DURATIONAL CHARACTERISTICS OF ORAL READING:
EFFECTS OF SAMPLE SIZE ON TEST-RETEST RELIABILITY

2.7

1226-5600

by

PATRICIA H. FLANAGAN

B. S., Western Michigan University, 1953

————————————

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

Approved by:

_____
Major Professor

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The phenomena of stuttering has been studied and researched
for nearly forty years and thousands of books and articles
written on the subject. The results of all this activity, how-
ever, have added little definitive information to the subject
of stuttering or disfluency. In recognition of this fact,
Beech and Fransella (1968) comment ". . . stuttering research
is confounded and bedevilled by a lack of solid and reliable
evidence, and an abundance of contradictory and conflicting
experimental results." They further comment that this state of
affairs may be the result of a lack of agreement on the defini-
tion of disfluency as well as the variety of methods utilized
to measure the phenomena.

The most commonly used methods of measuring stuttering have
been observer definitions of disfluency and observer ratings of
severity. Johnson (1961) devised the Iowa Speech Disfluency
Test in which tape recordings were made under several conditions
of speaking and oral reading. The recordings were then listened
to and the observer categorized the judged disfluencies into
eight types. The observer was allowed as many replays as was
necessary to categorize the disfluencies. The categories were
(1) interjections, (2) part-word repetitions, (3) word repeti-
tions, (4) phrase repetitions, (5) revisions, (6) incomplete
phrases, (7) broken words, and (8) prolonged sounds. The rate
of speaking and oral reading were also computed for each speaker.

The data were then compared to norms for normal speakers. Sander (1961) tested the reliability of the Iowa Speech Disfluency Test and found reliability coefficients for total disfluencies and speaking rate were approximately .90 for a sample of 40 stutterers.

Minifie and Cooker (1964) devised a disfluency index which considered the number of syllables uttered divided by the reading rate for a set reading passage. The data from this study suggested reliability of this procedure for discriminating fluent from disfluent speakers. Recordings of speech samples made under several conditions were compared to typescripts of the passage and the disfluencies were then categorized on two or more dimensions. On a test-retest basis, reliability coefficients above .90 were obtained for total disfluency count and rate of speech.

Measurements of the severity of stuttering have been made by asking observers to rate the severity of the disfluency on a rating scale. Cullinan, Prather and Williams (1963) made a comparison of the results of utilizing several types of rating scales with twenty-seven, twenty second segments representing stuttering severity from mild to severe in nature. The samples were rated twice on seven types of rating scales. This study showed that the reliability coefficient resulting from one judge using one rating of the speech sample was too low to be used as an individual predictor. The reliability did increase, however, when a single judge rated the same speech sample three or more

times and used a mean of these ratings, or if the mean of single ratings from a number of judges was used.

An examination of the studies measuring the number of disfluencies and ratings of severity suggest several factors to be considered. First, the amount of time required to administer and score the Iowa Speech Disfluency Test is excessive. Speech pathologists have reported that three to four hours is required for this task, presenting a considerable drain on time if the test is to be administered monthly as Johnson suggests, or if it is to be used with a number of subjects for research purposes. Second, although the studies showed acceptable reliability coefficients, it might be noted that this reliability was obtained by counting total disfluencies in each sample and dividing by the total number of words spoken. The acceptable reliability coefficients for the rating scales were obtained by averaging the ratings of three or more judges. It would be reassuring if acceptable reliability coefficients could be generated without resorting to rather gross methods to obtain the acceptable reliability. Third, the studies utilizing observer judgments suffer from lack of definitive information about what exactly is being measured. No information is brought forth which describes precisely what are the physical parameters of the behavior. When an observer judges that a disfluency has occurred, he finds it difficult to describe exactly how he came to this judgment and it is equally difficult for another observer to listen to the same passage and identify the same disfluency

in the same manner. Ayllon and Azrin (1969) devised the Dimensions of Behavior Rule which they consider to be of paramount importance in psychological research. The rule states "Describe the behavior in specific terms that require the minimum of interpretation." They also state, "Specification of the physical dimensions of the response is so critical that one might say that no behavior should be designed as a target unless this rule can be met."

While observer ratings and definitions of stuttering have not provided definitive measures of disfluency, they have been useful in providing information which consistently demonstrates findings relative to a physical correlate of stuttering. Bloodstein (1944) investigated the relationship between oral reading rate and stuttering severity. The study reported a strong relationship between reading rate and disfluency, and between reading rate and duration of disfluency. The study also reported that the non-stuttering reading rate of the stutterers was slower than for the normal speakers. Johnson (1961) and Sander (1961) also report high correlations between judged frequency of disfluencies and reading time. Wendall and Cole (1961) investigated whether or not listeners responded to any cues in a person's speech, other than actual disfluencies, to identify them as having been spoken by a stutterer. The sentences to be judged had all disfluencies removed and were carefully matched with those of a non-stutterer. The listeners were able to discriminate between the speech samples for stutterers

and non-stutterers as well as judging the stutterers to have poorer speaking rates.

These studies suggest the increased reading rate for stutterers indicated that stutterers do something during speech which occupies time. Time is a physical dimension of speech which is measurable in an objective fashion and would appear to meet the Dimensions of Behavior Rule. In pursuit of this fact, researchers have studied the differences between stutterers and normal speakers in terms of the durations of speech and silence events during on-going speech. This line of research views on-going speech as a series of speech events separated by silence events. Both of these dimensions of speech vary in duration and can be measured on an oscilliscope or by automated electronic devices. These data may then be assigned to arbitrary class intervals which permit construction of frequency distributions for speech and silence events.

In 1958, Roe and Derbyshire reported systematic differences between stutterers and normal speakers for the distribution of silence intervals during conversational speech. The stutterers' speech showed fewer intervals of short durations of silence and more intervals of longer durations of silence than did the normal speakers. Flanagan (1960) replicated the Roe and Derbyshire study using an automatic analyzing device and obtained further information concerning the frequency distributions of silence events which occurred during oral reading by fluent and disfluent subjects, relating these data to disfluency frequency

and reading time. The frequency distributions of silence events for the disfluent subjects were plotted individually, thus allowing each subject to be compared to himself on several trials and to a composite frequency distribution graph of the fluent speakers. While the frequency distribution of silence events for the fluent speakers did not shift from trial to trial, those of the disfluent subjects differed quantitatively between trials, among subjects and from the composite curves for the fluent subjects. In accordance with the findings reported by Roe and Derbyshire, the disfluent speakers had fewer silence events below 120 milliseconds than did the fluent speakers. Curves for disfluent speakers with the highest disfluency percentages were the most dissimilar from the fluent curves, and as the percentage of disfluency approximated that of the fluent speakers, the frequency distributions more nearly approximated the fluent subjects' data.

Minifie (1963) investigated the use of an automatic durational analysis of speech to describe the durational characteristics of speech under different conditions of oral reading or speaking performance, and to accumulate data pertinent to the statistical distribution of the durational characteristics of various modes of normal speaking or reading to use as a comparison to deviant speaking behavior. His findings indicated that changes in reading rate were relative to changes in the duration of the silence intervals of speech rather than to the durations of speech intervals.

To this point it has been demonstrated that the silence intervals of disfluent speech tended to differentiate it from fluent speech and that it was the silence intervals during reading time which were measurably different during instructions to alter reading rate. The problem of obtaining reliable estimates of the durational characteristics of normal speech remained however.

Flanagan (1966) conducted research to establish variance estimates for frequency distributions of speech and silence events for college age speakers. Reliability estimates of these dimensions were obtained on a test-retest basis. Sixty-four college age students and twelve stutterers, ages eighteen to twenty-nine, served as subjects for this research. Individual tape recordings of each subject reading a standardized reading passage on two trials, separated by 24 hours, were obtained. The tapes were analyzed by an electronic multi-class time analyzer with fifteen class intervals, ranging in durations from 50 milliseconds to 750 milliseconds and above. The voice operated relay and multiple class time analyzer were adjusted to exclude events with durations of less than 50 milliseconds. The tape recordings were analyzed for frequency distributions of speech events and frequency distributions of silence events. The resulting mean frequency distributions of the normal speakers for both speech events and silence events yielded L shaped curves, indicating for both, a high rate of events of short durations and a relatively lower rate of events of longer

durations. Five class intervals of speech events and nine
class intervals of silence events were found to significantly
discriminate between the group of fluent speakers and the
stutterers as a group. On a test-retest basis for the individ-
ual fluent speakers, however, the frequency distributions
obtained for speech events and silence events yielded low to
moderate test-retest coefficients of reliability. While these
data indicated to the author that there was some tendency for
the subjects to approximate their vocal response patterns, the
correlations were not of sufficient magnitude for reliable
prediction of individual response patterns. The observed lack
of reliability prompted Flanagan to question the adequacy of the
speech sample size per subject. The mean reading time for the
fluent subjects on Trial I was 95 seconds and inasmuch as the
class intervals with higher frequency counts had greater reli-
ability than did the class intervals with lower frequency
counts, it was possible that increased reliability could be
obtained by increasing the speech sample size being analyzed.

Thoms (1969) studied the test-retest reliability of the
durational characteristics of speech and silence events as a
function of sample length during spontaneous speech. Frequency
distributions for the speech of ten normal male speakers were
derived from recorded speech displayed on a cathode ray oscillo-
scope. Speech and silence events were photographed, measured
and assigned by durations into class intervals. The number of
speech or silence events was then sub-totaled within time units

of 30 seconds, culminating with five minutes. Percentages of individual test-retest reliability were measured at each class interval for each 30 second time unit. Median values were computed by ranking individual percentages of reliability at each class interval for units of one minute. Mean reliability for speech events peaked at 3.5 minutes and at 3.9 minutes for silence events. A tendency was reportedly observed for the reliability to increase with an increase in sample size and then to stabilize when the point of highest reliability was achieved.

The purpose of this research is to assess the test-retest reliability of the distribution of the durational properties of speech and silence events of normal speakers under conditions of oral reading, as a function of speech sample size. The research by Flanagan (1966) prompted the question as to whether or not acceptable test-retest reliability could have been obtained with his subjects had the speech sample size been larger. Thoms (1969) reported that the reliability of speakers during spontaneous speech increased as a function of speech sample size and stabilized when the point of highest reliability was achieved.

Assuming that acceptable test-retest reliability could be obtained under the conditions of oral reading by normal speakers, the technique of analyzing the frequency of the durations of speech and silence events could be a useful tool in rapidly and reliably evaluating a physically specifiable dimension of speech behavior which co-varies with observer defined disfluency rate and severity measure.

CHAPTER II

PROCEDURE

The procedures for this research will be described in
reference to selection of subjects, recording procedures,
stimulus material, apparatus and measurement.

SELECTION OF SUBJECTS:

Ten male college students from Freshman Oral Communication
classes at Kansas State University were subjects for this
research. They ranged in age from 18 to 21 years and were in
their first or second year of college. Eight of the subjects
were native Kansans, one a native of New York, and one a native
of Massachusetts. None had previous training in oral reading
or public speaking. The subjects were screened to insure nor-
mal hearing. None of the subjects was found to have or reported
ever having a speech defect.

RECORDING PROCEDURES:

The subject was seated at a desk across from the experi-
menter in a sound treated room. The subject was asked for
information such as his name, major in college, hometown, and
leisure time activities. The purpose of this questioning was
to aid him in adapting to the experimental setting and to the
experimenter. He was then asked to read the "Rainbow Passage"
(Fairbanks, 1940). During this reading, a monitor in the
adjacent room adjusted the VU meter on the recorder to the
intensity of the subject's speech. This allowed for the

subject to adapt to the recording process during oral reading.
The subject was then instructed that he would be reading from a
pack of cards and that he should read as he "ordinarily would."
He was also cautioned to remain a constant distance (18 inches)
from the microphone. The experimenter then handed him the
appropriate pack of cards for Trial I. The experimenter read
the subject's name, age and date into the microphone and after
a five second pause the subject began reading from the cards
and continued until all the cards in the pack had been read.
When the trial was finished, he made an appointment for Trial II
which was separated from Trial I by at least 48 hours. Identi-
cal conditions prevailed during Trial II. The purpose of the
research was reported to the subjects prior to their participa-
tion in this study by showing them an oscilloscope and explain-
ing that the examiner was studying the wave forms of different
speakers.

STIMULUS MATERIAL:

Sixty-one major subject areas from the World Book Encyclo-
pedia (1965) comprised the stimulus material for the oral
reading. The criteria for selecting this material were as
follows:

1. The material must be descriptive and non-
technical in nature.

2. The material must be regarded as easily read
by this class of subjects.

3. The material must be amenable to presentation
in a random order.

The material was judged to be descriptive and non-technical by virtue of being selected from introductory paragraphs in an encyclopedia which is by its nature general and non-technical. Each paragraph was approximately 60 words in length. The information contained in these paragraphs was judged to be of a calibre which could be read easily by this population of college males. The paragraphs were subsequently reproduced individually on 4 x 6 cards, thus enabling the order to be changed between subjects.

A total sample of 4,001 words were presented on 61 individual cards. The cards were first arranged in alphabetical order by subject matter, designated odd or even and then separated into two packs. The packs contained 2,000 and 2,001 words respectively. The packs were then shuffled. For Trial I, odd numbered subjects used the odd numbered pack and the even numbered subjects used the even numbered pack. The order was reversed for Trial II. Each pack was shuffled again between subjects. See the Appendix for a reproduction of these paragraphs.

APPARATUS:

The oral reading sample from each subject was detected by a EV665 microphone located in a sound treated room and recorded by an Ampex A6500-2 recorder located and operated in an adjacent room. Ambient noise was controlled by recording from outside the double walled sound treated room. Differences in speech intensity for each subject were adjusted by regulating the

volume of the tape recorder to enable the VU meter to peak at zero. The tape recordings were played to a Hewlett-Packard 132A Oscilloscope and the CRO sweeps were photographed with an attached polaroid camera. The sweep time of the CRO was set to record at 500 millisceonds per centimeter to enable the 9.4 centimeters of the scope to register 4.7 seconds of connected speech. A hand operated switch introduced external synchronizing voltage to start each sweep and open the camera shutter. The same switch was also used to stop the sweep and close the shutter. The tape was interrupted between sweeps to insure a continuous sample. Between sweeps, the spot remained off the screen, the shutter was closed and the polaroid camera moved along its bed. The procedure was repeated until 11 sweeps were recorded on a film plate. A new film plate was then inserted into the camera and the process initiated again. Each photograph consisted of eleven sweeps of 4.7 seconds each, with a sequential sample of 51.7 seconds per film plate. The entire five minute speech sample was portrayed on a series of photographs in sequence. A reference level was set for each tape to insure that only the speech of the subject, and not background noise, was being photographed. Amplification of the horizontal axis of the CRO spot was set at 20 millivolts a centimeter. When each tape was started, the amplification level was visually finely tuned so that only the subject's speech caused the CRO spot to disrupt the horizontal tracing, not constant tape noise. These fine tuning procedures preceded each photographing session.

MEASUREMENT:

The durations of speech or silence events were divided into eight class intervals. The class intervals, expressed in milliseconds, were of the following durations; x - 99 ms, 100 - 199 ms, 200 - 299 ms, 300 - 399 ms, 400 - 499 ms, 500 - 599 ms, 600 - 699 ms, and 700 - x ms. The first and last intervals were open-ended. A smoothing filter caused a small but undetermined delay for the x - 99 ms interval. A known signal of 30 milliseconds was observable as a deflection of the CRO on the scope face. Research by Minifie (1963) and Flanagan (1966) showed the 700 - x ms interval to be of asymptotic value with respect to the variables in question.

Before the photographs were measured, the following corrections for time were made. In the event that a sweep had been initiated or terminated during an interval which was less than 700 ms, these portions were eliminated from the count by subtracting their values from 4.7 seconds. The times of each sweep were then combined into 30 second periods. The 30 second sub-totals culminated at 300 seconds or five minutes. All computations were double checked with an adding machine.

The durations of speech and silence events were measured by hand with a metric scaled ruler and tabulated according to class interval. This process required two people, one measuring and the other recording the event in the appropriate class interval on a data sheet. Reliability of these measures was established and differences of less than two percent were noted by test-retest procedures for five photographs.

CHAPTER III

RESULTS

The purpose of this research was to obtain frequency
distributions of speech and silence events which occurred dur-
ing oral reading by normal speakers and to test the reliability
of these findings on a test-retest basis as a function of
speech sample size. The results of this research will be
presented in reference to (1) Comparing the frequency distribu-
tions for class intervals of speech and silence events obtained
in the present research with similar data obtained from previous
research; (2) The reliability of the frequency distributions of
speech events as a function of speech sample size; and (3) The
reliability of the frequency distributions of silence events
as a function of speech sample size.

COMPARISON OF FREQUENCY DISTRIBUTIONS:

Since the purpose of this present research was to establish
reliability as a function of speech sample size, it was of
interest to compare the frequency distributions of speech and
silence events from this study with those of a previous study
(Flanagan, 1966). Point by point comparison is not possible
because this research differed from Flanagan's research in
several aspects. Flanagan used an automated device for measuring
and categorizing durations of speech and silence events which
eliminated events of less than 50 ms, while the apparatus in the
present study treated intervals of less than that duration,

specifically, known signals of 30 ms were observable. The
triggering threshold for counting an event in the Flanagan
research was set at 5 dB above the average noise level as
measured by a volt meter, while the present researcher fine
tuned the apparatus visually to trigger "just above noise
level." Flanagan's subjects read a set 300 word passage with a
mean reading time of 95 seconds, while the subjects in the
present research read a total of 2,000 words with frequency
distributions being constructed for each 30 second time period.
For purposes of comparison with the previous research, the
means from the first 90 seconds of oral reading were used.
Also, the Flanagan research utilized 15 class intervals as com-
pared to eight in the present research. Despite the differences
noted between the two studies, an impression of the comparabil-
ity of frequency distributions can be obtained by plotting the
frequency distributions from the two studies graphically.
Figure 1 presents the means of distributions for speech events
for both studies. Both curves show negative slope with positive
acceleration. The inflection point of the frequency distribu-
tion for speech in the present study occurs sooner in time and
appears to be more gradual than for Flanagan's subjects. To
attain an assessment of the degree of the similarity between
curves, a Pearson product-moment correlation was computed. The
resulting correlation coefficient was .951 which suggests a
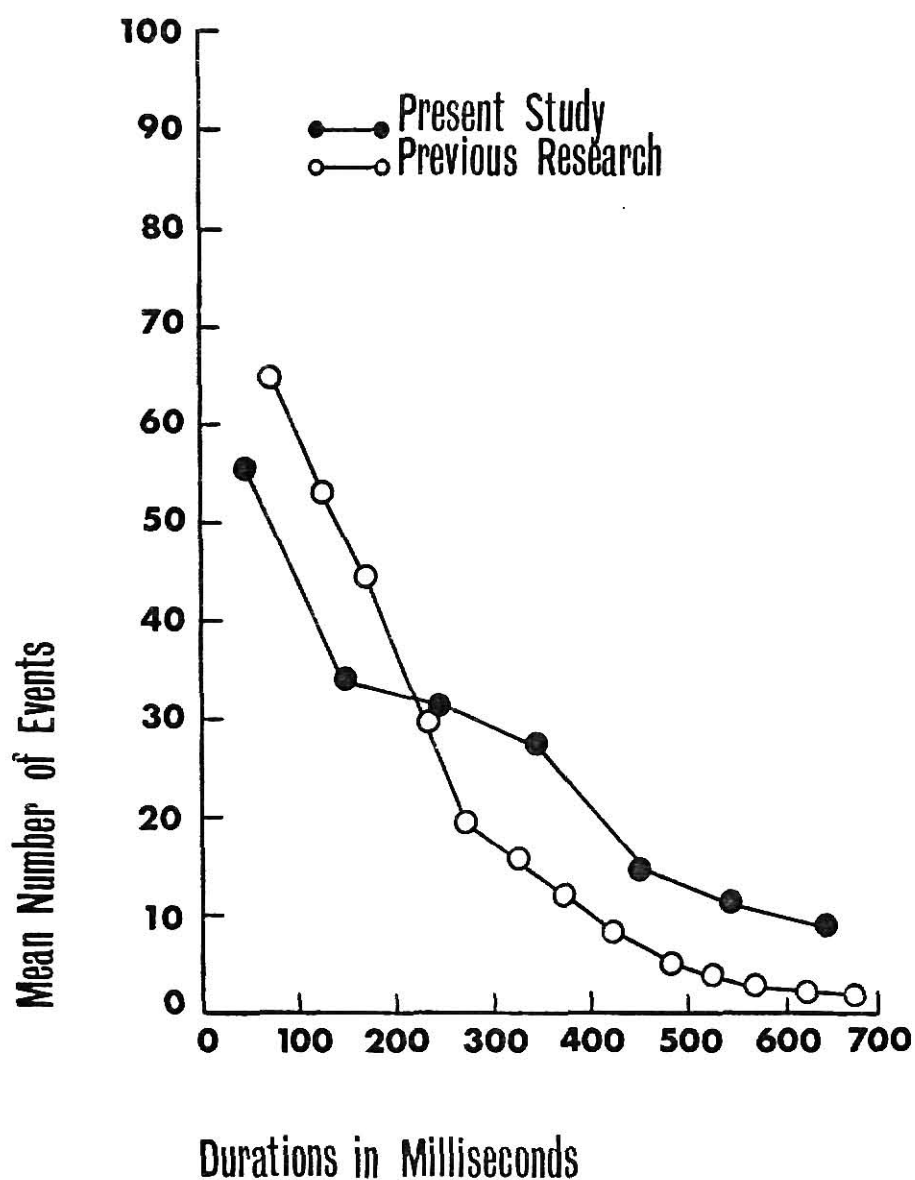high degree of similarity between the curves.

Figure I.
Comparison of Distribution of Speech Events of Present Study with those
of Previous Research

Figure 2 presents the mean frequency distributions by
class interval for silence events for both studies. Slope and
acceleration of these two curves appear to be nearly identical.
One notable difference is that the present research found
approximately one-third more mean silence events below 100 ms
in duration. This is perhaps in part due to the less restric-
tive interval for events below 100 ms. Pearson product-moment
correlation for these two curves was .952, again suggesting a
high degree of similarity between curves.

RELIABILITY OF SPEECH EVENTS:

For the purposes of this research, speech events were
those vocalizations which transcended the noise level of the
apparatus. The speech events were recorded during oral reading,
displayed on a cathode ray oscilloscope and photographed as
displayed. The speech events were then measured and tabulated
by duration into one of eight class intervals. The events for
each class interval were then sub-totaled at one minute sampling
periods up to five minutes. Individual test-retest reliability
percentage was determined for each class interval by dividing
the smaller frequency count by the larger count. The median
percentages of individual reliability were subsequently deter-
mined from them by class interval and sample size. Table I
presents the median percentages of test-retest reliability
between Trials I and II for speech events occurring at each
class interval for speech sampling periods of between one and
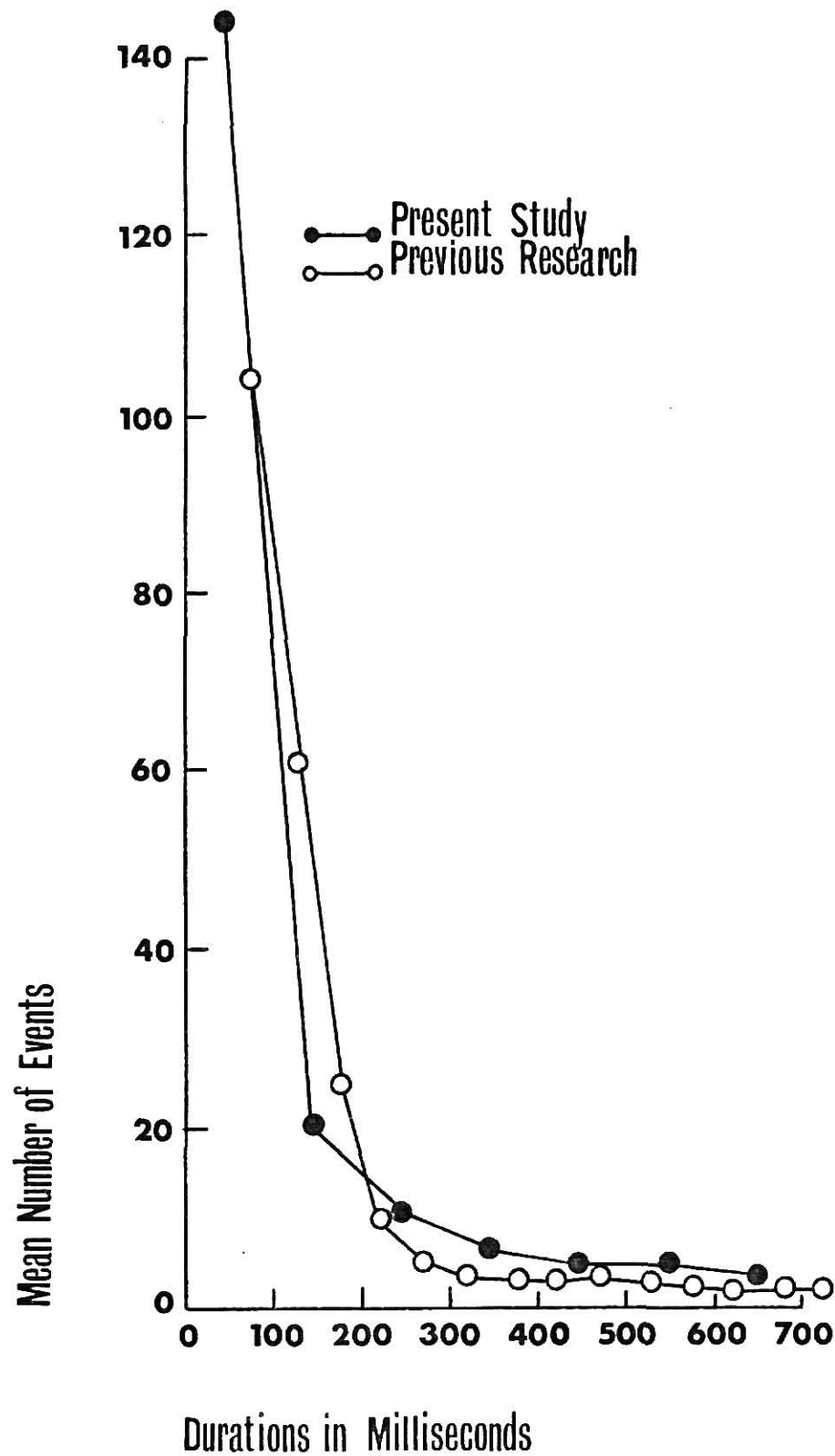
Figure 2.
Comparison of Distribution of Silence Events of Present Study with those
of Previous Research

TABLE I

SPEECH EVENTS: MEDIAN PERCENTAGE OF AGREEMENT BETWEEN
TRIALS I AND II AT ONE MINUTE SAMPLING PERIODS

Time in Milli-seconds	Median Agreement Sample Size in Minutes					Increase in Agreement
	1	2	3	4	5	
x - 99	77.5	71.5	73.0	77.0	86.0	yes
100 - 199	68.5	80.5	75.5	84.0	84.0	yes
200 - 299	61.0	66.5	77.5	79.0	74.5	yes
300 - 399	84.0	87.0	81.5	78.0	80.5	no
400 - 499	75.0	75.0	87.5	89.0	86.5	yes
500 - 599	77.5	71.0	72.0	73.5	77.0	no
600 - .699	57.0	75.0	74.5	85.0	89.5	yes
700 - x	66.5	81.5	76.5	86.5	89.0	yes
Median Percentages of Agreement across class intervals	72.3	75.0	76.0	81.5	85.0	

five minutes, and the median percentages of agreement across class intervals for these same sampling periods.

The percentage of agreement between Trials I and II for speech events, across class intervals, increased from 72.3 percent at the one minute sampling period to 85.0 percent at the five minute sampling period. The median percentages of agreement for the two, three, and four minute sampling periods are 75 percent, 76 percent and 81.5 percent respectively.

Within class intervals, when comparing the one minute and five minute samples, six of the eight class intervals showed an increase in median percentage of agreement between Trials I and II. The increases were 8.5 percent (x - 99 ms), 15.5 percent (100 - 199 ms), 13.5 percent (200 - 299 ms), 11.5 percent (400 - 499 ms), 32.5 percent (600 - 699 ms), and 22.5 percent (700 - x ms). The class intervals commencing with 200 and 400 milliseconds, showed peak median percentage of agreement before the five minute sampling period. The class interval commencing with 200 ms showed an agreement of 77.5 percent at the three minute sampling period and 79.0 percent at the four minute sampling period. The class interval commencing with 400 ms showed 87.5 percent agreement at the three minute sampling period and 89.0 percent agreement at the four minute sampling period. The class intervals commencing with less than 100, 100, 600 and 700 milliseconds, showed their highest median percentage of agreement at the five minute sampling period. Two class intervals, 300 - 399 ms and 500 - 599 ms did not show an

increase in the median percentage of agreement as the sampling
period increased from one to five minutes. The class interval
commencing with 300 ms appears to be in a steady state, peaking
at two minutes of sampling and then the reliability appears to
be dragging sideways. The class interval commencing with 500 ms
suggests that at this class interval, sample size does not
increase or diminish median reliability between trials.

RELIABILITY OF SILENCE EVENTS:

In this study, silence events were obtained, recorded,
measured and the data treated in the same manner as were the
speech events. Table II presents the median percentages of
test-retest reliability between Trials I and II for silence
events occurring at each class interval for speech sampling
periods of between one and five minutes, and the median percent-
ages of agreement across class intervals for these same sampling
periods.

The percentage of agreement between Trials I and II for
silence events, across class intervals, increased from 61.5
percent at the one minute sampling period to 79.7 percent at
the five minute sampling period. The median percentages of
agreement for the two, three, and four minute sampling periods
are 69.8 percent, 68.8 percent and 75.3 percent respectively.

Within class intervals, when comparing the one minute and
five minute samples, seven of the eight class intervals showed
an increase in median percentage of agreement between Trials I
and II. The increases were 5.0 percent (x - 99 ms), 12.0

TABLE II

SILENCE EVENTS: MEDIAN PERCENTAGE OF AGREEMENT BETWEEN
TRIALS I AND II AT ONE MINUTE SAMPLING PERIODS

| Time in Milli-seconds | Median Agreement | | | | | Increase in Agreement |
| | Sample Size in Minutes | | | | | |
	1	2	3	4	5	
x - 99	79.5	80.5	80.0	84.0	84.5	yes
100 - 199	61.0	77.0	68.5	70.5	73.0	yes
200 - 299	75.0	67.0	64.5	71.5	74.0	no
300 - 399	62.0	68.5	79.0	77.0	82.0	yes
400 - 499	50.0	71.0	68.0	74.0	79.0	yes
500 - 599	50.0	65.5	69.0	76.5	78.5	yes
600 - .699	36.5	58.5	52.0	72.0	72.0	yes
700 - x	71.0	73.0	83.0	85.0	85.0	yes
Median Percentages of Agreement across class intervals	61.5	69.8	68.8	75.3	79.7	

percent (100 - 199 ms), 10.0 percent (300 - 399 ms), 29.0 per-
cent (400 - 499 ms), 28.5 percent (500 - 599 ms), 35.5 percent
(600 - 699 ms), and 14.0 percent (700 - x ms). The class
intervals commencing with 100, 200, 600 and 700 milliseconds,
showed peak median agreement before the five minute sampling
period. The class interval commencing with 100 ms showed 77.0
percent agreement at the two minute sampling period. The class
interval commencing with 600 ms showed 72.0 percent agreement
at the four minute sampling period. The class interval commenc-
ing with 700 ms showed 85.0 percent agreement at the four minute
sampling period also. The class intervals commencing with x,
300, 400 and 500 milliseconds showed the highest percentage of
agreement at the five minute sampling period. The 200 - 299 ms
class interval showed highest median agreement at the one minute
sampling period. After this point, the median percentage of
agreement decreased and then increased to within 1.0 percent of
the agreement attained at the one minute sampling period.

CHAPTER IV

DISCUSSION

The intent of this study was to determine whether the
reliability of the durational characteristics of speech was
functionally related to speech sample size. As a general rule
this study found that the reliability of the frequency distri-
butions of speech and silence events did show a progressive
increase as a function of speech sample size. For speech
events, across class intervals, the median reliability for a
one minute speech sampling period was 72.3 percent agreement
and 85.0 percent for the five minute sampling period. The
median percentages of agreement for the two, three and four
minute sampling periods are 75 percent, 76 percent and 81.5
percent respectively, indicating that the median reliability
progressively increased between the extreme data points. Treat-
ing the class intervals for speech events separately, six of
the eight class intervals confirmed the major trend, showing
increasing reliability as a function of sample size. Both of
the class intervals (300 - 399 ms and 500 - 599 ms) which
failed to follow the major trend, appeared to vascillate
rather than showing a trend in either direction.

For silence events, across class intervals, the median
reliability for the one minute sampling period was 61.5 percent
agreement and 79.7 percent agreement for the five minute sampling
period. The percentages of agreement for the two, three and

four minute sampling periods are 69.8 percent, 68.8 percent and 75.3 percent respectively, again showing that median reliability progressively increased between the extreme data points. When the class intervals for silence events were treated separately, only one class interval (200 - 299 ms) failed to confirm the major trend. This class interval appeared to show neither decreasing nor increasing reliability with increased speech sample size.

The findings of this research are in substantial agreement with previous research which investigated test-retest reliability of the durational characteristics of spontaneous speech as a function of sample size (Thoms, 1969). The research on spontaneous speech showed that for speech events the median agreement between trials, across class intervals, at the one minute time unit was 72.6 percent and 80.1 percent at the five minute time unit. When each class interval was considered separately, all intervals confirmed the major trend. For silence events in the Thoms research, across class intervals, the median percentages of agreement were 68.3 percent at the one minute time unit and 79.3 percent at the five minute time unit. Again, the individual class intervals all confirmed the major trend. The similarities in the results between the present research and the Thoms study are more remarkable than are the differences. The present study found a median agreement percentage agreement between trials at the five minute sampling period of 85.0 percent for speech events as compared to 80.1 percent in the Thoms

study. Median percent of agreement for silence events were
almost identical at the five minute sampling period with 79.7
percent for the present study and 79.3 percent in the Thoms
study. The individual eight class intervals for speech and
silence events in the Thoms study each confirmed the major
trend. The individual class intervals in the present study
which did not follow the major trend, may be attributed to
either the nature of the oral reading task as opposed to
spontaneous speaking or to spuriously high median reliability
during the smaller sampling periods for oral reading (Speech,
300 - 399 ms, 84.0 percent at one minute, Speech 500 - 599 ms,
77.5 percent at one minute, and Silence 200 - 299 ms, 75.0 at
one minute). Sufficient data are not available at this time to
reconcile these observed differences.

The findings of this research confirm Flanagan's (1966)
conclusion that the sample size he utilized, mean reading time
of 95 seconds, did not result in acceptable reliability for
comparison of individuals. In the present study, looking at
the one and two minute sampling periods of speech events across
class intervals, reliabilities of 72.3 percent and 75.0 percent
respectively and silence events with reliabilities of 61.5 per-
cent and 69.8 percent respectively were observed. Flanagan's
conclusion was based on observed median correlations between
trials which were .270 for speech events and .297 for silence
events. Neither Flanagan's correlations nor the median percent-
ages of agreement in this study are of sufficient magnitude for
comparison of individuals.

At this point the question arises as to what is an acceptable reliability for the practical utilization of data on the durational characteristics of speech and silence events for comparing individual stutterers to norms for fluent speakers. The question is probably best answered by determining how finely the clinician wishes to discriminate between subsequent samples of the disfluent speech of the same client. It appears safe to assume that measurement procedures which yield an error of approximately 30 percent are not useful except for extremely gross comparisons. Hence, the one and two minute speech samples are inadequate. An error of 10 percent (90 percent agreement) is generally regarded as acceptable. Practical considerations and the capriciousness of some behavioral phenomena, frequently force clinicians to utilize procedures with percentages of agreement as low as 80 percent. Viewing the present data for a five minute speech sample, the observed percentage of reliability appears to be within the limits of practical use. However, it should be emphasized that the observed reliability percentages at the five minute sampling period did not appear to be at asymptotic values. See Figure 3. Extending a best fit line through the existing data points would suggest that increased reliability could be obtained by further increasing the speech sample size. By extension of the best fit line to 90 percent criteria, the suggested sample size for speech events would be obtained at six and one-half minutes and at seven and one-half minutes for silence events.
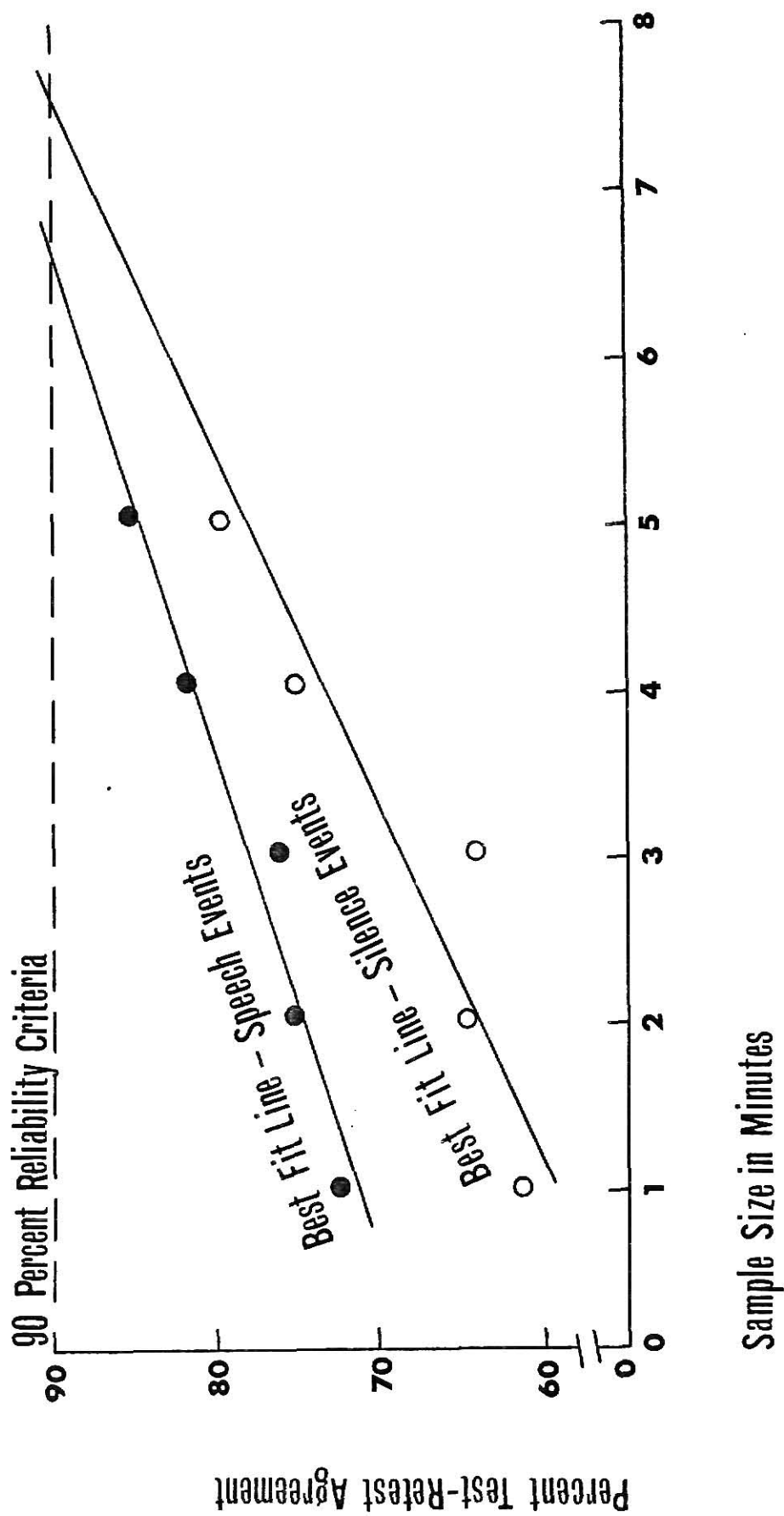
29

Figure 3. Test-Retest Reliability for Durational Characteristics of Speech and Silence Events, Plotted as a Function of Sample Size, Measured in Minutes

The seven to eight minute sampling period projected by the best fit line would have to be established as a fact by future research to exclude the possibility of plateaus and the growth in reliability which occurred between the two and three minute sampling periods.

Presuming that the intent of future research was to extend the speech sample size and attain improved reliability, it would not be practical to utilize the same measuring techniques as were used in this study. In this study, the experimenter expended in excess of twenty hours time photographing, measuring, tabulating and reducing the data for each subject. The amount of time per subject was directly proportional to the speech sample size being analyzed. Obviously, such a time consuming process would not be practical in the clinical environment. With the advances in electronic and computer technology, similar data concerning the durational characteristics of speech could be collected and analyzed more expediently. By way of comparison, Flanagan's (1966) study used an automated device which analyzed the speech sample in the time required to play the sample twice (once for speech events and once for silence events) to the multi-class analyzer.

This research has provided evidence that the reliability of measurements of the frequency distributions of speech and silence events are functionally related to the length of the speech sample. Data was obtained which indicates that the five minute speech samples or oral reading yield agreements on a

test-retest basis of 80 percent or better. Projection of these data, based on this functional relationship between speech sample size and reliability, indicate that samples of seven to eight minutes in length would yield test-retest agreements of 90 percent. Assuming that subsequent research establishes the speech sample size necessary for acceptable reliability, norms of this dimension of vocal behavior could then be established for fluent speakers. These norms could then serve as a basis for comparison with the data from disfluent speakers. There is a strong indication from the studies by Roe and Derbyshire (1958), Flanagan (1960, 1966), and Love (1969) which indicate that measures of this kind covary with observer defined measurements of disfluency.

Utilizing measurements of the durations of speech and silence events would provide the speech pathologist with information concerning a measurable physical dimension of an individual's speech which is not subject to observer bias.

SUMMARY

The purpose of this research was to assess the test-
retest reliability of the distribution of the durational
properties of speech and silence events of normal speakers,
under conditions of oral reading, as a function of speech
sample size. Durations of speech and silence events during
on-going speech have proved useful as a physical parameter of
speech which are measurable and not subject to observer bias.
Previous research has posed the question of whether or not
increased test-retest reliability could be obtained by length-
ening the speech sample size being analyzed.

Ten young adult males served as subjects for this research.
Two samples of oral reading were obtained from each subject.
The samples were equivalent and obtained at least 48 hours
apart.

The samples of oral reading were recorded, displayed on a
cathode ray oscilloscope, photographed and durations of speech
and silence events subsequently measured and tabulated by their
durations into eight class intervals. The class intervals were
x - 99 ms, 100 - 199 ms, 200 - 299 ms, 300 - 399 ms, 400 - 499
ms, 500 - 599 ms, 600 - 699 ms, and 700 - x ms. The number of
speech or silence events for each duration were sub-totaled
within speech sampling periods of one, two, three, four and
five minutes. The percentage of individual test-retest reli-
ability was measured at each class interval for each speech
sampling period. Median values for test-retest reliability

were then obtained for each class interval and across class intervals for each speech sampling period.

This research provided evidence that the reliability of measurements of the frequency distributions of speech and silence events during oral reading are functionally related to the length of the speech sample. The median reliabilities for speech events across class intervals increased progressively from the one minute sampling period to the five minute sampling period. The median reliabilities for silence events also increased progressively from the one minute sampling period to the five minute sampling period.

The research suggests that a five minute speech sample size does yield acceptable test-retest reliability if a minimum of 80 percent agreement between trials is acceptable. A best fit line projected from the obtained data points suggests that 90 percent agreement could be achieved by increasing the speech sample size to seven or eight minutes.

The discussion of this research considered practical applications of the durational analysis of speech and implications for further research.

REFERENCES

Ayllon, T. and Azrin, N. The Token Economy. Meredith Corporation (1968), Chap. 3.

Beech, H. R. and Fransella, Fay. Research and Experiment in Stuttering. New York: Pergamon Press (1968), Chap. 2.

Bloodstein, O. Studies in the psychology of stuttering: XIX. The relationship between oral reading rate and severity of stuttering. J. Speech Dis., 9, 161-173 (1944).

Cullinan, W., Prather, E. and Williams, D. Comparison of procedures for scaling severity of stuttering. J. Speech Hearing Res., 6, 187-194 (1963).

Fairbanks, G. Voice and Articulation Drillbook. New York: Harper (1940).

Flanagan, B. The distribution of pause durations in speech. Southern Illinois University, TR-4, Sept. 1960, on AFCRC Contrad AF 19(604) - 6127.

Flanagan, B. An analysis of the durational aspects of connected speech with reference to stuttering. Unpublished doctoral dissertation, University of Florida, 1966.

Johnson, W. Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and non-stutterers. J. Speech Hearing Dis., Monograph Suppl., 7, 1-20 (1961).

Johnson, W., Darley, E. L. and Spriestersbach, D. C. Diagnostic Methods in Speech Pathology. New York: Harper & Row (1963).

Love, Laura R. and Jeffress, Lloyd A. Identification of Brief Pauses in the Fluent Speech of Stutterers and Nonstutterers. J. Speech Hearing Res., 14, 229-240 (1971).

Minifie, F. An analysis of the durational aspects of connected speech samples by means of an electronic speech duration analyzer. Ph.D. dissertation, Univ. of Iowa (1963).

Minifie, F. and Cooker, H. A disfluency index. J. Speech Hearing Dis., 29, 189-192 (1964).

Roe, A. and Derbyshire, A. Quantifying periods of silence and sound in speech. Paper read at 1958 Convention of the American Speech and Hearing Association (1958).

Sander, E. Reliability of the Iowa Speech Disfluency Test. J. Speech Hearing Dis., Monograph Suppl., 7, 21-30 (1961).

Thoms, Janice U. Durational Characgeristics of Spontaneous Speech: Effects of Sample Size on Test-retest Reliability. Unpublished Master's thesis, Kansas State University (1969).

Wendahl, R. and Cole, J. Identification of stuttering during relatively fluent speech. J. Speech Hearing Res., 4, 281-286 (1961).

World Book Encyclopedia, Field Enterprises Educational Corp., Chicago (1965).

APPENDIX

Airplane is man's fastest means of traveling from place to place. A jet airliner can carry passengers halfway around the world, from Chicago to Calcutta, India, in about 15 hours. By refueling in the air, airplanes can fly completely around the world without stopping. Airplanes can fly faster than sound and can zoom many miles above the earth.

Alphabet is the series of letters used in writing a language. The name means exactly what the term ABC means as a name for the 26 letters of our alphabet. *The word comes from alpha and beta, the first two letters of the Greek alphabet.*

Bird is an animal with feathers. Birds live in all parts of the world. *Some make their homes in the cold places near the North Pole.* Others live in the hot, green jungles of South America and Africa. Some birds live in fields and some in the mountains. Others stay near water. *The colors of some birds brighten the places where they live.* The feathers of other birds are dull and help to hide them. All birds have wings, but some, such as the ostrich, cannot fly.

Books are seen and used almost everywhere today. They are familiar friends or useful tools. But there was a time when only a *very few learned persons ever saw or read books.* Printed books were made in China and Japan about 1,000 years ago, but the people of Europe did not know about them. The first printed books in Europe appeared about 500 years ago.

Books written especially for children are a relatively new kind of literature. In fact, such literature is little more

than 100 years old. Before about 1850, writers tried to teach
and improve children. The only books for children taught
lessons on how to behave and what to believe. They also con-
tained lessons on the ideas of history and science at that
time.

Man found the earliest type of bridge ready-made. It was
a fallen log lying across a ravine. When man felt the need for
a better stream crossing than the fallen log, he rolled another
log beside the first one. This made a wider roadway, and
saved him many duckings. By placing two logs side by side he
could pull a sled or cart across the stream. His next step was
to lay down several more logs, and fill in the cracks to make
the roadway smooth.

Canada is the second largest country in the world. Only
Russia covers more area than this great northern neighbor of
the United States. Canada is slightly larger than Europe, and
about 176,000 square miles larger than the United States. But
only about one-tenth as many people live in Canada as in the
United States. This is because much of northern Canada is an
empty wilderness of forests and frozen wasteland. Most
Canadians live within 200 miles of their country's 4,000-mile
southern border.

Color surrounds us everywhere. We see it in the sky, in
the oceans, in the rocks, and in all plants and animals. Color
adds beauty to our clothing and food. It makes our homes,
schools, and offices attractive. We see it in paintings and

photographs, in books and magazines, and in motion pictures and television shows. Advertisements and posters attract our attention with bright colors.

Dog has been "man's best friend" for thousands of years. These friendly, obedient animals serve people throughout the world in work, play, and sport. Dogs live near the Eskimo's igloo, in jungle villages, in farm homes, and in city apartments. About 25,000,000 dogs live in the United States.

Most people think of an egg only as a kind of food produced by chickens. But all birds lay eggs, although few besides chicken eggs are used as *human food*. And many other kinds of animals produce eggs, including worms, fishes, reptiles, insects, and mammals such as the sow and the cow. In all ranks of animal life above the very lowest, it is impossible for young animals to be produced except from eggs.

Elementary school is a school for children from 5 or 6 to 12 or 14 years old. Most elementary schools have either six or eight grades. A grade corresponds roughly to a year's work in school. Many elementary schools include kindergartens for children who are 5 years old. Pittsburgh and some other cities have primary-unit schools that include a kindergarten and the first three grades. In these cities, larger intermediate schools include fourth, fifth, and sixth grades for older children.

Erosion is the wearing away of earth or rock. This natural process works slowly but surely. In hundreds of thousands of year, erosion can wear away a mountain until it is

level with the plain. The surface of the earth is always being changed and modified by erosion.

A nation's flag is a stirring sight as it flies in the wind. Its bright colors and striking design stand for the country's land, its people, its government, and its ideals. A country's flag can stir people to joy, to courage, and to sacrifice. Many men and women have died to protect their national flags from dishonor and disgrace. Every citizen should know how to honor his nations' flag.

Folklore in its broadest meaning includes all the customs, beliefs, and traditions that people have handed down from generation to generation. Folklore in this sense refers to the stories, arts and crafts, dances, music, religion, and games a people have. It means the way they express their joy at a child's birth or a friend's marriage, or their sorrow at some-one's death. It may even include their customs in farming and running a household, and the way they govern themselves.

Freedom is the condition of being free. Few persons have ever been completely free. Nearly everyone has always been restrained from doing what he likes by the rights of other people, by laws and customs, and by his own physical and mental limitations.

Georgia is the largest state east of the Mississippi River. It was the last founded and farthest south of the original 13 American colonies. James Oglethorpe brought the first settlers only 42 years before the Revolutionary War and

named the colony for King George II of England. Georgia has the nickname of the Empire State of the South, because of its size and industry. Atlanta is the capital and largest city.

Glue is a sticky substance made from the skin, bones, and connective tissues of animals. There are three types of true glue: (1) hide or skin glue; (2) bone glue; and (3) fish glue. All are impure forms of gelatin. Other materials are sometimes called glues. But technically they are usually gums, cements, or adhesives.

The government of the United States represents, serves, and protects the American people both at home and abroad. From the nation's capital in Washington D.C., its activities and influence reach every part of the world.

Grinding and polishing are two important manufacturing processes. Grinding uses abrasives to remove material, and polishing uses them to smooth surfaces. Grinding probably ranks as the earliest of all manufacturing processes. Its use goes back to prehistoric times, when ancient man shaped stone tools by rubbing them against hard, abrasive stones or other kinds of abrasive materials.

Hippopotamus is a big land and water animal. It lives in the swamps, streams, and marshes of tropical Africa. Next to the elephant, it is the heaviest of all land mammals. A big rhinoceros has a larger frame than a hippopotamus, but may weigh less. The hippopotamus' name means river horse. It is not related to the horse, however, but to the hog.

Horse has been one of man's most useful animals for thousands of years. Horses once provided the fastest and surest way to travel on land. Hunters on horseback chased animals and killed them for food or for sport. Soldiers charged into battle on sturdy war horses. The pioneers used horses when they settled the American West in the days of stagecoaches, covered wagons, and the pony express.

India is the seventh largest country in the world and the largest democracy in Asia. It has more people than any other country except China, its communist neighbor to the north. About one of every seven persons in the world lives in India. The country's population is greater than that of all the nations of Africa and South America together.

Irish Wolfhound is the tallest of all dogs. It stands at least 32 inches high, and weighs from 105 to 140 pounds. But it has the streamlined look of a greyhound. In Ireland, the dog was used for hunting wolves and elk. In the United States, it has been trained to hunt timber wolves and coyotes. The dog's rough, wiry coat may be gray, brindle, red, black, pure white, or fawn. The Irish wolfhound may have the same coloring as a deerhound.

Jacks is the name of a children's game played with small metal objects called jacks. The object of the game is to pick up the jacks from the ground while tossing and catching a ball or another jack. Children usually play jacks with a ball and six jacks. Players take turns trying to do a number of stunts without a miss.

January, the first month of the year, is named for Janus, a Roman god. According to legend, Numa Ponpilius added January and February to the end of the 10-month Roman calendar, about 700 B.C. He gave the month 30 days. In 46 B.C., Julius Caesar added a day and made January the first month. The Norsemen named the month for Thor, their god of thunder and storms. The Anglo-Saxons called it Wolfmonth because wolves came into the villages in winter in search of food.

Jew's-Harp is a small musical instrument used mainly in folk music and by children. It consists of a flexible metallic reed at one end of a curved metal frame. The other end of the reed is tapered and bent forward at a right angle. The player holds the metal frame against his teeth and makes the reed vibrate by hitting its forked end with his free hand. He produces different notes by changing the size and shape of his mouth cavity.

Kangaroo is the largest marsupial (an animal that carries its young in a pouch in the abdomen). Kangaroos are native to Australia and neighboring islands. There are many different kinds and sizes of kangaroos. The smallest, called Wallabies, are as small as a rabbit. The largest, called the great kangaroos, may grow 7 feet tall and weigh about 200 pounds. Fossil bones left in rocks show that one kangaroo that lived long ago grew twice as large as any living kind of kangaroo.

One type of king snake is called milk snake because farmers once believed that it took milk from cows. Today,

scientists know that no snake is physically able to take milk from a cow. But any snake might drink milk it finds in a pail, mistaking it for water. Like other king snakes, the milk snake helps man by killing harmful rodents, such as rats and mice. It often comes into barnyards to look for nesting places of the rodents.

Knife is probably the most useful of all the common tools used by man. The knife was one of the first tools developed by the cave man. He sharpened and ground thin pieces of flint and quartz to make knives for skinning animals and cutting meat.

Language is the main means of communication between peoples. But so many different languages have developed that language has often been a barrier rather than an aid to understanding between peoples. For many years, men have dreamed of setting up an international, universal language which all people could speak and understand.

Life is one of the great mysteries of the universe. It has no simple definition. We can easily recognize most things as either living or nonliving. We know that a rabbit is alive, but a rock is not. We identify living things by certain activities that nonliving things do not perform. For example, living things grow, reproduce themselves, and require food. Living things also differ from nonliving things in the way the substances that compose them are put together.

Manufacturers and farmers produce goods chiefly for sale. When customers with plenty of money or credit stand ready to

buy the products of farms and factories, business is good. New factories are built to meet growing demands. The building of a factory takes large sums of money from people who have savings or profits to invest, and gives it to workers who want to spend it for goods. In this way, the effort to meet a growing demand actually makes the demand itself steadily greater.

Mosquitoes are among the deadliest enemies of man and domestic animals. Some mosquitoes spread such deadly diseases as malaria and yellow fever. In the far North, great swarms of mosquitoes may sting animals to death. And in many places in the United States, persons out of doors suffer from annoying and painful mosquito bites.

Napoleon I crowned himself Emperor of the French and created an empire that covered most of western and central Europe. He was the greatest military genius of his time. Napoleon's armies crushed one foe after another until he seemed invincible. For nearly 20 years, many European nations fought him.

Nervous system is made up of the brain, the spinal cord, and the nerves that extend throughout the body. The various sense organs contain specialized nerve endings that respond to stimuli, and tell us what is going on in the world around us. Nerves run from the brain and spinal cord to the muscles and organs of the body. They enable us to react and adjust to our environment.

Northwest territory was a vast tract of land lying north of the Ohio River, west of Pennsylvania, and east of the Mississippi River. It extended to the northern limits of the United States. The states of Ohio, Indiana, Illinois, Michigan, Wisconsin, and part of Minnesota east of the Mississippi were carved out of the Northwest Territory.

Ocean is the great body of water that covers more than 70 per cent of the earth's surface. The world is really one huge ocean, broken here and there by islands that we call continents. The ocean is so vast that you could sail across it for days without seeing land. It is also deep. In some areas, the bottom of the ocean lies more than six miles below the surface.

Offset is a printing process in which the printing is done first on the rubber surface of a rotating cylinder. The impression is then transferred to paper by the pressure of other cylinders. The term offset describes the printing, or offsetting, of the ink from the rubber. In recent years, offset has grown more rapidly in popularity than any other printing process.

Pearl is one of the most valuable gems. Large, perfectly shaped pearls rank in value with the most precious stones. But pearls differ from other gems. Most gems are minerals that are mined from beneath the earth. But pearls are formed inside the shells of oysters. Mineral gems are hard and usually reflect light. But pearls are rather soft, and absorb as well as reflect light.

Persons who want to be physical therapists can take one of two kinds of educational programs. One program, for high-school graduates, leads to a bachelor's degree. Studies include courses in humanities and sciences. It also includes professional subjects, such as the theory and practice of physical therapy. The other program is for college graduates. It consists of 12 to 18 months of intensive professional education.

Port is a place where ships and boats load and unload passengers and cargoes. Large bustling ports have buildings and equipment for receiving, storing, and reshipping goods. Such facilities include wharves, warehouses, tugs, ferries, mechanical loaders and unloaders, and railroad and truck transportation.

Every day, hundreds of ships enter and leave ports throughout the world. They cross the oceans carrying passengers and cargo from country to country. The ships that sail the seas provide man with one of his most important kinds of transportation.

When anything has size, weight, number, mass, or volume which may be measured, increased, or lessened, it is said to have quantity. The term may also mean a certain or considerable amount, as in the sentence, "Quantities of shells are found on the shore."

Quince is a shrub or small tree that is related to the apple and pear. It is grown mainly for its fruit. The quince fruit is shaped somewhat like a pear and has a golden-yellow color. It has a fragrant smell.

Razor is a cutting instrument used to remove hair from the skin. The men of ancient Egypt used razors, because they liked clean-shaven faces better than bearded ones. Julius Caesar and many men of his time were entirely clean shaven.

Rice is one of the world's most important food crops. About half the people of the world eat this valuable grain as their chief food. Most of these people live in Asia, where rice is even more important than wheat is to the people of Europe and the United States. Many Asians eat rice three times a day, and often have little else to eat. People in India and Japan eat an average of from a half to two-thirds of a pound daily.

Rugs and carpets are floor coverings used in homes and other buildings. They are useful for many reasons. They add warmth and comfort to a room. They soften noises and protect floors. They also add beauty. A carefully chosen rug or carpet not only adds its own color and loveliness, but also sets the feeling and tone for a whole room. It can fit together various furniture styles. It can blend the colors used in decorating the room. It can add to the beauty of other furnishings.

Space travel is man's greatest adventure--the chance to explore the moon, the planets, and the stars. Giant rockets carry explorers into space. A roaring blast of orange flame lifts the rocket from the ground. Climbing into the blue sky, the rocket leaves a white trail. Then it speeds out of sight

into space, where the sky is always black and the stars always shine.

Telegraph was the first method used to send messages by electricity. At one time, most telegraph messages were sent by tapping out words letter by letter with a telegraph key. The telegraph changed the dots and dashes of the Morse code into electrical impulses and sent them over telegraph wires. For more than 40 years in the 1800's, the telegraph was man's fastest means of long-distance communication.

Toad is a tailless animal that breeds in water but spends most of its time on land. Toads make up one of the large families of tailless amphibians. They make their homes wherever the summer is warm and there is fresh water in which to breed.

Unemployment is the stage of being out of work. It applies to persons who are normally members of the labor force and are able and willing to work full or part time. It applies both to men and women who are looking for work, and to those who are temporarily laid off their jobs and are not looking for work.

Valentine's day is celebrated on February 14 as a festival of romance and affection. People send greeting cards called valentines to their sweethearts, their friends, and members of their families. Verses on many valentines contain tender thoughts. Other valentines may include humorous pictures and sayings. But almost all ask, "Be My Valentine."

Viking was a member of the Scandinavian bands of sea rovers who launched a series of devastating and successful raids on England, France, Germany, Ireland, Italy, and Spain between the 700's and the 1100's. The vikings also explored and settled Greenland and Iceland. They tried to establish settlements in America, but did not succeed.

Wart is a horny growth on the surface of the skin. Warts may appear anywhere, in a wide range of shapes, sizes, and numbers. Flat warts that grow on the sole of the foot look like corns and hurt like tacks. Warts on a man's face may form little bear-like projections. In moist parts of the body, warts may grow into masses like tiny cauliflowers. Warts can even appear on the lips or tongue.

Weather includes all the daily changes in temperature, wind, moisture, and air pressure. It affects everyone. Today's weather may make us feel hot or cold. We may get soaking wet in a sudden shower, or have to struggle through deep snow. Bright sunshine may make the day cheerful and happy. Dark, dull clouds may make us sad and unhappy. Too much rain can cause floods. Too little rain may kill farm crops.

Wheat is the world's most important grain crop. Wheat kernels, or seeds, are ground into flour to make bread and other products. These wheat products are the main food of hundreds of millions of people throughout the world. As a result, wheat covers more of the earth's surface than any other food crop. The world's wheat farmers grow about 9,000,000,000

bushels of wheat a year. This wheat would fill a freight train stretching one and third times around the world.

Wherever men have lived together, they have found it necessary to develop rules of conduct. They need rules for the settlement of disputes. They also need rules for the organization of their governments. Law is the set of rules which the government enforces through its police, its courts, and its other agencies.

The discovery of X rays was one of the most important events in modern physics, because of the information these rays supply concerning the nature and properties of matter, and their effects upon living tissue. The rays cannot be seen, but they can penetrate the human body, heavy metal castings, or the latticework of atoms inside solids.

Yale University is a privately endowed, nonsectarian school in New Haven, Conn. Chartered in 1701, Yale is the third oldest university in the United States. Only Harvard University and the College of William and Mary are older. The undergraduate school known as Yale College, is open to men only. But Yale admits women to the graduate and professional schools. Yale students come from all 50 states; Washington D.C.; Puerto Rico; The Virgin Islands; and about 90 other countries.

The yeast with which people are most familiar is a substance that bakers put in dough to make it rise. This yeast contains a mass of tiny, one-celled plants called yeasts.

Yeasts are among the simplest kinds of plants. Like mushrooms, they belong to the group of plants called fungi.

Zipper is a term often used to mean any kind of slide fastener. These fasteners have two edges of teeth and hollows which fit into each other very snugly. A slide draws the two edges together and meshes the teeth into the hollows. The edges remain fastened until they are released again by drawing the slide back over them, unmeshing the teeth. Slide fasteners are used to fasten clothing, luggage, brief cases, and similar articles.

Zoo is a collection of different kinds of living animals. A large zoo has birds, snakes, mammals, and other kinds of animals from all parts of the world. Most zoo animals live in cages. Others live in open areas that look like their natural surroundings.

DURATIONAL CHARACTERISTICS OF ORAL READING:
EFFECTS OF SAMPLE SIZE ON TEST-RETEST RELIABILITY

by

PATRICIA H. FLANAGAN

B. S., Western Michigan University, 1953

_____

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

The purpose of this research was to assess the test-retest reliability of the distribution of the durational properties of speech and silence events of normal speakers, under conditions of oral reading, as a function of speech sample size. Durations of speech and silence events during on-going speech have proved useful as a physical parameter of speech which are measurable and not subject to observer bias. Previous research has posed the question of whether or not increased test-retest reliability could be obtained by lengthening the speech sample size being analyzed.

Ten young adult males served as subjects for this research. Two samples of oral reading were obtained from each subject. The samples were equivalent and obtained at least 48 hours apart.

The samples of oral reading were recorded, displayed on a cathode ray oscilloscope, photographed and durations of speech and silence events subsequently measured and tabulated by their durations into eight class intervals. The class intervals were x - 99 ms, 100 - 199 ms, 200 - 299 ms, 300 - 399 ms, 400 - 499 ms, 500 - 599 ms, 600 - 699 ms, and 700 - x ms. The number of speech or silence events for each duration were sub-totaled within speech sampling periods of one, two, three, four and five minutes. The percentage of individual test-retest reliability was measured at each class interval for each speech sampling period. Median values for test-retest reliability

were then obtained for each class interval and across class intervals for each speech sampling period.

This research provided evidence that the reliability of measurements of the frequency distributions of speech and silence events during oral reading are functionally related to the length of the speech sample. The median reliabilities for speech events across class intervals increased progressively from the one minute sampling period to the five minute sampling period. The median reliabilities for silence events also increased progressively from the one minute sampling period to the five minute sampling period.

The research suggests that a five minute speech sample size does yield acceptable test-retest reliability if a minimum of 80 percent agreement between trials is acceptable. A best fit line projected from the obtained data points suggests that 90 percent agreement could be achieved by increasing the speech sample size to seven or eight minutes.

The discussion of this research considered practical applications of the durational analysis of speech and implications for further research.