FINAL REPORT

PROJECT NO. A-483

ENGINEERING INVESTIGATION AND STUDY OF AUTOMATIC VOICE INTELLIGIBILITY COMPUTATION METHODS

By

D. W. Robertson and C. W. Stuckey

- o - o - o - o - o -

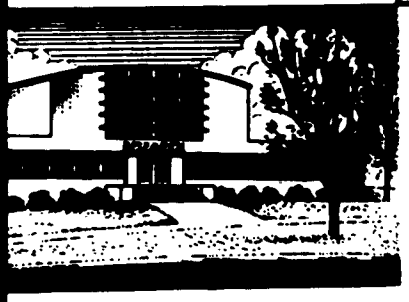
CONTRACT NO. AF-30(602)-2150

- o - o - o - o - o -

1 February 1962

Prepared for

Rome Air Development Center
Air Research and Development Command
United States Air Force
Griffiss Air Force Base, New York



Engineering Experiment Station
Georgia Institute of Technology
Atlanta, Georgia

RADC-TR-62-95

ENGINEERING INVESTIGATION AND STUDY OF AUTOMATIC VOICE INTELLIGIBILITY COMPUTATION METHODS

By

D. W. Robertson and C. W. Stuckey

ENGINEERING EXPERIMENT STATION
of the Georgia Institute of Technology
Atlanta, Georgia

FINAL REPORT

PROJECT NO. A-483

Contract No. AF 30(602)-2150

Prepared for

Rome Air Development Center
Air Research and Development Command
United States Air Force
Griffiss Air Force Base, New York

FOREWORD

This is the Final Report covering work performed over a period of nine months under Contract No. AF 30(602)-2150, Supplemental Agreement No. 1 entitled "GEL Intelligibility Meter Study." The report deals with an investigation and study of electronic techniques for automatically determining the intelligibility of speech corrupted by various interferences and contains data resulting from laboratory experiments performed at Georgia Tech.

One technical note and a "Final Report for Item I" were previously published during the course of the project. The technical note dealt with the development of a method to evaluate learning in articulation testing, and the report dealt with the evaluation of the GEL Speech Systems Test Set.

The program was conducted under the general supervision of W. B. Wrigley, Head, Communications Branch. In addition to the authors, staff members O. B. Francis and Guy H. Smith, Jr. also participated in the work described in this report.

Respectfully submitted:

D. W. Robertson

D. W. Robertson
Project Director

Approved:

M. W. Long

M. W. Long, Chief
Electronics Division

ABSTRACT

A technique is described for electronically determining the degradation in intelligibility that occurs when speech is corrupted by added noise. Primary attention is given to the extraction of the relative time positions of the speech waveform zero crossings, the computation of bivariate correlation coefficients from the zero crossings, and the comparison of the coefficients with articulation scores at various S/N ratios.

Although the quantity of experimental data is limited, the results of the investigation indicate that the technique is useful in providing a measure of intelligibility comparable to that obtained by a listening team.

TABLE OF CONTENTS

This report contains 70 pages.

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

1. INTRODUCTION

1.1 Background

In order to predict the degradation in performance of voice communication systems subject to interference, and in order to evaluate different system designs with respect to interference susceptibility, it is necessary to have some quantitative measure of performance. Basic to a meaningful measure of performance for voice systems is the intelligibility of the message at the receiver output.

A number of tests have been developed which permit a quantitative measure of intelligibility to be obtained. The generally accepted standard articulation scoring test provides one measure of intelligibility. In this test, a preselected group of syllables, words, or sentences is presented orally to a panel of listeners who record what they hear. The results expressed in per cent correct and commonly known as the "articulation score" (AS) provide the measure of intelligibility. Because it is subjective, this test yields highly variable results in individual cases. Therefore, a relatively large number of listeners must be used in order to obtain statistically meaningful results. The proper conduct of such a test is tedious and time consuming. This situation is aggravated by the necessity of a fairly lengthy training program for the listeners.

The disadvantages of articulation scores have been recognized by several agencies engaged in interference studies. In particular, the General Electronic Laboratories, through a Signal Corps Contract[1], has developed the GEL Speech Systems Test Set to replace, in special cases, the group of trained human listeners. This test set measures the degradation in understandability

of a speech sample that passes through a noisy intelligence channel by comparing the original and degraded speech signals and by computing a running short-time amplitude correlation function. This function, after being integrated over the test period, is presented in the form of a meter reading called the "pattern correspondence index" (PCI). The instrument is designed with the intention that the PCI would be monotonically related to the AS obtained under identical conditions. The test set would, in effect, serve as a standard listener capable of providing a synthesized listener score.

Prior work on this contract included an engineering investigation, study, and evaluation of the GEL Test Set method of measuring speech intelligibility.[2]

Although the GEL instrument was found to be satisfactory for a number of interference conditions, the results were not valid for various signal distortions. Further, the method required that the input and output of the system under test be physically present in real time at the input of the test instrument.

1.2 Purpose and Scope

The purpose of the work reported herein was to conduct an engineering investigation and study directed towards the development of a nonobserver technique for determining the retained intelligibility of speech after passage through a communications system.

The scope of the investigation included the use of mathematical models and computational methods for approximating the ability of human listeners to measure intelligibility and the development and test of electronic techniques for implementing likely models or methods.

1.3 Approach

The primary effort of the investigation was directed to the identification of various parameters associated with the electrical replica of the acoustical speech waveform which appeared to contain the basic intelligibility, to the extraction of these parameters in a form amenable to the application of statistical techniques, and to the relation of the results to those obtained by an articulation team for identical interference conditions. From this comparison, the ability of a particular parameter to provide a measure of intelligibility is determined.

2. DISCUSSION

2.1. Intelligence Carrying Elements of Speech

In the search for a nonobserver means for measuring the degradation in intelligibility that occurs when speech is passed through a communications system, it appears necessary to approximate the decision mechanism actually employed by a human listener. The listener receives a signal and is forced to decide which of a class of possible transmitted words it most closely resembles. This means he determines the "distances" from the observed signal to the members of the vocabulary of the possible transmitted ones and selects the "nearest." The relative distance between a received signal and a vocabulary word is then the basis of selection.

The objective sought then is a reasonable approximation to the actual distance function used by the human listener. Apparently the best approach to obtaining a good approximation of the true distance function is to select likely candidates and compare them on a statistical basis with the results from experiments using articulation teams. One criterion in the choice of candidates is the selection of particular parameters which are closely related to the speech attributes that are thought to carry the intelligence.

It is generally concluded that the essential elements of speech intelligibility are contained in the instantaneous spectral envelope of the speech wave. Further, the contrast between the high rate of information contained in the complex trace of the original wave and the limited information-rate ability of the ear (8,000 bits per second) indicate that hearing incorporates a short-time averaging process.[3]

The sound spectrogram, or sonogram, is essentially a pictorial representation of the short-time average power spectrum.[4] The spectrogram may also be

considered under certain limitations as a running short-time representation of the Fourier transform of the original speech wave in which the part representing phase is omitted, and the magnitude is squared. Such approximations appear valid since the ear is relatively insensitive to phase and amplitude insofar as they are related to speech intelligibility. The spectrogram has thus provided a convenient working representation for the analysis of speech and its invariants. In fact, speech of good intelligibility has been synthesized directly from artificial or "painted" spectrograms.[5]

Similarly, the vocoder may be considered as extracting the short-time average frequency components on a running time basis.

The successful utilization of these techniques substantiates that the intelligence is carried in the spectral envelope and further that it may be extracted or synthesized on a short-time running average basis.

The analysis and study of spectrograms have resulted in the discovery of certain regions of maximum intensity whose relative time varying positions are independent of the pitch and emotional qualities of an individual speaker or a number of speakers. These regions or formants can be extracted, and they are in fact represented by the "painted" sonograms used to artificially create speech.

Licklider and Pollack[6] demonstrated that speech intelligibility remains even when the original wave is subjected to severe amplitude clipping and that the intelligibility is retained in the zero crossings of the speech wave. Zhang[7] and Crater[8] have established the close relationships that exist between the running power spectrum, clipped speech, and the time varying correlation function. They conclude that each representation contains the speech invariants.

Chang further indicates that the major formant movements may be closely approximated by obtaining the running average of the number of zero crossings in the original and differentiated wave.

Since it contains the intelligence carrying attributes of speech, the "running power spectrum" appears to have considerable merit as a choice for a distance function approximation. However, as implemented by the GEL Speech Systems Test Set, this choice was only useful with linear systems and interferences.

Another possibility is the pattern of zero crossings of the speech waveform. The good intelligibility of infinitely clipped speech lends credence to this choice.

A study of various zero crossing parameters indicated that additional knowledge of the crossover information should be obtained before it could be considered for use either alone or in conjunction with other attributes for the purpose of measuring speech degradation.

Consider the waveforms of normal and infinitely clipped speech as shown in Figure 1 (a) and (b). If the zero axis crossing positions are defined in some manner, such as differentiation and full wave rectification of the rectangular clipped speech, the resulting pulse position markers, as shown in Figure 1 (c), may be used to actuate a bistable device and the speech may be reconstructed with little loss in intelligibility. Since the intelligibility is essentially retained with infinitely clipped and reconstructed speech, it is apparent that the intelligibility must be contained in the relative time positions of the zero crossings. Dukes[9] has shown that the high degree of intelligibility is to be expected since the spectral energy content of the clipped speech wave approximates on the average, that of the original speech.

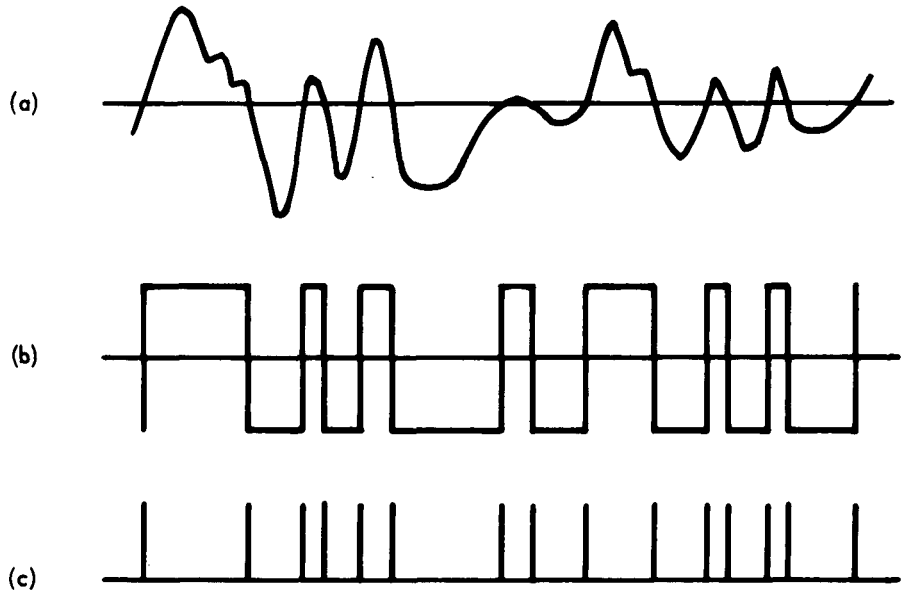


Figure 1. Relative Time Positions of Zero Crossings.
(a) Normal Speech.
(b) Clipped Speech.
(c) Clipped Speech After Differentiation and Rectification.

The retention of the intelligibility and the tractable nature of the pulse representation of the zero crossings provide a convenient approach to the statistical investigation of methods for specifying speech degradation.

2.2 Zero Crossover Investigation

2.2.1 Test Material. The Harvard PB-50 Words, List No. 1, were used in the crossover tests. There are 20 Harvard PB-50 word lists, each consisting of 50 unrelated monosyllabic words and each approximating the relative frequency of phoneme occurrence in conversational American English.[10]

The words, which were spoken by a local radio announcer, were recorded on one track of a dual channel tape.

The average power over the duration of each word was measured in arbitrary units, using an electronic power meter which was designed at Georgia Tech specifically for this purpose;[11] relative powers in decibels were then computed using the smallest word power as a reference. By properly presetting an attenuator, the words were re-recorded with equal average powers. These re-recorded words are referred to as "leveled power words."

Band-limited white noise derived from a General Radio Type 1390-A Random Noise Generator was recorded on the second track of the dual channel tape at a level which yielded a signal-to-noise ratio of unity. Equal level 1,000-cps reference tones were recorded on both tracks of the tape to serve as guides for setting signal-to-noise ratios.

Continuous tape loops were made from a copy of the PB-50-1 master tape. Each loop contained one of the "leveled power words" on one track and band-limited white noise on the other. A loop from a section containing the 1,000-cycle test tone permitted the 0-db signal-to-noise output to be established on playback. A negative timing pulse was recorded on the word track of each loop, approximately 250 milliseconds prior to the beginning of the initial word sounds. This timing pulse served to actuate the crossover counter control circuitry and served as the reference point for the selection of desired word segments.

The PB-50-1 word group was selected for use in the crossover tests because of the existence of considerable articulation score data for a number of interference conditions.[2] In addition to the use of the PB-50-1 words, a selected list of "CVC Words" (nonsense syllables) were prepared for use on

the project. Although some articulation score data were obtained (see Section 2.3.2 of this report), the time schedule did not permit crossover data to be collected for these word units.

2.2.2 Test Equipment. Figure 2 is a block diagram of the system used in obtaining the zero crossover counts. Calibrated attenuators (Nos. 2 and 3 of Figure 2) permitted various speech and noise signal levels to be established prior to being combined in the bridge mixer. A 200- to 4,000-cycle band-pass filter is used to confine the mixed signal to the frequency band necessary for good voice intelligibility.

It has been established that the original speech power spectrum is least altered, for infinitely clipped or crossover reconstructed speech, if the energy concentrations at each frequency are all of the same general magnitude.[8] In order to equalize the energy concentration, a pre-emphasis network was used which had a response that was the inverse of the frequency rms pressure spectrum curve as established by Dunn[12] for conversational English. The response curve for the band-pass filter and the pre-emphasis network combination is presented in Figure 3.

The filtered and pre-emphasized speech-noise signal is fed to the crossover detector. A schematic diagram of the detector is shown in Figure 4. After two preliminary stages of amplification and clipping, the speech waveform actuates a Schmidt trigger circuit. The square wave output of the Schmidt trigger is differentiated and the resulting pulses are passed through a full wave rectifier. These pulses then trigger a one-shot multivibrator which produces constant-amplitude, constant-width output pulses. These output pulses, which define the relative time positions of the speech zero

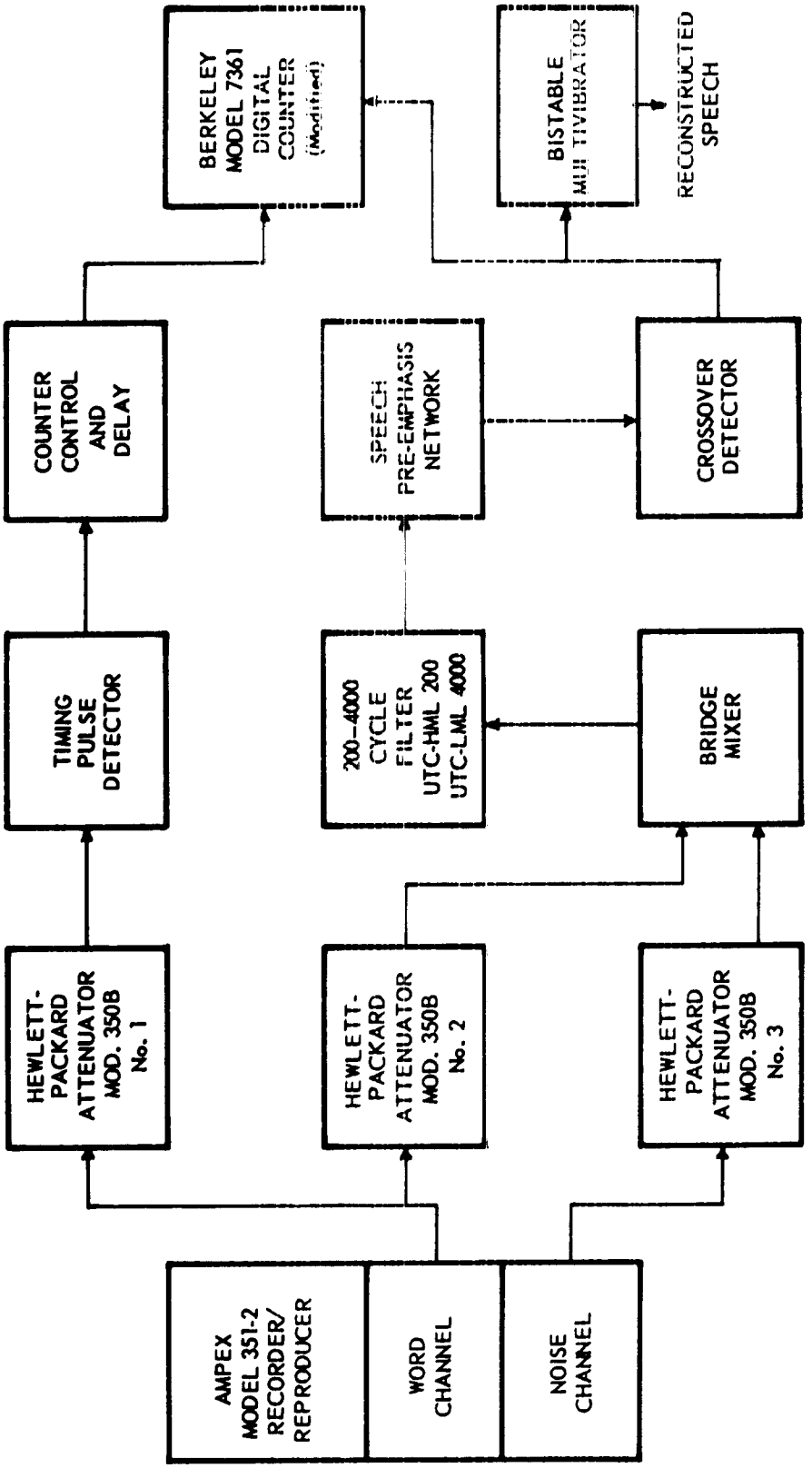


Figure 2. Block Diagram of the Crossover Counter System.

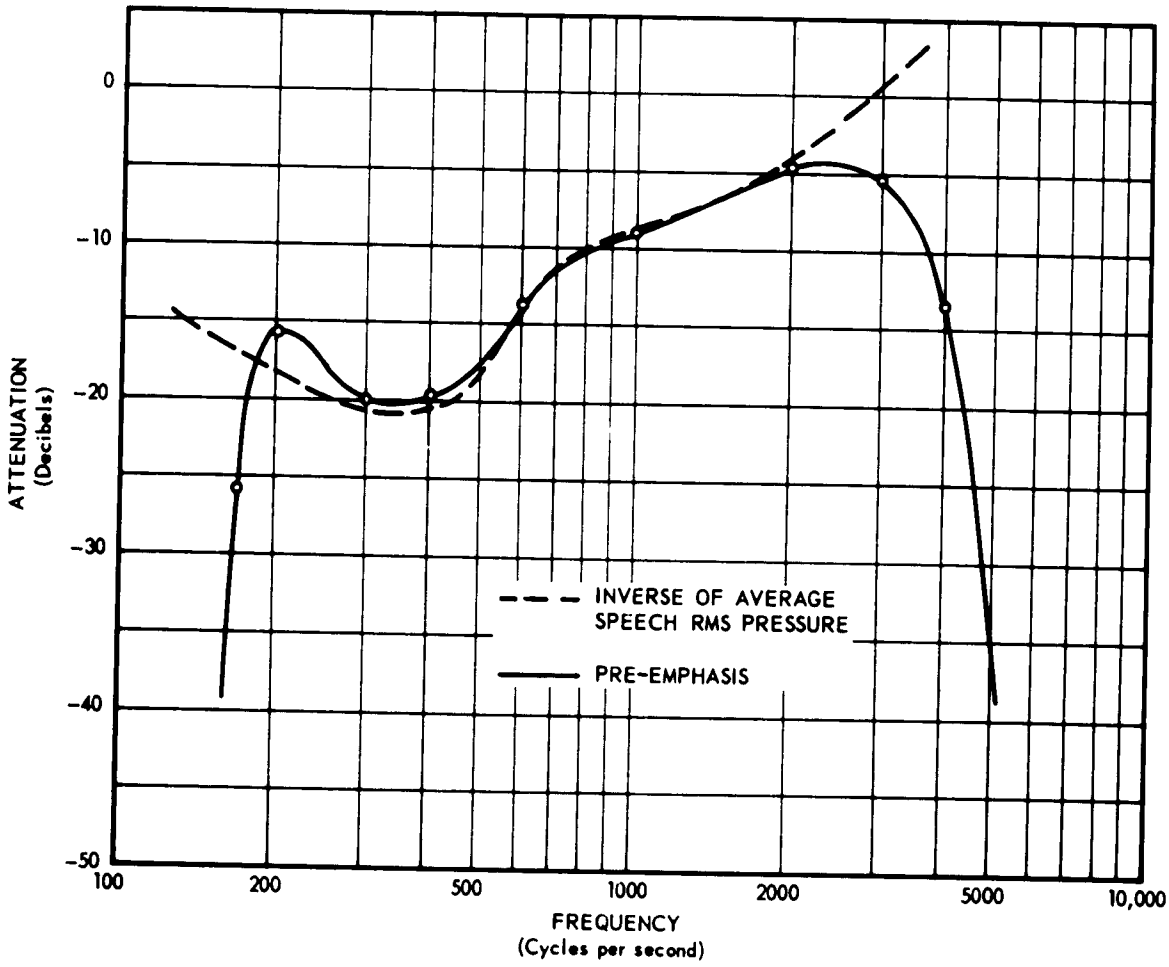


Figure 3. Bandpass Characteristic of Pre-emphasis Network.

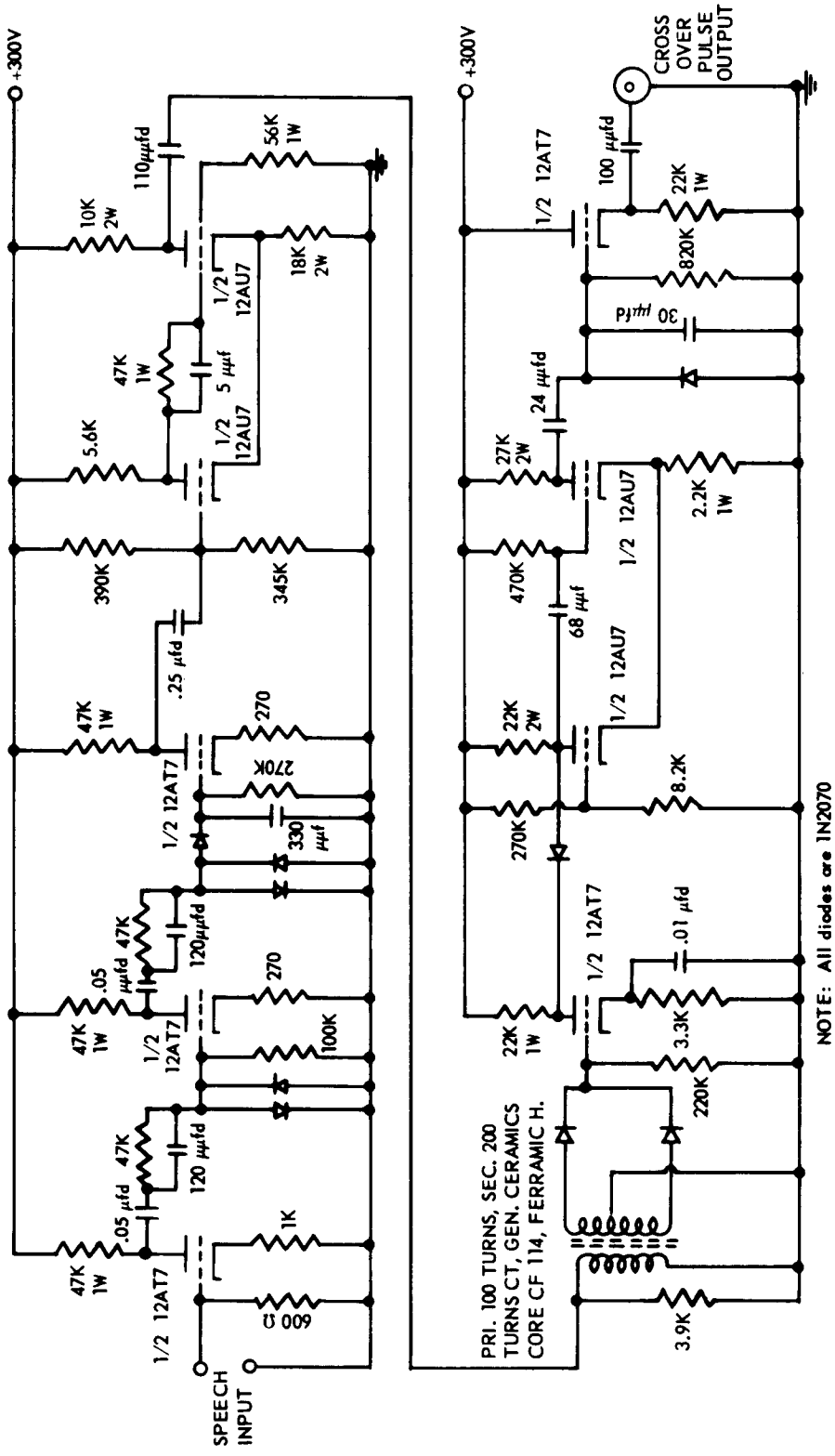


Figure 4. Schematic Diagram of Crossover Detector.

crossings, are designated as crossover pulses and are fed to the counter circuitry for data collection.

In order to determine the quality and intelligibility that were retained in the crossover information, a number of listener tapes were made of PB-50-1 words reconstructed from the zero crossover pulses. The reconstructed speech sounds are obtained from a bistable multivibrator actuated by the constant-amplitude, constant-width crossover pulses. The rectangular multivibrator output is fed directly to an Ampex 351 Recorder/Reproducer and recorded on tape. Although the amplifier and tape frequency response provide some band limiting, the speech is further constrained on playback by a 200- to 8,000-cps band-pass filter which is located just prior to the listener earphone distribution line.

A sample of the voice signal, containing the negative timing pulse, is fed through Attenuator No. 1 to the timing pulse detector. The attenuator allows the timing signal level to be independently set for each tape loop. The timing pulse detector, a duplicate of the crossover detector, provides an initial output pulse in time coincidence with the timing pulse and initiates the counter control action.

The instrumentation for obtaining the zero crossover counts is centered around a Berkeley Model 7361 Preset Universal EPUT and Timer. Modifications and auxiliary control circuitry permit the desired time segments to be selected.

In the normal "E/UT X N" mode of operation of the Berkeley 7361, the 0.1- or 1.0-millisecond period clock pulses are fed to a digital divider. The division ratio may be selected so as to obtain pulses separated by some multiple (1 to 9,999) of the clock pulse interval. These pulses actuate the

internal gate control circuitry, opening a count gate for the length of time separating successive pulses. During this interval "input" pulses pass through the gate and are counted. The pulse count is displayed in the form of illuminated decimal digits. At the end of the display duration, which is continuously variable from 0.1 to 10 seconds, all circuitry is reset and the cycle is repeated.

In order to obtain a count of the speech crossover pulses which occur during a selected word time segment, two additional capabilities are required: (1) a method for relating or synchronizing in time the word timing pulse and the crystal controlled clock pulses, and (2) a timing capability for providing a selected delay from the word timing pulse to the opening of the count gate. The counter control circuitry performs these functions.

A functional block diagram of the counter and counter control system is shown in Figure 5, and a timing diagram for the system is depicted in Figure 6. The tube numbers of Figure 5 are the Berkeley Model 7361 designations and the broken line boxes indicate the locally constructed auxiliary circuits, whose schematic diagrams are presented in Figures 7 and 8.

The timing pulse from the word track of the tape loop initiates the counter sequence by opening a gate through which the basic 100-kc clock pulses pass. This action relates the word timing pulse to the initial clock pulse to within $\pm$ 10 microseconds. These pulses are subsequently divided to produce the timing increment pulses (choice of 1.0- or 0.1-millisecond periods) which serve as the inputs to both the external preset and the internal selective digital dividers.

The preset divider emits a counter start pulse delayed from the initial clock pulse in steps of 10 by any desired number up to 9,990 of the increment

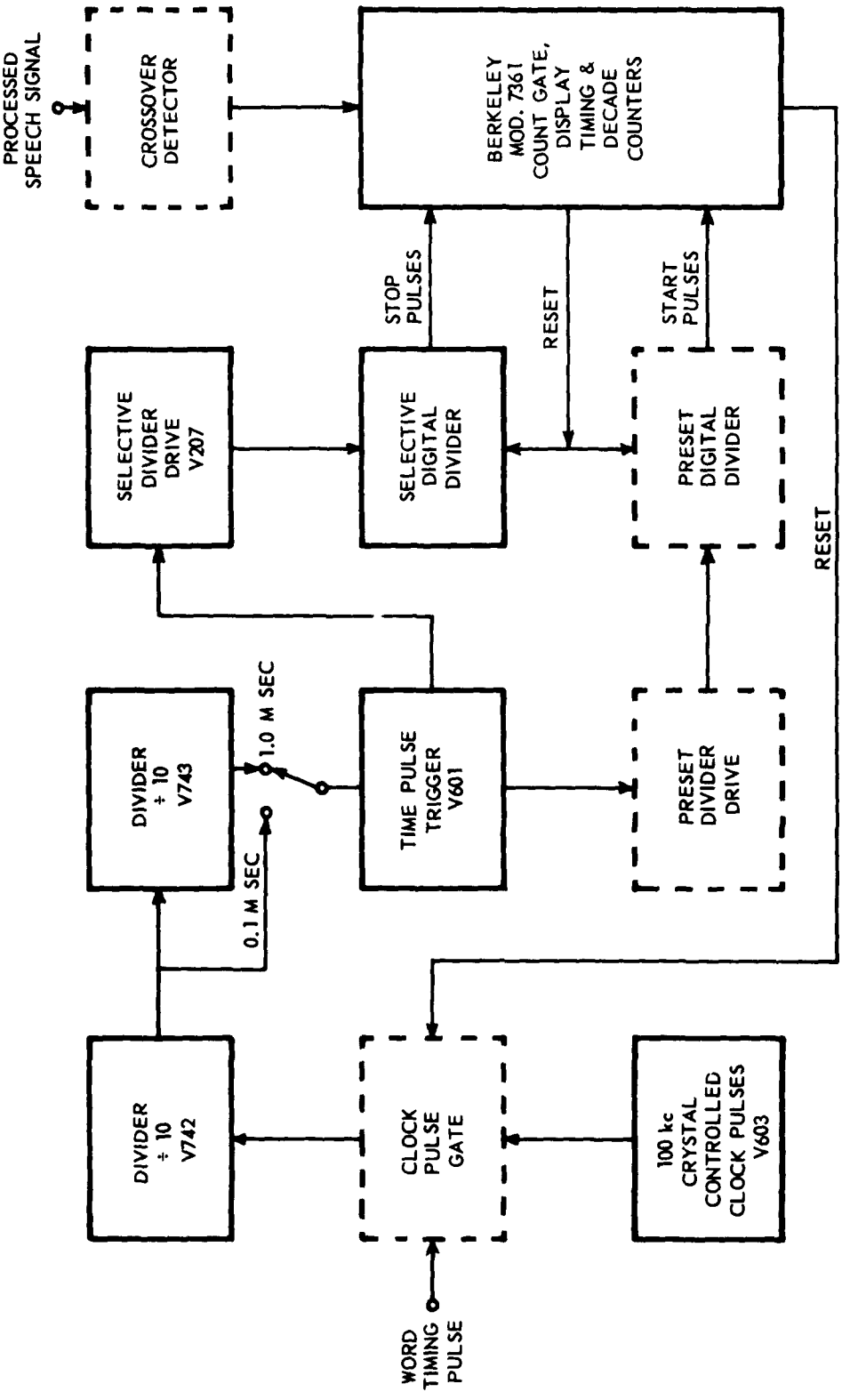


Figure 5. Functional Block Diagram of Counter and Counter Control.

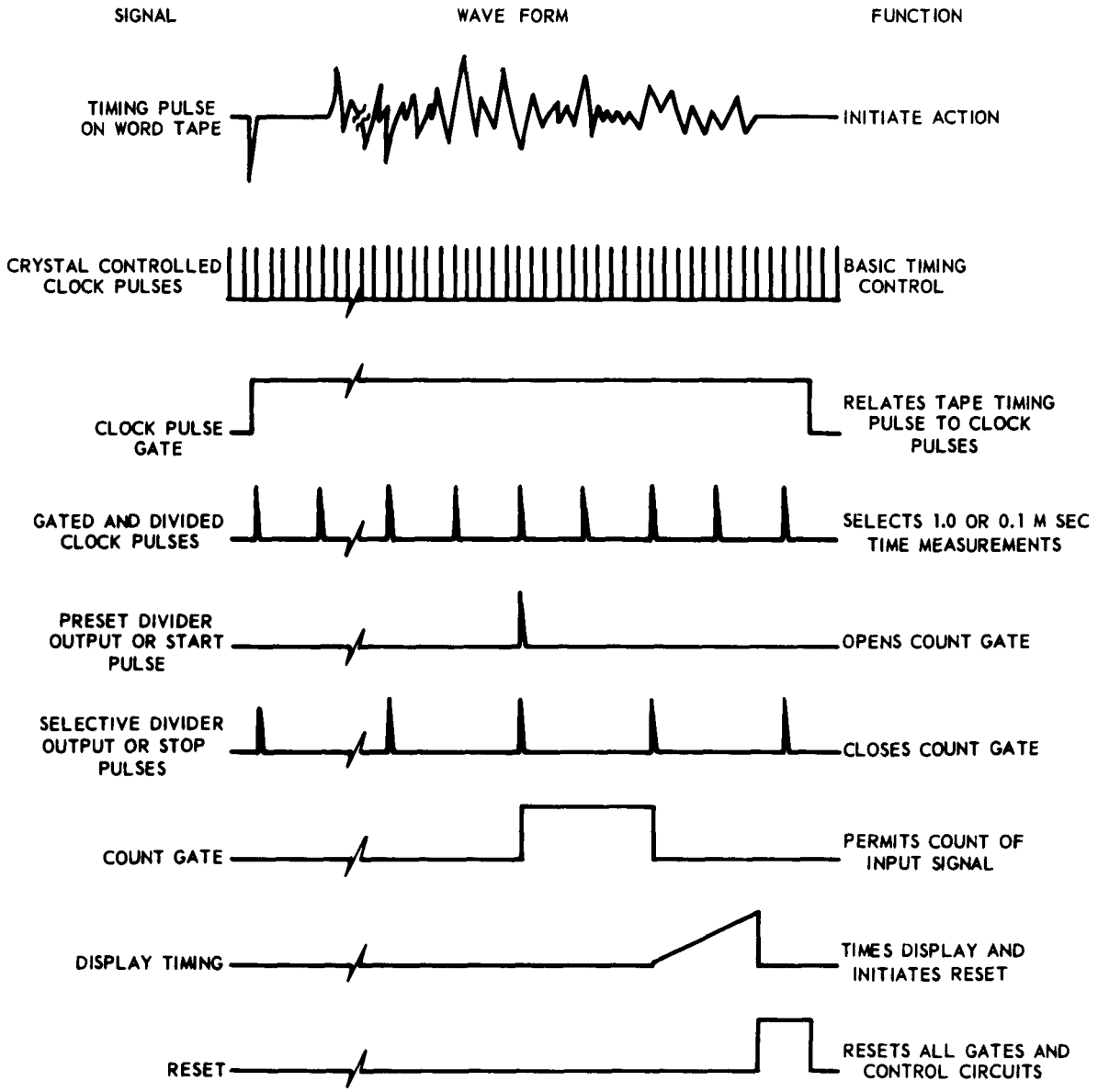


Figure 6. Counter and Counter Control Timing Diagram.

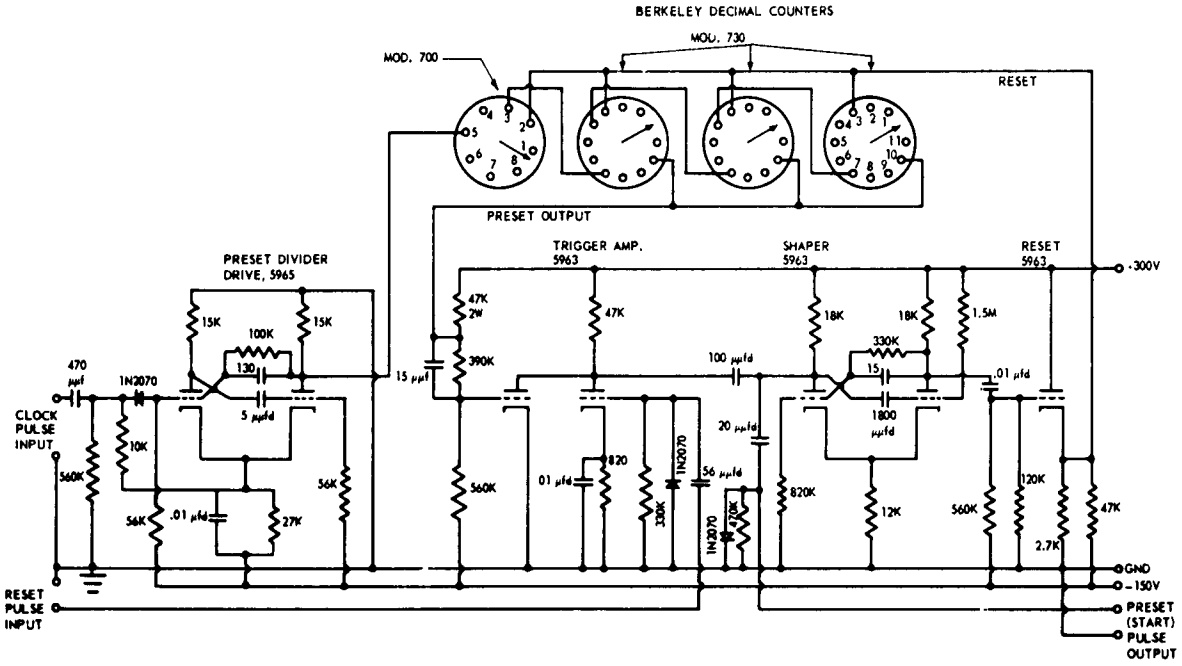


Figure 7. Preset Digital Divider and Driver.

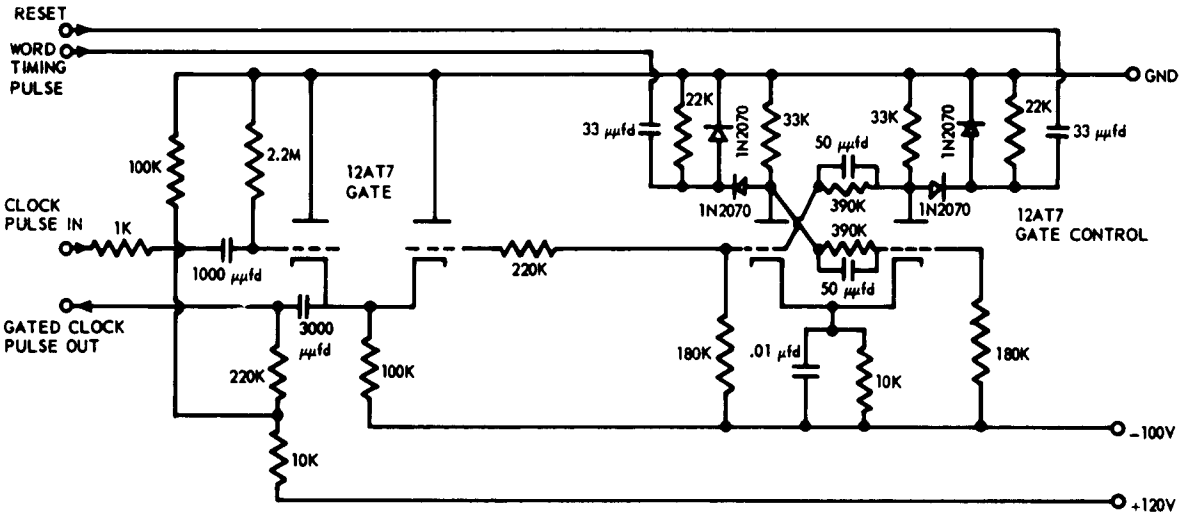


Figure 8. Clock Pulse Gate.

input pulse. Thus a digital delay, ranging from 1 to 999 milliseconds for the 0.1-millisecond pulses or from 10 to 9,990 milliseconds for the 1.0-millisecond pulses, is obtained for the purpose of opening the crossover count gate.

The internal digital divider of the Berkeley 7361 counter acts as a variable scaler which emits output pulses having periods from 0.1 to 9,999 milliseconds, as determined by the setting of four rotary switches. Although these outputs or counter "stop" pulses are continuously produced, they have no effect prior to the opening of the crossover count gate. However, the first stop pulse that occurs after the count gate is opened serves to close the gate, stop the crossover count, and initiate the display time action.

At the culmination of the count display, a reset pulse is generated which closes the clock pulse gate and returns all circuitry to the initial state and thus completes the counter sequence. When the timing pulse from the tape loop occurs again, the cycle is repeated.

In normal operations, the selective divider period is set to give the desired segment count duration, for example 10 milliseconds. The delay from the timing pulse, as determined by the preset divider push buttons, is then increased until the initial word crossover counts occur. The resulting count for the initial word segment is recorded, the preset divider is manually stepped for an additional delay of 10 milliseconds, and the second 10-millisecond segment count is recorded. This procedure is progressively continued until the crossover count for each 10-millisecond segment of the word is obtained.

2.2.3 Test Results. From a number of intrinsic zero crossing parameters, the following are of particular interest:

(1) The number of crossings, N.

(2) The crossover rate, $R = dN/dt$.

(3) The average rate of occurrence of crossovers, R_{av} = N/sec.

(4) The time rate of change of the rate of occurrence, $dR/dt = d^2N/dt^2$.

In addition to these, a number of subparameters exist, which may prove to be of importance. Of those listed, it was felt that the rate of occurrence, the time rate of change of occurrence, or a combination of these two, may provide the relationship sought. This study is primarily concerned with an investigation of $R = dN/dt$.

The criterion which was used to determine that the parameters N and R_{av} did not carry the intelligence was a plot of the parameter probability densities as a function of a selected continuum. For test words appearing in the clear (AS of 100 per cent at $S/N = +\infty$), the probability density plots should not overlap on the continuum. As the S/N ratio is decreased, the density plots of the words should tend to overlap such that at very low (or negative) S/N ratios (AS of 0), the density plots should occupy approximately identical positions on the continuum and become indistinguishable. This basic S/N relationship which is the criterion for choosing a valid continuum appears to be a necessary, if not a sufficient, condition for establishing the existence of a relationship between a particular zero crossing parameter and the intelligence carrying attribute of speech.

Zero crossover data for the parameters N and R_{av} were collected from the PB-50-1 words in the clear. The probability density functions representing ten successive counts for a number of the PB-50-1 words are shown in Figures 9 and 10. The density functions are plotted as triangles for simplicity. The triangle peak occurs at the mean value, and the skirt extremities are located at

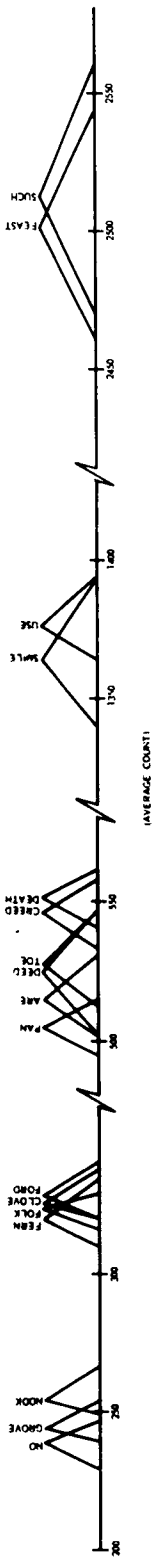


Figure 9. Continuum for Average Crossover Counts.

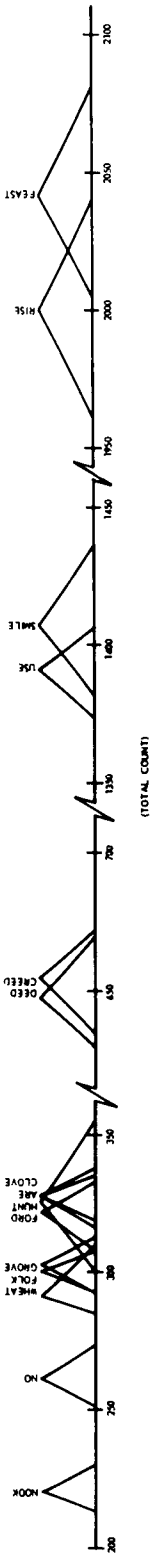


Figure 10. Continuum for Total Crossover Counts.

the low and high recorded counts. As shown, the overlapping and bunching of plots revealed that neither of these two parameters contained in itself the desired relationship. The particular data represented by the curves were collected early in the project, prior to several equipment changes which permitted lower threshold levels and, consequently, larger total counts. However, the results are representative and consistent with later observations.

The test instrumentation capability for obtaining the digital zero crossover information permitted the crossover rate and the change of rate to be approximated as follows:

$$R = dN/dt \doteq \Delta N/\Delta t \text{ and,}$$

$$d^2N/dt^2 = dR/dt \doteq \Delta R/\Delta t,$$

where $\Delta t \geq 0.1$ millisecond.

It has been found empirically[13] that the speech formant information may be sufficiently preserved if a sampling rate of 35 cps is utilized. This sampling rate should also be more than adequate for the voicing information, since the voicing excitation varies at syllabic rates, i.e. less than 15 cps. Voiced sounds are restricted to a large extent by the 4,000-cps high frequency cutoff of the speech band-pass filter. Also, it has been determined that the intelligibility may be essentially preserved if a 7-millisecond sample of each 21-millisecond period is transmitted and repeated for the remaining 14 milliseconds.[14] This infers that Δt counting periods of less than 30 milliseconds should be reasonable. Although some of the early data was accumulated using a Δt period of 20 milliseconds, a period of 10 milliseconds was adopted as the most practical in that it appeared to be adequate for approxi-

mating the instantaneous rate (dN/dt) without overextending the already lengthy time period required to obtain complete data for individual words.

If the level at which the speech zero crossovers are defined is set too low, the thermal and background noise will give rise to additional zero crossings. When the signal is reconstructed, the amplitude from the noise crossovers will be equal to that of the speech and the relative power of the noise will be enhanced. In addition to reducing the intelligibility, the resulting harshness is distracting to the listener. Conversely, if the level is set too high, only the stronger signals give rise to zero crossings and a considerable reduction in intelligibility results. It would appear that the establishment of a level somewhat above the background noise would give the best results.

For the work reported herein, the crossover "threshold" level was defined at a point 3 db above the level at which sporadic crossover pulses (3-4 per second) were obtained for the quiet periods between words. Since the gain of the systems was set for each word through use of a common test tone, a single setting sufficed for all words. Once established, the level also served as a convenient reference for establishing relative peak signal excursions and crossover levels. Since the test words were power leveled, the peak excursions were not greatly different and were in general 40 to 50 db above the threshold level.

For a S/N ratio of 0 db, the articulation score curve of Figure 11 shows that the intelligibility of reconstructed speech is not overly sensitive to various crossover levels near the threshold level. Although specific articulation score data were not obtained for the case of speech in the clear, it

was the consensus of the investigators that no noticeable change in the intelligibility or naturalness was apparent for crossover levels 10 to 15 db above the threshold level. There were indications, however, that a level somewhat below the background noise may possibly give rise to speech of higher intelligibility and naturalness. Certainly under such conditions, the abruptness of the reconstructed speech is not so apparent.

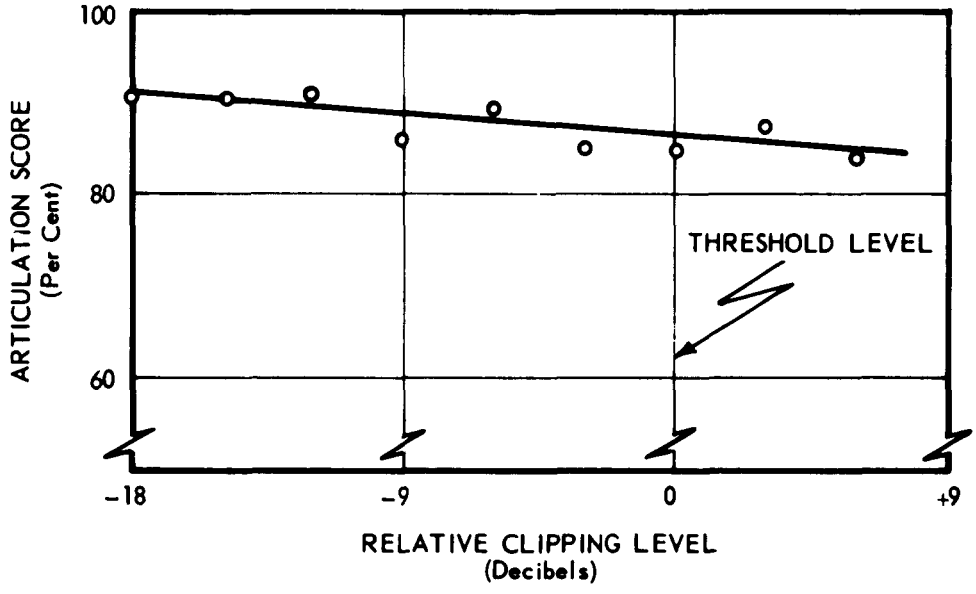


Figure 11. Effect of Clipping Level on AS for Reconstructed Speech at a Zero db S/N Ratio.

Further evidence of this effect is apparent in the AS peak that results from reconstructed speech containing moderate amounts of noise. As shown by Figure 12, the highest AS was obtained at S/N ratios in the vicinity of

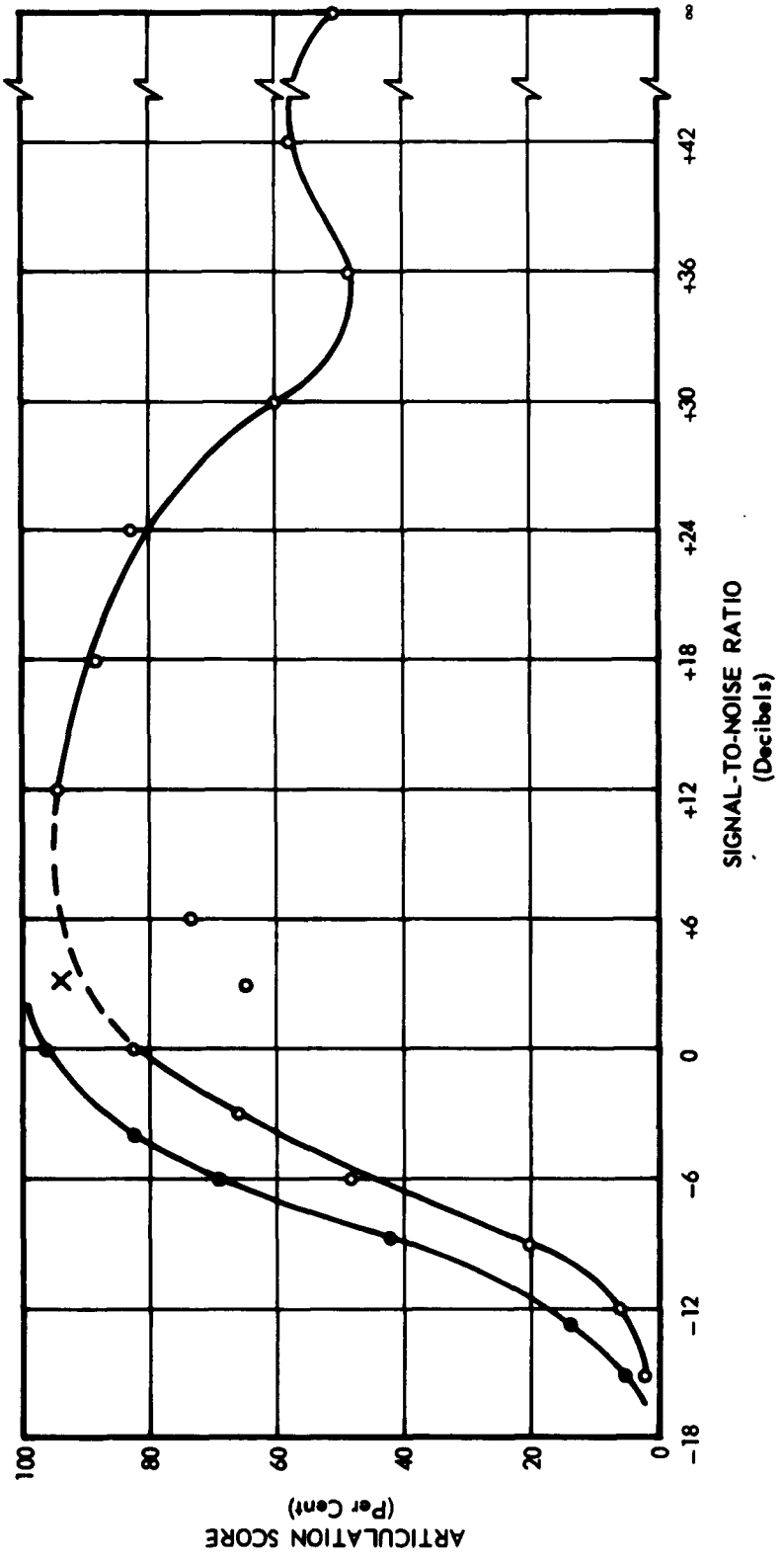


Figure 12. AS for Reconstructed Speech.

+12.* As the S/N ratio was increased, the intelligibility decreased until a 50 to 60 per cent plateau was reached at a S/N ratio of approximately +30 db. Probably this effect is closely related to the known technique of mixing an audio tone with speech (prior to clipping) to improve the quality of clipped speech.[6]

Figure 12 also compares the reconstructed speech articulation scores with the scores for the original speech. For equal intelligibility, the curves indicate that the S/N ratio required for the reconstructed speech is consistently 3 db higher than that of the original speech. It is known that the elimination of redundancy, in the transmission of intelligence through an interfering environment, has the same effect as decreasing the S/N ratio, thereby decreasing the intelligibility. Since the redundancy existing in normal speech sounds is decreased when zero crossing information alone is considered, an increase in the required S/N ratio would be expected for equal intelligibility. Also, some decrease in scores undoubtedly results from the existence of additional frequency components in the rectangular speech representation.

On the other hand, if the threshold level is established below the background noise level, the reconstructed speech scores may be higher as suggested above and the AS curve might exhibit an even closer relationship to the original speech curve.

It should be noted that a basic difference exists between speech reconstructed from zero crossings at the established threshold level and clipped

* The points at +3 db and +6 db of Figure 12 were questionable because of equipment difficulties during preparation of the listener tapes, and this portion of the curve was extrapolated as shown by the dotted line. Data collected from one repetition and six listeners of a reconstructed speech tape made under slightly different conditions produced an AS of 94 per cent at a +3-db S/N ratio. This point is shown by the "cross" in the figure.

speech as it is normally defined. Clipped speech retains considerable background noise since only those waveform excursions which exceed the clipping level are removed. Reconstructed or infinitely clipped speech retains only those signals which cross the threshold (triggering) level of the Schmidt or multivibrator circuit; thus signal variations above and below the threshold are eliminated from consideration. Licklider and Pollack[6] obtained articulation scores of almost 100 per cent for clipped speech in the clear, as compared to the 51 per cent score of Figure 12 for reconstructed speech in the clear ($S/N = +\infty$). The need for an additional investigation of the effects of the background noise on the intelligibility and naturalness of clipped and reconstructed speech is evident. The investigation should also consider the effects of added tones prior to clipping. It is known that under certain conditions the tone acts as a carrier whose zero crossings are position modulated by the speech.[15]

The rate of zero crossings as a function of time is shown in Figure 13 for the word "rub" at three S/N ratios. The curves were obtained by determining the total number of zero crossings that occurred during each successive 10-millisecond interval. The successive rates (number of crossings per 10 milliseconds) were plotted as a function of time

As seen in Figure 13, the addition of sufficient noise for a 20-db S/N ratio only slightly disturbed the pattern of rate of zero crossings for this word; however, when equal signal and noise power were mixed, the resulting pattern of rate of zero crossings for the mixed signal only vaguely resembles the pattern for the speech in the clear. While a graphic representation of the type shown in Figure 13 gives the investigator a qualitative view of the disturbance of the pure speech rate of zero crossing pattern that takes place when noise is added, it does not readily allow him to determine how much or to what extent the pattern is disturbed.

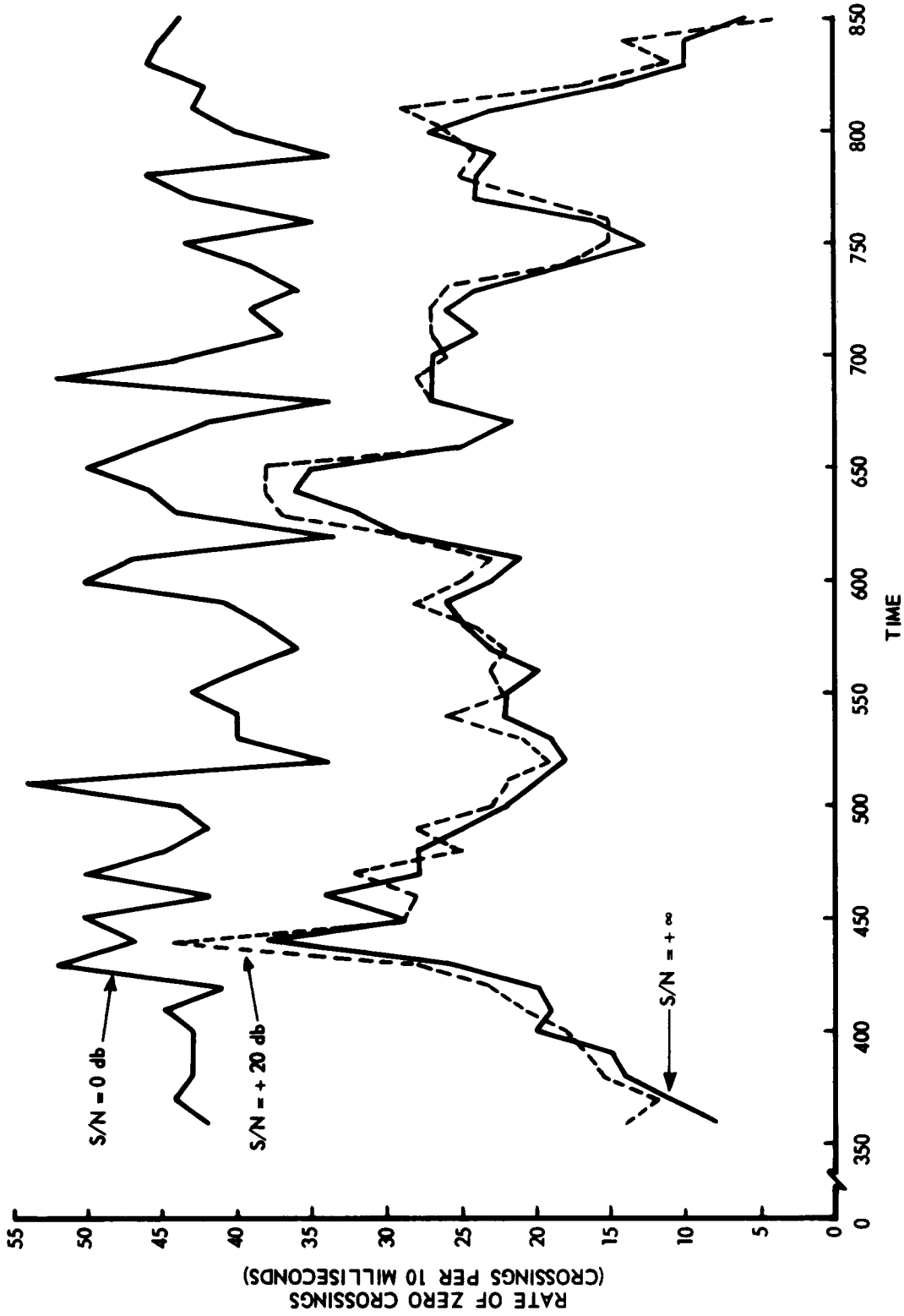


Figure 13. Rate of Zero Crossings as a Function of Time for "Rub" at Three S/N Ratios.

An alternate method of viewing the disturbance of the rate of zero crossings pattern is shown in Figures 14 and 15. Figure 14 is the scatter diagram of the rate of zero crossings for "rub" at a S/N ratio of +20 db. This diagram was prepared from the graph of Figure 13. For each 10-millisecond interval, the rate of crossings was plotted for the word in the clear versus the rate for the word at a S/N ratio of +20 db. Figure 15 was prepared in a similar manner, using the rates for the word at a S/N ratio of 0 db in the place of those for the word at a +20-db ratio. In the scatter diagrams, the rates of zero crossings for the word in the clear will be called the x variables and the rates for the word to which noise has been added will be called the y variables.

For the scatter diagram in which "rub" at +20-db S/N ratio was used (Figure 14), there is a tendency for small values of x to be associated with small values of y. In fact, the general trend of the scatter is that of a straight line. Had a scatter diagram been made of "rub" in the clear versus itself (e.g. no added noise), the points would all lie on a straight line. For the scatter diagram in which "rub" at 0-db S/N ratio was used (Figure 15), there is no marked tendency for small values of x to be associated with small values of y, and vice versa. A straight line approximation to this data would be of necessity very crude. Thus, the extent to which the pattern of rates of zero crossings of the word in the clear is disturbed by the addition of noise is given by the degree of nonlinearity introduced into the scatter diagram of the rates of crossings. For any given S/N ratio, it would be desirable to measure the degree to which the x and corresponding y variables are linearly related.

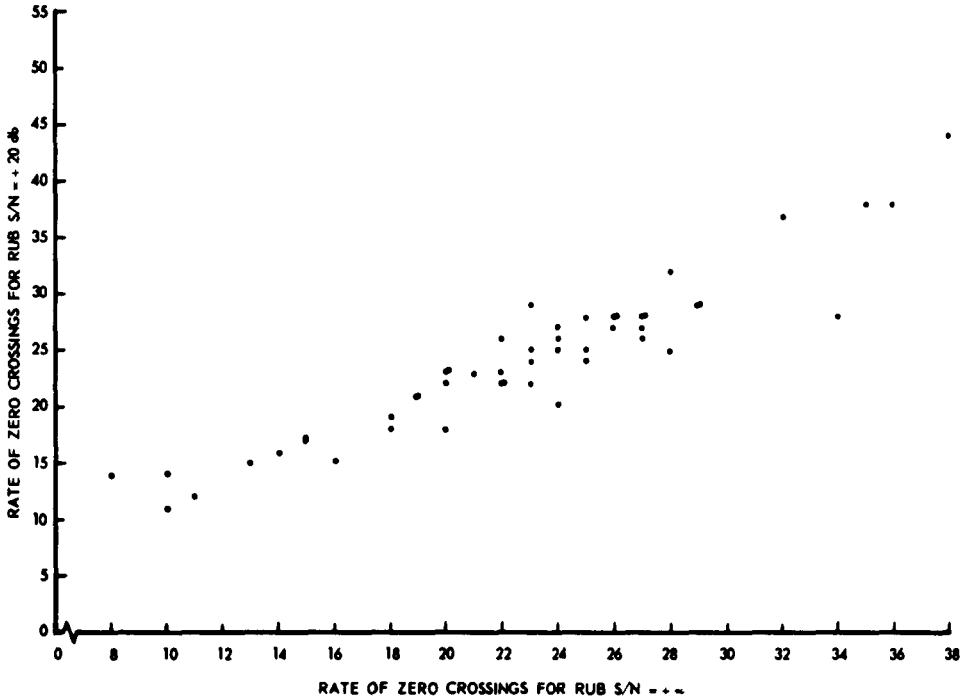


Figure 14. Scatter Diagram for "Rub" in the Clear vs. "Rub" at S/N = +20 db.

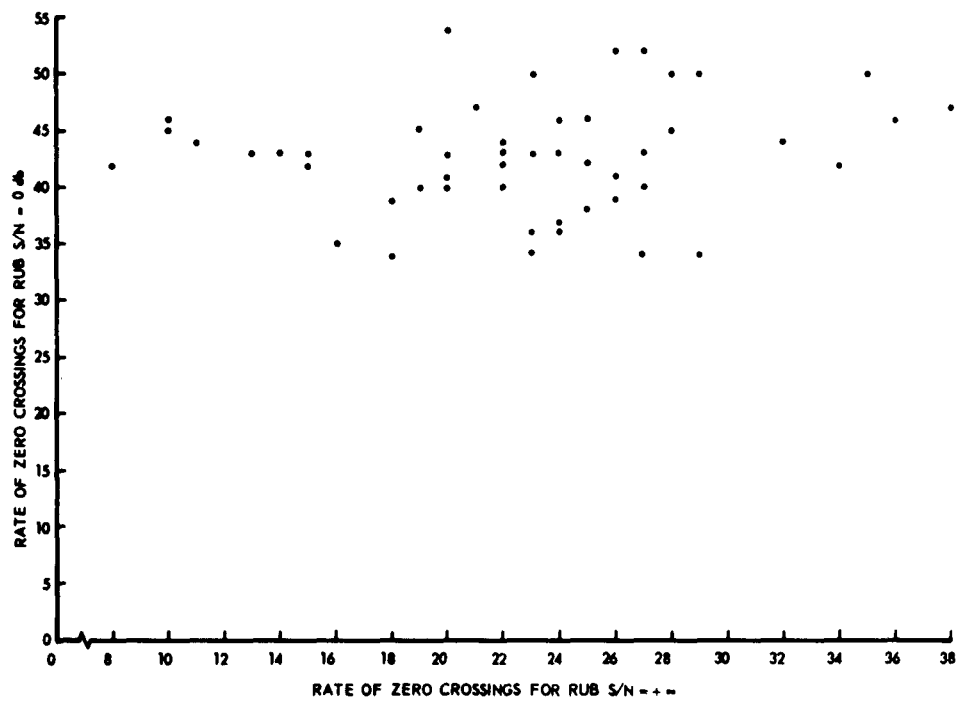


Figure 15. Scatter Diagram for "Rub" in the Clear vs. "Rub" at S/N = 0 db.

Such a measure should be independent of the choice of origin of the variables. For example, a scatter diagram whose points approximate a straight line to the same extent as those in Figure 14, but whose y values are in general larger (thereby giving the same scatter higher up on the diagram) should give the same measure of linearity as do the points in Figure 14. This property can be realized by using the variables x_i and y_i in the form $(x_i - \bar{x})$ and $(y_i - \bar{y})$, where $\bar{x}$ and $\bar{y}$ are the means of the x_i and y_i.

The measure should also be independent of the scale of measurement used for x and y. If, for example, 20-millisecond time intervals had been used in computing the rates of zero crossings, the rates would have been in terms of number of crossings per 20 milliseconds, and so would have been about twice as large. If the same degree of linearity had existed in the scatter diagrams when the 20-millisecond intervals were used, the desired measure of this linearity should so indicate. This independence of scale may be accomplished by dividing the variables x and y by their respective sample standard deviations s_x and s_y.[17]

The measure having both of the aforementioned properties may be constructed by using the variables x_i and y_i in the form

$$u_i = \frac{x_i - \bar{x}}{s_x} \quad \text{and} \quad v_i = \frac{y_i - \bar{y}}{s_y} .$$

The u_i and v_i are called the standard sample units of x_i and y_i, respectively. A typical scatter diagram of the points (u_i, v_i) is shown in Figure 16. This figure is for illustration only and does not represent a mapping of the points (u_i, v_i) of the scatter diagrams of Figures 14 and 15.

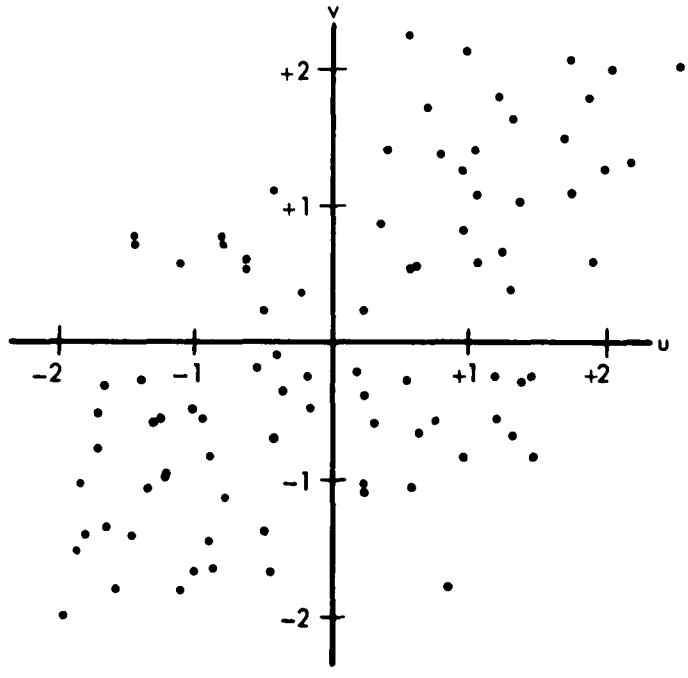


Figure 16. Scatter Diagram for Standardized Units.

If the points (x_i, y_i) on the scatter diagram tend to lie in a straight line with a positive slope as do those in Figure 14, the corresponding points (u_i, v_i) will tend to lie in the first and third quadrants. Furthermore, the points in the first and third quadrants will tend to have larger coordinates than those in the second and fourth quadrant. If the sum

$$q = \sum_{i=1}^{n} u_i v_i$$

is formed, the terms that are contributed by the points in the first and third quadrants will be positive, while those in the second and fourth quadrants will be negative. Thus, a large positive value of q indicates a strong

linear trend in the scatter diagram. If the points (x_i, y_i) on the scatter diagram do not exhibit linearity (such as those in Figure 15), the corresponding points (u_i, v_i) will lie more or less equally in all four quadrants. The positive terms in the sum q will tend to be offset by the negative terms, and the total will be close to zero. Therefore, a small value of q indicates very little linear relationship between the variables x and y. A large negative value of the sum q also indicates a strong linear relationship; however, the line formed by the corresponding points (x_i, y_i) would have a negative slope.

Thus, it would seem that the degree to which the variables x and y were linearly related can be determined by the size of the sum q. However, it can readily be seen that if the number of points (u_i, v_i) making up the sum were doubled (and the degree of linearity remained the same), the sum would also approximately double. It is, therefore, necessary to divide this sum by n, the number of points in the scatter diagram.

The sum

$$\frac{1}{n}\sum_{i=1}^{n} u_i v_i$$

is then the desired measure of linearity between the variables x and y. This sum is called the correlation coefficient and is defined in terms of the original measurements by the following formula:

$$\text{Correlation Coefficient} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n\, s_x s_y} .$$

The correlation coefficients were computed for the scatter diagrams of Figures 14 and 15. The coefficient for "rub" in the clear correlated against "rub" at a +20-db S/N ratio was found to be 0.938 The coefficient for "rub" in the clear correlated against "rub" at a 0-db S/N ratio was found to be 0.150.

It is unnecessary to draw scatter diagrams in order to compute correlation coefficients, as the formula above can be applied directly to the data. An IBM 650 Computer was programmed to process the data and obtain the correlation coefficients. Details of this program are presented in the Appendix.

Zero crossover data for 10-millisecond time intervals were collected for a number of PB-50-1 words at various signal-to-noise levels for the purpose of evaluating the bivariate correlation coefficient technique for measuring the extent to which the zero crossings are disturbed by noise, and establishing the relationship between the intelligibility of speech and the zero crossing rate in the presence of various quantities of noise. The zero crossings for the word in the clear were compared to those for the word at each S/N ratio of interest Perfect correlation is indicated by a coefficient of one (which would be obtained if the word in the clear were correlated against itself); no correlation is indicated by a coefficient of zero (which might be obtained if the word in the clear were correlated against pure noise). Thus, the extent to which the zero crossings of the word in the clear have been disturbed is given for each S/N ratio as a coefficient whose value lies between one and zero.

It is only natural to obtain data using the word in the clear as the dependent variable, since the noise is normally considered as corrupting the word. However, valid results should also result if the word is considered as

corrupting the noise (for a given noise sample). Further, some additional insight may be gained by using the noise as the dependent variable since the articulation scores are obtained primarily for negative signal-to-noise ratios. The intelligibility of the word would then become proportional to (1 - r), where r is the correlation coefficient for the noise case. Figure 17 presents the correlation coefficient curve for several words using noise as the dependent variable and Figure 18 shows similar curves with the word in the clear as the dependent variable.

It is not necessary to use the word in the clear as the dependent variable in computing the extent to which the zero crossings have been disturbed by noise. For example, the zero crossings for that S/N at which the articulation team barely scores 100 per cent might be used as the dependent variable, and all lower S/N ratios correlated against it. Other possibilities for the dependent variable also exist. For reconstructed PB-50-1 words in white noise, the articulation team scores 94 per cent at a -3-db S/N ratio. This ratio, which was the point, to the nearest 3 db, at which the articulation team barely scored 100 per cent for the reconstructed speech, was selected for the dependent variable and the bivariate correlation coefficients for a number of words at various S/N ratios were computed.

Figures 19 and 20 present the individual correlation coefficient curves for these words and Figure 21 compares the composite curve of all words tested with the articulation score curve for reconstructed speech.

As shown by Figure 21, a close monotonic relationship exists between the correlation coefficient curve and the speech articulation score curves.

Several factors should be considered in interpreting the correlation coefficient results. One consideration results from the establishment of the

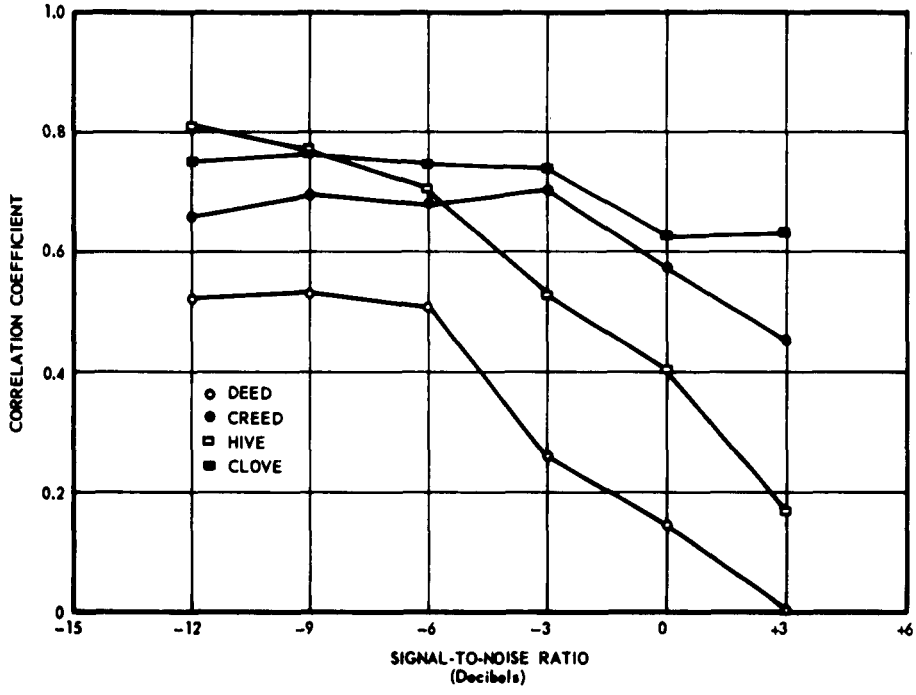


Figure 17. Correlation Coefficients for a Dependent Variable of - ∞ S/N Ratio.

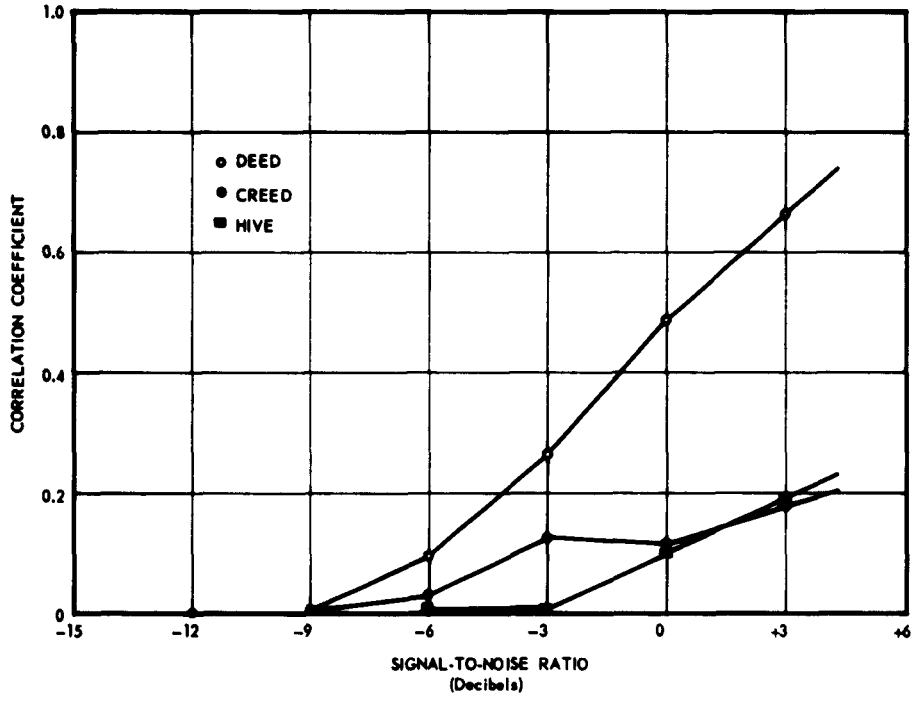


Figure 18. Correlation Coefficients for a Dependent Variable of + ∞ S/N Ratio.

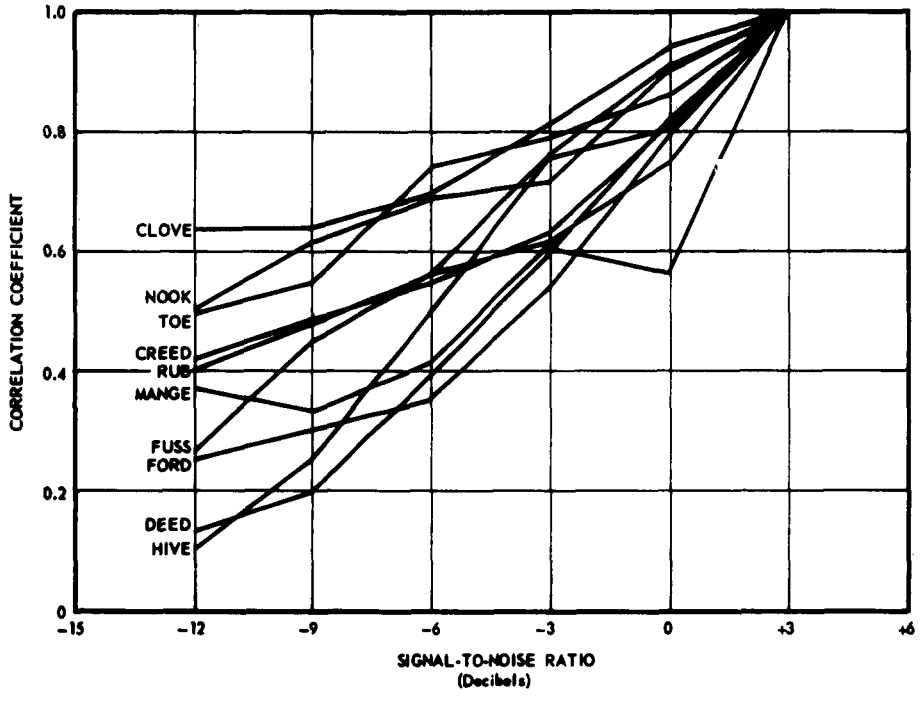


Figure 19. Correlation Coefficients for Individual PB-50-1 Words (Group I).

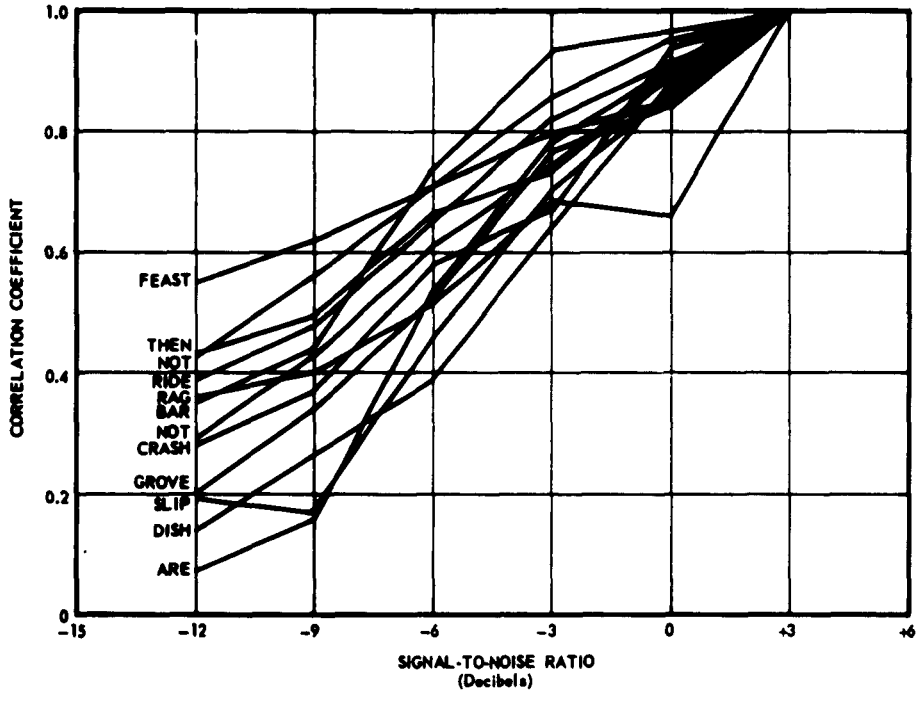


Figure 20. Correlation Coefficients for Individual PB-50-1 Words (Group II).

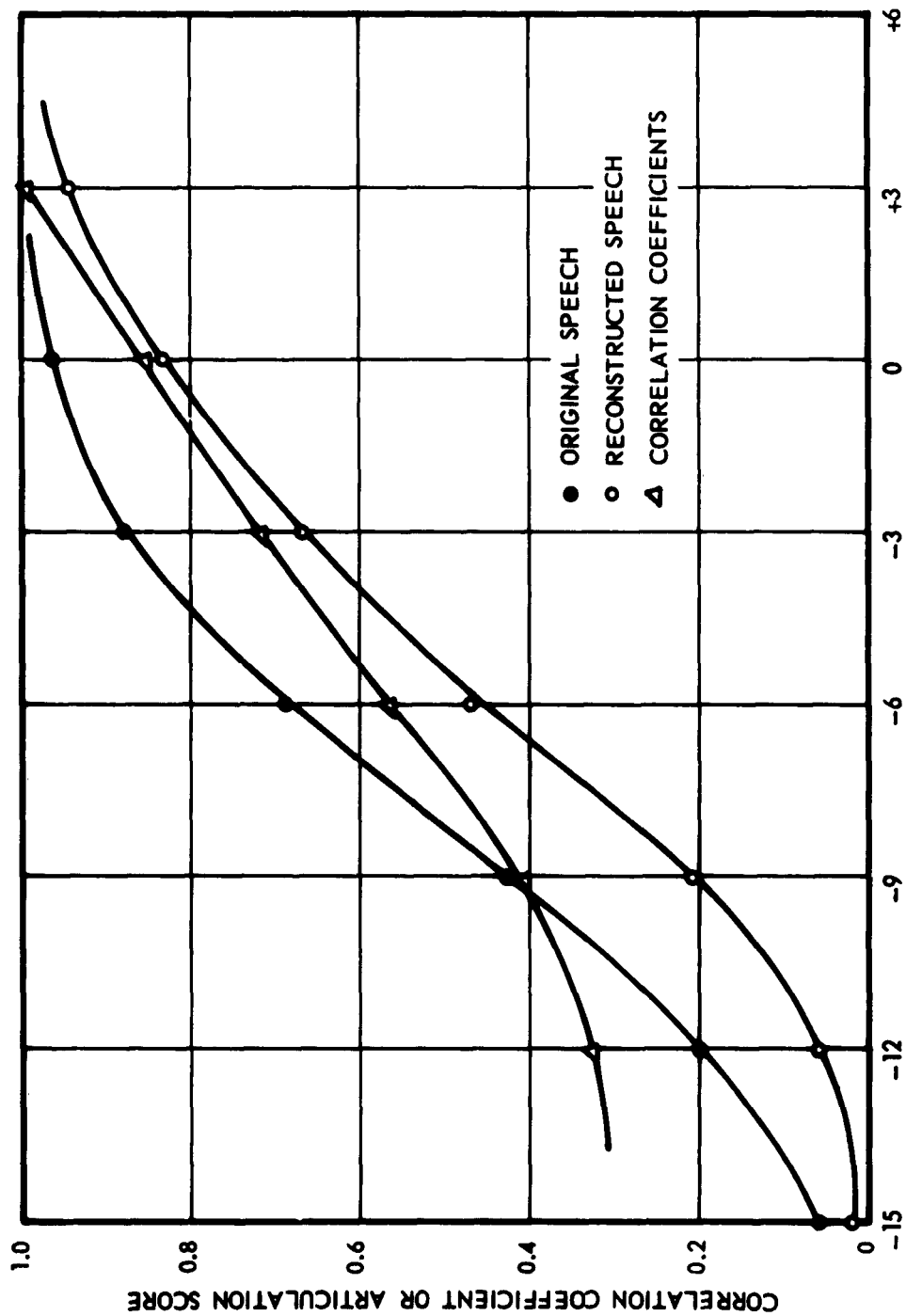


Figure 21. Comparison of AS and Correlation Coefficients for a Number of PB-50-1 Words.

+3-db S/N ratio as the point at which the articulation team just scores 100 per cent for reconstructed speech. Initially this appeared to be a reasonable assumption, but it was later concluded that a 100 per cent articulation score "to the nearest 3 db" was too great. A better method would be to find the individual S/N ratios for each word to the nearest 1 db at which the team scored 100 per cent for that word. Thus, each word would be correlated at various values of S/N against itself at the lowest S/N ratio at which it was completely intelligible. Unfortunately, the time required to obtain the information necessary for using this better method was not available. The lowest S/N ratio for 100 per cent intelligibility for most of the 50 words is greater than +3 db, but less than +6 db (and, therefore, within the 3-db tolerance). A few of the words have 100 per cent intelligibility ratios below +3 db, and it is suspected that some words have theirs above +6 db. In general, +3 db is too low for the majority of the words; thus, the composite correlation coefficients curve for all words is biased high, as shown in Figure 21. It should be noted that by using the individual ratios method outlined, a nonlinear S/N ratio transformation would occur, and the curve in Figure 21 would not only be shifted to the right, but its characteristic shape and slope would also be changed.

A second consideration involves the selection of a particular threshold level for defining time positions of the zero crossings. As discussed earlier, the available data implies that an optimum threshold level may exist and that the data reported herein may deviate from that obtained at the optimum level.

Still another consideration is the use of a fixed noise sample for each word. The correlation coefficients for each word apply only to that particular noise sample that had been recorded on the second track of the dual channel tape

loop. The extent to which other noise samples would produce equivalent correlation coefficient curves remains undetermined as does the effect of different talkers.

At the outset of the investigation, the instrumentation philosophy dictated a versatile capability for providing detailed data on a number of cross-over parameters for a wide range of conditions. Although the system served this purpose well, it was not well suited for the purpose of collecting the large quantities of specific data needed for a comprehensive statistical analysis. The disadvantage resulted from the time required to collect the desired data. The system required the 2-3 second tape loop for each word to be cycled at least once for each 10-millisecond time segment at each S/N ratio. The handicap in obtaining data for a PB-50 word group over a range of S/N ratios for several interferences or various threshold levels is obvious.

Consideration was given to several techniques for automatically recording the data or directly transmitting it to a computer, but economic factors voided this approach. Consequently, the data were recorded manually from the counter display and transferred to punched cards for input to the computer. As a result, the data collection phase was restricted to that required in the primary task of ascertaining the existence and nature of the relationship between the zero crossing rate and speech degradation. In spite of this limitation, data were obtained for about one-half of the PB-50-1 word group.

Although the limited number of words for which correlation coefficient data were obtained represents too small a sample of the PB-50-1 word group to permit definite conclusions to be drawn, the close monotonic relationships of Figure 21 are considered indicative of the capability of the technique for measuring the degradation of speech.

2.3. Articulation Testing

2.3.1. Listener Training. Prior to the initiation of work on Supplemental Agreement No. 1, several of the student listening team members were lost because of graduation. Four new members and one alternate were selected from some 35 applicants to bring team membership back to six. As in the past, the choice was made primarily on the basis of hearing acuity, availability, and grades.

The training of the new articulation team members on the PB-50-1 word list was in accordance with the procedure set forth in Technical Note No. 1 of this contract.[16] The tapes used to train the previous team were also used in training the new members. These tapes contain randomized, leveled-power words imbedded in band-limited white noise at S/N ratios of 0, -3, -6, -9, -12, and -15 db.

After 20 repetitions of the training material, an analysis of the data was made. The analysis considers the S/N ratio effect, N; the listener effect, L; the repetition effect, R; their interactions, NL, NR, and LR; and the experimental error. This analysis revealed that learning was still continuing, but at a slow rate. Another analysis was made after 24 repetitions of the training material; the results of this analysis are given in Table I.

The repetition effect (R) and all first order interactions involving it were found to be not significant at the 1% level; therefore, the decision was made that the team was fully trained to the adopted criterion and training was stopped.

The previous team required approximately 20 hours of training as compared to approximately 24 hours for the new members. However, the previous team was given a more intensive pretraining familiarization with the speech material and speaker s voice than were the new members. The pretraining listener

TABLE I.

RESULTS OF THE ANALYSIS OF ARTICULATION SCORES
COLLECTED FROM REPETITIONS 19 THROUGH 24

Source	Degrees of Freedom	Sum of Squares	Mean Square	Variance Ratio	Significant
N	4	14.8924	3.72310	254.83	Yes
L	4	0.3464	0.08660	39.54	Yes
R	5	0.0160	0.00320	1.46	No
NL	16	0.2338	0.01461	6.67	Yes
NR	20	0.0412	0.00206	0.94	No
LR	20	0.0546	0.00273	1.25	No
Experimental Error	80	0.1752	0.00219		
TOTAL	149	15.7596			

familiarization was given to the previous team while the actual training tapes were being prepared. These tapes were in existence when the new team members were chosen and only a few hours were spent in familiarization prior to the start of training. Thus, the actual over-all training times seem to be in very close agreement.

An analysis was made to determine the extent of the memorization of the ordering of the speech material used in training the new articulation team members. Four randomizations of the speech material used in training the new members and four new randomizations with which they were completely unfamiliar were presented to them six times each.

Analysis of the resultant scores showed that the mean squares of the N, L, R, and NL sources of variance were significant at the one per cent level. The significance of the L source of variance is to be expected; the significance of the N and NL sources of variance indicated that memorization of the ordering of the material had taken place. Duncan's multiple range test was used to determine the extent of memorization.[18] The results of the multiple range test on the mean articulation scores of the signal-to-noise ratio effect (N) are given in Table III. In this table, any two articulation scores not underscored by the same line are significantly different.

TABLE III.

RESULTS OF THE MULTIPLE RANGE TEST ON THE MEAN ARTICULATION SCORES OF THE S/N RATIO EFFECT

	Nonrepeated Randomizations				Repeated Randomizations			
Randomization N .	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Articulation Score (%)	53.5	54.0	55.2	59.7	60.3	63.3	66.6	67.2

Table II indicates that the orderings of the randomizations used in training the team were memorized; a similar test of the NL interaction effect showed that one or more of the randomizations used in training had been memorized by each of the listeners. Thus, these randomizations were not used in subsequent tests with this team.

A multiple range test was used to determine the cause of significance of the R source of variance. The results of this test are given in Table IV.

TABLE III.

RESULTS OF THE MULTIPLE RANGE TEST ON THE MEAN ARTICULATION SCORE OF THE REPETITION EFFECT

Repetition No.	1	2	6	3	5	4
Mean Articulation Score (%)	56.6	57.6	60.4	61.4	61.5	62.4

Table III indicates that the significance of the R source of variance was not caused by learning. It can only be concluded that the significance of the R source of variance was caused by some unassignable effect not included in the mathematical model, such as external testing room noise.

Upon completion of training the articulation team, the articulation scores of the new members were compared to those of the remaining old members and, subsequently, to the previous team. From this, an attempt was made to formulate a procedure whereby a continuing measure of confidence could be assigned to a team in which several members are occasionally replaced.

Comparisons of the scores were made for a number of interference conditions including white noise, positive tilted noise, negative tilted noise, chopped noise, diphasic speech, and chopped speech. An examination of the data indicated that the old and new teams were representative samples of the same population. The articulation curves from the two teams have very nearly the same magnitudes and demonstrate identical characteristic shapes. The maximum difference in observed articulation scores for any of the types of interference conditions tested was less than ten per cent. The observed homogeneity of the two teams is not surprising, since identical criteria were used in choosing the 1960 and 1961 Georgia Tech teams.

Under such conditions, it is difficult indeed to establish a valid method for relating replacement members to an existing team. The parameters necessary for the prediction of the articulation scoring capability of a new member (s) could not in general be determined without the benefit of data from one or more distinctly different populations.

2.3.2 "CVC" Tests. Special test tapes of "CVC" nonsense syllables were furnished by RADC during the contract period. Data from these speech units were desired because of their greater sensitivity to formant peak ratios and their ability to provide controlled information regarding distinctive features of the various phonemes and phoneme classes.

These "CVC" words were re-recorded with band-limited white noise on parallel tracks of magnetic tape. A 1000-cps test tone was recorded on each track preceding the noise and words. The tone and noise levels were set to give a 0-db S/N ratio when the tone levels of each track were set equal on playback. Various S/N ratios are obtained by inserting the required attenuation in either the word or noise channel prior to mixing.

No attempt was made to power level the individual words as was done in earlier tests using PB-50-1 words. The establishment of a precise S/N ratio for each word was not felt to be economically justifiable in light of the immediate purpose of the tests. However, the relative S/N ratios for the entire group of "CVC" tests are accurate since calibrated attenuators are used to establish the various S/N ratios. Further, if more exact S/N ratios for each word or each test group are required, they may be obtained from the recording at a later date.

The original purpose of administering the "CVC" tests to the articulation team was to obtain data on a controlled group of phonemes. If the same "CVC"

tests were presented to the team, using crossover reconstructed "CVC" syllables, an analysis could be made which would indicate which consonant groups were made less intelligible by the removal of the amplitude information. In such an analysis, the data presented herein would be used as a control. Correlation coefficients for the reconstructed "CVC" lists could be computed and comparisons to the control data made as was done for the PB-50 words. The information so obtained would have an advantage by providing controlled data on specific phonemes and phonetic classes. Unfortunately, there was not sufficient time available to carry this analysis to a conclusion.

However, articulation scores were obtained for the original "CVC" consonants as a function of S/N ratio. These data can be used to determine the effects of the addition of varying amounts of noise on the retained intelligibility of certain consonant classes

The results of "CVC" tests are presented in Figures 22 and 23. Figure 22 shows the articulation scores for the stop consonants (both voiced and voiceless) as a function of S/N ratio (solid curve) and the scores for the fricative consonants (both voiced and voiceless) as a function of S/N ratio (dashed curve). Although the data are far from conclusive, it appears that the stop consonants were degraded to a slightly lesser extent by the noise than were the fricative consonants.

Figure 23 shows the articulation scores for the voiceless consonants (both stop and fricative) as a function of S/N ratio (solid curve) and the scores for the voiced consonants (both stop and fricative) as a function of S/N ratio (dashed curve). The difference between the two curves is not large enough to be strongly conclusive as to whether the voiced or voiceless sounds were degraded more by the noise.

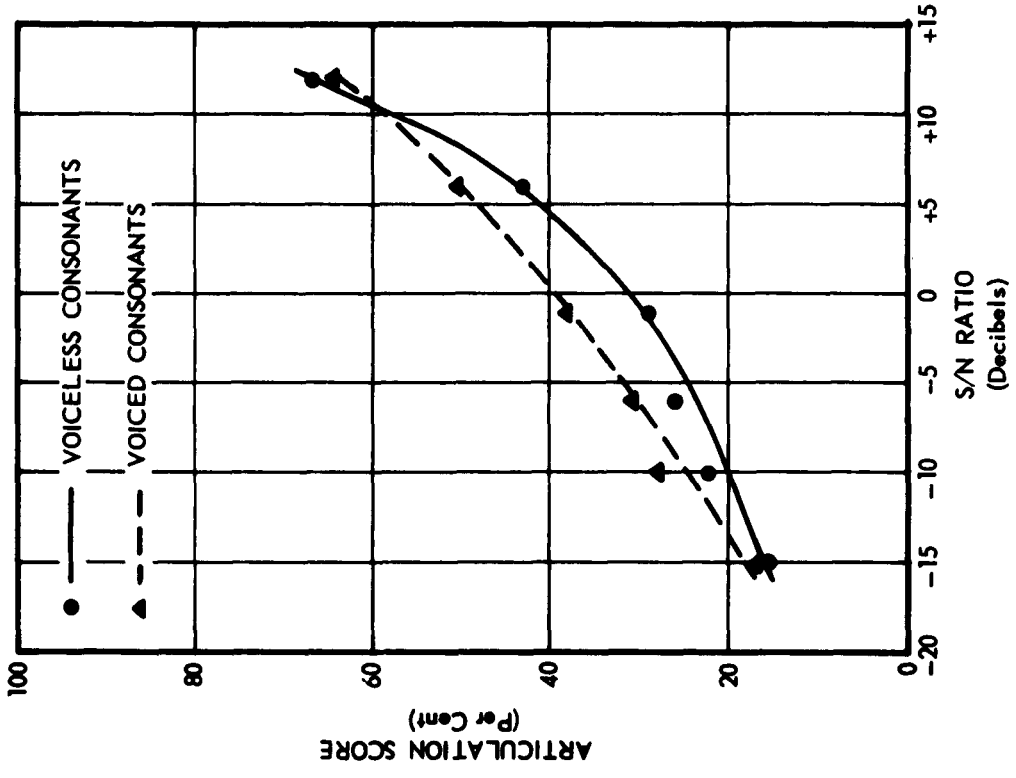


Figure 23. AS as a Function of S/N Ratio for Voiced and Voiceless Consonants.

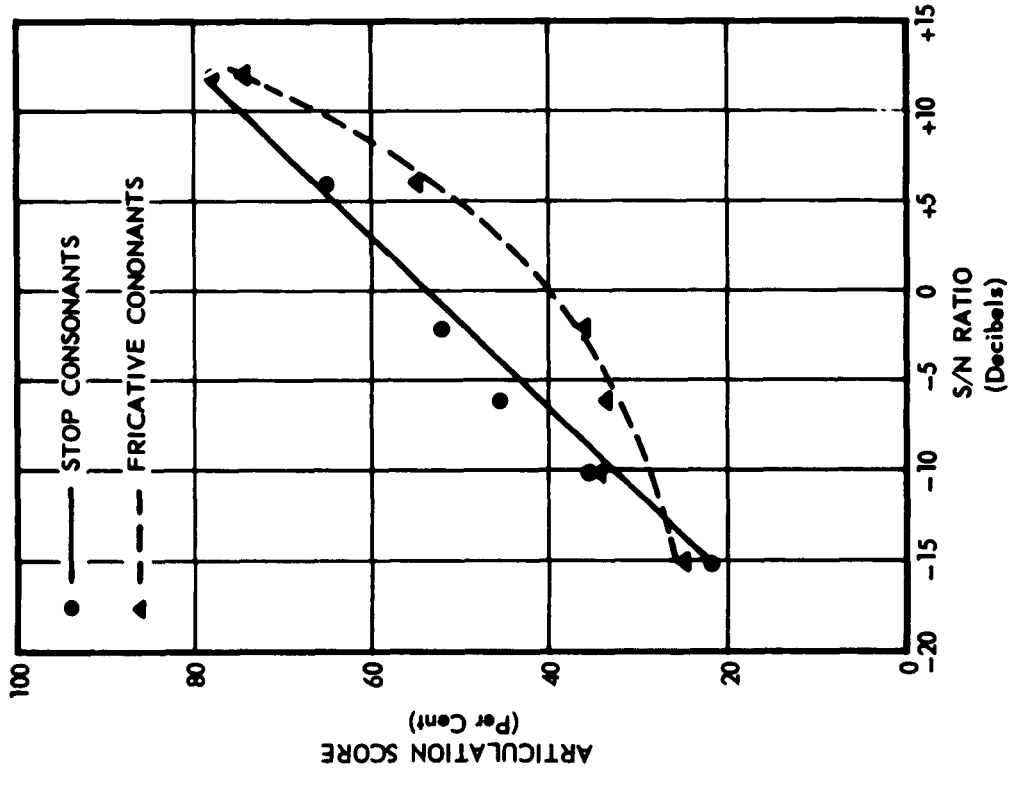


Figure 22. AS as a Function of S/N Ratio for Stop and Fricative Consonants.

While the foregoing data do not give strong indications as to which specific consonant groups are degraded more by the addition of noise, this should in no way detract from the original purpose of administering the "CVC" tests to the articulation team.

Figure 24 shows the articulation scores for a consonant group made up of nasals, voiced fricatives, liquids, and glides, plotted as a function of S/N ratio. A comparison of this figure with Figures 22 and 23 indicates that these sounds were not degraded by the noise as much as were the voiced and voiceless stop and fricative consonants

2.4. Miscellaneous Tests

2.4.1. Vocabulary Size Relationship. At the culmination of earlier work involving the validation of the GEL Speech Systems Test Set, it was concluded that the transform curves necessary to relate the GEL results to the Georgia Tech articulation team were, "... essentially the same transformations that independent investigators have established for relating the various test vocabularies used in articulation testing."[2] At the time, only 50 word articulation score data were available from the Georgia Tech team. Consequently, the transform curves were compared with 50- to 1,000-word relationships obtained by other investigators. Recently the Georgia Tech team was subjected to a 1,000-word (PB-50-1 through PB-50-20) test.* Figure 25 compares the results of this test with those obtained for 50 words. Figures 26 and 27 present

* The results consist in part of data collected in connection with the Doctoral research program of G. B. Hawthorne, Jr. of the Georgia Tech School of Electrical Engineering.

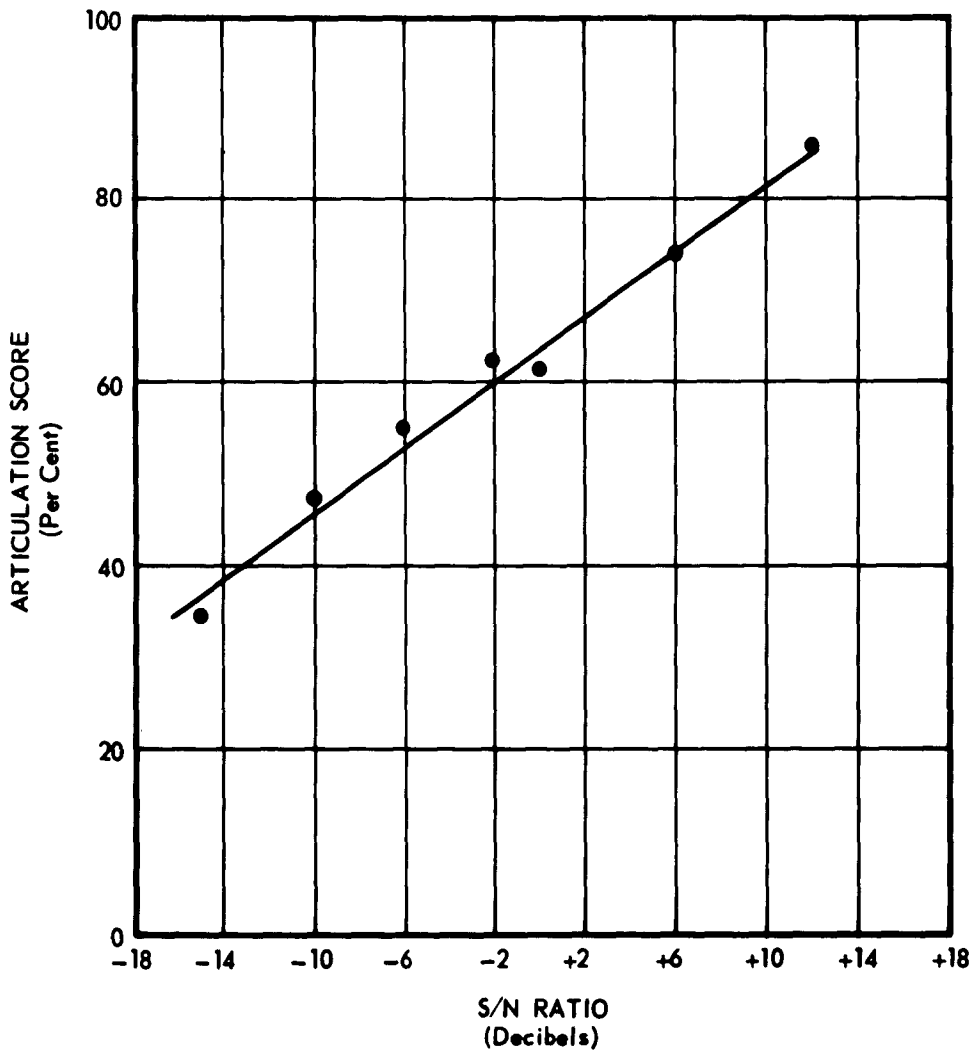


Figure 24. AS as a Function of S/N Ratio for a Consonant Group Composed of Nasals, Voiced Fricatives, Liquids, and Glides.

the articulation score and signal-to-noise relationship curves derived from Figure 25. The original transform curves used in the GEL investigation are also shown for comparison. These curves further substantiate that the GEL transform curves are essentially the same curves that depict the relationship between 50- and 1,000-word articulation test data.

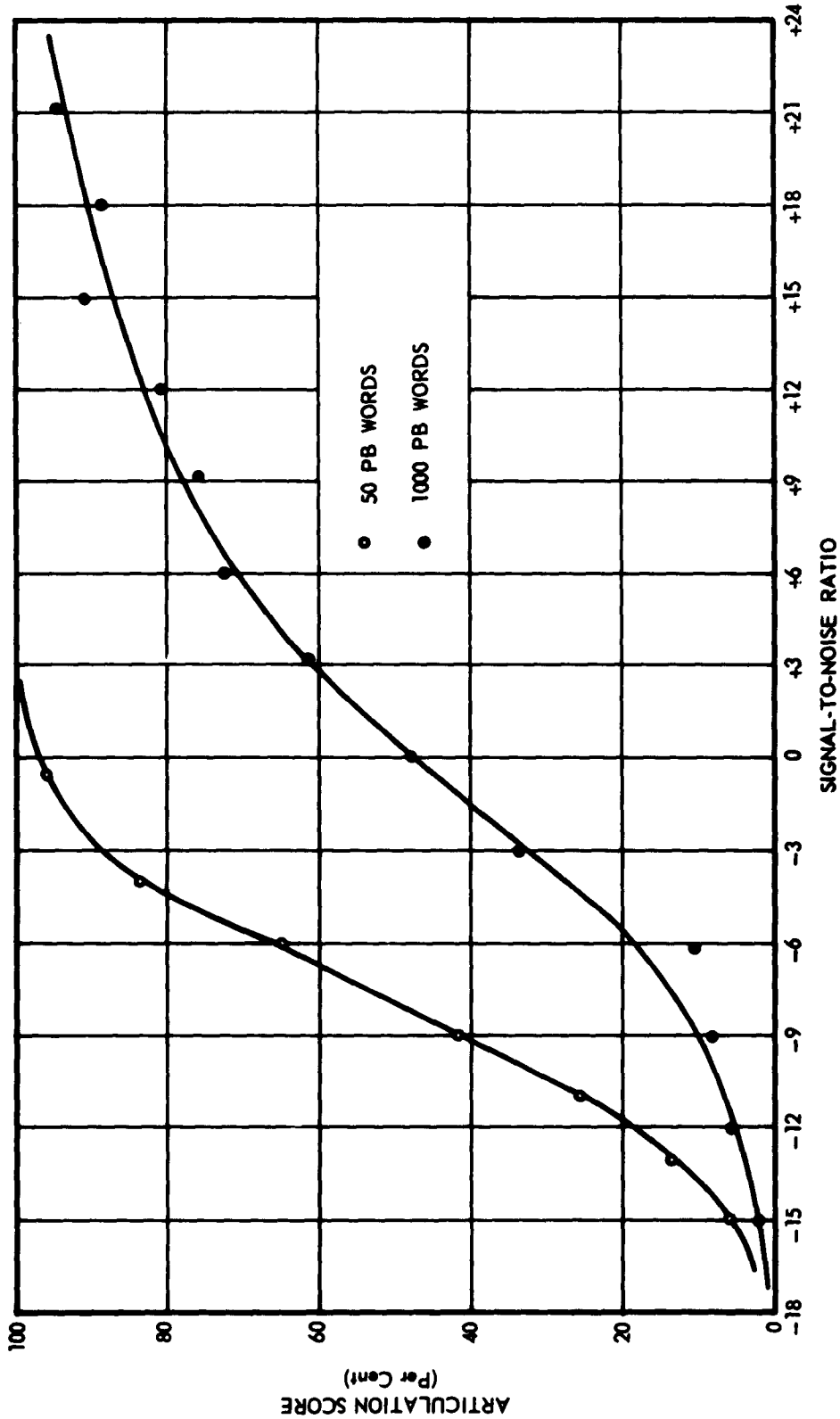


Figure 25. AS as a Function of S/N Ratio for 50 and 1000 PB Words.

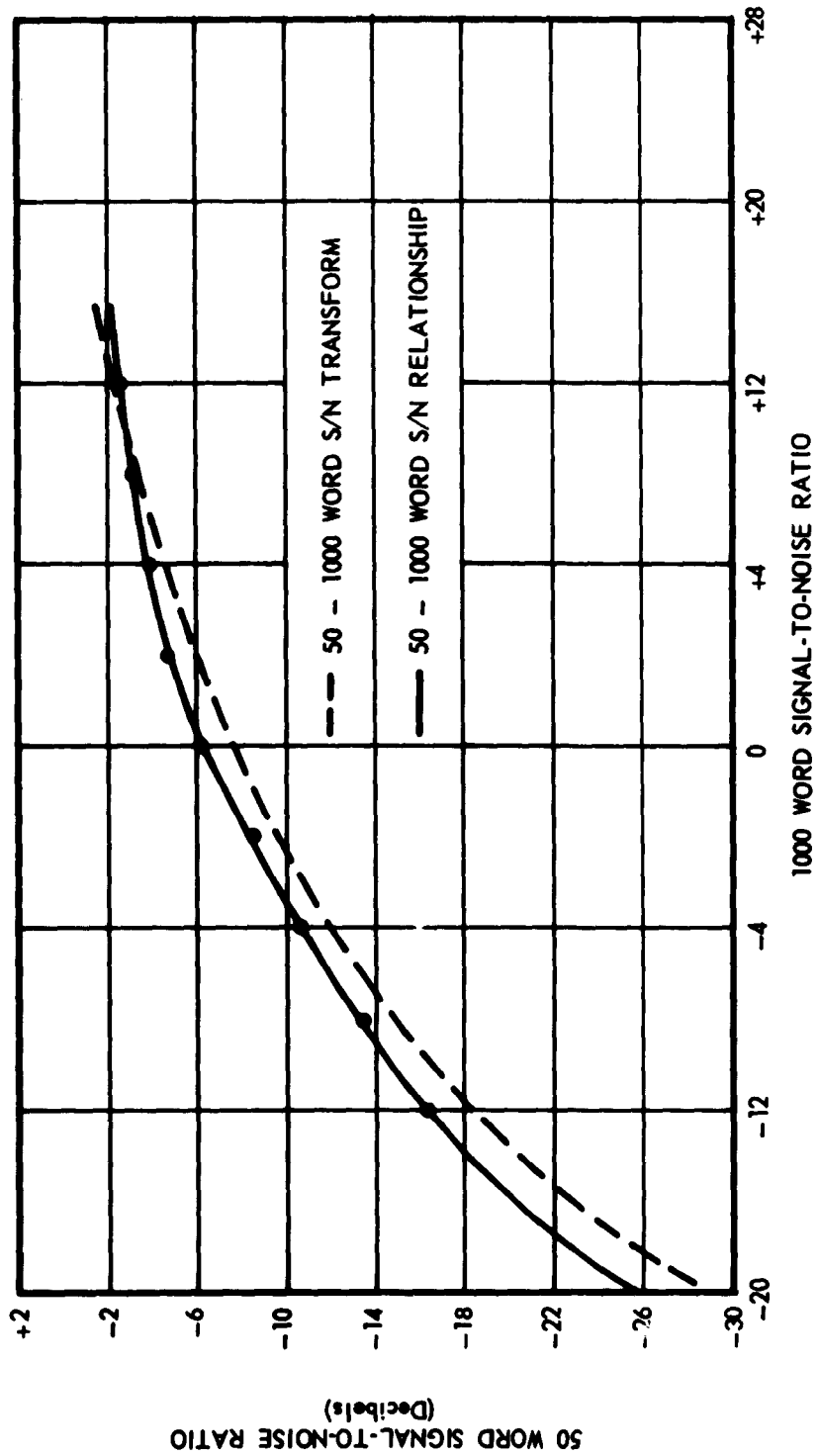


Figure 26. Comparison of the Georgia Tech Articulation Team 50 to 1000 Word S/N Relationship with the GEL S/N Transform Curve.

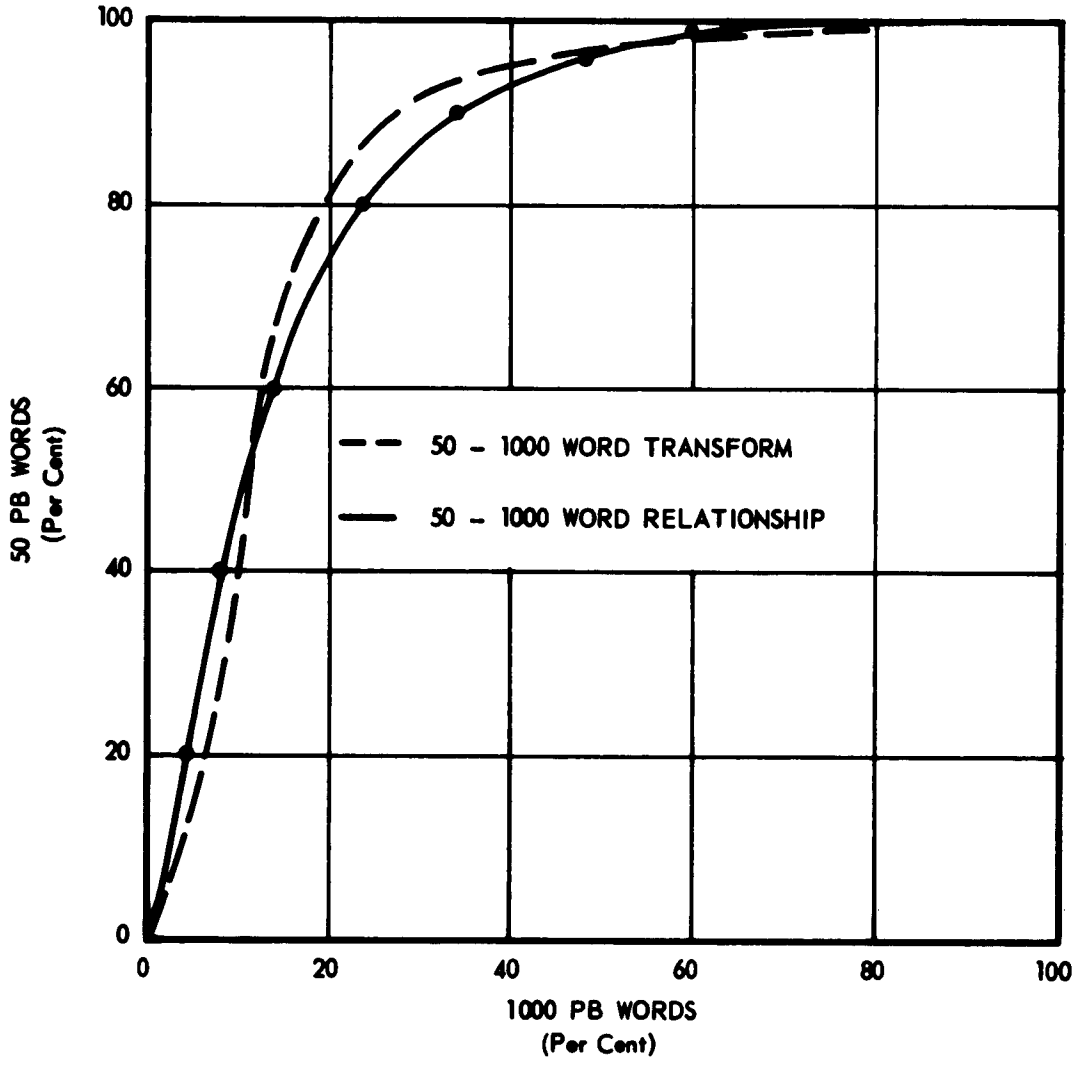


Figure 27. Comparison of the Georgia Tech Articulation Scores with the GEL Transform Curve.

2.4.2 "On-Line" Mixing. Data were collected from the team for white noise interference, using a new "on-line" mixing facility; this was done primarily to determine if the new system was compatible with the old method of using prerecorded tapes. Figure 28 shows AS as a function of S/N for both the prerecorded tapes and "on-line" mixing.

As seen from this figure, the method of "on-line" mixing yields slightly higher articulation scores for a given S/N. This probably results from the addition of an 8-kc low-pass filter, whereas the prerecorded tapes allowed noise up to 12 kc to reach the earphones. It would appear that an S/N ratio correction of about 1 db would compensate for the additional filtering. The "on-line" mixing would then yield the same results as the prerecorded tapes, while greatly reducing the number of required input tapes.

2.4.3 Threshold Effects. In the past, the total sound level of the word plus noise going into the listeners' earphones has been held constant from run to run as a function of the S/N ratio being used. An alternative method of holding the level of the noise between words constant from run to run was used to determine if the noise level between words and, consequently, the threshold just prior to the word influenced the articulation scores. Using this method the level of the signal plus noise varied from run to run as a function of S/N ratio. The results of this investigation are also given in Figure 28.

As seen from this figure, essentially the same results are obtained by holding the noise level constant from run to run, as by holding the signal plus noise level constant.

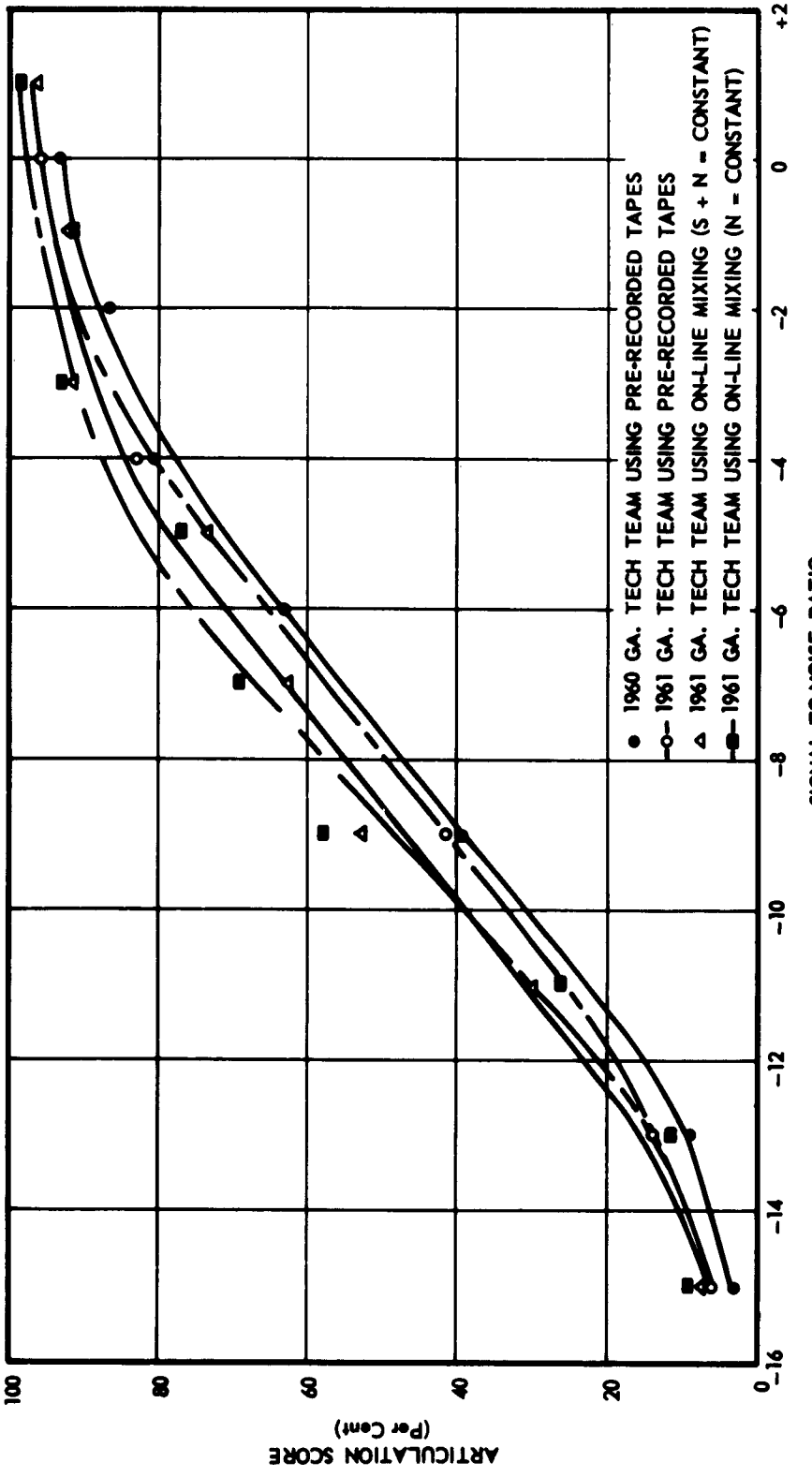


Figure 28. AS as a Function of S/N Ratio for Several Presentation Types.

3. CONCLUSIONS AND RECOMMENDATIONS

The retention of the intelligibility and the tractable nature of the pulse representation of the speech zero crossing rate provide a convenient approach to the development of electronic techniques for automatically determining speech degradation. Moreover, the close monotonic relationship between the crossover rate, correlation coefficient curve and the articulation score curve for a limited number of PB-50-1 test words indicates that the technique may be capable of providing a measure of intelligibility equivalent to that obtained by an articulation team.

An analysis of the total number and the average rate of zero crossings substantiated the belief that these two parameters were not an analog of the intelligence carrying attributes of speech.

Articulation team scores for speech reconstructed from zero crossing information shows that, under certain conditions, increased intelligibility results when moderate amounts of noise are mixed with the speech prior to the extraction of the zero crossing information.

Instrumentation design criteria for implementing the crossover technique should await a more comprehensive investigation of the utility of the technique for a larger representative set of interference conditions. If such an investigation is considered, proper emphasis should be given to the time element involved in the collection and reduction of large quantities of data.

Two additional tasks for further work in this general area are recommended; first, a study to determine the underlying cause of the increased intelligibility noted for reconstructed speech when noise is added to the

original speech, and second, the establishment of an optimum "threshold level" for defining the relative time positions of the zero crossings. Intuitive observations suggest that these two tasks are closely related and probably should be the subject of a single investigation.

4. BIBLIOGRAPHY

(1) General Electronic Laboratories Incorporated, Cambridge, Massachusetts, "Anti-Jamming Testers," Final Technical Report, Contract No. DA 36-039 SC-72788 (July 1958).

(2) Robertson, D. W. and Stuckey, C. W., "Investigation and Evaluation of the GEL Speech System Test Set," Final Technical Report for Item I, Engineering Experiment Station, Georgia Institute of Technology, Contract No. AF 30(602)-2150 (February 1961).

(3) Jacobson, H., "Information and the Human Ear," J. Acoust. Soc. Amer., 23, 463-471 (July 1951).

(4) Potter, R. K., et al, "Visible Speech," D. Van Nostrand Co., Inc., New York, N. Y. (1957).

(5) Cooper, F. S., et al, "Some Experiments on the Perception of Synthetic Speech Sounds," J. Acoust. Soc. Amer. 24, 597-606 (November 1952)

(6) Licklider, J. C. R. and Pollack, I., "Effects of Differentiation, Integration, and Infinite Peak Clipping Upon the Intelligibility of Speech," J. Acoust. Soc. Amer. 20, 42 (January 1948).

(7) Chang, Sze-Hou, et al, "Representations of Speech Sounds and Some of Their Statistical Properties," Proc. of I.R.E. 39, 147-153 (February 1951).

(8) Crater, T. V., "Communication with Clipped Speech Signals," Ph.D. Thesis, Northwestern University, Evanston, Ill. (September 1953).

(9) Dukes, J. M. J., "The Effect of Severe Amplitude Limitation on Certain Types of Random Signals: A Clue to the Intelligibility of Infinitely Clipped Speech," Journal of IEE, Part C, 88-102 (1955).

(10) Egan, J. P., Laryngoscope 58, 955-991 (1949).

(11) Hawthorne, C. B., et al, "Performance of Communications Systems in the Presence of Interference," Final Report, Vol. I, Engineering Experiment Station, Georgia Institute of Technology, Contract AF 30(602)-1789 (1959).

(12) Dunn, H. K. and White, S. D., "Statistical Measurements on Conversational Speech," J. Acoust. Soc. Amer. II, No. 3 (January 1940).

(13) David, E. E., "Signal Theory in Speech Transmission," IRE Transactions on Circuit Theory, Vol. CT-3 (December 1956).

(14) Paxton, R. K., "A Practical Implementation of Reiterated Speech Concepts," Seventh National Communication Symposium Record, Utica, N. Y. (October 1961).

(15) Warren, W. B., "Single Sideband Generator," Invention Disclosure No. 2040, Engineering Experiment Station, Georgia Institute of Technology, Contract AF 30(602)-1789 (March 1954).

(16) Stuckey, C. W., "A Method to Evaluate Learning in Articulation Testing," Technical Note No. 1, Engineering Experiment Station, Georgia Institute of Technology, Contract AF 30(602)-2150 (1960)

(17) Hoel, P. G., "Introduction to Mathematical Statistics," Second Edition, John Wiley & Sons, New York, Chapter 7 (1954).

(18) Duncan, D. B., Biometrics 2, 1-42 (1955).

5. APPENDIX--CORRELATION COEFFICIENTS COMPUTER PROGRAM

The following IBM 650 computer program and instructions were used in making the analyses described in Section 2.2.3. This program is for use in computing bivariate correlation coefficients for which there are 200 or fewer pairs of measurements for each set of two variables. The number of sets is unlimited by the program.

Input Card Format

There are eight types of cards necessary to compute correlation coefficients. They are:

1. The Bell General Purpose System Program. This program is available in the IBM 650 library at most computer centers.
2. A problem number card. This card may be used conveniently as a code card by the operator, since the digits punched into it appear on every output data card. The problem number card is prepared by leaving columns 1 through 76 blank. The problem number is placed in columns 77 through 80. If one or more of these four columns is not needed for the problem number, it should contain a zero.
3. The correlation coefficients program which consists of 51 cards. A listing of this program appears in Table IV.
4. The input data cards for the dependent or "x" variable. This is the variable against which all the independent variables are correlated. The "x" input data are punched five to a card and are loaded into the computer starting at address 201. An "x" input data card is prepared as follows:

   (a) Columns 1 and 2 are blank.

TABLE IV.

CORRELATION COEFFICIENTS PROGRAM LISTING

Card No.	Program	Card No.	Program
01	+6009601009	27	+9800005000
02	+6014600014	28	+9100110000
03	+6019600019	29	+2201201000
04	+6025600025	30	+1607000607
05	+6031600031	31	+8101000005
06	+6037600037	32	+7201900608
07	+9800001000	33	+9800006000
08	+7000610610	34	+9100110000
09	+7000401400	35	+2401401000
10	+7201900604	36	+1608000608
11	+9800002000	37	+8101000006
12	+9100010000	38	+2602606611
13	+1604201604	39	+2604605612
14	+8101000002	40	-1611612613
15	+7201900605	41	+2602607614
16	+9800003000	42	+2604604615
17	+9100010000	43	-1614615616
18	+1605401605	44	+2602608617
19	+8101000003	45	+2605605618
20	+7201900606	46	-1617618619
21	+9800004000	47	+2616619620
22	+9100110000	48	+0300620621
23	+2201401000	49	+3613621609
24	+1606000606	50	+7300609610
25	+8101000004	51	+8000000001
26	+7201900607		

Note: Columns 1 and 2 are blank. The card number is punched into columns 3 and 4. Columns 5, 6, 7, and 8 are blank. Column 9 of each card must contain a plus (+); column 10 of each card must contain a one (1). The sign of the program is punched into column 11, and the program is punched into columns 12-21. All other columns are blank.

(b) Columns 3 and 4 contain the "x" input data card number, starting with 01 for the first card, 02 for the second card, etc.

(c) Columns 5 and 6 are blank.

(d) Columns 7, 8, and 9 contain the data address, starting with 201 for the first card, 206 for the second card, 211 for the third card, etc. Since five pieces of data are punched into each card, the address of the $(N+1)^{st}$ card is 5 higher than that of the N^{th} card.

(e) Column 10 contains the number of pieces of data punched into the card. Column 10 will contain a 5 punch on all "x" input data cards except possibly the last card. If fewer than five pieces of data are required on the last card, column 10 should contain a 1, 2, 3, or 4, corresponding to the number of pieces of data required.

(f) Columns 11 through 65 contain the five pieces of data; columns 11 through 21 contain the first piece of data; 22 through 32 the second, 33 through 43 the third, 44 through 54 the fourth, and 55 through 65 the fifth. The first column for each piece of data (columns 11, 22, 33, 44, 55) contains a plus or minus, depending upon the sign for that piece of data. The next eight columns (columns 12-19, 23-30, 34-41, 45-52, 56-63) contain the number, and the last two columns (columns 20-21, 31-32, 42-43, 53-54, 64-65) contain the exponent. The size of the exponent is arbitrarily increased by 50, so that an exponent of 52 means 10^2, while an exponent of 47 means 10^{-3}. Thus,*

+ 57.5 is written +5750000051

- 1.2 is written -1200000050

+0.375 is written +3750000049

* A zero is written as +0000000000.

Columns 66 through 80 are blank on all "x" input data cards. If, for example, the last "x" input data card contains only three pieces of data, column 10 should contain a 3 punch and columns 44 through 80 should be blank. A typical "x" input data card is shown in Figure 29.

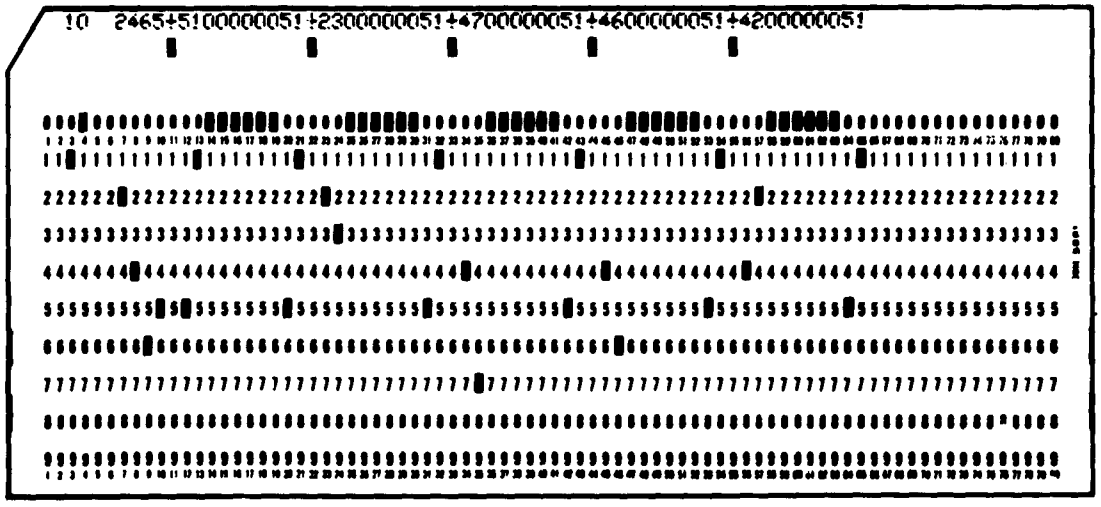


Figure 29. A Typical "x" Input Data Card Containing the Positive Numbers 51, 23, 47, 46, and 42.

5. The data number card. This card tells the computer the number of pairs of dependent and independent observations to be used in computing the correlation coefficient. It is prepared as follows:

   (a) Columns 1 and 2 are blank.

   (b) Columns 3 and 4 contain the card number. The number of this

card should be one greater than the last "x" input data card. For example, if 17 "x" input data cards were required, the data number card should be numbered 18.

(c) Columns 5 and 6 are blank.

(d) Columns 7-10 contain the numbers 6003.

(e) Columns 11, 22, and 33 contain a plus sign.

(f) Columns 12-15 contain zeros.

(g) Columns 16-18 contain the number of pieces of "x" input data (and, hence, the number of pairs of dependent and independent observations). This number should be written in decimal form; thus, one hundred is written 100, but fifty is written 050.

(h) Columns 19-21 and 23-29 contain zeros.

(i) Columns 30-32 again contain the number of observations written in decimal form.

(j) Columns 34-43 contain the number of observations written in floating point form; therefore. one hundred is written as 1000000052, and fifty is written as 5000000051, as outlined in subsection (f) of the previous section.

(k) Columns 44-80 are blank.

A typical data number card is shown in Figure 30.

6. A start card. This card is prepared by placing a zero in columns 7, 8, 9, and 10. All other columns should be left blank.

7. The "y" input data code cards. One such card is prepared for each set of independent or "y" variables. These cards are prepared as follows:

(a) Columns 1 through 3 are blank.

(b) Column 4 contains a one (1).

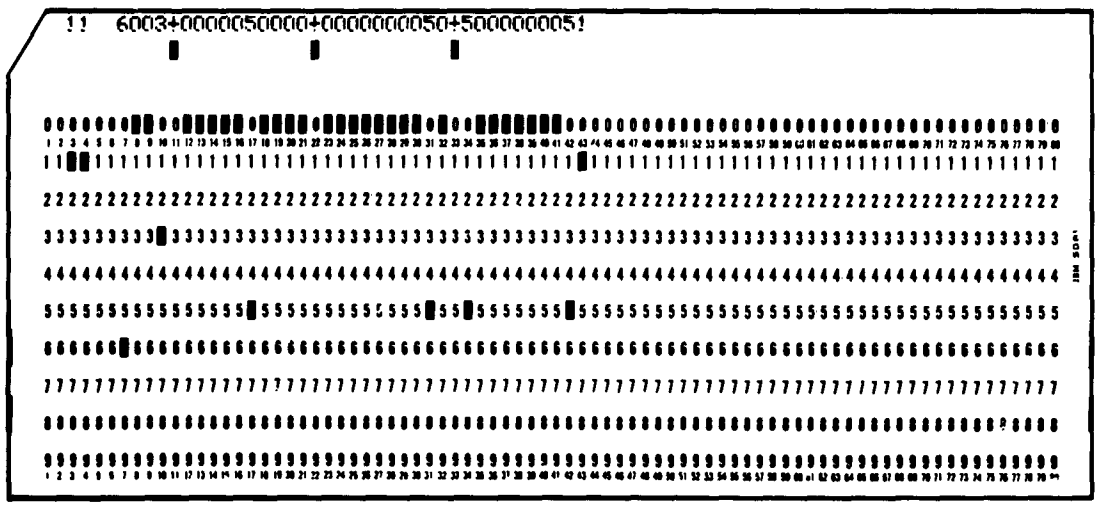


Figure 30. A Typical Data Number Card for an Analysis Containing Fifty Observations.

(c) Columns 5 and 6 are blank.

(d) Columns 7, 8, 9, and 10 contain the numbers 6101.

(e) Columns 11-21 contain the code for the particular set of "y" data that follows the code card. Any numerical system of coding may be used, and the code will be punched only into the output data card associated with the particular set of "y" data. Column 11 must contain either a plus or minus sign. Columns 12-21 may contain digits only (no signs).

(f) Columns 22-80 are blank.

8. The "y" or independent variable input data cards. There is no limit to the number of sets of "y" variable cards that the computer can handle,

since the computer reads and computes only one set at a time. Each set of "y" input data are punched five to a card and loaded into the computer starting at address 401. They are prepared as follows:

(a) Columns 1 and 2 are blank.

(b) Columns 3 and 4 contain the "y" input data card number. These numbers start with 02 for the first card, 03 for the second, etc.; card number one is the "y" input data code card which will precede the data cards when the problem is put in the computer.

(c) Columns 5 and 6 are blank.

(d) Columns 7, 8, and 9 contain the data address. Starting at 401 for the first card, 406 for the second, 411 for the third, etc.

(e) Column 10 contains the number of pieces of data punched into the card, as outlined in section 4(e).

(f) Columns 11 through 65 contain the five pieces of data and are prepared as outlined in section 4(f). EACH "Y" PIECE OF DATA MUST BE PLACED IN A STORAGE POSITION CORRESPONDING TO ITS PAIRED PIECE OF "X" DATA. Thus, the "y" data punched into address 401 must belong to the same pair as the "x" data punched into address 201 on the "x" input data cards. Likewise, the "y" data in addresses 402, 403, 404, etc. must be the respective pair mate of the "x" data in 202, 203, 204, etc.

(g) Columns 66 through 80 are blank on all "y" input data cards. A typical "y" input data card is shown in Figure 31.

Operating Instructions

The following is a detailed list of instructions for loading and computing the bivariate correlation coefficients.

1. Put the Bell General Purpose control panel in the read punch unit.
2. Place blank cards in the punch hopper (top right hopper) of the read punch unit and press the punch start button. This will load the output punch. (Install cards face down, 12 edge first.)

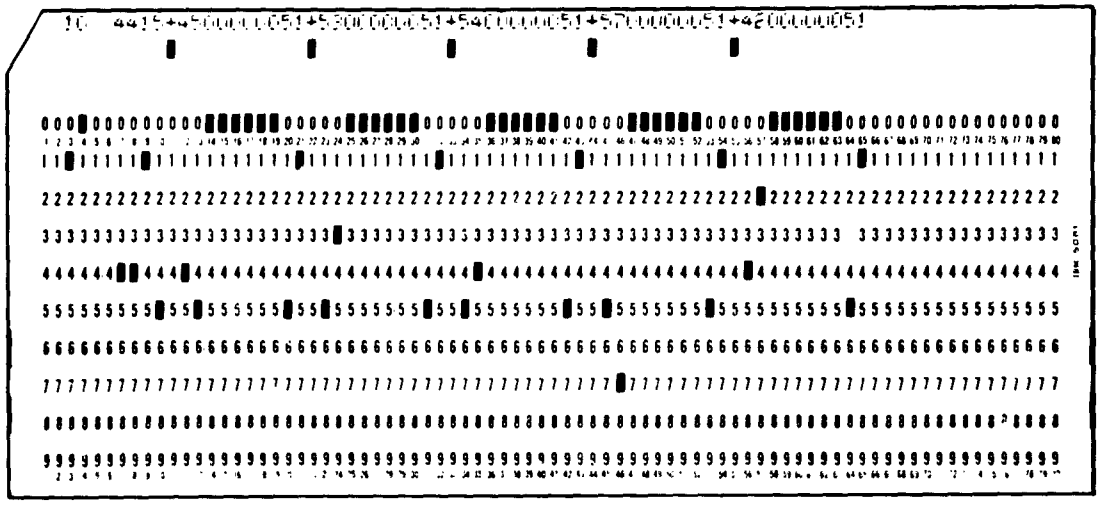


Figure 31. A Typical "y" Input Data Card.

3. Place the following cards in the read hopper (top left hopper) of the read punch unit in the order in which they are listed below:
   (a) Bell General Purpose System Program.
   (b) Problem number card.
   (c) Correlation coefficients program.
   (d) "x" input data cards.
   (e) Data number card.

(f) Start card.

(g) "y" input data card sets, each of which consists of a "y" input data code card followed by the associated "y" input data card. (Install cards face down, 12 edge first.)

4. Set the control console switches as follows:

Storage Entry	=	70 1951 1333
Sign	=	+
Programmed	=	Stop
Half Cycle	=	Run
Address Selection	=	Anything
Control	=	Run
Display	=	Upper Accummulator
Overflow	=	Stop
Error	=	Stop

5. Press the Computer Reset button on the control console unit.

6. Press the Program Start button on the control console unit.

7. Press the Read Start button on the left of the read punch unit. The machine will read cards until the first set of "y" input data cards have been read. One output card will be punched during the reading time; this card is an identification card only and does not contain information pertinent to any correlation coefficients. The machine will stop reading cards when the first set of "y" input data have been read; the coefficient of correlation between the "x" input data and the first set of "y" input data will then be computed and punched out. The second set of "y" input data will then be read, and the correlation coefficient between it and the "x" input data will be computed

and punched out. This process will continue until the last card leaves the read hopper.

8. Press the End of File button on the read punch unit. The computer will calculate and punch out the correlation coefficient between the "x" input data and the last set of "y" input data. The machine will stop with the input-output checking light lit.
9. Remove the blank cards from the punch hopper and hold the Punch Start button down until the machine has cycled three or more times.
10. Remove the data cards from the punch stacker; the first and last cards will be blank.
11. Hold the Read Start button down until all of the program deck cards are in the read stacker and remove the program deck.

This completes the use of the computer with the correlation coefficients program.

Output Data Format

The output data are punched out in the order that the sets of "y" input data are placed in the read hopper. The first card punched is an identification card and may be discarded. The pertinent output data cards have the following format.

Columns 22 through 32 contain the "y" input data code word. This word indicates which particular set of "y" input data was used in computing the correlation coefficient punched into the card.

Columns 11 through 21 contain the correlation coefficient. The sign of the coefficient is given in Column 11 (a twelfth level punch for plus and an eleventh level punch for minus) and the significant digits are given in

floating point form in columns 12 through 19. The exponent is given in columns 20 and 21. (Floating point form is explained in Section 4(f) under Input Data Format).

Columns 77 through 80 contain the problem number. All other columns may be disregarded.

A typical output data card is shown in Figure 32. The correlation coefficient for this card is +0.7868...., the code word is -0302065122, and the problem number is 0203.

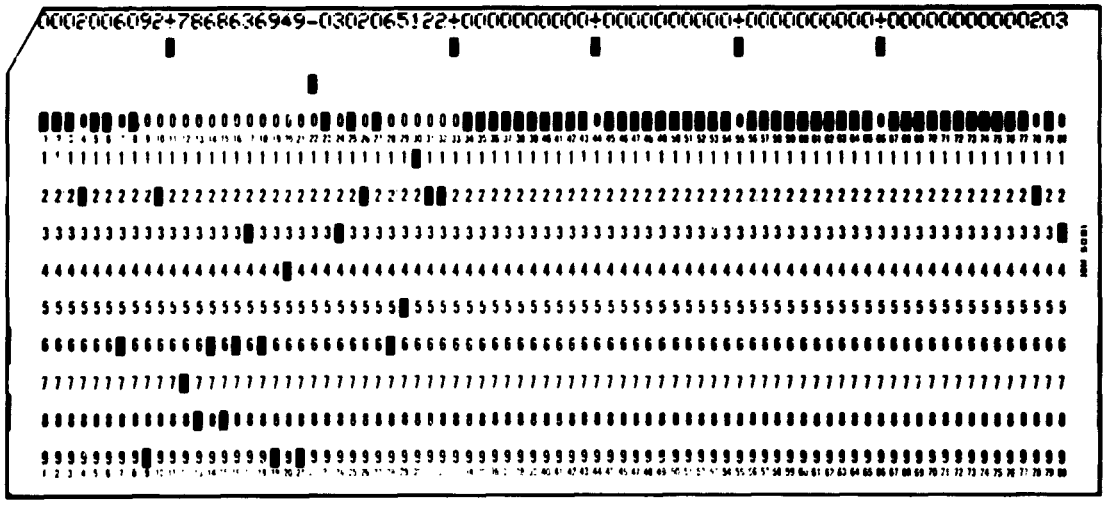


Figure 32. A Typical Output Data Card.

Modifications

It may be desirable for some applications to have certain terms that are used in computing the correlation coefficients; provision has been made to

extract the following terms:

$$\sum_{i=1}^{n} x_i, \quad \sum_{i=1}^{n} y_i, \quad \sum_{i=1}^{n} x_i y_i, \quad \sum_{i=1}^{n} (x_i)^2, \quad \text{and} \quad \sum_{i=1}^{n} (y_i)^2,$$

where x and y have the meaning given them in the sections on Input and Output Data Format.

To extract this information, the following modification of the correlation coefficients program is necessary:

Change column 18 of card number 50 of the program from the digit 9 to the digit 4. (Card 50 is the next-to-the-last card in the correlations coefficients program deck.)

After this modification is made, two output cards will be punched for each set of "y" input data. The first card will have the following format:*

Columns 1 through 10 may be disregarded.

Columns 11 through 21 contain $\sum_{i=1}^{n} x_i$ in floating point form.

Columns 22 through 32 contain $\sum_{i=1}^{n} y_i$ in floating point form.

Columns 33 through 43 contain $\sum_{i=1}^{n} x_i y_i$ in floating point form.

Columns 44 through 54 contain $\sum_{i=1}^{n} (x_i)^2$ in floating point form.

* Note: Card 1 of each pair does not contain the "y" input data code word; for this reason, care must be taken to keep the cards in correct order.

Columns 55 through 65 contain $\sum_{i=1}^{n} (y_i)^2$ in floating point form.

Columns 77 through 80 contain the problem number.

The second card in each pair is a typical output data card and is interpreted as outlined in the previous section.

DISTRIBUTION LIST

Nr. of Copies	
1	RADC (Project Engineer)
1	RAAPT
1	RAALD
1	ROZMSTT
1	RAIS, Mr. Malloy (For: Flt Lt Tanner)
1	Signal Corps Liaison Officer RADC (RAOL, Capt Norton) Griffiss AFB NY
1	AU (AUL) Maxwell AFB Ala
1	ASD (ASAPRD) Wright-Patterson AFB Ohio
1	Chief, Naval Research Lab ATTN: Code 2021 Wash 25 DC
1	Air Force Field Representative Naval Research Lab ATTN: Code 1010 Wash 25 DC
1	Commanding Officer USASRDL ATTN: SIGRA/SL-ADT Ft. Monmouth NJ
1	Chief, Bureau of Ships ATTN: Code 312 Main Navy Bldg Wash 25 DC
1	Office of the Chief Signal Officer Dept of the Army ATTN: SIGRD Wash 25 DC
1	AFPR Rockland Br GE Co PO Box 91 Cincinnati 15 Ohio
1	Chief, AF Section MAAG Germany Box 810 APO 80 New York NY

Nr. of Copies	
1	AFSC (SCSE) Andrews AFB Wash 25 DC
10	ASTIA (TIPCA) Arlington Hall Station Arlington 12 Va
1	Electronics Research Directorate Air Force Cambridge Res. Labs Mr. Weiant Wathen-Dunn, CRRSV Bedford, Massachusetts
1	Elec. Research Directorate AF Cambridge Res. Labs Mr. Caldwell P. Smith Bedford, Mass.
1	Bolt, Beranek and Newman, Inc. Dr. J. C. R. Licklider 50 Moulton St. Cambridge 38, Mass.
1	Bolt, Beranek & Newman Inc. Dr. K. D. Kryter 50 Moulton St. Cambridge 38, Mass.
1	Electronic Systems Div. OAL Dr. Stephen Stuntz, ESRHI Bedford, Mass.
1	Res Lab of Electronics Mass. Inst. of Tech. Dr. K. N. Stevens 77 Massachusetts Ave. Cambridge 38, Mass.
1	RADC (RASCH) Dr. R. C. Christman Griffiss AFB, NY
1	RADC (RASCH) G. Renaud Griffiss AFB, NY
10	RADC (RAOPA) Captain J. Griffiths Griffiss AFB, NY Remainder to ASTIA