers used in Part 1 (Tape 1) are also presented in Table 4, along with *t* tests between means for Tape 1 and Tape 2. The results of these *t* tests must be interpreted in light of (a) the strength of relationship between measures and (b) the marked heterogeneity of variances for the two distributions of measures of time. Despite these limitations the conclusion was drawn that the experimental speakers in Part 2 took more time to speak 200 words, on the average, than the experimental speakers in Part 1. An explanation for this difference might be that many tape-recorded samples of speech displaying relatively extreme degrees of severity of stuttering, in the judgment of the investigator, could not be included as part of Tape 1 because they included less than 200 words. Somewhat longer tape-recorded samples of speech were obtained, however, from the speakers used in Part 2. It would appear, therefore, that severe stutterers who took more time to speak 200 words were more likely to be included as speakers in Part 2 than Part 1.

9′. *Listeners.* The hypothetical population of listeners to which the results of Part 2 were to be generalized was defined by means of the same operations that were performed in selecting the experimental listeners for Part 1. This meant that the hypothetical population of listeners in Part 2 was composed of persons with the problem of stuttering, speech clinicians, and laymen reputed to have normal speech, all of whom were attending the University of Iowa.

10′. *Selection of Listeners.* Volunteers to participate in the listening sessions were solicited from the course entitled Introduction to Speech Pathology and Audiology and from persons with the problem of stuttering participating in therapy in the University of Iowa Speech Clinic. Student clinicians enrolled in the Stuttering Practicum were also asked to participate. An important difference between the experimental operations for selecting the listeners in Part 1 and Part 2 is to be noted. Listeners in Part 2 were told that they would be paid for listening. The decision to pay raters was made in view of the number of hours of listening time required of them. In the investigator's opinion, it was the prospect of being paid that prompted many subjects to volunteer to participate in the listening sessions. Fewer subjects, however, appeared at the announced times than were present for the listening sessions in Part 1, when the listeners were not paid.

11′. *Experimental Listeners* (L2). Forty listeners served as the experimental listeners in Part 2. Group 1 was composed of 13 persons with the problem of stuttering; Group 2 consisted of 14 student clinicians enrolled in the Stuttering Practicum; and Group 3 was composed of 13 listeners, who, as in Part 1, were selected from the Introduction to Speech Pathology and Audiology class and were considered to constitute the 'laymen' with normal speech in the present study.

TABLE 5. Mean median scale values of severity of stuttering obtained from listeners (N = 48) in Part 1 and listeners (N = 40) in Part 2.

Listener Subgroups	*Listeners Part 1*	*Listeners Part 2*	*Combined*
Stutterers	3.95	3.96	3.96
Clinicians	3.88	3.64	3.76
Laymen	3.59	3.81	3.70
Combined	3.81	3.80	

In order to compare the responses of the listeners employed in Part 1 (L1) and those of the listeners in Part 2 (L2), the investigator asked the latter group to rate the 50 tape-recorded samples of speech used in Part 1 with respect to severity of stuttering on a nine-point equal-appearing intervals scale. Ratings obtained from L1 and L2 were compared in several ways. Listeners agreement for L2 evaluating Tape 1 was estimated by means of intraclass correlations. For Groups 1, 2, 3, and the total of 40 listeners, the coefficients were, respectively, .80, .82, .84, and .81, which were very similar to the coefficients computed for L1 evaluating Tape 1. Table 5 contains the mean median scale values for each subgroup for both sets of listeners, and these means were evaluated by using an analysis of variance, a summary of which is shown in Table 6. The difference between the means of L1 and L2, for combined subgroups, was nonsignificant, the difference being only .01 in magnitude. The differences among the three groups of listeners (stutterers, clinicians, and laymen), for L1 and L2 combined, were also evaluated and were found to be nonsignificant. The interaction between the two sets of listeners (L1 and L2) and among the three subgroups of listeners (G1, G2, and G3) was found to be statistically nonsignificant. In addition, median scale values of severity of stuttering were computed for L1 and L2 by combining subgroups; that is, scale values were based on ratings made by all 48

TABLE 6. Summary of analysis of variance employed to evaluate the differences in median ratings of severity of stuttering between listeners in Part 1 and Part 2 and among subgroups of listeners (stutterers, clinicians, and laymen) in evaluating 50 samples of speech used in Part 1.

Source	*df*	*ms*	*F*
Listeners (A)	1	.01	<1
Subgroups (B)	2	1.84	<1
AB	2	1.26	<1
w	294	4.89	
Total	299		

and 40 listeners, respectively. The mean median scale values for L1 and L2 were 3.80 and 3.83. The difference of .03 between these means was evaluated by a *t* test and, as expected, was found to be nonsignificant. The correlation between these two sets of scale values was .99.

TABLE 7. Summary of results of linear regression procedures for listeners in Part 1 and Part 2 evaluating Tape 1.*

	Multiple R	*Regression Weights*		*Standard Error of Estimate*
		Repetitions	*Time*	
Part 1	.89	.71	.26	.99
Part 2	.88	.69	.27	1.06

Zero Order Correlations

	Severity vs Repetitions	*Severity vs time*	*Repetitions vs time*
Part 1	.87	.68	.59
Part 2	.85	.67	.59

*In these analyses, the dependent variable was rated severity of stuttering, and the independent variables were frequency of repetitions, as defined in the text, and time, in seconds, required to speak 200 words.

As a further check on the comparability of ratings obtained from the two sets of listeners, a linear regression analysis was performed using scale values of severity of stuttering from L2 as the dependent variable and the two disfluency measures, repetitions and time, as the independent variables. The multiple correlation coefficient was .88, and the regression weights in unit deviation form were .69 for repetitions and .27 for time, both being significantly different from zero at the 1% level. Table 7 contains a summary of the values computed in the regression analyses of the median ratings of the listeners used in Part 1 and Part 2 evaluating Tape 1, and it can be seen that the results of the two analyses closely approximated each other.

12′. *Listening Task.* The listening task in Part 2 was essentially the same as that described in Part 1 with the following exceptions. Only two listening sessions were held in Part 2 and the members of all three subgroups of listeners participated at random in either session. In addition, the listeners in Part 2 did not hear segments representing the extremes of severity of stuttering, since they had heard them previously when evaluating Tape 1 (see section 11′). The instructions for the listening sessions in Part 2 were read aloud by the investigator rather than being recorded on the tape. Approximately three weeks intervened between the two listening sessions for L2, that is, between the evaluations of Tape 1 and Tape 2. Finally, of the 40 listeners in L2 who evaluated Tape 1, only 38 attended the listening sessions for Tape 2, one clinician and one layman being absent.

13′. *Ratings of Severity of Stuttering.* Nineteen-hundred ratings of severity of stuttering were obtained from the 38 listeners, and the mean of all these ratings was 4.66. The means for stutterers, clinicians, and laymen were 4.99, 4.47, and 4.50, respectively.

14′. *Numerical Analysis.* Intraclass correlations for L2 evaluating Tape 2 were .83, .75, .86, and .81 for G1, G2, G3, and all 38 listeners combined, respectively. Table 8 contains a summary of all the intraclass correlations obtained in both parts of the present study.

TABLE 8. Summary of intraclass correlations estimating the agreement among listeners in rating severity of stuttering.

	Stutterers	*Clinicians*	*Laymen*	*Total*
Listeners, Part 1 Tape 1	.79	.83	.87	.83
Listeners, Part 2 Tape 1	.80	.82	.84	.81
Listeners, Part 2 Tape 2	.83	.75	.86	.81

TABLE 9. Summary of analysis of variance employed in evaluating the differences in ratings of severity of stuttering among the three subgroups of listeners in Part 2 evaluating Tape 2.

Source	*df*	*ms*	*F*
Groups	2	5.62	1.47*
within	147	5.54	
Total	149		

*Not significant; $F_{.20} = 1.63$, $df = 2, 120$.

15′. *Scale Values of Severity of Stuttering.* In order to evaluate the differences among the median ratings of Tape 2 obtained from the three subgroups of listeners (L2) an analysis of variance (Lindquist (*14*) simple randomized design) was performed. The mean median scale values of severity of stuttering for G1, G2, and G3 were respectively, 5.03, 4.42, and 4.48. The analysis of variance indicated no significant differences among the three subgroups of listeners, and the results of this analysis are presented in Table 9.[11] As before, the scale values computed for all 38 listeners served as the dependent variable in the linear regression analysis.

16′. *Linear Regression Analysis.* The two independent variables in the multiple linear regression analysis were frequency of repetitions and time. Similar computational procedures were used as described in Part 1.

17′. *Multiple Correlations, Prediction Equations, Estimates of Error, and Intercorrelations.* Using median scale values of severity of stuttering obtained from L2 as the dependent variable and the two measures of disfluency, frequency of repetitions and time, a multiple correlation coefficient of .87 was computed. The regression weights in deviation form were .72 for repetitions and .18 for time.[12] The standard error of estimate was 1.14, with the prediction equation in raw score form as follows: $Y_{\text{severity}} = .0676\ X_{\text{repetitions}} + .0033\ X_{\text{time}} + 2.1224$. The zero order correlations were: .86 between severity and repetitions, .75 between severity and time, and .79 between repetitions and time, all coefficients being significantly different from zero at the 1% level.[13]

Data gathered thus far were re-evaluated by combining Tape 1 and Tape 2, making a total of 100 tape-recorded samples of speech available for analysis. Ratings by L2 were used as the dependent variable and a new linear regression analysis computed.[14] Additional analyses were made by

[11]The three groups of listeners were more homogeneous than the experimenter desired; and although the differences among the groups were in the expected direction, that is, persons with the problem of stuttering giving higher ratings of severity of stuttering than clinicians or laymen, the statistically nonsignificant differences among mean ratings of severity of stuttering should have been anticipated.

[12]For $\beta_{\text{repetitions}}$, $t = 6.21$; for β_{time}, $t = 1.52$.

[13]For $df = 48$, $r = .36$ is necessary for significance at the 1% level.

[14]The prediction equation in raw score form based on 100 samples of speech was: $Y_{\text{severity}} = .0571\ X_{\text{repetitions}} + .0058\ X_{\text{time}} + 1.7839$

TABLE 10. Descriptive disfluency data pertaining to all 100 speech samples combined, to 50 speech samples with the lowest median scale values of severity of stuttering, and to 50 speech samples with the highest median scale values of severity of stuttering.

	Severity		*Repetitions*		*Time*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
100 Speech Samples	4.23	2.30	26.49	26.94	161.52	108.51
50 Samples with lowest severity ratings	2.27	.80	8.60	6.77	109.58	24.52
50 Samples with highest severity ratings	6.19	1.52	44.38	27.67	213.48	132.47

TABLE 11. Summary of results of linear regression procedures employing all 100 speech samples, 50 speech samples with lowest ratings of severity of stuttering, and 50 speech samples with highest ratings of severity of stuttering.*

	Multiple R	*Regression Weights*		*Standard Error of Estimate*
		Repetitions	*Time*	
100 samples	.87	.67	.27	1.09
50 samples with lowest severity ratings	.85	.73	.27	.42
50 samples with highest severity ratings	.84	.49	.47	.83

Zero Order Correlations

	Severity vs Repetitions	*Severity vs time*	*Repetitions vs time*
100 samples	.85	.72	.67
50 samples with lowest severity ratings	.81	.49	.31
50 samples with highest severity ratings	.74	.73	.54

*In these analyses, the dependent variable was rated severity of stuttering and the independent variables were frequency of repetitions, as defined in the text, and time in seconds required to speak 200 words.

dividing the 100 tape-recorded samples of speech into two groups. The first set of 50 samples consisted of those with the lowest median ratings of severity of stuttering and the second set of 50 were those with the highest median ratings. The cut-off point was approximately a median rating of 4.00. A linear regression analysis was made of the values in each of these two sets of data. Descriptive data pertaining to the ratings of severity of stuttering, frequency of repetitions, and time used in these three analyses are to be found in Table 10. Table 11 contains a summary of the results obtained from these three regression analyses.

The zero order correlations between severity and repetitions (.85) and between severity and time (.72), for Tape 1 and Tape 2 combined, were tested for departure from linearity. Both tests indicated a degree of departure from linearity significant at the 1% level.[15]

18′. *Interpretation of Results.* The assumptions presented in section 1′ of Part 2 were re-evaluated on the basis

[15]For repetitions: $\eta = .88$, $F = 2.98$, $df = 7, 91$. For time: $\eta = .82$, $F = 5.86$, $df = 7, 91$.

of the data in Part 1 and Part 2. Assumptions implying that an evaluation or rating of severity of stuttering represents an interaction between a speaker and a listener, and that the listener's evaluation of severity of stuttering is highly associated with certain aspects of the speaker's fluency and disfluency were still to be considered valid.

The two groups of experimental listeners employed in the two parts of the present experiment responded with essentially identical median ratings of severity of stuttering when evaluating the tape-recorded samples of speech used in Part 1. The relatively small difference between the mean median ratings of these two groups of listeners and the high correlation between their median scale values (.99) appeared to indicate a high degree of stability in the group measures obtained by means of the procedures described in the present experiment.

The experimental speakers who participated in Part 1 differed from those in Part 2 with respect to rated severity of stuttering and the measure of time. These group differences appeared to be due to a bias in the sampling procedure which resulted in fewer individuals with large values for the measure of time being included in Part 1 of the experiment; and an additional result of this bias appeared to be that the speakers in Part 2 received generally higher ratings of severity of stuttering than did those in Part 1.

The use of linear regression procedures to estimate the accuracy of predictions of median ratings of severity of stuttering is to be questioned on the basis of the data obtained. Although the standard error of estimate of 1.09, obtained when both samples of speakers were combined, may be considered an indication of reasonably good prediction, this measure of error is appropriate only so far as assumptions of homoscedasticity and linearity can be satisfied. In the present study these assumptions appeared not to be fully met.

When the 100 tape-recorded samples of speech were divided into two groups with respect to rated severity of stuttering and the regression analysis performed for each group separately, the zero order correlations and regression weights indicated that the assumptions of homoscedasticity and linearity could not be made; that is, the relationship between the measures of disfluency and time and between each of the measures and rated severity of stuttering was not the same throughout all levels of rated severity of stuttering. Frequency of repetitions, as defined in the present study, and the measure of time used were reacted to differentially by listeners in evaluating severity of stuttering. At the lower or mild levels of severity of stuttering the variable of time is hardly an important factor, while at the higher or more severe levels of severity the variable of time is as important as that of frequency of repetitions in predicting median ratings of severity of stuttering.

The continued used of frequency of repetitions and time as the predictors of median ratings of severity of stuttering is to be considered in relation to the following factors: (a) the high multiple correlation of .87 and the reasonably small standard error of estimate of 1.09, and (b) the apparent failure to satisfy the assumptions of linearity and homoscedasticity. These assumptions, of course, need only be made in

TABLE 12. Summary of measures of disfluency and time and ratings of severity of stuttering for Part 1 and Part 2.

	Severity		Repetitions		Time	
	Mean	SD	Mean	SD	Mean	SD
50 Speakers, Tape 1	3.80	2.19*	24.92	28.73	133.98	82.03
50 Speakers, Tape 2	4.64	2.33	28.04	24.94	189.06	123.70
100 Speakers, Tapes 1 and 2	4.23	2.30	26.49	26.94	161.52	108.51
50 Speakers with lowest severity ratings	2.27	.80	8.60	6.77	109.58	24.52
50 Speakers with highest severity ratings	6.19	1.52	44.38	27.67	213.48	132.47

*Based on listeners in Part. 2

TABLE 13. Summary of results obtained in Part 1 and Part 2 by means of linear regression procedures with rated severity of stuttering (S) as the dependent variable and frequency of repetitions (R) and time (T) as the independent variables.

	R	β_R	β_T	r_{SR}	r_{ST}	r_{RT}	SE
Listeners, Part 1, Tape 1 (N = 50)	.89	.71	.26	.87	.68	.59	.99
Listeners, Part 2, Tape 1 (N = 50)	.88	.69	.27	.85	.67	.59	1.06
Listeners, Part 2, Tape 2 (N = 50)	.87	.72	.18*	.86	.75	.79	1.14
Listeners, Part 2, Tapes 1 and 2 (N = 100)	.87	.67	.27	.85	.72	.67	1.09
50 Samples with lowest severity ratings	.85	.73	.27	.81	.49	.31*	.42
50 Samples with highest severity ratings	.84	.49	.47	.74	.73	.54	.83

*Not significantly different from zero at the 1% level.

order to test the statistical significance of the correlation coefficients obtained, and the sample multiple correlations, intercorrelations, etc., do describe the observed data. A related purpose of the present experiment, however, was to develop a measure of severity of stuttering based on an analysis of disfluency that could be employed with other speakers in the future, and the continued use of these two disfluency variables, repetitions and time, and the corresponding prediction equation is not entirely endorsed.

The preceding interpretation has been made with reference to all the relevant data thus far obtained in the studies here reported. These data are summarized in Tables 12 and 13.

Discussion

The relative lack of precision of a prediction of a rating of severity of stuttering, using the prediction equation based on data obtained from all 100 speakers, is a function of two factors: (a) the nonlinearity of the relationships, since linear regression procedures were employed, and (b) the failure of the independent variables to affect or to be related to the listener's evaluation of severity of stuttering.

In order to consider the relative weights of these two factors as they might affect the precision of prediction, the predicted values of severity of stuttering and the discrepancy between the obtained and predicted values of severity of stuttering were computed. Ignoring the direction of the discrepancies, the mean discrepancy was .91, the standard deviation of the discrepancies was .64, and the range extended from .10 to 2.81.

The 10 samples of recorded speech for which the discrepancies between predicted and obtained values were largest were considered in detail. Table 14 contains relevant data. It can be seen that all 10 of the predicted values were underestimates. These samples were not at either extreme with respect to rated severity of stuttering, for it is within this range that the predictions might be expected to be reasonably accurate. With respect to the low end of the scale of rated severity of stuttering, 11 samples fell within the range of 1.0 to 1.5; and since the added constant in the prediction equation was 1.78, all the predicted values for these particular 11 samples would necessarily be overestimates. It is interesting to note that although no repetitions were observed in three of these 11 samples the obtained ratings for these three samples

TABLE 14. Summary of measures related to tape-recorded speech samples with largest discrepancies between predicted and obtained median ratings of severity of stuttering.

Subject	*Obtained Rating*	*Predicted Rating*	*Discrepancy*	*Repetitions*	*Time*
Tape 1					
5	6.27	3.99	2.28	26	125
37	5.91	3.87	2.04	26	103
38	7.50	4.85	2.65	43	105
39	4.90	2.95	1.95	8	122
Tape 2					
5	5.75	3.41	2.34	17	113
6	7.81	5.72	2.09	49	196
14	5.60	3.58	2.02	13	182
15	7.82	5.01	2.81	29	271
23	8.86	6.46	2.40	56	255
42	6.12	3.72	2.40	22	188

TABLE 15. Summary of measures for those tape-recorded speech samples having observed median ratings of severity of stuttering greater than 8.00.

Observed Rating	*Predicted Rating*	*Discrepancy*	*Repetitions*	*Time*
9.00	8.15	—.85	51	595
9.00	10.81	1.81	77	799
8.94	10.83	1.89	126	320
8.86	6.46	—2.40	56	255
8.59	8.84	.25	96	272
8.46	9.72	1.26	100	384
8.23	10.14	1.91	106	398
8.22	7.78	—.44	69	354

were 1.17, 1.35, and 1.50. It appears that some listeners make a rating of severity of stuttering greater than one even though other listeners make ratings of one (no stuttering). On the other end of the scale of rated severity of stuttering, three large positive discrepancies (overestimates) were noted for three samples displaying obtained ratings of 8.94, 8.23, and 9.00. There were other samples, however, with similar obtained ratings of 8.59, 9.00, and 8.22 for which there were considerably smaller discrepancies of —.25, —.85, and —.44, respectively. Table 15 contains relevant data for samples with observed ratings greater than 8.00. There does not appear to be any consistent pattern of discrepancy between predicted and obtained ratings of severity for these samples. Of the four largest discrepancies, the largest represents an under-

TABLE 16. A comparison among tape-recorded speech samples with relatively large discrepancies between predicted and obtained median ratings of severity of stuttering and samples having similar frequencies of repetition and relatively small discrepancies between predicted and obtained median ratings of severity of stuttering.

Discrepancy	*Repetitions*	*Obtained Rating*	*Obtained ratings of other samples with similar frequencies of repetition and relatively small discrepancies*
2.81	29	7.82	5.00, 4.92
2.65	43	7.50	6.00, 6.20
2.40	56	8.86	6.12, 6.57, 7.34
2.40	22	6.12	4.00, 3.08, 2.77
2.34	17	5.75	3.82, 3.81
2.28	26	6.27	3.50, 4.58, 3.38
2.09	49	7.81	5.14, 5.69
2.04	26	5.91	3.50, 4.58, 3.38
2.02	13	5.60	3.29, 2.97, 3.89
1.95	8	4.90	2.50, 3.00

estimate and the next three indicate overestimates.

The 10 samples for which there were the largest discrepancies between predicted and obtained severity ratings were compared to other samples involving similar frequencies of repetition and smaller discrepancies. Time was not considered in this connection because of its smaller predictive potential throughout all the regression procedures. Table 16 presents data for these 10 samples as compared to other samples with smaller discrepancies and similar frequencies of repetition. The obtained ratings of severity of stuttering were higher for all of the 10 samples for which there were large discrepancies than for the comparable samples for which the discrepancies were smaller. It would appear that the speakers represented by these 10 samples displayed one or more dimensions of fluency or disfluency not adequately represented in a frequency count of repetitions, as defined in the text, and the time required to speak 200 words. It is to be concluded that most of the discrepancy between the observed and predicted ratings for these 10 speakers was due to the lack of relationship between the fluency variables considered in the present study and rated severity of stuttering, and not to the nonlinearity of the zero order correlations.

In order to explore this possibility more fully the experimenter listened to the 10 tape-recorded samples of speech with the largest discrepancies between obtained and predicted ratings of severity of stuttering. There appeared to be no highly specific discernible forms of speech behavior common to all 10 speakers. Instead, each speech sample, in the opinion of the investigator, was characterized by some unusual and distinctive pattern of speaking. In general, however, the patterns of disfluency seemed to be one of the following: (a) long, drawn-out repetitions of syllables or prolongations of sounds, (b) very rapid repetition of syllables, (c) audible manifestations of excessive tension, and (d) some combination of two or all of these patterns. One speaker who exhibited relatively few syllable repetitions spoke with severe harshness throughout his tape-recorded sample. It may be possible that the listeners were influenced in their evaluations by the speaker's unpleasant voice quality as well as, or even rather than, his degree of disfluency. Another speaker exhibited two patterns of speaking which may have accounted for a higher obtained than predicted rating of severity of stuttering; his speech was characterized by a rapidly rising pitch level during syllable repetitions and a general lack of intelligibility due to a regional dialect. The speaker with the largest discrepancy between the obtained and predicted rating of severity of stuttering ($D = 2.81$) exhibited many phrase revisions and word repetitions, types of disfluency not taken into account in the present analysis, as well as many rapid and extended repetitions of sounds. Another tape-recorded sample of speech was of relatively poor quality, and the listeners may have been reacting in part to the relatively low degree of intelligibility of the speech. The failure to obtain accurate predictions of severity ratings for these 10 samples may be substantially explained by reference to the fact that the method of analysis employed in the present study did not allow for due consideration of certain of their distinctive characteristics.

For the reason just mentioned, the regression equation based on 100 samples of speech (see Part 2, section 17′) is not recommended for predicting a single speaker's median rating of severity of stuttering. It is suggested that the regression equation may be used to estimate median ratings of severity of stuttering when investigating the effects of different experimental treatments on the speech fluency of randomly selected groups of speakers.

Another possible application suggested by the results of the present study involves using the disfluency category labeled 'repetitions' (number of words involving part-word repetitions, prolongations, broken utterances, and apparent unusual stress or tension) as an operational definition of speech disfluency; and instead of counting 'moments of stuttering' the observer would attempt to be more descriptive and tally frequency of 'repetitions,' as here defined.

Summary

The present study was designed to estimate the accuracy with which a rating of severity of stuttering in a 200-word tape-recorded sample of speech can be predicted from measurements of the disfluency of the speech and its rate of utterance. Speech samples, each 200 words long, obtained from each of 100 male speakers with the problem of stuttering, were analyzed in terms of frequency counts of each of a number of types of disfluency and the time in seconds required to speak the 200 words. Each speech sample was also rated by listeners on a nine-point equal-appearing intervals scale of severity of stuttering. The speakers and listeners were divided into several subgroups for the distinctive purposes of the various aspects of the analysis. Linear regression procedures were employed using rated severity of stuttering as the dependent variable and the various measurements of disfluency and the values of speaking time as the independent variables. The reliability of the investigator in making the measurements of disfluency and the degree of agreement among listeners in rating severity of stuttering were considered.

The following conclusions were drawn on the basis of the linear regression analyses and other observations:

(a) The types of speech disfluency that appear to be associated with ratings of severity of stuttering are syllable or sound repetitions, sound prolongations, broken words, and words involving apparent unusual stress or tension.

(b) The regression equation based on all 100 samples of speech was:

$$Y_1 = .0571\ X_r + .0058\ X_t + 1.7839$$

in which Y_1 = predicted rating of severity of stuttering, X_r = frequency of 'repetitions,' as defined in the present study, and X_t = number of seconds needed to speak 200 words. The multiple correlation coefficient was .87 and the standard error of estimate was 1.09.

(c) Prediction of an individual rating of severity of stuttering with the above regression equation is not endorsed because the assumptions of linearity and homoscedasticity, necessary for this kind of prediction, cannot be adequately satisfied. Predictions of severity of stuttering to be employed on a group basis would be considered under certain experimental conditions.

(d) The largest discrepancies between obtained and predicted median ratings of severity of stuttering were associated with speakers whose obtained ratings of severity of stuttering appear to be based on types of disfluency not considered in the present analysis.

References

1. Bloodstein, O., Studies in the psychology of stuttering: XIX. The relationship between oral reading rate and severity of stuttering. *J. Speech Dis.*, 9, 1944, 161-173.
2. Bloodstein, O., Jaeger, W., and Tureen, J., A study of the diagnosis of stuttering by parents of stutterers and nonstutterers. *J. Speech Hearing Dis.*, 17, 1952, 308-315.
3. Boehmler, R. M., Listener responses to nonfluencies. *J. Speech Hearing Res.*, 1, 1958, 132-141.
4. Branscom, Margaret E., Hughes, Jeannette, and Oxtoby, Eloise T., Studies of nonfluency in the speech of preschool children, in Wendell Johnson (ed.), *Stuttering in Children and Adults*. Minneapolis: University of Minnesota Press, 1955.
5. Davis, Dorothy M., The relation of repetitions in the speech of young children to certain measures of language maturity and situational factors, Part I. *J. Speech Dis.*, 4, 1939, 303-318.
6. Davis, Dorothy M., The relation of repetitions in the speech of young children to certain measures of language maturity and situational factors, Part II and Part III. *J. Speech Dis.*, 5, 1940, 235-246.
7. Duffy, R. J., Quantitative data on the speech nonfluencies of adult female stutterers. M.A. thesis, University of Iowa, 1957.
8. Egland, G. O., Repetitions and prolongations in the speech of stuttering and nonstuttering children, in Wendell Johnson (ed.), *Stuttering in Children and Adults*. Minneapolis: University of Minnesota Press, 1955.
9. Giolas, T., and Williams, D., Children's reactions to nonfluencies in adult speech. *J. Speech Hearing Res.*, 1, 1958, 86-93.
10. Guilford, J. P., *Psychometric Methods*. New York: McGraw-Hill, 1954.
11. Johnson, W., Measurements of oral reading and speaking rate and disfluency of college-age male and female stutterers and nonstutterers. *J. Speech Hearing Dis.*, Monograph Supplement 7, 1961, 1-20.
12. Johnson, W., *Stuttering in Children and Adults*. Minneapolis: University of Minnesota Press, 1955.
13. Kools, J., Analysis of recorded speech samples. Summarized in chapter 8 in Wendell Johnson, *The Onset of Stuttering*. Minneapolis: University of Minnesota Press, 1959.
14. Lindquist, E. F., *Design and Analysis of Experiments in Psychology and Education*. Boston: Houghton Mifflin, 1953.
15. Mann, Mary B., Nonfluencies in the oral reading of stutterers and nonstutterers of elementary school age, in Wendell Johnson (ed.), *Stuttering in Children and Adults*. Minneapolis: University of Minnesota Press, 1955.
16. Sander, E. K., Reliability of the Iowa Speech Disfluency Test. *J. Speech Hearing Dis.*, Monograph Supplement 7, 1961, 21-30.
17. Sherman, Dorothy, and Trotter, W. D., Correlation between two measures of the severity of stuttering. *J. Speech Hearing Dis.*, 21, 1956, 426-429.
18. Sherman, Dorothy, Young, M. A., and Gough, K., Comparison of three measures of stuttering severity. *Proceedings of Iowa Acad. Science*, 65, 1958, 381-384.
19. Tuthill, C. E., A quantitative study of extensional meaning with special reference to stuttering. *Speech Monographs*, 13, 1946, 81-98.
20. Williams, D., and Kent, Louise, Listener evaluations of speech interruptions. *J. Speech Hearing Res.*, 1, 1958, 124-131.
21. Winitz, H., Repetitions in children's speech in the first two years of life. *J. Speech Hearing Dis.*, Monograph Supplement 7, 1961, 55-62.

Repetitions in the Vocalizations and Speech of Children in the First Two Years of Life

HARRIS WINITZ

This study deals with an investigation of the speech sound repetitions in the vocalizations and speech of children in the first two years of life. The data for this investigation were gathered by Irwin and his associates (*5*, *6*, 7, *8*, *9*, *10*, *11*), who recently conducted a longitudinal investigation of the speech sound utterances of 95 white children during the first two and one-half years of life.[1]

The subjects were from Iowa City homes and were considered to be physically normal. The median birth weight was seven and one-half pounds. With few exceptions the children were from monolingual homes. According to Irwin (*10*), 'These infants were from middle class homes, the parents being professional, business, clerical, and some laboring people.' The subjects were divided approximately equally with respect to sex and family constellation. With respect to family constellation 'only children' were defined as having no older siblings, while 'infants with siblings' had one or more older siblings. The age interval between the child and his nearest sibling was not considered. As the study progressed, some of the families moved away so that fewer subjects were available at the older age levels. Irwin calls 'most continuous data' those data obtained from the 35 subjects who were most continuously followed throughout the two and one-half year period. He refers to the data obtained from all 95 subjects as 'all data.'

The data were collected in the following manner. The spontaneous speech sounds of the infants (no attempt was made by the recorder to stimulate the child verbally) were transcribed on paper using the abbreviated International Phonetic Alphabet described by Fairbanks (*4*). Three consonants were added (not listed by Fairbanks): /x/, /ç/ and /ʔ/. Lip smacking sounds and 'clicking' sounds were omitted in the transcriptions. The sampling unit was 30 breaths; that is, the sounds which were made by an infant on 30 exhalations, not necessarily consecutive, were recorded.

Transcriptions were limited to noncrying sounds. A transcription of 30

[1]The data for this investigation were made available by Dr. O. C. Irwin, then Professor of Psychology, Iowa Child Welfare Research Station, University of Iowa, who also provided technical supervision for the analysis of the data presented in this report. The data for the first year of life were reported by Han P. Chen (2) in a doctoral dissertation directed by Dr. Irwin. Professor Irwin is Research Professor, Institute of Logopedics, University of Wichita.

Harris Winitz (Ph.D., University of Iowa, 1959) is Assistant Professor of Speech Pathology, University of Kansas. This article is based on an M.A. thesis completed under the direction of Professors Wendell Johnson and Orvis C. Irwin. This study was supported, in part, by the Louis W. and Maud Hill Family Foundation and by Grant RD 319 of the Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

breaths was called a 'record'; each record took approximately 30 minutes to collect. This sampling procedure was found to involve high observer agreement (*9*).

All the data were collected during the afternoon after the child had his noon meal. No attempt was made to control the infant's position during the making of a record, although almost all were sitting or held in an upright position. Usually one parent, the mother, was present. The data were grouped and reported in 15 two-month intervals, beginning with months one and two and ending with months 29 and 30. Age level one refers to months one and two, age level two refers to months three and four, etc.

Procedure

Since the original investigation did not restrict the recording of sounds to consecutive breaths, an analysis of repeated sounds and sound patterns was by necessity limited to the repetitions occurring in each instance within a single breath unit. In addition, an examination of the data indicated that in many instances only one or two sounds were uttered on an exhalation, precluding the possibility of repetition occurring in the case of one sound, and limiting the possibility of repetition in the case of two sounds to the two utterances of the same sound.

In this investigation breath patterns that contained repetition were regarded as *repeated breaths* and breath patterns that contained no repetitions as *nonrepeated breaths*. Thus, in this analysis the maximum number of repeated breaths that could occur on a given infant's record was 30. The number of nonrepeated breaths was always the converse of the number of repeated breaths. For example, a child who had 16 of his breaths classified as containing at least some kind of sound repetition would have 14 breaths that did not contain repetition.

TABLE 1. Number of subjects, means, standard deviations, and ranges of breaths containing repetition in the speech of infants at age levels one (months one and two) through twelve (months 23 and 24). Each mean is based on a total of 30 breaths.

Age Level	*Subjects*	*Records*	*Mean*	*SD**	*Range**
1	62	125	7.4		
2	80	181	7.4		
3	75	166	8.9		
4	64	170	9.4		
5	62	147	10.1		
6	62	149	12.1		
7	58	58	8.4	4.8	2-23
8	55	55	8.5	4.6	0-23
9	49	49	8.6	4.5	2-22
10	43	43	7.0	3.6	2-14
11	38	38	6.0	3.6	0-16
12	31	31	5.3	3.6	0-16

*Not reported by Chen (*2*).

Subjects. The number of subjects and the number of records analyzed are given in Table 1. In the first year of life approximately three records were gathered per infant, the analysis being in terms of the weighted average of the available records. In the second year of life only one record per subject was analyzed. When more than one record was available for a child, the record used for analysis was selected randomly. The sexes were grouped in this study.

Classification of Breath Patterns. Each of the breaths of each record was classified in one of the indicated categories.

Vowel Patterns. The following categories were used for classifying the breaths consisting of vowels only. The vowel patterns were divided into five repeated and nonrepeated patterns:

(a) Nonrepeated patterns are patterns in which vowel or vowel patterns are not repeated, such as /ɛɪʌ/.

(b) Partial repetition of vowels are patterns in which one or more, but not all, of the vowels appear two or more times in succession, such as /ɛɛɪ/.

(c) Complete repetition of vowels are patterns in which only one vowel occurs and it is produced repetitiously, such as /ɛɛɛ/.

(d) Syllabic repetition of vowels is a pattern in which the same vowels occur more than once as part of a repeated syllable, such as /ɛʌɛʌ/.[2]

(e) Varying syllabic repetition of vowels is a pattern in which a vowel recurs as part of a repeated syllable in which one or more of the other phonemes vary as the syllable is repeated, such as /ɛʌɛɪ/.[2]

Mixed Patterns. The following categories were used for classifying the breaths that contained both vowel and consonant phonemes. The mixed patterns were divided into the following repeated and nonrepeated patterns.

(f) Nonrepeated patterns are patterns in which sounds or sound patterns are not repeated, such as /bʌdim/.

(g) Phonemic repetition of sounds is a pattern in which one or more, but not all, of the sounds appear two or more times in succession, such as /bɪɪ/ or /ssɛ/.

(h) Syllabic repetition of sounds is a pattern in which the same sounds occur more than once as part of a repeated syllable, such as /bɑbɑ/.

(i) Syllabic repetition of sounds with an added phoneme is a pattern in which the same sounds occur more than once as part of a repeated syllable, and, in addition an added phoneme is attached to the final syllable, such as /bɑbɑm/.[3]

(j) Varying syllabic repetition of sounds is a pattern in which a sound recurs as part of a repeated syllable in which one or more of the other phonemes

[2]This classification was not included by Chen in his analysis, presumably because a repetition of this type did not occur in the first year of life.

[3]This classification was not included by Chen in his analysis, presumably because a repetition of this type did not occur in the first year of life.

vary as the syllable is repeated, such as /babi/.

(k) Varying syllabic repetition of sounds with an added phoneme is a pattern in which a sound recurs as part of a repeated syllable in which one or more of the other phonemes vary as the syllable is repeated, and in addition an added phoneme is attached to the final syllable, such as /babim/.[4]

(l) Mixture repetition of sounds is a pattern consisting of one of the following combinations:
  (1) Phonemic repetition and syllabic repetition, such as /ɪɪhihi/.
  (2) Phonemic repetition and varying syllabic repetition, such as /ɪɪhiha/.
  (3) Syllabic repetition and varying syllabic repetition, such as /bababi/.

(m) Two-syllable repetition of sounds is a pattern in which the same syllables occur more than once as part of a combination of repeated syllables, such as /bɪzubɪzu/.[5]

In the analysis diphthongs, blends, and affricates were considered by Winitz (*16*) as single units or 'phonemes' since these combinations of sounds tend to be heard as elemental units.

Linguistic Patterns. The breath patterns containing words were identified as linguistic patterns. Linguistic patterns are composed of standard words, word approximations, and mixed patterns (standard words and word approximations).

The terms *word approximation* and *standard word* have been defined by McCurry and Irwin (*15*). 'A word approximation is defined as a phonetic pattern which is interpreted by the observers at the time of the transcription as an attempt by the infant to pronounce a standard word. The word approximation is further delimited as a phonetic pattern in which one or more of the phonetic elements of the standard word, either vowel or consonantal elements, are present. This means that some elements of the standard pattern are omitted, and other elements are substituted or added.' They have defined a *standard word* 'in terms of its phonetic listing in Kenyon and Knott's *Pronouncing Dictionary of American English*.'

The following exceptions have been made with respect to the definition of a standard word. The present investigators felt the following exceptions were justified because many words used in the home by parents and learned by children are not of the exact phonetic structure as listed by Kenyon and Knott. The following are examples of such exceptions:

The following form of daddy was accepted: /dædi/.
The following forms of mother were accepted /mami/, /mama/, /mamʌ/, /ma/, /mam/, and /mamə/.
The following forms of baby were accepted: /beɪbi/, and /bebi/.
Words, as doggy, choo-choo, bow wow, tick-tick, moo moo, etc., were considered standard words, when they were of the correct phonetic

[4]This classification was not included by Chen in his analysis.
[5]This classification was not included by Chen in his analysis.

structure. If these words were not of the correct phonetic structure, they were considered word approximations.

Certain words commonly found in infant utterances were not considered as instances of repetition. Examples of these are: tick-tick, choo-choo, bow wow, moo moo, mama.

The linguistic patterns were divided into the following patterns:[6]

(n) Standard word patterns are patterns in which there is at least one standard word, such as /bɔl/.

(o) Standard word repeated patterns are patterns in which two or more standard words, which mean the same, are present and appear in succession, such as /bɔl bɔl/.

(p) Word approximation patterns are patterns in which there is at least one word approximation, such as /bɔ/.

(q) Word approximation repeated patterns are patterns in which two or more word approximations, which mean the same, are present and appear in succession, such as /bɔbɔ/.

(r) Mixed patterns are patterns in which there is at least one word approximation and one standard word, such as /bɔʌp/.

(s) Mixed repeated patterns are patterns in which a word approximation and a standard word, which mean the same, are present and appear in succession, such as /bɔbɔl/.

Reliability. In order to establish scorer reliability Winitz (*16*) selected 100 records at random and classified each breath twice (with an average interval of approximately one month) according to the previously indicated categories. The formula used was: Agreement Index $= a/(a + d)$ in which $a =$ agreements and $d =$ disagreements. There were 2939 agreements and 61 disagreements. The obtained index of agreement was 98%.

Results and Discussion

The means, standard deviations and ranges for all categories of repeated breaths combined for the first two years of life are shown in Table 1; the means for each repeated and nonrepeated category for all age levels is presented in Table 2. It is apparent from the analysis of the speech sound productions of children during the first two years of life that repetitions are characteristic of their infant vocalizations and early speech. There appears to be a peak of breath patterns containing repetition at the age of one year and a gradual decrease from the end of the first year through the second. This decrement may be attributed, in part, to the large number of words uttered in the second year of life. Generally a child at this age level does not utter more than one or two words on an exhalation. Thus, the decreasing number of repetitions (exhalations characterized by repetitions) may possibly be an artifact of the method employed. Some slight differences between the methods used by Chen and Winitz in classifying breaths might also have accounted for some of the differences.

Stuttering. The results of this study indicate that repetitions are typical of sound production during the first two years of life and do not begin with the

[6]These classifications were not included by Chen in his analysis.

TABLE 2. Means for each repeated and nonrepeated category in the speech of infants at age levels one (months one and two) to 12 (months 23 and 24). Each mean is based on a total of 30 breaths.

Categories	*Age Levels*											
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Vowels												
a*	15.7	14.4	12.3	11.6	9.7	7.1	7.7	5.9	5.6	4.8	4.3	3.1
b	1.1	1.4	1.7	1.6	1.5	1.0	.4	.3	.6	.3	.1	.1
c	3.7	2.3	2.4	2.0	2.1	1.5	1.7	1.4	1.4	0	0	0
d†							0	0	0	0	0	0
e							0	0	0	0	0	0
Mixed												
f	6.8	8.0	8.3	8.0	8.7	9.3	13.2	14.2	13.6	13.8	12.6	12.9
g	1.3	1.8	2.0	2.0	1.8	2.4	1.1	1.4	1.0	.9	.4	.3
h	1.0	1.2	1.5	2.2	2.3	3.2	2.5	2.3	2.1	2.2	1.7	1.5
i							.3	.3	.4	.3	.4	.3
j	.2	.6	.9	1.2	1.7	2.8	1.6	2.0	2.0	1.7	1.8	1.8
k							.3	.3	.3	.4	.3	.4
l	.1	.2	.4	.5	.7	1.2	.4	.4	.5	.4	.4	.1
m							0	0	0	0	0	0
Linguistic												
n							.1	.4	.9	1.2	2.7	2.7
o							0	0	0	0	0	.1
p							.5	1.0	1.2	3.0	3.8	5.2
q							0	.1	0	.7	0	0
r							0	0	.1	.3	.7	.9
s							0	0	0	0	0	0

*a, no repetition; b, partial repetition; c, complete repetition; d, syllabic repetition; e, varying syllabic repetition; f, no repetition; g, phonemic repetition; h, syllabic repetition; i, syllabic repetition with added phoneme; j, varying syllabic repetition; k, varying syllabic repetition with added phoneme; l, mixed sound repetition; m, two syllable repetition; n, standard word; o, standard word repetition; p, word approximation; q, word approximation repetition; r, mixed word; s, mixed word repetition. See text for more complete descriptions.

†Chen (*2*) did not consider categories d, e, i, k, and m through s.

advent of linguistic or codified speech. This finding would appear to have particular meaning in relation to Johnson's theory of the onset of stuttering (*14*).

Knowing that stuttering is almost always originally diagnosed by a parent or layman and that the original diagnosis is usually made when the child is between the ages of two and four years (*3, 12, 14*), one might be tempted to assume that significant amounts of repetition also begin to occur between the ages of two and four, or at the time the child begins to say words. In fact, parents sometimes report that their child began to 'stutter' (to be repetitious) when he began to 'talk.' The results of this study indicate, contrary to such reports, that repetitions occur in the vocalizations and speech of infancy and early childhood from the first month through the first two years. Johnson and his co-workers have shown that repetitions in speech continue to be common between the ages of two and eight years (*12, 14*). It appears, then, that parents who report that what they mean by the beginning of stuttering is the same as the beginning of repetitions in their child's speech, and that this occurred when the child 'started to talk' or at the age of three or four years, are implying that they had not noticed the earlier repetitions and that only at some point in the course of the child's speech development did they perceive them and evaluate them as worthy of attention and concern.

Language Development. It has been noted by students of language development that the repetition of sounds and sound patterns may play an important role in language learning. Berry and Eisenson (*1*) report as follows:

> Lalling, which usually begins during the second six months of the child's life, may be defined as the repetition of *heard* sounds or sound combinations. The great significance of lalling is that hearing and sound production have become associated. The seemingly endless repetition of 'ba-ba' or 'ma-ma' or 'gub-gub' affords the child, if not all listeners, considerable satisfaction.

An examination of Tables 1 and 2 would seem to support this hypothesis. A noticeable increase in repeated breath patterns is observed from age levels three to six. More intensive investigations of the phenomenon of infant speech sound repetition might lead to improved understanding of the processes of speech and language development.

Summary

The repetition of speech sounds in the vocalizations and speech of children in the first two years of life was investigated. The data were provided by research conducted by Professor O. C. Irwin in the Iowa Child Welfare Research Station. The results of this study indicate that from one-fourth to one-third of the vocalized breath exhalations during the first two years contain repetitions of sounds or sound patterns.

The findings of this study would appear to have particular relevance to Johnson's theory of the onset of stuttering. The data indicate that repetitions do not begin to occur at the time children begin to employ conventional codified spoken language, or at some later age. Where they are reported by certain parents as having originally appeared at this time, generally between the ages of two and five years, it would seem necessary to assume that the parents are reporting the beginning, not of repetitiveness in the child's speech, but of their perceptual and evaluative responsiveness to it.

References

1. Berry, M. F. and Eisenson, J., *Speech Disorders; Principles and Practices of Therapy*. New York: Appleton-Century-Crofts, Inc., 1956.
2. Chen, Han P., Speech development during the first year of life, a quantitative study. Ph.D. dissertation, University of Iowa. 1946.
3. Darley, F. L., The relationship of parental attitudes and adjustments to the development of stuttering. In W. Johnson (Ed.), *Stuttering in Children and Adults*. Minneapolis: Minnesota Press, 1955.
4. Fairbanks, G., *Voice and Articulation Drillbook*. New York: Harper & Brothers, 1940.
5. Irwin, O. C., Infant speech: consonantal sounds according to manner of articulation. *J. Speech Dis.*, 12, 1947, 402-404.
6. Irwin, O. C., Infant speech: consonantal sounds according to place of articulation. *J. Speech Dis.*, 12, 1947, 397-401.

7. Irwin, O. C., Infant speech: development of vowel sounds. *J. Speech Hearing Dis.*, 13, 1948, 31-34.
8. Irwin, O. C., Infant speech: variability and the problem of diagnosis. *J. Speech Dis.*, 12, 1947, 287-289.
9. Irwin, O. C., Reliability of infant speech sound data. *J. Speech Dis.*, 10, 1945, 227-235.
10. Irwin, O. C. and Chen, Han P., Development of speech during infancy: curve of phonemic types. *J. exp. Psychol.*, 36, 1946, 431-436.
11. Irwin, O. C. and Chen, Han P., Infant speech sounds and intelligence. *J. Speech Dis.*, 10, 1945, 293-296.
12. Johnson, W., A study of the onset and development of stuttering. In *Stuttering in Children and Adults*. Minneapolis: W. Johnson (Ed.), Univ. Minnesota Press, 1955.
13. Johnson, W., Measurement of oral reading and speaking rate and disfluency of college-age male and female stutterers and nonstutterers. *J. Speech Hearing Dis.*, Monograph Supplement 7, 1961, 1-20.
14. Johnson, W., and Associates, *The Onset of Stuttering*. Minneapolis: Univ. Minnesota Press, 1959.
15. McCurry, W. H., and Irwin, O. C., A study of word approximations in the spontaneous speech of infants. *J. Speech Hearing Dis.*, 18, 1953, 133-139.
16. Winitz, H., A quantitative study of the repetitions of children's speech in the second year of life. M.A. thesis, Univ. of Iowa, 1956.

A Study of the Speech Behavior of Stutterers and Nonstutterers under Normal and Delayed Auditory Feedback

JAMES N. NEELLEY

The performance of stutterers under delayed auditory feedback (DAF) has not been studied extensively. Ham (*6*), as a part of a larger study, examined the effects of DAF on normal speakers and stutterers. Under the four different delay conditions which he used, stutterers were significantly different from the normal group on only one measure under only one delay condition, average speech power at 0.1 second delay. Neely (*18*) investigated the effect of DAF on the rate and sound pressure level of the speech of stutterers following oral practice at three different delay times. Each delay time was used as a practice time, each practice session being followed by testing at all three delay times. In the main, the stutterers performed essentially like a group of nonstutterers which Neely had studied in a preliminary study. These studies by Ham and Neely seem to suggest that under DAF stutterers may perform like nonstutterers.

Speech produced under the condition of DAF is characterized by repetitions of sounds and syllables and a slowing down of speaking rate. Perhaps it was these effects which prompted Lee (*15*) to employ the term 'artificial stutter' in referring to the disturbances caused by temporal alterations in feedback. Not only did Lee use the word 'stuttering' in talking about the phenomena he observed, but he seemed to imply a connection between DAF and clinical stuttering, at least the type he referred to as 'phoneme repeating' stuttering. Chase (*2*) expressed the belief that the facilitating effect of delay on the repetition of speech sounds might be responsible for 'some types of stuttering.' Azzi (*1*) stated that the stutterer may have a 'deficient nervous circuit' which creates, it would seem the author meant, a condition in the stutterer which can delay feedback. Stromsta

---

James N. Neelley (Ph.D., University of Iowa, 1960) is Assistant Director, Speech and Hearing Clinic, University of Kansas. This article is based on a doctoral dissertation completed in the Department of Speech Pathology and Audiology at the University of Iowa under the direction of Professor Wendell Johnson. The assistance of Dr. Martin Young in connection with statistical analyses made in preparation of the dissertation is acknowledged with appreciation. The findings reported here both supplement and qualify the conclusions of the dissertation. These findings include the results obtained by comparison of the stutterers' speech behavior under normal and auditory feedback, and by application to the differences between specified speech samples of tests appropriate for nonhomogeneous and skewed distributions. These tests were applied and related analyses and evaluation of the data were carried out with the indispensable counsel and generous assistance of Professor Merle W. Tate.

This study was supported, in part, by the Louis W. and Maud Hill Family Foundation and by Grant RD 319 of the Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

(*23*) has reported data which he feels suggest that there are differences between stutterers and nonstutterers with respect to the phase angle of bone-conducted sound, the phase angle in the stutterers of his study being such that a condition comparable to a delay in external sidetones might be inferred to exist within them. In brief, some observers have noted or implied similarities between stuttering and speech produced under DAF and some have entertained the hypothesis that stuttering may be related to a built-in system providing more than normal delay.

There are no published accounts of attempts to compare the speech of stutterers and nonstutterers under conditions of DAF when the criteria are those which Fairbanks (*3*) has called omission, substitution, addition, and correct word rate (CWR). Likewise, there have been no attempts to compare the two groups under DAF with respect to listener ratings of speech disturbance.

Further, if stuttering behavior is related to the factors which cause speech disturbances in normal speakers under DAF, behavioral trends might be found in the speech of nonstutterers produced under DAF which would resemble certain stuttering phenomena. This is to say, if the speech of stutterers produced under the condition of normal feedback were compared with the speech of nonstutterers under the condition of DAF, such a comparison would indicate similarities or differences in behavior. The measures to be considered in comparing stuttering and the effects of DAF are: (a) adaptation, (b) consistency, and (c) certain listener data.

Adaptation in stuttering, as originally described by Johnson and Knott (*14*), refers to a reduction in the frequency of the stuttering response as a passage is read orally several times consecutively. This reduction amounts, on the average, to 20% when a passage is read two times, to almost 40% when read three times, to over 40% when read four times, and to approximately 50% when read five times (*13, 21*). This adaptation phenomenon has been confirmed repeatedly in experimental studies.

The usual procedure followed in measuring adaptation is that as the subject reads from a copy of the reading passage, the experimenter, using a duplicate copy of the reading passage, marks the words on which stuttering is judged to have been done during each reading. The number of consecutive readings is usually five. Percentage of adaptation is calculated by the formula, $(x\text{-}y)(100)/x$, where x is the number of stuttered words in Reading 1 and y is the number of stuttered words in a specified reading subsequent to Reading 1.

Tiffany and Hanley (*25*) in an investigation of adaptation to delayed sidetone had 20 subjects, who had been rated on *general speech effectiveness*, read a 45-word prose passage 12 times with a feedback delay time of 0.18 second. One week later the same subjects read the same material again under identical conditions. There did not seem to be significant adaptation in terms of 'fluency breaks' during either set of 12 readings. However, the second set of 12 readings had about 60% fewer breaks than the first set which had been obtained a week earlier. The authors said that readers may learn to avoid the fluency breaks when speaking under DAF, but that the reduced reading rate may remain practically constant.

Another observation, originally made by Johnson and Knott (*14*), is that there is a tendency for the words stuttered in any given reading of a series of readings of a passage to be among those stuttered in previous readings. Johnson and Knott called this phenomenon the consistency effect. The percentage of consistency between two readings of a passage is calculated by the formula, (Number of words stuttered in Reading X that were previously stuttered in Reading Y/ Number of words stuttered in Reading X) 100.

Johnson and Inness (*13*) reported a mean consistency of 66% in the combined Johnson and Knott (*14*) and Maddox (*17*) data when each reading was compared with the one before it. This observation, that approximately two-thirds of the words stuttered in a given reading will have been stuttered in the previous reading, has been reaffirmed in many experiments.

The last of the measures to be considered involves certain listener data. If speech samples from stutterers and nonstutterers, the former speaking with NAF and the latter with DAF, are randomized as to order and presented to listeners, can the listeners distinguish the samples of the stutterers from those of the nonstutterers? In addition, can a group of listeners determine when a stutterer is reading under NAF and when he is reading with DAF? If the required discriminations can be made with a degree of correctness better than chance would permit, it would seem to be indicated that stuttering and the effects of DAF are perceptibly different.

It may be assumed that stutterers are probably able to make a substantial contribution toward an answer to the question, 'Is speaking with delayed auditory feedback different from or similar to stuttering?' Requesting stutterers to discuss the similarities and dissimilarities between their stuttering and their speech under DAF provides a peculiarly direct experiential comparison of these two types of speech behavior.

Problem

The purposes of the present study were:

(1) To compare the speech behavior of stutterers and nonstutterers under the condition of delayed auditory feedback with reference to
   (a) frequency of omissions, substitutions, and additions
   (b) correct word rate expressed as number of words spoken per second without error in articulation or pronunciation
   (c) listener ratings on a speech disturbance scale

(2) To compare the speech of stutterers under the condition of normal auditory feedback and the speech of nonstutterers under the condition of delayed auditory feedback with reference to
   (a) the decrement in frequency of error words (words spoken with specified distortion or error) over several readings of a passage (adaptation effect)

(b) the tendency to make errors on the same words from reading to reading of the same passage (consistency effect)
(c) listeners' ratings of the degree of disturbance in recorded speech samples
(d) listeners' discriminations of stutterers' speech samples under normal auditory feedback and nonstutterers' speech samples under delayed auditory feedback
(e) listeners' discriminations of stutterers' speech samples under normal auditory feedback and under delayed auditory feedback

(3) To obtain from stutterers verbal reports of their comparisons of their speech under delayed auditory feedback and their stuttered speech under normal feedback conditions.

(4) To compare the speech of stutterers under the conditions of normal auditory feedback and delayed auditory feedback with reference to
(a) frequency of omissions, substitutions, and additions
(b) correct word rate expressed as number of words spoken per second without error in articulation or pronunciation
(c) the decrement in frequency of error words (words spoken with specified distortion or error) over several readings of a passage (adaptation effect)
(d) the tendency to make errors on the same words from reading to reading of the same passage (consistency effect)

Procedure

Subjects. One group of subjects was composed of 23 male adults, ranging in age from 17 to 36 years, who were considered by themselves and their associates to be stutterers. Another group of subjects was made up of 23 male adults, ranging in age from 19 to 32 years, who were considered by themselves and their associates to be normal speakers. None of the subjects had a speech problem aside from stuttering in the group of stutterers. All subjects had normal hearing and had had no previous experience with DAF.

Equipment and Recording Procedure. The equipment used in the experiment consisted of (1) an adjustable dental chair with head rest, (2) a Turner Dynamic Microphone, Model 221, (3) an Ampex Magnetic Tape Recorder, Model 35OC, (4) a feedback amplifier built by the experimenter, and (5) a pair of PDR-8 earphones. The subject did the experimental readings while sitting in the dental chair, the head rest being adjusted so that he could maintain a distance of 6 inches from the microphone. The Ampex tape recorder was used to record the speech samples under both feedback conditions, and this recorder was also used to delay the feedback in the DAF experimental condition, the delay being approximately 0.14 second.

At the first experimental reading session the subject first read a 100-word practice passage in order to permit the experimenter to adjust the record level. Then with normal feedback and without the earphones on, the subject read the 100-word experimental passage five times.

Twenty-four hours later the subject returned to do the five readings of the same experimental passage with DAF. The decision to use the same reading passage at the second experimental reading was justified by previous demonstrations of the phenomenon of spontaneous recovery (*4, 9*). Before the experimental readings were done, however, the subject again read the practice passage without the earphones on. As the subject read the passage, record level and playback level were adjusted so that there was peaking between -10 and -7 VU at both the record setting and the playback setting.

After this adjustment was made, the subject read the practice passage with the earphones on, with no delay in the feedback, and with the attenuator on the feedback amplifier set at its maximum setting (least amount of attenuation). (Appropriate measurements showed that with the record and playback levels set as described above and with the feedback amplifier at the maximum setting, the level of the speech in the earphones was approximately 75 db above normal threshold for spondee words.) As recommended by Schwartz (*19*), the subject was asked to adjust the potentiometer of the feedback amplifier as he read until the sound of his voice was at 'a loud but comfortable level.' This setting was maintained for all subsequent readings.

Then the subject read the passage one more time, this reading with DAF. Fairbanks (*3*) and Schwartz (*19*) permitted this pre-experimental experience and it seemed necessary to the present experimenter to permit it because of the initial 'shock' of DAF to the naive subject.

After this reading of the practice passage, the subject read the experimental passage five times under DAF, following which the stuttering subjects were interviewed regarding their experience in speaking with DAF. These interviews were tape recorded.

Abstracting the Data from the Tapes. The instances of omissions, substitutions, and additions for the appropriate experimental readings were ascertained by following generally the method of Fairbanks (*3*), with modifications by the present experimenter. First, a broad phonetic transcription of the 100-word experimental reading passage was made. It was broad in that some of the words in the passage had more than one permissible articulation; for example, the third word in the passage, *really*, and the fifth word, *member*, each had four permissible articulation patterns. Any deviations in the speech of the subject from the standard phonetic transcription constituted an instance of error. All phonetic elements in the reference transcription were regarded as potential sources of error, but when the same type of error occurred in consecutive elements, only one instance of error was counted. Errors of different types in consecutive phonetic elements were counted as separate instances, however. Insertions, regardless of the length of the inserted material, were each counted as one instance. In cases where two errors occurred (e.g., a repeated syllable with one phonetic element omitted), only one instance of error was counted. An *error word* was a word which contained any portion of an instance of error; conversely, a *correct word* was a word which contained no portion of an instance of error. The ratio of correct words to total reading time in seconds furnished the *correct word rate*.

These data were obtained from all five readings of the passage by stutterers under NAF and DAF, and from all five readings of the passage by nonstutterers under DAF. The experimenter listened as many times as necessary to each recording requiring analysis and marked the instances of error and categorized them on a mimeographed copy of the reading text.

To measure his reliability in marking instances of errors and loci of errors, the experimenter randomly selected five of the stuttering subjects and again analyzed their five experimental readings under NAF and DAF in the manner described above. At least a week intervened between the first and second analyses. The index of agreement, according to Tuthill's *(26)* formula, was .95 for the NAF readings; for the DAF readings it was .90; for all readings together it was .93.

The reading time for the appropriate experimental readings was determined through stop-watch measurements of the recorded speech samples.

Three tapes were assembled for presentation to listening groups. One tape was assembled which contained the first 57 words of the first reading of each stutterer and nonstutterer under DAF. The order of the recordings was randomized. A group of 25 listeners was instructed to rate each of the 46 speech samples on a nine-point 'speech disturbance' scale, point *one* on the scale representing 'little speech disturbance' and point *nine* representing 'great speech disturbance.'

A second tape was assembled which contained the speech of stutterers under NAF and the speech of nonstutterers under DAF. The speech sample for each stutterer was the first 57 words of the first reading under NAF and the speech sample for each nonstutterer was the first 57 words of the first reading under DAF. The order of the speech samples was randomized. The tape was played to 18 listeners who had had clinical experience with stutterers and who were unacquainted with the speakers represented on the tape. The listeners were told that some of the speakers they were to hear would be stutterers while others would be nonstutterers reading under DAF. The listeners were asked to attempt to distinguish speech samples of the stutterers from those of the nonstutterers.

Forty-six speech samples, two from each of the 23 stutterers, were assembled on the third tape. One sample for each subject consisted of the first 57 words of the first reading under NAF and the other sample of the first 57 words of the first reading under DAF. The two speech samples of each stutterer were contiguous on the tape, but the order of the two samples and the order of the 23 stutterers were randomized. This tape was played to 20 listeners who were instructed to judge whether the second of the two readings of each subject was done under NAF or DAF.

Verbatim transcriptions were made of the interviews with the stuttering subjects, but only portions of the interviews were subjected to study.

Results

Stutterers and Nonstutterers under

TABLE 1. Medians and means (in parentheses) of distributions of omissions, substitutions, and additions of 23 stutterers and 23 nonstutterers in five DAF readings.

	Reading					
	1	*2*	*3*	*4*	*5*	*Total*
Stutterers						
Omissions	.46	.27	.22	.22	.38	1.64
	(.56)	(.52)	(.52)	(.35)	(.65)	(2.60)
Substitutions	.32	.22	.27	.32	.27	1.67
	(.74)	(.48)	(.48)	(.56)	(.43)	(2.60)
Additions	2.15	2.15	2.11	2.15	2.38	9.14
	(5.70)	(6.13)	(5.35)	(6.00)	(5.87)	(29.04)
Nonstutterers						
Omissions	.69	.32	.46	.18	.32	2.78
	(.91)	(.52)	(.74)	(.30)	(.56)	(3.03)
Substitutions	.38	.27	.22	.10	.10	1.42
	(.65)	(.52)	(.43)	(.30)	(.30)	(2.20)
Additions	3.06	2.71	3.68	1.96	2.38	13.43
	(4.00)	(4.26)	(4.43)	(3.39)	(3.65)	(19.74)

DAF. The stuttering and nonstuttering groups were compared with reference to the number of instances of omissions, substitutions, and additions, and with regard to correct word rate.

Omissions, Substitutions, and Additions. The medians and means of the distributions of omissions, substitutions, and additions of stutterers and nonstutterers in five readings under DAF are shown in Table 1. All distributions for both groups were severely skewed to the right. Twenty-three of the 24 distributions of omissions and substitutions had modes at zero. Nonparametric tests[1] indicated no significant differences between distributions with the exception of additions. The distributions of total additions and their quartiles are shown in Table 2. A chi square test for differences between the two distributions, with frequencies grouped in the classes 0-4, 5-9, 10-14, 15-29, and 30-257, indicated significance at the 1% level. It may be tentatively concluded that, although stutterers as a group make fewer additions than nonstutterers, the two groups are not comparable.[2]

Correct Word Rate. The distributions, means, and variances of correct word rate of stutterers and nonstutterers in five readings under DAF are shown in Table 3. When the rates were

[1]Nonparametric tests for differences between the speech samples of stutterers and nonstutterers under DAF included the rank-sum and median tests. The nature and application of these tests, as well as that of the sign and signed-rank tests referred to later, are discussed in Siegel (*22*) and Tate and Clelland (*24*).

[2]In a later section it will be seen that 15 of the 23 stutterers made fewer additions under DAF than under NAF. Thus, whether or not the groups are comparable here, the effect of DAF on additions of stutterers is clearly different from its effect on additions of nonstutterers. Since no nonstutterer made fewer additions under DAF, the proportions to be tested for difference are 15/23 and 0/23. These are significantly different at the .1% level.

TABLE 2. Distributions of total number of additions of 23 stutterers and 23 nonstutterers in five readings under DAF.

Number of Additions	*Stutterers*	*Non-stutterers*
0-4	5	4
5-9	7	2
10-14	3	7
15-19		1
20-24		5
25-29		1
30-34	3	1
35-39		
40-44	1	
45-49	1	
50-54		
55-59		
60-64		1
65-69	2	
.		
.		
.		
95-99		1
.		
.		
.		
255-259	1	
Q_1	5.04	8.88
Mdn	9.14	13.43
Q_3	33.25	22.75

classified for conventional analysis of variance, the differences between means of stutterers and nonstutterers and between means of the five readings were significant at the 10% level. However, analysis of covariance, with correct word rate under NAF the control variable, resulted in reduction of the difference between means of stutterers and nonstutterers to a clearly insignificant level ($P > .25$). If there is a real difference between mean correct word rate of stutterers and nonstutterers in reading under DAF, it apparently is accounted for by the difference in correct word rate under NAF.

However, the inference must be qualified. In a later section it will be seen that five stutterers showed higher correct word rate under DAF than under NAF. Further, the two groups were unequally variable under DAF. The distributions in Table 3 suggest that correct word rates of stutterers are more variable than those of nonstutterers. The difference in variances of the distributions of means is significant, by the F test, at the 4% level.

The within-self variability is also significantly greater for stutterers. The distributions of ranges of correct word rates are shown in Table 4. The distributions are significantly different at the 7% level, according to the nonparametric median test.

The differences in variability, although significant at the levels stated, are small, and the difference between means without control of initial frequency is small. It may be concluded that stutterers and nonstutterers, as groups, are not very different in correct word rate under DAF.

Listeners' Ratings of Speech Disturbance. In order to compare the stuttering and nonstuttering groups under DAF with regard to listener ratings on the speech disturbance scale, the median of the 25 ratings for each stuttering and nonstuttering subject was obtained and the mean of these median ratings was calculated. The mean for the stuttering group was 4.00 and the mean for the nonstuttering group was 3.70. The value of t which was calculated from the data was .42, $df = 44$; this t was not significant at the 10% level, suggesting that the speech disturbance under DAF, as measured

TABLE 3. Distributions, means, and variances of correct word rate of 23 stutterers and 23 nonstutterers in five readings under DAF.

		Frequency											
		Stutterers						Nonstutterers					
CWR	*Reading*	*1*	*2*	*3*	*4*	*5*	*Mean*	*1*	*2*	*3*	*4*	*5*	*Mean*
0- .24		1	1										
.25- .49				1	1	3	1						
.50- .74		1	1	1	2		1						
.75- .99		1		2	1	1	2				1	1	
1.00-1.24		4	4	3	1	1	2	2	1	1	1		1
1.25-1.49		2	2	2	3	4	3	1	2	2	2	3	3
1.50-1.74		3	3	2	2	1	1	5	4	3	2	2	3
1.75-1.99		2	2	4	3	2	3	3	4	5	3	3	3
2.00-2.24		3	5	2	1	3	4	3	4	5	4	6	5
2.25-2.49		2	2	2	4	5	2	6	4	2	4	3	3
2.50-2.74		3	2	3	4	1	3	2	3	4	4	2	4
2.75-2.99		1		1		1			1	1	2	3	1
3.00-3.24			1		1	1	1	1					
Mean*		1.71	1.72	1.72	1.78	1.77	1.74	2.00	2.03	2.04	2.07	2.06	2.04
Variance*		.452	.453	.468	.610	.590	.483	.224	.209	.199	.304	.248	.223

*Computed from ungrouped correct word rates

TABLE 4. Distributions of ranges of correct word rates of 23 stutterers and 23 nonstutterers in five readings under DAF.

	Frequency	
Range	*Stutterers*	*Nonstutterers*
.04-.11	2	
.12-.19	1	4
.20-.27	4	10
.28-.35	7	4
.36-.43	3	4
.44-.51	3	
.52-.59		
.60-.67		1
.68-.75	1	
.76-.83	1	
.84-.91		
.92-.99	1	
Median	.326	.255

by this scale, was perceived by the listeners to be essentially the same in the two speech groups.

Stutterers Under NAF and Nonstutterers Under DAF. Decrement of Frequency of Error Words over Five Readings of the Passage (Adaptation Effect). Using *error words,* as defined above, as the measure of DAF disturbance and stuttering 'disturbance,' the adaptation of nonstutterers under DAF was compared with the adaptation of stutterers under NAF.

Percentage of adaptation was calculated by the formula, $(x - y)(100)/x$, where x is the total number of error words in Reading 1 and y is the total number of error words in any specified reading subsequent to Reading 1.

The decrement of error words for stutterers under NAF and nonstutterers under DAF is expressed in Table 5 in percentage of adaptation over Readings 1 and 2, 1 and 3, 1 and 4, and

TABLE 5. Percentage of adaptation for specified readings for stutterers under NAF and nonstutterers under DAF.

	Readings			
	1-2	*1-3*	*1-4*	*1-5*
Stutterers, NAF	20.8	26.9	36.7	45.4
Nonstutterers, DAF	12.9	6.1	27.3	20.4

1 and 5. The difference between mean percentage adaptation in the two groups over the 1-5 readings, where adaptation is usually observed, is highly significant with $P < .005$ according to Lord's (*16*) shortcut *t* test.[3] Adaptation of stutterers under NAF is significantly greater than that of nonstutterers under DAF.

Consistency of Error Words. Using *error words* as evidence of speech disturbance, the speech of nonstutterers under DAF was compared with the speech of stutterers under normal feedback with respect to the percentage of consistency in the words on which errors were made from reading to reading of the same material.

TABLE 6. Percentages of consistency between specified readings for stutterers under NAF and nonstutterers under DAF.

	Readings			
	1-2	*2-3*	*3-4*	*4-5*
Stutterers, NAF	55.5	45.1	52.1	52.1
Nonstutterers, DAF	20.9	21.8	29.2	29.5

The percentage of consistency was calculated for Readings 1 and 2, Readings 2 and 3, Readings 3 and 4, and Readings 4 and 5. Table 6 summarizes these data. The difference between percentage consistency in the two groups is highly significant, P from the shortcut *t* test being less than .001 in each of the four comparisons. Stutterers are significantly more consistent under NAF than nonstutterers under DAF.

Analysis of Listener Judgments. Judgments were made by 18 listeners as to whether each of 46 speech samples was that of a stutterer (speaking under NAF) or a nonstutterer (speaking under DAF). The 18 listeners made a total of 828 judgments in evaluating the speech samples of the 46 speakers. Of the 828 judgments, 79% were correct. The binomial test[4] described by Walker and Lev (*23*) was used to ascertain whether 79% was significantly different from chance. The obtained *z* was 3.92, indicating that the correct judgments were significantly different from chance beyond the 1% level.

For the nonstuttering group 92% of the judgments were correct; for the stuttering group 66%. The former majority is significant at the 1% level and the latter at the 5% level according to the binomial test.

It is apparent that some samples were incorrectly judged. Of the 177 errors in judgment, 142, or 81%, were made by misclassifying speech samples of stutterers under NAF as nonstutterers under DAF, and only 35, or 19%, were made by mistaking the DAF speech of nonstutterers for the NAF speech of stutterers. The judges were much more inclined to mistake stuttering for the DAF speech of nonstutterers

[3]Tate and Clelland (*24*) illustrate several applications of the shortcut *t* test for both independent and related samples.

[4]The simple binomial test is not altogether appropriate here, since the population of judgments is stratified by listener and by speech sample. However, the error in using the binomial test is on the conservative side.

than to mistake the DAF speech of nonstutterers for stuttering.

An analysis was also made of data obtained by means of the tape in which the listeners were asked to judge whether the second of two contiguous speech samples was produced under NAF or DAF, the speakers being stutterers. The listeners were correct 93% of the time in judgments of speech samples produced under NAF, 94% of the time in judgments of speech samples produced under DAF, and 93% of the time in total judgments.

The Stutterers' Experience with DAF. The interviews with the stutterers following their experience with DAF were directed toward getting them to compare their own speech behavior which they called stuttering with their speech behavior experienced under DAF. The interview question of primary interest was, 'Is the kind of speech behavior you experienced under DAF the same kind of behavior as stuttering?'

Of the 23 subjects, 18 or 78% stated that they clearly felt the disturbances in their overt speech behavior under DAF were different from their stuttering behavior. The 95% confidence interval for the true proportion is .55 — .93, and it may be concluded that a majority of stutterers similar to those of the sample in hand would report clearly felt differences. The differences were described and explained in various ways by the subjects.

One subject, Subject 17, said that he felt that there was 'no difference' between his usual stuttering behavior and speech produced under DAF. The subject did notice, however, that his speech was different from its usual form in that it was 'kind of slow and drawn out' under DAF. It was the subjective impression of the experimenter that DAF affected only speaking time in this subject, not fluency. Although there was disfluent behavior in the subject's speech under DAF, it seemed to the experimenter that this behavior was essentially identical to the subject's usual manner of stuttering; therefore, his affirmative answer to the question, 'Is the kind of speech behavior you experienced under DAF the same kind of behavior as stuttering?' was not surprising. It is to be accepted, even so, with the qualification that a difference in the time dimension of speech was observed and reported by the subject.

Subject 2 observed himself performing speech behavior under DAF that he had never heard himself doing before; the experimenter, who knew the subject and his speech behavior well, concurred. Nevertheless, the subject seemed to consider this behavior, which the experimenter considered to be delay-caused behavior, to be a component of his habitual stuttering which he had never heard before and had heard on this occasion because of the amplified feedback through the headphones.

Subject 5 was inconsistent in his answers to the interviewer's questions and his responses were difficult to interpret. In general, the subject stated that speaking under delay was a novel experience, but it was difficult for him to isolate the specific factors which made it a novel experience for him. The experimenter thought he observed repetitions under DAF which were not in the stuttering repertoire of this subject.

Subject 11 was so unaffected by delay, with the exception of the time factor, that the experimenter's question was not very meaningful. The subject felt that under DAF he had done nothing that resembled stuttering.

Subject 19 believed that he had performed some unusual behavior under DAF, but he was not able to differentiate this behavior in a clearcut way from stuttering. The best interpretation the experimenter could make of this subject's responses was that the subject heard behavior which sounded like stuttering, but not like his stuttering; to the extent that it sounded like stuttering to him, the subject reacted to it as though it was stuttering.

In summary, 18 of the 23 subjects stated unequivocally that they recognized differences between their usual stuttering and their DAF speech behavior, one subject reported that he could not recognize such differences except for a change in the time dimension of his speech under DAF, and the responses of four subjects were difficult to interpret for various reasons, but indicated that they made certain differentiations between their stuttering and their speech under the DAF condition.

Stutterers Under NAF and DAF. The speech of stutterers under the conditions of NAF and DAF was compared with regard to several measures.

Omissions, Substitutions, and Additions. The medians and means of the distributions of omissions, substitutions, and additions of stutterers in five readings under NAF and five under DAF are shown in Table 7, and the totals for each stutterer are listed in Table 8.

Seven of the differences in total omissions under NAF and DAF are 0; the remainder are negative. The probability, from the sign test, for the null

TABLE 7. Medians and means (in parentheses) of distributions of omissions, substitutions, and additions for 23 stutterers in five NAF and five DAF readings.

Condition	*Reading* 1	2	3	4	5	*Total*
NAF						
Omissions	.11	.11	.05	.05	.05	.14
	(.17)	(.22)	(.09)	(.13)	(.09)	(.70)
Substitutions	.18	.22	.22	.18	.22	.46
	(.43)	(.43)	(.39)	(.35)	(.48)	(2.08)
Additions	6.25	5.00	3.92	3.61	3.33	21.17
	(11.35)	(8.56)	(7.78)	(6.83)	(5.96)	(40.48)
DAF						
Omissions	.46	.27	.22	.22	.38	1.64
	(.56)	(.52)	(.52)	(.35)	(.65)	(2.60)
Substitutions	.32	.22	.27	.32	.27	1.67
	(.74)	(.48)	(.48)	(.56)	(.43)	(2.60)
Additions	2.15	2.15	2.11	2.15	2.38	9.14
	(5.70)	(6.13)	(5.35)	(6.00)	(5.87)	(29.04)

TABLE 8. Total omissions, total substitutions, and total additions of 23 stutterers in five readings under NAF and in five readings under DAF.

Subject	*Omissions*		*Substitutions*		*Additions*	
	NAF	*DAF*	*NAF*	*DAF*	*NAF*	*DAF*
1	0	0	2	0	58	257
2	0	2	0	1	39	31
3	1	3	0	1	1	2
4	0	1	0	1	25	0
5	0	3	4	2	36	7
6	0	0	3	2	3	0
7	0	7	5	6	1	14
8	0	7	1	6	26	33
9	0	3	1	0	17	9
10	0	1	0	1	1	0
11	5	9	2	0	5	7
12	3	7	17	12	8	6
13	0	0	0	2	13	4
14	2	2	0	3	12	10
15	0	1	0	2	256	6
16	0	0	5	5	138	49
17	0	0	0	0	73	68
18	0	3	6	9	90	42
19	0	0	0	0	59	65
20	0	3	0	3	10	34
21	5	6	2	1	40	5
22	0	1	0	4	0	11
23	0	1	0	1	20	8

hypothesis is less than .001. Stutterers make highly significantly more omissions under DAF than NAF.

Three of the differences for substitutions are 0, 13 negative, and 7 positive. The majority of stutterers make more substitutions under DAF than NAF, but the median difference is not significant at the 10% level according to the sign test and the ordinarily more powerful signed-rank test.

The effect of DAF on additions is in the opposite direction. The medians of the NAF and DAF distributions are 21.17 and 9.14, respectively. However, they are not significantly different at the 10% level by either the sign or signed-rank test. The most striking features about the distribution of differences are the relatively great variability and the fact that 15 stutterers made fewer additions, while eight made more under DAF.

TABLE 9. Mean correct word rate in correct words per second for 23 stutterers in the five readings under NAF and the five readings under DAF.

	Reading					
	1	*2*	*3*	*4*	*5*	*Total*
Stutterers NAF	1.93	2.18	2.25	2.25	2.32	2.18
Stutterers DAF	1.71	1.72	1.72	1.78	1.75	1.74

TABLE 10. Summary of analysis of variance of correct word rates classified by reading, condition, and subject.

Source	*df*	*ss*	*ms*	*F**	*F†.05*
Readings (A)	4	1.42	.36	7.20	2.45
Conditions (B)	1	11.28	11.28	8.95	4.30
Subjects (C)	22	120.33			
AB	4	.75	.19	4.75	2.45
AC	88	3.95	.05		
BC	22	27.63	1.26	31.5	1.70
ABC	88	3.44	.04		
Total	229	168.80			

**F* ratios: ms_{AB}/ms_{ABC}; ms_{BC}/ms_{ABC}; ms_A/ms_{AC}; ms_B/ms_{BC}.
†*F*.05 is the tabled value for the nearest given *df*.

Correct Word Rate. The mean correct word rate (CWR) in correct words per second under NAF and DAF are shown in Table 9. The results of an analysis of variance of the observed rates classified by Reading, Condition, and Subject are summarized in Table 10.

Both the Reading-Condition and the Condition-Subject interactions are significant, the latter being particularly marked; and the meaning of the significance of the main effects is clouded.[5] In general, DAF significantly lowers the rate of stutterers, but the generalization must be qualified. Examination of the individual CWR measurements for the 23 subjects showed that 18 of the subjects had a lower rate under DAF than they had under NAF, but that five of the subjects had a higher CWR under DAF than they had under NAF. There were no obvious homogeneous factors among the five subjects who increased their CWR under DAF. For example, two of these five subjects read with 'slight stuttering' under NAF, one with 'average stuttering,' and two with 'severe stuttering.' As a group, the stutterers were significantly more variable in correct word rate under NAF than under DAF ($P < .05$).[6]

[5]The test of main effects reported in Table 10 in neither logical nor dependable where interaction is so marked. However, the nonparametric signed-rank test showed differences in mean correct word rate under the two conditions significant at the 1% level.

TABLE 11. Percentage of adaptation for specified readings for stutterers under NAF and DAF.

	Readings			
	1-2	*1-3*	*1-4*	*1-5*
Stutterers NAF	20.8	26.8	36.7	45.4
Stutterers DAF	0	6.0	2.0	0

Decrement of Frequency of Error Words over Five Readings of the Passage (Adaptation Effect). The percentage of adaptation for the five readings of stutterers under DAF was calculated as described above. Table 11 gives the adaptation percentages for the NAF and DAF readings. Although average adaptation is significantly greater under NAF, any generalization must be qualified. In the four comparisons, seven, six, three, and two stutterers showed negative or zero adaptation under both conditions; five, five, six, and seven showed greater adaptation under DAF.

Consistency of Error Words. The percentage of consistency for the five DAF readings was calculated as described above. Table 12 gives the consistency percentages for the NAF and DAF readings. Average consistency

[6]Nonstutterers were significantly more variable at the 5% level under DAF.

TABLE 12. Percentages of consistency between specified readings for stutterers under NAF and DAF.

	Readings			
	1-2	*2-3*	*3-4*	*4-5*
Stutterers NAF	55.5	45.1	52.1	52.1
Stutterers DAF	37.7	39.3	33.1	32.7

is significantly greater under NAF, but, as in the case of adaptation, any generalization must be qualified. In the four comparisons, seven, seven, five, and six stutterers showed no consistency under either conditions; five, seven, five, and six were more consistent under DAF than under NAF.

Discussion

There are five questions to be considered. In reading under DAF, (1) Do stutterers and nonstutterers show equivalent articulatory errors and correct word rates, (2) Are stutterers and nonstutterers affected similarly in articulatory error and correct word rate, (3) Are the effects on stutterers homogeneous, (4) Are the effects on nonstutterers homogeneous, and (5) Is DAF speech behavior like stuttering behavior? The fourth question can be disposed of immediately. No nonstutterer showed fewer articulatory errors or higher word rate under DAF than under NAF.

Since stutterers are initially different from nonstutterers in articulatory error and correct word rate, an affirmative answer to the first question obviously would mean a negative answer to the second and vice versa, provided the effects on the two groups were homogeneous. The effects on stutterers were not homogeneous, however, and the first (and second) question cannot be answered without reservation. At the moment, group differences and similarities will be considered without reference to individual stutterer response.

The difference in average correct word rate of the two groups under DAF was of borderline significance, and the difference in variability of rate, although significant, was small. With respect to the articulatory errors of omissions, substitutions, and additions, the two groups differed significantly only in additions. It is evident that in correct word rate, omissions, and substitutions the performances of stutterers and nonstutterers, as groups, are not very different. These near similarities are in agreement with Ham's finding (*6*) that in the main stutterers reacted to DAF in somewhat the same way as nonstutterers.

Not only did the two groups appear to behave essentially alike with respect to omissions, substitutions and correct word rate; listeners could not detect differences between the speech disturbances in the two groups.

This finding that stutterers and nonstutterers perform similarly in some respects under DAF, in both a quantitative and perceptual sense, is of considerable importance as it relates to other findings about the similarities of stutterers and nonstutterers. Much of the basic research on stuttering in recent years has involved comparison of stutterers and nonstutterers with regard to various developmental and physical dimensions and behavioral criteria. The objective of such research was the detection of differences, if such differences existed, between stutterers and nonstutterers that might be related to the onset and development of the stuttering problem. The work of Johnson and his associates (*11*) indicated that when young stutterers were compared with young nonstutterers with respect to basic health data and various criteria of development, including speech development, the two groups appeared to be essentially similar. Goodstein's (*5*) and Sheehan's (*20*) reviews of the literature on the personality characteristics of stutterers and nonstutterers concluded with expressions of the belief that no outstanding differences between the two groups have been demonstrated. Likewise, Hill (7, *8*), in his reviews of studies designed to reveal physiological and biochemical differences between the two speech groups, concluded that no significant differences had been discovered. Williams' (*28*) myographic study provided no evidence to support the proposition that stuttering behavior is a symptom of neuromuscular disturbance or abnormality in stutterers. As of today it appears that there are no substantial differences between the stuttering and nonstuttering groups in these respects. The findings of the present study add to the growing list of ways in which stutterers and nonstutterers are observed not to differ significantly. Whatever the interactions and uncertainties in the present data, it seems clear that an adequate account of stuttering behavior—or the more comprehensive stuttering problem—is not to be found in the auditory feedback mechanisms.

One of the most interesting and possibly important findings of the present study was the lack of homogeneity of DAF effects on stutterers. A substantial minority of stutterers responded in respect to additions, substitutions, correct word rate, adaptation, and consistency in a direction opposite to that of the majority. Some of the differences were dramatic. For example, the total additions of one stutterer under NAF and DAF were 58 and 257, respectively; those of another, 256 and 6. The mean correct word rates of one stutterer were .30 under NAF and 1.86 under DAF; those of another, 1.26 and .28.

The differences invite further study. It may be that the DAF data of stutterers are so variable that they are essentially unrepeatable. Or it may be that responses under DAF in fact separate stutterers into two or more subgroups, or distribute them along a clinically important continuum. If this turns out to be the case, reading under DAF may prove to be a valuable tool in achieving a diagnostically more clear and differentiating description of the speech behavior of speakers classified as stutterers. It is reasonable to suppose, assuming reliable data, that the speech problems of stutterers who show, say, higher word rates under DAF than under NAF may be in some way different from those who show lower rates.

A considerable segment of this experiment was an exploratory investigation of the similarity or dissimilarity of stuttering behavior and DAF behavior. Some interesting observations should be noted.

Insofar as the decrement of frequency of error words was concerned, the stutterers in this experiment reading under NAF manifested a decrement. In fact, the error word adaptation percentages over the five readings under NAF were quite similar to stuttering adaptation percentages quoted in the literature. When the trend line of decrement of error words for nonstutterers under DAF was inspected and compared with that for stutterers under NAF, it appeared that the lines were dissimilar in several respects. While decrement of error words for stutterers over the five readings under NAF was somewhat linear, the decrement for the nonstutterers under DAF was rather erratic and inconsistent. The nonstutterers had almost as many error words on the third reading as they had on the first, and they had more errors on the fifth reading than on the fourth, not a characteristic of the stuttering group. Also it should be noted that by the fifth reading the stutterers under NAF had reduced the number of error words by 45%, while the reduction by the nonstutterers under DAF was only 20%. The difference was highly significant.

The percentage of consistency of error words shown by the stutterers under NAF in this experiment was not as high as had been reported for stuttering in the literature; nevertheless, the stutterers in this experiment, with a mean percentage of consistency under NAF of 51%, were about twice as consistent as the nonstutterers under DAF, who had a mean percentage of consistency of 25%, a highly significant difference.

That stuttering behavior and DAF speech behavior are presumed to sound alike has probably been the strongest basis for wondering if the one might have something in common with the other. However, listeners in the present experiment seemed to agree that stuttering, within the bounds of the experimental listening conditions employed, sounded different from DAF speech behavior. In the condition in which the listeners were asked to judge whether the speech samples were those of stutterers speaking under NAF or nonstutterers speaking under DAF, the listeners were significantly successful in making correct judgments. A greater percentage of the errors in judgment was made by misclassifying speech samples of stutterers under NAF as those of nonstutterers under DAF, although the proportion of the judgments of speech samples of stutterers that were correct was higher than would be expected by chance. This kind of error in judgment tended to occur most often on the speech samples of relatively fluent stutterers, as if the assumption of the judges was that the speaker must be a nonstutterer if he presents no unusual speech behavior. In the condition in which the listeners were asked to judge whether the second of two contiguous speech samples was produced under NAF or DAF, the speakers being stutterers, 93% of the judgments were correct. The data from these two listening conditions would seem to indicate that when the auditory aspects of stuttering behavior and DAF behavior are compared in more than a casual way, they sound more different than alike.

The work of Johnson (*11*) and his students suggests that it is not unusual for the term *stuttering*, which in the field of speech pathology is applied to a complex, yet distinctive, speech problem characterized in part by disfluent speech, to be employed as a label for other types of disfluent speech. The most inappropriate and disadvantageous usage of the word probably occurs when the generally observed disfluencies of childhood speech are referred to as 'stuttering' and reacted to as unusual or abnormal. Speech can be disfluent for several reasons, but disfluencies due to one cause may be qualitatively different from disfluencies due to another cause. The present data indicate that there are distinguishable differences between the speech behavior conventionally classified as stuttering and DAF speech behavior.

As a supplement to the objective findings of possible dissimilarity between stuttering and DAF speech behavior, the 23 stutterers in this experiment unanimously reported that there were experiential differences between the two speaking conditions, that their stuttering behavior is different from the experience of making errors under DAF. For 18 of the 23 subjects these reports of experiential differences were unequivocal.

The hypothesis that stuttering may be somehow related to a delay in auditory feedback because speech produced under conditions of delayed auditory feedback is assumed to behave like stuttering, to sound like stuttering, and to be an experience like stuttering, is discredited by the findings of this experiment.

Summary

Twenty-three adult stutterers and 23 adult nonstutterers read a 100-word passage five times under normal auditory feedback (NAF). Twenty-four hours later all subjects read the same passage five times under delayed auditory feedback (DAF). The delay time was 0.14 second.

The speech behavior of the two groups under DAF was studied with reference to the omission, substitution, and addition of sounds, and correct word rate (CWR) in seconds (*3*). There were no significant differences between the groups in omissions and substitutions. Although the stutterers made significantly fewer additions, the distributions of additions indicated that the groups were of doubtful comparability. The stutterers' correct word rate was lowei than the nonstutterers', but the difference was significant at only the 10% level.

Samples of speech produced under DAF were rated on a nine-point scale of 'speech disturbance.' The mean rating for the stuttering group was 4.00 and the mean rating for the nonstuttering group was 3.70. The difference between the means was not significant at the 10% level, suggesting that the 'speech disturbance' was perceived by the listeners to be essentially the same in the two groups.

These findings indicate that, in several important respects, the performances of stutterers and nonstutterers under DAF were not very different, if different at all.

The effects of DAF on substitutions, additions, correct word rate, adaptation, and consistency of stutterers were not homogeneous. It was suggested that the differences be studied further as a possible aid in achieving an increasingly clear and diagnostically discriminating description of the speech behavior of speakers classified as stutterers.

The speech behavior of stutterers under NAF was compared with the speech behavior of nonstutterers under

DAF with regard to the decrement of the frequency of error words over five readings of the passage (adaptation effect), the consistency of error words, and certain listener data. (An error word was a word in which any portion of an instance of omission, substitution, or addition of sounds occurred.) The error word adaptation percentages for the stutterers over the five readings under NAF were quite similar to stuttering adaptation percentages quoted in the literature. The error decrement for the nonstutterers under DAF, however, was erratic and significantly different from that of the stutterers under NAF. The mean percentage of consistency of error words for stutterers over the five readings under NAF was 51%, while that for the nonstutterers over the five readings under DAF was 25%, a significant difference.

Listeners were asked to distinguish the recorded speech samples of stutterers reading under NAF from those of nonstutterers reading under DAF. Of the 828 judgments made, 79% were correct; the proportion of correct judgments was significantly different, at the 1% level, from that to be expected by chance. Listeners were also asked to determine whether the stuttering subjects represented on a prepared tape were reading under NAF or DAF. The listeners were 93% correct in these judgments. These findings indicate that the auditorily perceptible aspects of stuttering behavior and speech behavior under DAF are different.

All stuttering subjects were interviewed regarding their experience speaking under DAF. Eighteen of the 23 subjects stated unequivocally that they recognized differences between their usual stuttering and their DAF speech behavior. One subject reported that he could not recognize such differences except for a change in the time dimension of his speech under DAF. The responses of four subjects were difficult to interpret for various reasons, but indicated that they made certain differentiations between their stuttering and their speech under the DAF condition. The stutterers' feeling that these were two different speaking conditions was substantiated in part by an objective comparison of their speech under NAF and DAF.

According to the measures used in this experiment, there are differences between stuttering behavior and speech behavior under the condition of delayed auditory feedback. The hypothesis that stuttering may be somehow related to a delay in auditory feedback, on the ground that speech produced under conditions of delayed auditory feedback appears to behave like stuttering, to sound like stuttering, and to be an experience like stuttering, is not supported by the findings of this experiment.

References

1. Azzi, A., Gli impieghi della voce ritardata in fonetica (The uses of the delayed voice in phonetics). *Rivista di Audiologia Practica*, 2, 1952, 75-85.
2. Chase, R. A., The effect of delayed auditory feedback on the repetition of speech sounds. *J. Speech Hearing Dis.*, 23, 1958, 583-590.
3. Fairbanks, G., Selective vocal effects of delayed auditory feedback. *J. Speech Hearing Dis.*, 20, 1955, 333-346.
4. Frick, J. V., Spontaneous recovery of stuttering behavior as a function of the degree of adaptation. M.A. thesis, University of Iowa, 1949.
5. Goodstein, L. D., Functional speech disorders and personality: a survey of the

research. *J. Speech Hearing Res.*, 1, 1958, 359-379.

6. Ham, R. E., Certain effects on speech of alterations in the auditory feedback of normal and defective speakers. Ph.D. thesis, Purdue Univ., 1956.
7. Hill, H., Stuttering: I. A critical review and evaluation of biochemical investigations. *J. Speech Dis.*, 9, 1944, 245-261.
8. Hill, H., Stuttering II. A review and integration of physiological data. *J. Speech Dis.*, 9, 1944, 289-324.
9. Jamison, Dorothy J., Spontaneous recovery of the stuttering response as a function of the time following adaptation. In Wendell Johnson (Ed.), *Stuttering in Children and Adults*. Minneapolis: Univ. of Minnesota Press, 1955.
10. Johnson, W., *People in Quandaries*. New York: Harper, 1946.
11. Johnson, W. and associates, *The Onset of Stuttering*. Minneapolis: Univ. of Minnesota Press, 1959.
12. Johnson, W., Brown, S. F., Curtis, J. F., Edney, C. W., and Keaster, J. *Speech Handicapped School Children*, rev. ed. New York: Harper, 1956.
13. Johnson, W. and Inness, Marjorie, Studies in the psychology of stuttering: XIII. A statistical analysis of the adaptation and consistency effects in relation to stuttering. *J. Speech Dis.*, 4, 1939, 79-86.
14. Johnson, W. and Knott, J. R., Studies in the psychology of stuttering: I. The distribution of moments of stuttering in successive readings of the same material. *J. Speech Dis.*, 2, 1937, 17-19.
15. Lee, B. S., Artificial stutter. *J. Speech Hearing Dis.*, 16, 1951, 53-55.
16. Lord, E., The use of range in place of standard deviation in the *t* test. *Biometrika*, 34, 1947, 41-67.
17. Maddox, J. S., Studies in the psychology of stuttering: VIII. The role of visual cues in the precipitation of moments of stuttering. *J. Speech Dis.*, 3, 1938, 90-94.
18. Neely, K. H., The effect of oral practice in the presence of different conditions of side-tone upon the rate and sound pressure level of the speech of a group of stutterers. Ph.D. thesis, Ohio State Univ., 1951.
19. Schwartz, R. J., Vocal responses to delayed feedback in congenitally blind adults. Ph.D. thesis, Purdue Univ., 1957.
20. Sheehan, J. G., Projection studies of stuttering. *J. Speech Hearing Dis.*, 23, 1958, 18-25.
21. Shulman, E., Factors influencing the variability of stuttering. In Wendell Johnson (Ed.), *Stuttering in Children and Adults*. Minneapolis: Univ. of Minnesota Press, 1955.
22. Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill, 1956.
23. Stromsta, C. P., A methodology related to the determination of the phase angle of bone-conducted speech sound energy of stutterers and nonstutterers. Ph.D. thesis, Ohio State University, 1956.
24. Tate, M. W. and Clelland, R. C., *Nonparametric and Shortcut Statistics*. Danville, Ill.: Interstate Printers and Publishers, 1957.
25. Tiffany, W. R., and Hanley, C. N., Adaptation to delayed sidetone. *J. Speech Hearing Dis.*, 21, 1956, 164-172.
26. Tuthill, C. E., A quantitative study of extensional meaning with special reference to stuttering. *Speech Monogr.*, 13, 1946, 81-98.
27. Walker, Helen and Lev, J., *Statistical Inference*. New York: Henry Holt, 1953.
28. Williams, D. E., Masseter muscle action potentials in stuttered and nonstuttered speech. *J. Speech Hearing Dis.*, 20, 1955, 242-261.