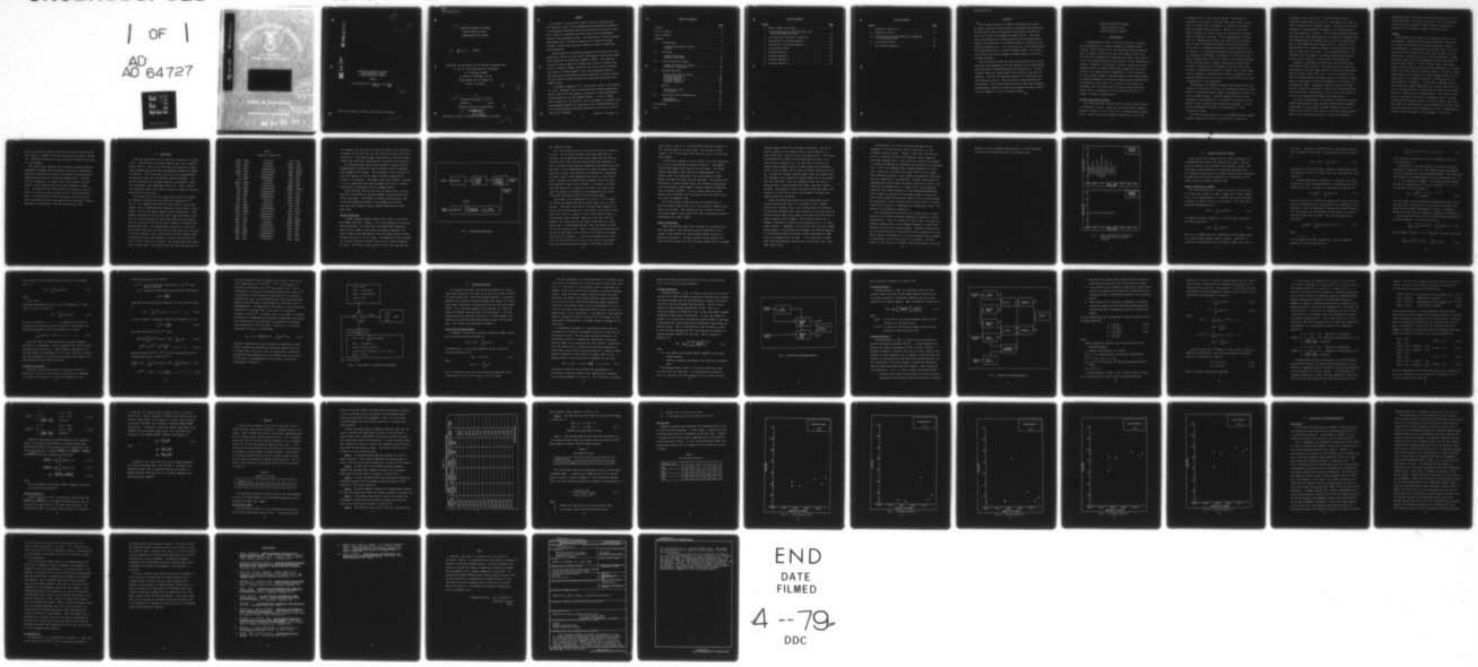
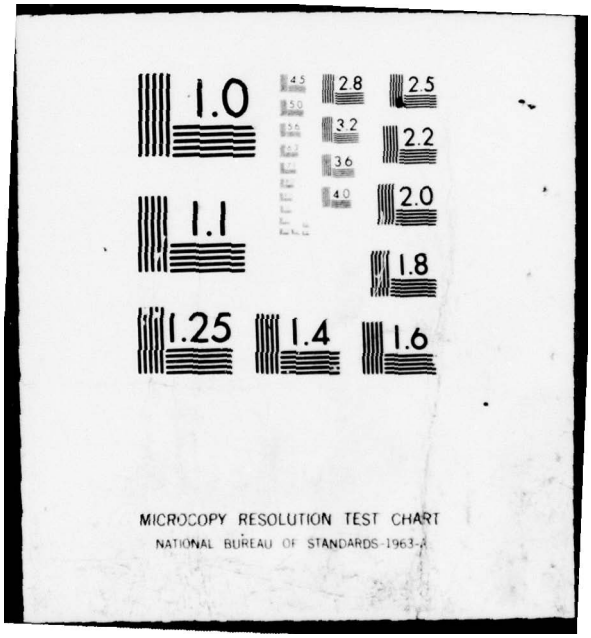
AD-A064 727 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 17/2
OBJECTIVE MEASURE OF SPEECH INTELLIGIBILITY USING LINEAR PREDIC--ETC(U)
DEC 78 D M OTTINGER

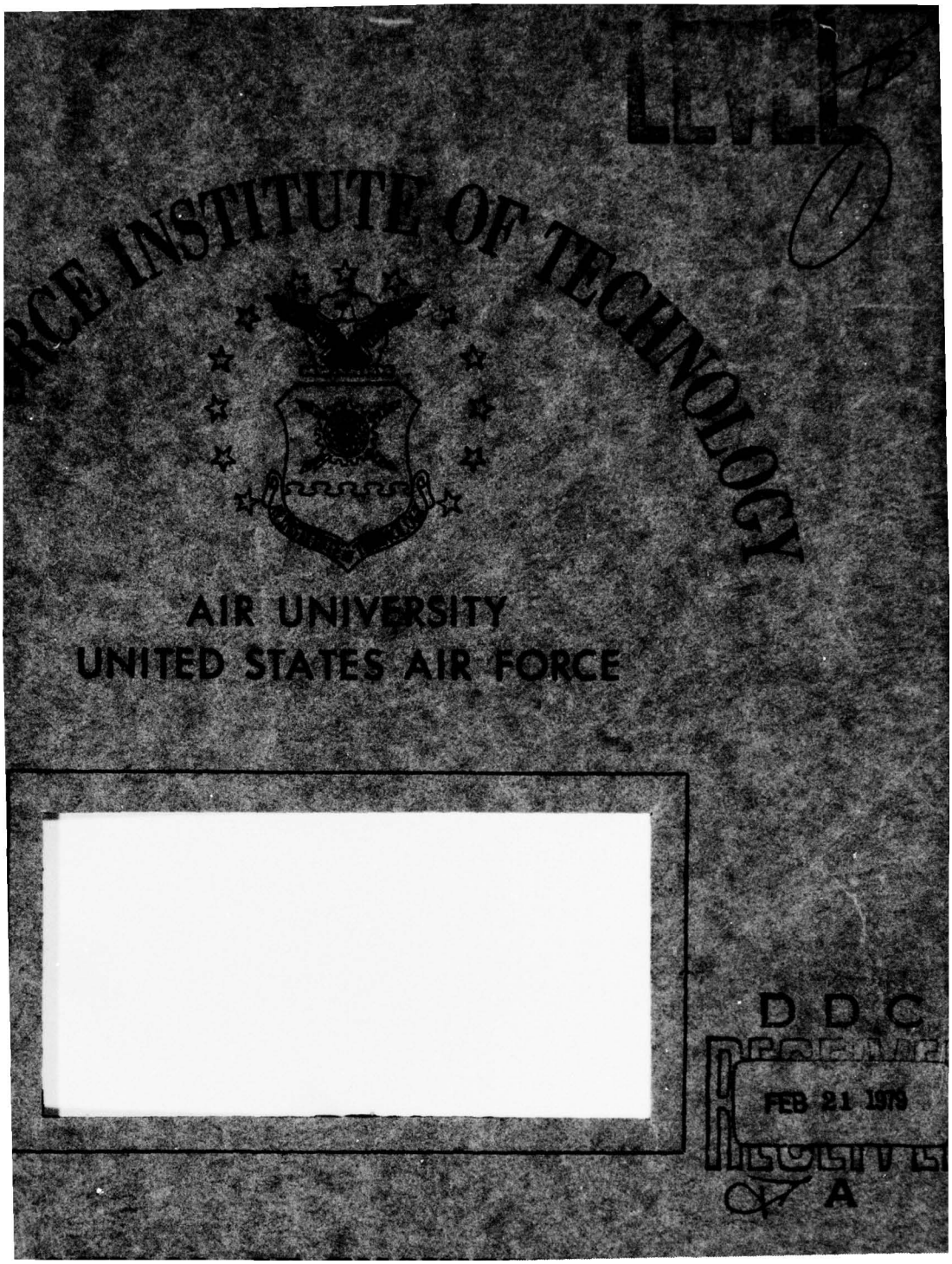
UNCLASSIFIED AFIT/GE/EE/78-35 NL

1 OF 1
AD
AD 64727



END
DATE FILMED
4 -- 79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A


RCE INSTITUTE OF TECHNOLOGY
AIR UNIVERSITY
UNITED STATES AIR FORCE
DDC
FEB 21 1978
A

OBJECTIVE MEASURE OF SPEECH INTELLIGIBILITY USING LINEAR PREDICTIVE CODING

THESIS

AFIT/GE/EE/78-35 Donald M. Ottinger
Captain USAF

Approved for public release; distribution unlimited.

(14)
AFIT/GE/EE/78-35

(6)

OBJECTIVE MEASURE OF SPEECH INTELLIGIBILITY USING LINEAR PREDICTIVE CODING

(9) Master's THESIS,

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air Training Command
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

(10) by
Donald M. Ottinger, Jr., B.S.
Captain USAF
Graduate Electrical Engineering

(11) December 1978

(12) 61p.

Approved for public release; distribution unlimited.

012225

Preface

The ability to objectively measure speech intelligibility has long been a goal of the communications-engineering community. A few automated techniques have been developed in the past years, but to date, no technique has fulfilled all the requirements desired of an automated system. The subjective scoring of speech intelligibility by trained listeners still remains the most reliable, though maybe the most expensive, means of measuring intelligibility.

Linear predictive coding has appeared on the horizon of communications theory of late, and in preliminary systems has proven quite effective in producing synthetic speech. The question arises, if linear predictive coding can be used to produce high quality synthetic speech, then why can't it be used to measure the quality of human speech? This study addresses itself to this question by developing objective measures of speech intelligibility based on linear predictive coding and measuring their effectiveness.

I am deeply indebted to Mr. William Hall and Mr. Dave McGrew for their invaluable help in processing the analog speech data and for the use of the computer resources of the Analog/Hybrid System Branch of the ASD Computer Center. I wish to thank Major Joseph Carl, my advisor, for his guidance, assistance, and encouragement during this study and to Mr. Richard McKinley of the Aerospace Medical Research Laboratory for the use of their audio test equipment.

Donald M. Ottinger, Jr.

Table of Contents

List of Figures

List of Tables

Abstract

Four distance measures of speech intelligibility based on linear predictive coding (LPC) are developed and evaluated. The data base used for evaluating the measures consisted of lists of 58 words from Diagnostic Rhyme Test IV. The lists were transmitted over a spread spectrum radio communications channel and subjected to 7 different levels of non-white, non-Gaussian jamming noise. The lists were all scored subjectively for intelligibility by a trained listener panel. The subjective scores were used to judge the effectiveness of the four distance measures.

The Articulation Index was also calculated for each of the word lists and compared to the LPC measures as to effectiveness and efficiency in measuring speech intelligibility. The Articulation Index was significantly more effective than the LPC measures. The best LPC measure provided 42% correlation with the subjective scores. The Articulation Index provided 69% correlation. The overhead associated with data tape alignment and parameter computation makes LPC measures extremely inefficient as compared to the Articulation Index.

OBJECTIVE MEASURE OF SPEECH INTELLIGIBILITY USING LINEAR PREDICTIVE CODING

I. Introduction

A continuing need exists within the military to measure the intelligibility of speech produced on communications systems. While methods exist for measuring system parameters such as signal-to-noise ratio and idle channel noise, very few tests are available for measuring the actual intelligibility of the speech produced at the receiver of the communication channel. Examples of situations in which a measure of speech intelligibility is needed include the comparative testing of similar voice communications equipments, on-line evaluation of voice channel quality, and measurement of the effectiveness of spectrum jamming in destroying communications capabilities. The purpose of this thesis is to explore a relatively new technique used for speech analysis called linear predictive coding (LPC) to determine if LPC can form the basis for a measure of speech intelligibility.

Present Measurement Systems

The oldest and most reliable test of speech intelligibility is subjective scoring. Subjective scoring involves trained speakers reading a list of words over a communications channel, while a panel of trained listeners subjectively scores the

intelligibility of the received speech. The method is extremely reliable due to the fact that actual human listeners are involved and no equipment is required to attempt to model the human hearing process. However, the fact that human listeners are used accounts for the numerous disadvantages of subjective scoring. If two communications systems are to be comparatively tested, the same group of listeners must be used to prevent distortion of the intelligibility scores due to a difference between the hearing abilities of two groups. If a significant number of intelligibility tests are required or the number of systems to be tested is quite large, considerable time must be spent in the testing process. If the listener group is large, considerable manhours and, therefore, expense will be expended in testing the systems. For tests evaluating the quality of speech over on-line communication channels or measuring the effect of jamming on disrupting communications, the use of a controlled listener group is impractical if not impossible.

The decrease the manhours, expense, and impracticality of subjective intelligibility tests, an automated system is needed that will accurately measure speech intelligibility without the use of listener groups. The most important quality of any automated system developed is the ability to produce the same results that a listener panel would have produced. To date, the only automated technique in widespread use is the Articulation Index.

The Articulation Index (AI) is an automated speech intelligibility measure that was first described by French and

Steinberg in 1947 (Ref 1:10). The Articulation Index is computed by measuring the signal-to-noise power ratio (SNR) in twenty separate audio frequency bands. The SNR value for each frequency is then weighted according to its contribution to the intelligibility of speech. The American National Standards Institute has established weights to be applied, dependent upon the communications environment and the type of distortion present in the communications channel (Ref 2). The sum of the weighted SNR's are scaled to produce an intelligibility score with a range from zero to one. An Articulation Index of one indicates that the speech is perfectly intelligible, while a value of zero indicates a total lack of intelligibility.

Hardware is presently available that can calculate the Articulation Index directly. One system which has been used extensively in military applications is the Voice Interference Analysis System (VIAS) (Ref 1:11). The system measures the SNR in fourteen separate frequency bands, as opposed to twenty bands as specified by French and Steinberg, to calculate the Articulation Index. Reasonably accurate results have been achieved using the equipment as long as the interfering noise present was white and Gaussian, and any other distortion present in the communication channel was known apriori.

The VIAS system appears to provide a system for evaluating speech intelligibility when testing communications equipment performance in the presence of known channel distortions. However, the system is not suited for on-line channel measurements or for studying the effects of real-time jamming of

communications where the type of distortions present are not known beforehand. The need, therefore, is for an automated system that can accurately predict speech intelligibility without prior knowledge of channel distortion types.

Purpose

The purpose of this thesis is to evaluate the use of a mathematical technique called linear predictive coding (LPC) as a basis for developing an objective measure of speech intelligibility. Linear predictive coding is not a new technique and can be traced back to the works of Gauss in 1795 (Ref 11:10). However, the use of linear prediction in communications theory only first appeared in 1949 in the works of Norbert Weiner (Ref 12). More recently, Saito and Itakura began applying linear prediction to the formulation of a human vocal tract model used to synthesize speech. The use of linear prediction for the synthesis of speech suggests that the technique might be successfully applied to the analysis of the intelligibility of speech. A study done by Hartman (Ref 8) has proven that linear prediction can, in fact, form the basis for an accurate measure of intelligibility when the distortion present is additive white Gaussian noise. In a similar study done by the Georgia Institute of Technology (Ref 4), several intelligibility measures based on LPC were tested. The tests were made on a communications system subjected to various types of distortion that could be expected to occur in communication channels and with digital voice equipment. Of all the

objective measures tested, the LPC based measures provided the best results. However, the data base used was severely limited and, therefore, resulted in large estimated standard deviations for the measures.

This thesis will evaluate the LPC based objective measures developed by Hartman and the Georgia Institute of Technology against the data base created by subjecting a spread spectrum communications system to non-white jamming noise. The data base was created by J.E. Bauer (Ref 5) and consists of monosyllabic words selected from the Harvard Diagnostic Rhyme Test. The data base has been subjectively scored for intelligibility as well as being scored by use of the Articulation Index. The evaluation of the LPC measures will be based on their correlation with the subjective scores and their relative advantage or disadvantage over the Articulation Index.

II. Data Base

The data base used in this study was created by J.E. Bauer (Ref 5) and modified to the form used in this test by Wayne R. Beeson (Ref 6). The core of the data base consists of fifty-eight rhyming word pairs from the Diagnostic Rhyme Test Number IV (DRT-IV). DRT-IV was used since the list is phonetically balanced and tests for six specific speech attributes. The speech attributes are voicing, nasality, sustenation, sibilation, graveness, and compactness (Ref 6:9). Table I shows the word pairs used in the data base and the specific attribute associated with each pair.

Four master lists of fifty-eight words each were created by randomly selecting one word from each rhyming pair of DRT-IV. Two speakers were used as test subjects, one a male subject with a southern accent (Arkansas) and the other a male subject with no noticeable regional accent (Minnesota). Each speaker recorded two of the four fifty-eight-word master lists. The lists were recorded on stereo audio tape with one channel used for each word list and the other channel for timing marks between each word. The timing marks consisted of a one kilohertz (kHz) sine wave, one-half second long, which was used both to cue the speaker to say a word and to provide a marker separate from the actual data channel to identify the interval of tape in which a word was recorded. The timing marks were spaced seven seconds apart, and Beeson found that the reaction time of

TABLE I

Diagnostic Rhyme Test

PEST - TEST	-(filler)-	FAN - PAN
VAULT - FAULT	-(voicing)-	CHOCK - JOCK
DUES - NEWS	-(nasality)-	NOTE - DOTE
VEE - BEE	-(sustention)-	TICK - THICK
THANK - SANK	-(sibilation)-	CARE - CHAIR
ROD - WAD	-(graveness)-	DONG - BONG
SO - SHOW	-(compactness)-	YOU - RUE
LID - RID	-(filler)-	REEK - LEAK
DENSE - TENSE	-(voicing)-	GAFF - CALF
BOSS - MOSS	-(nasality)-	BOMB - MOM
FOO - POOH	-(sustention)-	DOUGH - THOUGH
ZEE - THEE	-(sibilation)-	GILT - JILT
FAD - THAD	-(graveness)-	PENT - TENT
HOP - FOP	-(compactness)-	YAWL - WALL
ROW - LOW	-(filler)-	LOOT - ROOT
GIN - CHIN	-(voicing)-	VEAL - FEEL
BEND - MEND	-(nasality)-	NAB - DAB
CHAW - SHAW	-(sustention)-	BON - VON
JUICE - GOOSE	-(sibilation)-	SOLE - THOLE
PEAK - TEAK	-(graveness)-	THIN - FIN
BAT - GAT	-(compactness)-	KEG - PEG
ROCK - LOCK	-(filler)-	LONG - WRONG
GOAT - COAT	-(voicing)-	TUNE - DUNE
MIT - BIT	-(nasality)-	MEAT - BEAT
THEN - DEN	-(sustention)-	SHAD - CHAD
GAUZE - JAWS	-(sibilation)-	GOT - JOT
NOON - MOON	-(graveness)-	DOLE - BOWL
KEY - TEA	-(compactness)-	DILL - GILL
RAMP - LAMP	-(filler)-	LEND - REND

the speakers was such that the word was spoken (and, therefore, recorded) within the first two and one-half seconds after the timing mark. The master tapes represented the baseline speech signal from which all intelligibility distances were measured.

The baseline tapes were played through a spread spectrum communications system with seven different levels of jamming noise added to the signal. The recordings of the receiver output were labeled as to the signal-to-signal jamming ratio present in the system. The output tapes were labeled 1 through 7, with 1 signifying the lowest jamming level and 2 through 7 signifying an increasing level of jamming (Ref 5).

Since the data base was entirely in analog form, the data had to be converted to a digital format to allow digital computer processing. The analog to digital conversion was done by the Analog/Hybrid Branch of the Aeronautical Systems Division (ASD) Computer Center, Wright-Patterson Air Force Base, Ohio.

Analog Processing

A Comcor CI5000/6 analog computer was used to pre-process and sample the data. Figure 1 is a block diagram of the analog data processing. To insure that the analog data effectively used the full range of the analog-to-digital converter, the speech signals were amplified so that the peak-to-peak voltage swings were from -75 to +75 volts. The amplifiers were followed by a 4-pole Chebyshev low-pass filter with a cutoff frequency of 4 kHz. The output of the low-pass filter was then fed into

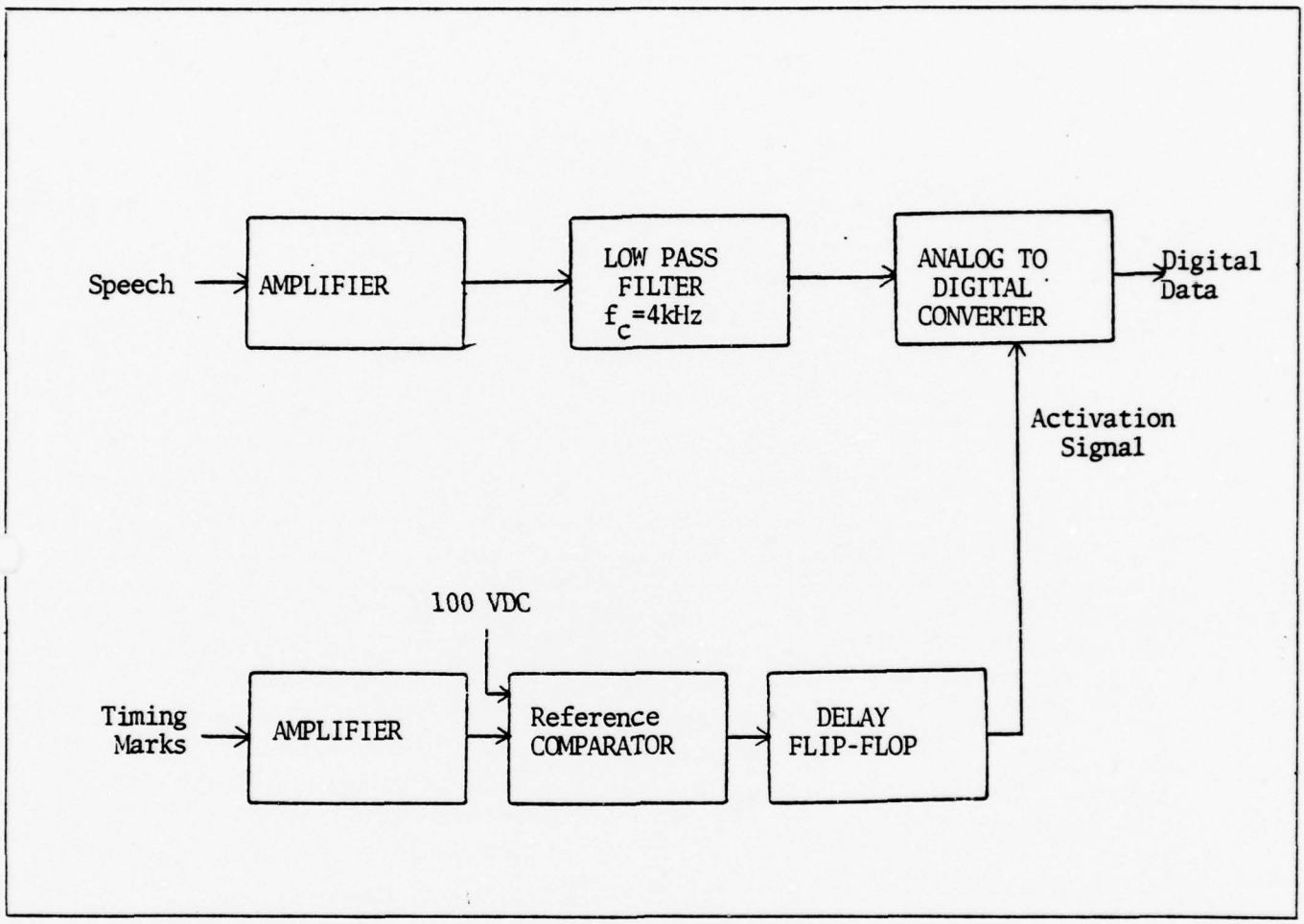


Fig 1. Analog Signal Processing

the sampling circuit.

The 1 kHz timing marks were used to activate the sampling circuit. As the speech waveform was being amplified and filtered, the timing marks were being amplified and detected. To detect the 1 kHz sine wave and activate the sampling circuit, a network consisting of a comparator and delay flip-flop was used. After being amplified to a peak voltage of approximately 125 volts, the sine wave was applied to a comparator. The other input of the comparator was tied to a 100 volt DC level. Whenever the input tone was greater than 100 volts, the comparator output was a logical 1 (+5 volts); any other time the output was a logical 0 (0 volts). As long as the tone was present, the output of the comparator was a pulse train with the same period as the sine wave input.

The output of the comparator was used as a clock signal to a delay flip-flop with the interval timer set at 2 milliseconds. The delay flip-flop is trailing edge triggered, so that on the trailing edge of a clock pulse, the output of the delay flip-flop is a logical 1 for a time interval equal to the interval timer setting. When the 1 kHz sine wave was present, the clock input to the delay flip-flop was a pulse train with a 1 millisecond period, twice the amount of time set on the interval timer. Thus, as long as the sine wave was present at the input to the circuit, the output of the delay flip-flop was a logical 1. Upon the occurrence of the last cycle of the sine wave, the pulse train input to the delay flip-flop would end and the output of the delay flip-flop

would drop to logical 0, 2 milliseconds after the presence of the trailing edge of the last pulse. The transition from logical 1 to 0 of the delay flip-flop was used to activate the data sampler.

The data was sampled at 8 kHz (Nyquist rate) and quantized to 12 bits by the analog to digital converter. The sampler took 20,480 samples each time it was activated. The 20,480 samples represent a time interval of approximately 2-1/2 seconds after the timing mark, the time interval in which Beeson observed that each word was recorded. The samples were converted to actual voltage values and stored on digital magnetic tape using a Xerox Sigma 7 digital computer integrated with the Comcor CI5000/6. In all, eleven word lists were samples (four baseline lists and seven noise corrupted lists) and stored on magnetic tape.

All of the words in the data base are monosyllabic so that the speech signal lasts for a time interval somewhat less than 2-1/2 seconds. The problem now was to detect which of the 20,480 sample values taken after each timing mark actually represented the speech signal.

Digital Processing

Since the baseline tapes were recorded in an almost noise free environment, the data words would be detected in the stream of data samples by using average energy criteria to establish thresholds. The data stream (20,480 samples) for each word was divided into 160 128-point windows and the average

squared sample value for each window calculated. The noise that is present in the sample values is due primarily to analog tape hiss, receiver noise, and quantization. This noise was assumed to be additive white Gaussian noise with a one-sided spectral height of N_o. By physically observing the average squared sample values of each window of the baseline words, it was evident that an upper threshold could be set for N_o. Anytime the average squared sample value was greater than the threshold, the presence of signal energy due to the spoken word was indicated. With the assumption that the noise is Gaussian with a flat spectrum, the signal detection scheme is optimum. The windows at which the word started and ended were stored on a disk file as well as the number of windows and samples each word contained.

Since the analog data tapes were re-synchronized every 7 seconds and only the first 2-1/2 seconds of the 7 second interval was actually used, it was assumed that the location of the noise corrupted data words would be at the same relative position within their data stream as the baseline words were in their respective data streams. The optimum receiver for a channel corrupted by additive white Gaussian noise is a correlation receiver. Therefore, to prove that the lists were indeed synchronized, a cross-correlation between the detected samples of the baseline word and the data stream containing the same word plus jamming noise was made. It was hoped that a sharp peak indicating maximum correlation, and therefore word alignment, would occur.

Unfortunately, the cross-correlation showed the two sequences to be uncorrelated, and no information on tape alignment could be gained. Figure 2 shows the cross-correlation of a baseline word with the receiver output signal at the lowest jamming level. Two factors may explain the absence of correlation between the baseline word and the noise corrupted word. First, the jamming noise, though intended to be additive, could have also had a non-linear effect on the signal. The correlation receiver can no longer be expected to work when the noise component is not additive. Second, spread spectrum communications involve many non-linear processes as well as the spreading and despreading of a signal over a wide bandwidth. Either the non-linear processes of the spreading/despreading could change the spectrum of the speech enough to cause zero correlation between the baseline signal and the received signal. The possibility of the occurrence of the last factor is strengthened by the results of computing the Articulation Index. The results are discussed in Chapter VI.

The synchronization provided by the timing marks is representative of the data gathering techniques available in field organizations. Thus, for purposes of this study, the tapes are assumed aligned within a close enough tolerance to objectively evaluate the use of LPC based measures. Whether the failure of the LPC based measures is associated with the inability to accurately measure intelligibility or with inadequate tape alignment is immaterial as far as this study is concerned. The main thrust of this study is to evaluate the effectiveness of the LPC

measures in an environment representative of the environment in which the Articulation Index is presently used.

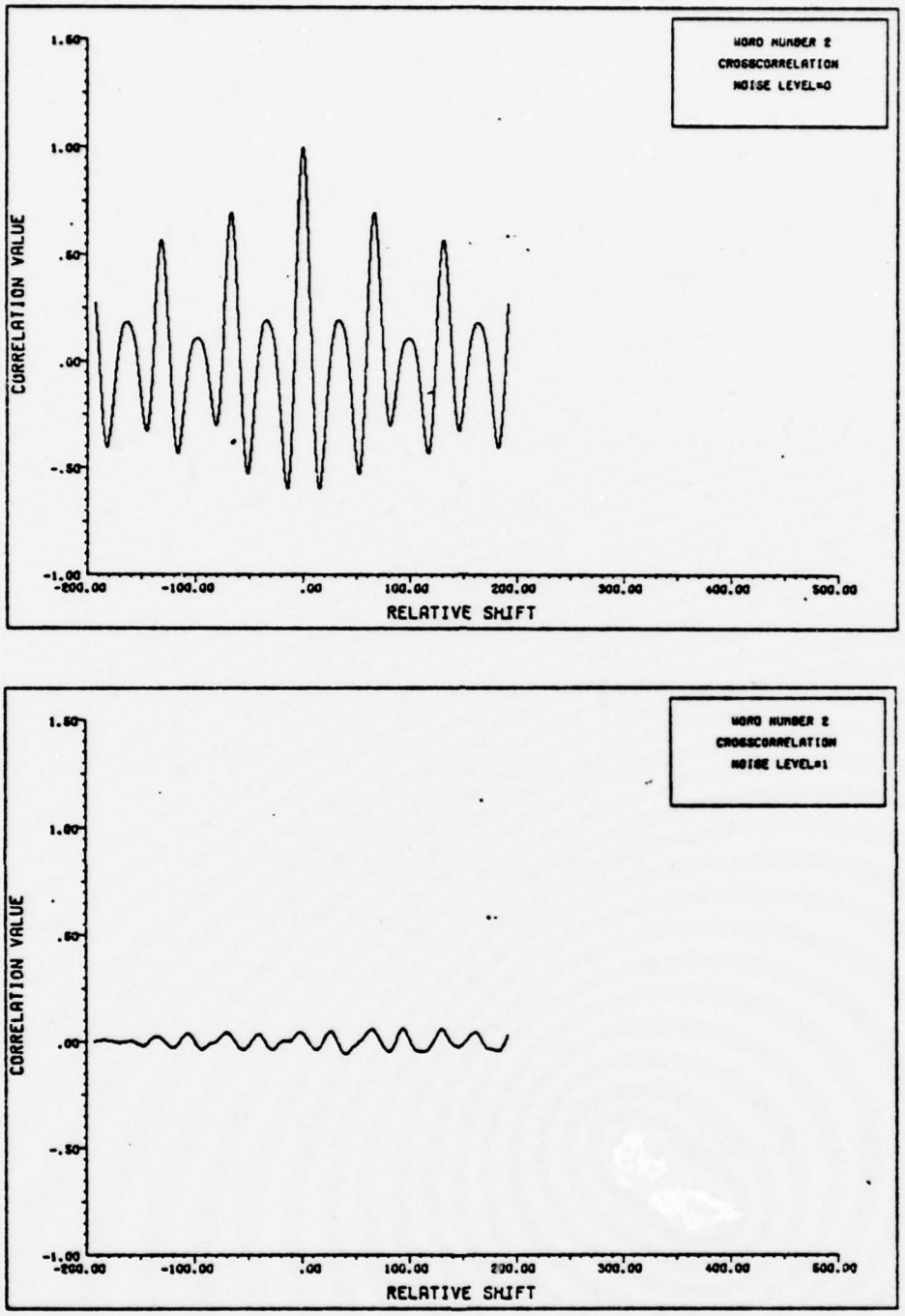


Fig 2. Cross-correlation of Baseline Word with Noise Corrupted Versions

III. Linear Predictive Coding

Linear predictive coding (LPC) has found widespread use in communications theory over the past few years. Specific areas of interest have included voice encoding, speaker identification, word recognition, and spectrum approximation. This chapter presents the basic theory behind linear prediction of speech and one solution algorithm for formulation of the linear prediction speech analysis model.

Linear Prediction of Speech

The linear prediction of speech is based on the idea that at a particular instant in time, a sample of a speech signal, S(nT), can be approximated by a weighted sum of the preceding P samples of speech, where P is an integer. This idea can be expressed mathematically as

$$S(nT) \simeq \sum_{i=1}^{P} a_i S(nT-iT) \qquad (3.1)$$

To simplify notation, Equation 3.1 is most often written in the form shown below

$$S(n) \simeq \sum_{i=1}^{P} a_i S(n-i) \qquad (3.2)$$

where it is assumed that S(m) represents the mth sample value of a speech signal sampled every T seconds. Equation 3.2 represents an approximation to the speech signal and, thus, is

not exact. The error between the exact speech sample during the nth sample interval and its approximation can be defined by

$$e(n) = S(n) - \sum_{i=1}^{P} a_i S(n-i) \qquad (3.3)$$

The goal is to find the weights (predictor coefficients) that will minimize the error in some sense over some specified time interval.

A common minimization technique is to minimize the total squared error over a defined interval. By defining the total squared error as E, the goal is to minimize the expression

$$E = \sum_{n} [S(n) - \sum_{i=1}^{P} a_i S(n-i)]^2 \qquad (3.4)$$

where the limits on n define the interval over which the error is to be minimized and are deliberately left undefined for now. Equation 3.4 can be minimized by taking the partial derivative of E with respect to each predictor coefficient, a_i, and setting the result equal to zero. The resulting equation is shown below.

$$\sum_{n} 2[S(n) - \sum_{k=1}^{P} a_k S(n-k)][-S(n-i)] = 0 \qquad (3.5)$$

where

$$i = 1, 2, \ldots, P$$

By rearranging terms and changing the order of summation, Equation 3.5 can be rewritten as:

$$\sum_{k=1}^{P} a_k \sum_{n} S(n-k)S(n-i) = -\sum_{n} S(n)S(n-i) \qquad (3.6)$$

At this point the limits on n must be defined in order to solve Equation 3.6.

The limits on n are specified by the choice of solution technique for Equation 3.6. Two common solution techniques are the Covariance and Autocorrelation Methods. The Covariance Method defines the minimization of E for an interval of n = 0, 1, ..., N-1 consecutive samples. The Autocorrelation Method defines the minimization of E for an interval $-\infty < n < +\infty$, but defines the speech signal as

$$S(n) = \begin{cases} S(n), & n = 0, 1, \ldots, N-1 \\ 0, & \text{otherwise} \end{cases} \qquad (3.7)$$

For this study, the Autocorrelation Method was chosen since it requires fewer calculations in the solution and insures the speech analysis model constructed is stable. A detailed analysis of both solution methods is contained in Reference 11.

Having specified the use of the Autocorrelation Method for solution, Equation 3.6 can be rewritten as

$$\sum_{k=1}^{P} a_k \sum_{n=-\infty}^{+\infty} S(n-k)S(n-i) = -\sum_{n=-\infty}^{+\infty} S(n)S(n-i) \qquad (3.8)$$

With a change of index, j = n-i, Equation 3.8 can be written as

$$\sum_{k=1}^{P} a_k \sum_{j=-\infty}^{+\infty} S(j+i-k)S(j) = -\sum_{j=-\infty}^{+\infty} S(j+i)S(j) \qquad (3.9)$$

The estimate of the autocorrelation function of the signal S(n) is

$$R(i) = \sum_{n=-\infty}^{+\infty} S(n)S(n+i) \qquad (3.10)$$

where

$$R(i) = R(-i)$$

Using the definition of S(n) as given in Equation 3.7, Equation 3.10 can be written as

$$R(i) = \sum_{n=0}^{N-1-i} S(n)S(n+i) \qquad (3.11)$$

For cases in which i = 1, 2, ..., P, Equation 3.11 will be defined as the short-term autocorrelation of the signal S(n). Substituting Equation 3.11 into Equation 3.9 yields

$$\sum_{k=1}^{P} a_k R(i-k) = -R(i), \quad i = 1, 2, \ldots, P \qquad (3.12)$$

Once the short-term autocorrelation has been computed, Equation 3.12 represents P linear equations that can be solved simultaneously for each a_k. Linear algebra techniques exist for efficiently solving Equation 3.12, but a recursive solution has been developed by Levinson that provides even greater computational efficiency (Ref 12:129-148).

Levinson's Algorithm

Levinson's algorithm provides a recursive solution to Equation 3.12 that is both simple and efficient to implement. To simplify the notation for recursive computation, the

following quantities are defined:

$A_i^{(P)}$ = the i^{th} predictor coefficient of the P^{th} order prediction model

$r(n)$ = normalized short-term autocorrelation coefficient

$$r(n) = \frac{R(n)}{R(0)}$$

Using the above definitions, Equation 3.12 can now be written as

$$-r(j) = \sum_{i-1}^{P} A_i^{(P)}\, r(i-j)\ ,\ j = 0, 1, \ldots, P \qquad (3.13)$$

To start Levinson's algorithm, define a new quantity, K_o, as

$$K_o^{(0)} = \frac{r(1)}{r(0)} \qquad (3.14)$$

and solve recursively for $k_i^{(P)}$ using

$$K_o^{(P)}\left[r(0) - \sum_{i=0}^{P-1} K_i^{(P-1)}\, r(P-i)\right] = r(P+1) - \sum_{i=1}^{P} K_{i-1} r(i) \qquad (3.15)$$

$$K_i^{(P)} = K_{i-1}^{(P-1)} - K_o^{(P)}\, K_{P-i}^{(P-1)}\ ,\ i = 1, 2, \ldots P \qquad (3.16)$$

Having calculated $K_i^{(P)}$, $A_i^{(P+1)}$ can be calculated from the following equations; define $A_o^{(0)} = 1$

$$A_{(P+1)}^{(P+1)}\left[r(0) - \sum_{i=0}^{P} K_j^{(P)}\, r(P+1-i)\right] = r(P-1) - \sum_{i=0}^{P} A_j^{(P)} r(P+1-i) \qquad (3.17)$$

$$A_i^{(P+1)} = A_i^{(P)} - K_i^{(P)}\, A_{(P+1)}^{(P+1)}\ ,\ i = 0, 1, \ldots, P \qquad (3.18)$$

Two vector quantities are generated as a result of the recursive computations, $A^{(P)}$ and $K^{(P)}$. $A^{(P)}$ is the vector of predictor coefficients for a P^{th} order model. $K^{(P)}$ can be interpreted as a vector of reflection coefficients; each $K_j^{(P)}$ is analogous to the reflection coefficients of a P-section transmission line. As mentioned earlier, the autocorrelation method allows the model to be checked for stability prior to implementation. In transmission line theory, if any reflection coefficient is greater than 1, the circuit is unstable. Likewise, if any $K_j^{(P)}$ is greater than 1, the model is unstable. In addition to producing the prediction coefficients and reflection coefficients, the algorithm also generates the minimum total squared error for the model. Define $E_o = 1$ and solve recursively for E_{P+1} as shown below.

$$E_{P+1} = E_P + A_{(P+1)}^{(P+1)}[R(P+1) - \sum_{i=0}^{P} K_i^{(P)} r(i)] \qquad (3.19)$$

Levinson's algorithm not only provides the prediction coefficients, reflection coefficients, and total squared error for a P order model, but also, since the algorithm is recursive, provides the same quantities for all models less than order P. A flow chart for implementation of Levinson's algorithm is illustrated in Figure 3.

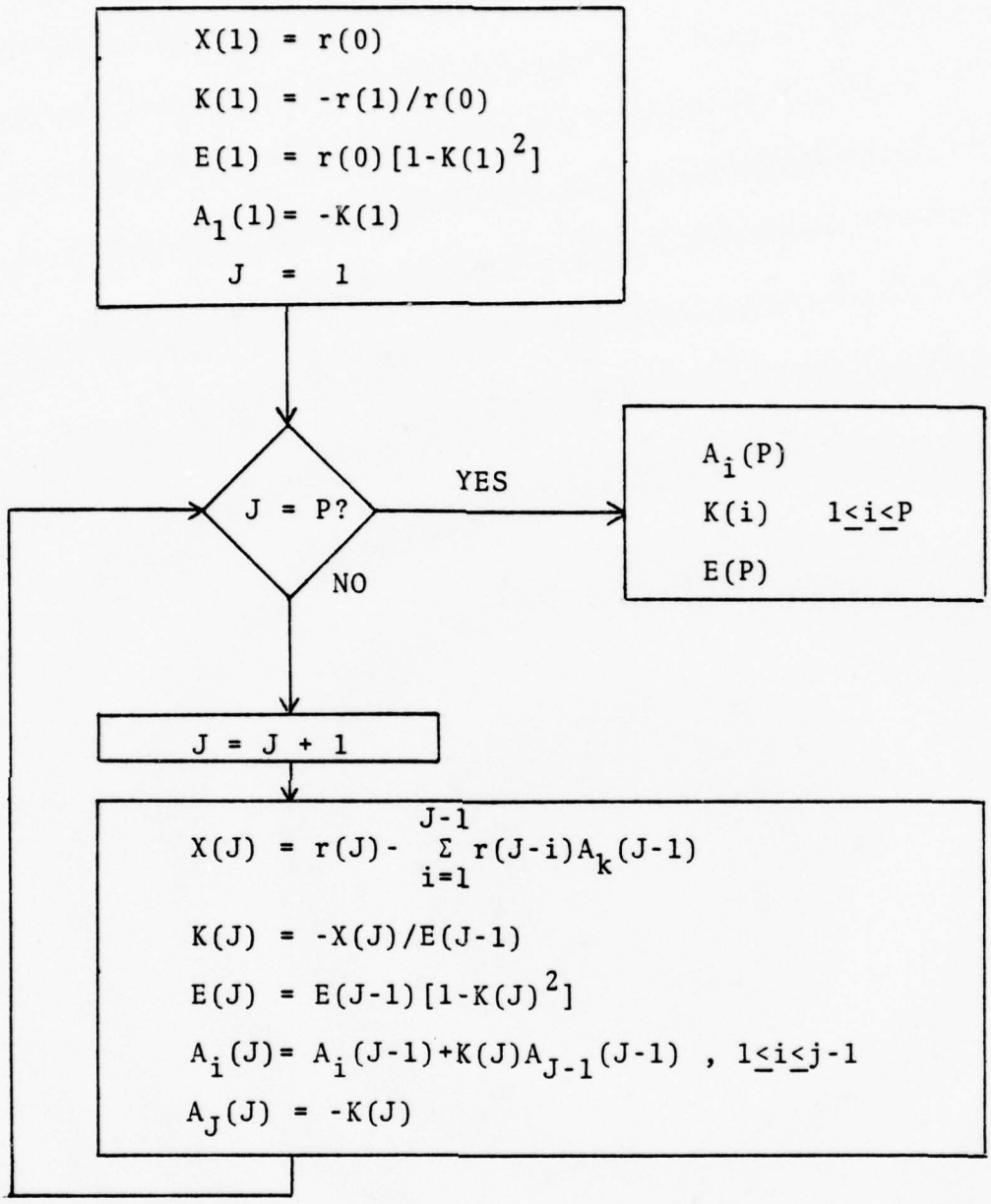


Fig 3. Flow Chart for Levinson's Algorithm

IV. Distance Measures

This chapter describes four objective measures of speech intelligibility which are based on a vocal tract model created by linear prediction. The term distance measure, as used here, indicates the relative distance between some aspect of a base-line speech signal and a distorted version of that same speech signal. For a distance measure to be valuable, it should be highly correlated with subjective scoring results. The four distance measures described here were all tested against the subjective scores and Articulation Index results of the data base. The results are contained in Chapter V.

Vocal Tract Analysis Model

In Chapter III the error between a predicted speech signal and the actual value was defined as

$$e(n) = S(n) - \sum_{i=1}^{P} a_i S(n-i) \qquad (4.1)$$

By defining $a_o = 1$, the above equation can be written in Z-transform notation as

$$E(Z) = S(Z)\, H(Z) \qquad (4.2)$$

where

$$H(Z) = 1 - \sum_{i=1}^{P} a_i Z^{-i}$$

H(Z) is defined as the vocal tract analysis model and can be interpreted as an all zero filter of the P^{th} order.

Fant has developed a very detailed model of the human vocal tract which is described as a time varying all pole filter (Ref 11:5-8). The filter is time varying since it must model the changes in the vocal tract which are made to produce different sounds. Fant has shown, however, that the vocal tract and, therefore, the model filter pole locations remain stationary for a period of 15-20 milliseconds during speech production. Thus H(z), the analysis model, can be interpreted as the inverse of the vocal tract model described by Fant, and must be updated every 15-20 milliseconds. The updating of the analysis model is required in order to account for the changes in the time domain characteristics of speech caused by changes in the vocal tract.

As described in Chapter II, the digitized data base was sectioned into 128-point rectangular windows for detection of the baseline words. The 128-point sections represent a 16 millisecond interval of speech and, therefore, a stationary analysis model can be developed for each section. Markel and Gray have shown that to preserve the spectral properties of speech when using linear prediction analysis, a tapered window should be applied to each section of speech (Ref 11:157). A Hamming window of the form shown in Equation 4.3 was used.

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) , \quad 0 \leq n \leq N-1 \qquad (4.3)$$

To analyze an entire word required the development of a collection of analysis models, each representing a separate 16 millisecond segment of the word. The collection of analysis

models describing the characteristics of the word formed the basis for all four distance measures.

Distance Measure 1

Distance Measure 1, DM1, is based on the ratio of the total squared error (TSE) produced by passing a baseline word through its analysis model and the TSE produced by passing a distorted version of the word through the same model. Figure 4 illustrates the block diagram description of DM1. S(n) is a 128-point segment of speech and S'(n) is the same speech segment corrupted by some type of distortion. As each new segment of speech is to be analyzed, the baseline speech signal is analyzed to determine the linear prediction coefficients that define the analysis model. DM1 is calculated for each 128-point section of the word and averaged over all word sections to produce an average distance measure for the word. DM1 is defined in Equation 4.4.

$$DM1 = \frac{1}{NS} \sum_{k=1}^{NS} \left[\sum_{n=0}^{128} \frac{[e_k(n)]^2}{[e'_k(n)]^2} \right]^{1/2} \qquad (4.4)$$

where

NS = the number of 128-point speech segments in the word being analyzed

' denotes a quantity associated with the noise corrupted word

The masimum value of DM1 is 1.0 and can occur only when S(n) and S'(n) are identical. As the distortion present in S'(n) is increased, the TSE produced by S'(n) should increase

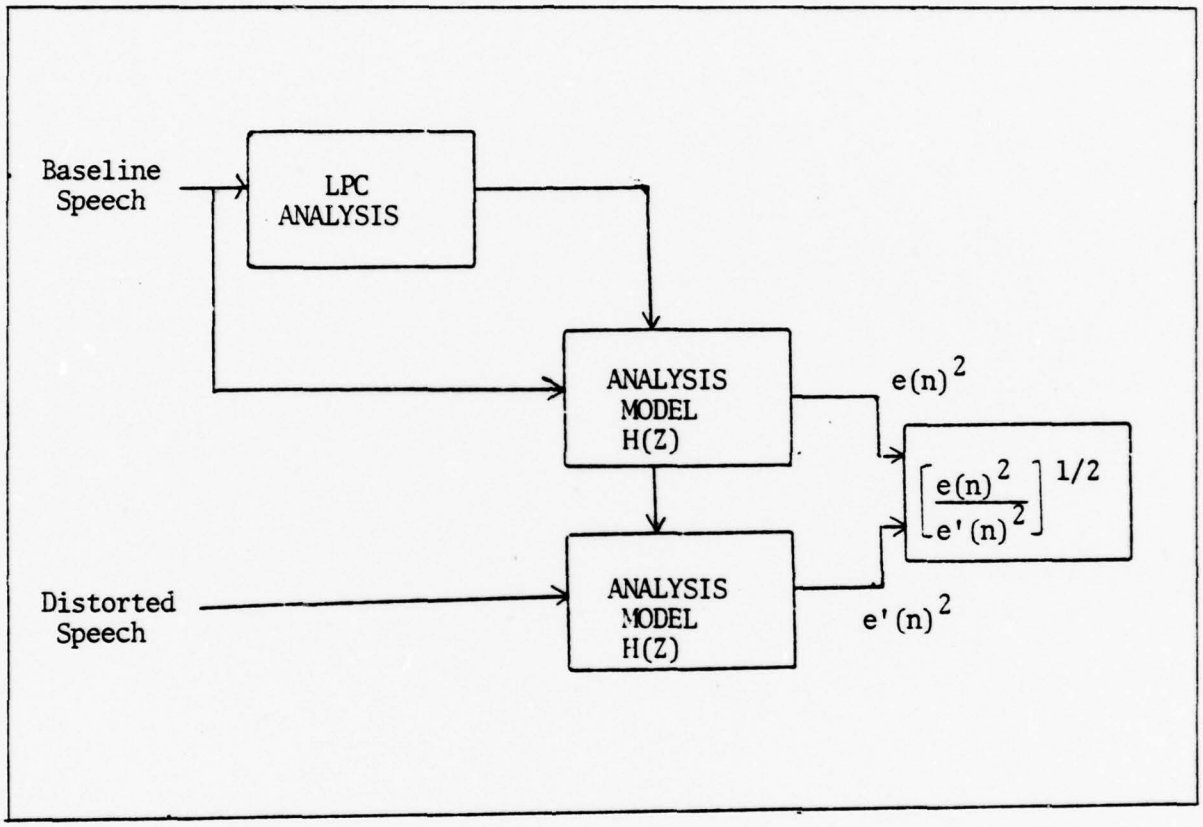


Fig 4. Calculation of Distance Measure 1

and, therefore, decrease the value of DM1.

Distance Measure 2

Distance Measure 2, DM2, is identical to DM1 with the exception that the ratio of the total squared errors of the two speech sections is normalized, based on the sum of the squares of the speech samples. DM2 is defined in Equation 4.5.

$$DM2 = \frac{1}{NS} \sum_{k=1}^{NS} \frac{R0'(k)}{R0(k)} \sum_{n=0}^{128} \frac{[e(n)]^2}{[e'(n)]^2} \qquad (4.5)$$

where

R0(k) is the sum of the squared sample values in data section k of the master word

R0'(k) is the sum of the squared sample values in data section k of the distorted word

Distance Measure 3

Distance Measure 3, DM3, is based on an intelligibility measure developed by Hartman (Ref 8). A block diagram illustrating the signal processing for DM3 is shown in Figure 5. Hartman's measure is based on the ratio of total squared error between two data sequences but, as opposed to DM1 and DM2, the measure is calculated from the short-term autocorrelation and LPC coefficients of the two signals, instead of creating an analysis model and processing both signals. DM3 is based on four quantities: E, E', D, and D', which are defined below.

E Minimum total squared error calculated from Levinson's algorithm by analyzing an undistorted section of speech

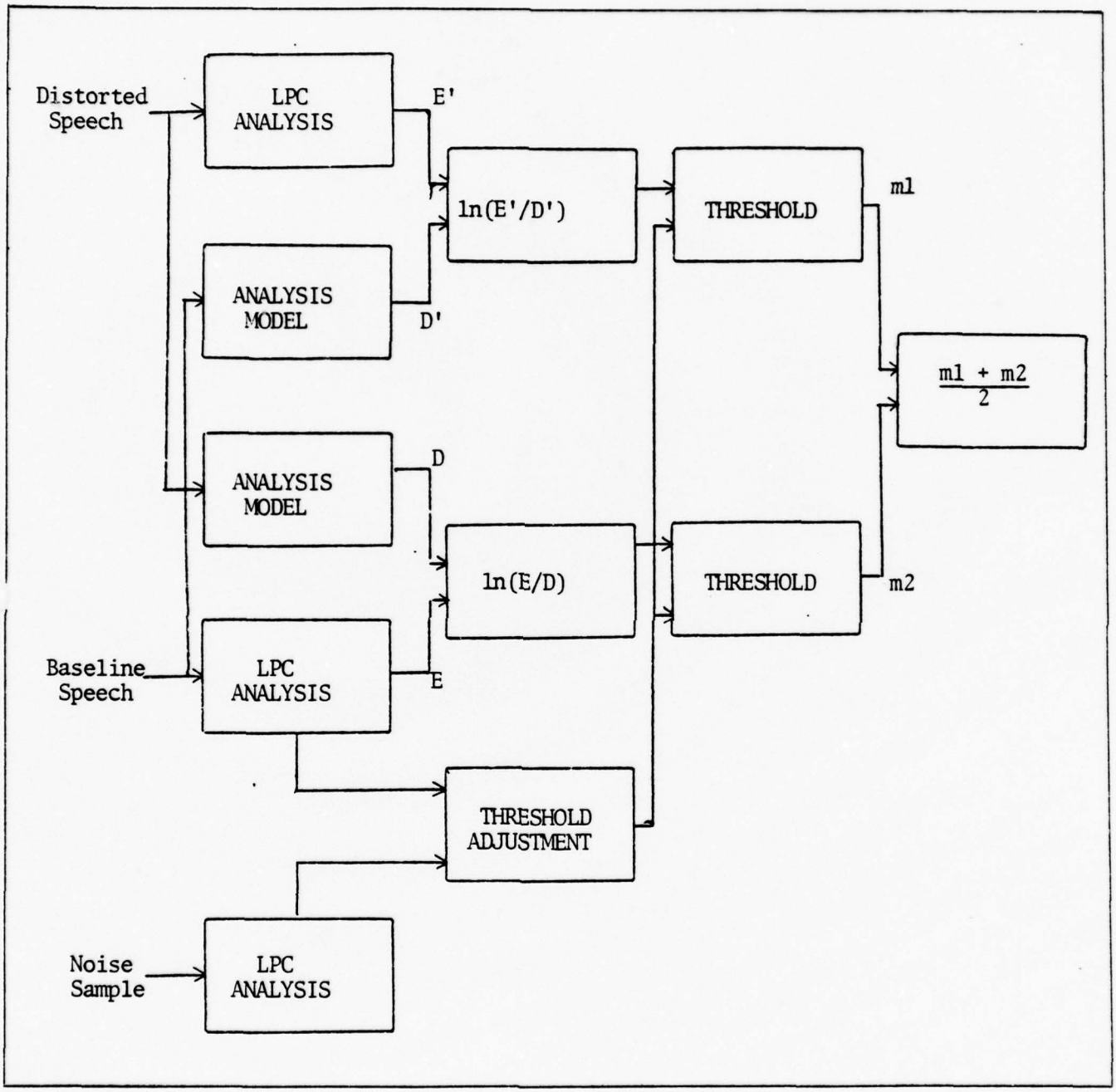


Fig 5. Calculation of Distance Measure 3

E' Minimum total squared error calculated from Levinson's algorithm by analyzing a distorted section of speech

D Total squared error created by comparing an undistorted section of speech with the predictor coefficients calculated from the corresponding section of the distorted speech

D' Total squared error created by comparing a distorted section of speech with the predictor coefficients calculated from the corresponding section of undistorted speech

All four quantities can be expressed in matrix notation by the following equations.

$$E = \underline{A}^T \underline{\underline{R}} \underline{A} \tag{4.6}$$

$$E' = \underline{A}'^T \underline{\underline{R}}' \underline{A}' \tag{4.7}$$

$$D = \underline{A}'^T \underline{\underline{R}} \underline{A}' \tag{4.8}$$

$$D' = \underline{A}^T \underline{\underline{R}}' \underline{A}$$

where

primed quantities indicate the value associated with distorted speech

T denotes the transpose of a vector

$\underline{A}^T$ is a P+1 element vector of predictor coefficients

$\underline{A}^T = (1, -a_1, -a_2, \ldots, -a_p)$

$\underline{\underline{R}}$ is a P+1 by P+1 matrix of short-term autocorrelation value

$R_{ij} = R(|i-j|)$

As stated earlier, E and E' are a direct result of Levinson's algorithm, but D and D' must be calculated using

Equations 4.8 and 4.9. However, computational time for D and D' can be saved by exploiting the symmetrical properties of the autocorrelation matrices. The $\underline{\underline{R}}$ matrix is structured such that the elements of each diagonal are equal so that D and D' can be calculated from Equations 4.10 and 4.11.

$$D = \sum_{i=0}^{P} g'(i)r(i) \tag{4.10}$$

$$D' = \sum_{i=0}^{P} g(i)r'(i) \tag{4.11}$$

where

$$g(i) = 2 \cdot \sum_{k=0}^{P} \alpha_k \alpha_{k+i} \quad , \; i=1, 2, \ldots P$$

$$g(0) = \sum_{k=0}^{P} \alpha_k^2$$

$$g'(i) = 2 \cdot \sum_{k=0}^{P} \alpha'_k \alpha'_{k+i} \quad , \; i=1, 2, \ldots P$$

$$g'(0) = \sum_{k=0}^{P} \alpha'^2_k$$

α_k and α'_k are the k^{th} elements of the $\underline{A}$ and $\underline{A}'$ vectors of predictor coefficients. Two distance measures are defined based on the ratios D'/E' and D/E.

$$E1 = \ln (D'/E') \tag{4.12}$$

$$E2 = \ln (D/E) \tag{4.13}$$

where ln denotes the natural logarithm.

To facilitate the comparison of E1 and E2 to the subjective scores, thresholds are established for E1 and E2 and the thresholded quantities are scaled to a range of 0 to 1. A value of 1 indicates that the speech is completely understandable, while a value of 0 indicates complete misunderstanding. Hartman established thresholds for determining whether a segment of speech is completely understood or completely misunderstood based on the work of Flanagan (Ref 7) and Sabur and Jayant (Ref 12). Initially, the thresholds were set so that a value for E1 or E2 greater than 2.46 indicated the speech was totally unintelligible. At the other end of the scale, a value less than 0.82 indicated that the speech was completely intelligible. An intelligibility metric was, therefore, defined by

$$\text{METRIC1} \begin{cases} = 1, \text{ if } E1 < 0.82 & \text{completely intelligible} \\ = 0, \text{ if } E1 > 2.46 & \text{completely unintelligible} \\ = \dfrac{2.46 - E1}{2.46 - 0.82}, & \text{otherwise} \end{cases} \quad (4.14)$$

$$\text{METRIC2} \begin{cases} = 1, \text{ if } E2 < 0.82 & \text{completely intelligible} \\ = 0, \text{ if } E2 > 2.46 & \text{completely unintelligible} \\ = \dfrac{2.46 - E2}{2.46 - 0.82}, & \text{otherwise} \end{cases} \quad (4.15)$$

Hartman found that the establishment of fixed thresholds proved unsatisfactory for predicting intelligibility and he, therefore, suggested that the thresholds be adjusted depending on the character of the noise present. In determining the threshold adjustments, eight 128-point segments of noise were taken from the data tapes and LPC analysis performed on the

samples. For the noise segments with the largest and smallest values of the sum of the squares of predictor coefficients, E1N and E2N were calculated as defined below

$$E1NH = \ln(D'/E') \quad \text{calculated for noise segment with} \tag{4.16}$$

$$E2NH = \ln(D/E) \quad \text{largest sum of squares} \tag{4.17}$$

$$E1NL = \ln(D'/E') \quad \text{calculated for noise segment with} \tag{4.18}$$

$$E2NL = \ln(D/E) \quad \text{smallest sum of squares} \tag{4.19}$$

where primed quantities in these equations indicate values associated with noise samples, and unprimed quantities indicate values associates with samples of the baseline word. E1NH and E1NL are averaged to produce E1N, and E2NH and E2NL are averaged to produce E2N. E1N and E2N are calculated for each segment of speech to be analyzed and the intelligibility thresholds are adjusted according to the following equations.

$$T1E1 = 0.82$$
$$T2E1 = 2.46 \quad \text{if } E1N < 2.46 \tag{4.20}$$

$$T1E1 = 0.82 + 0.82(E1N - 2.46)$$
$$T2E1 = 2.46 + 0.82(E1N - 2.46) \quad \text{if } E1N > 2.46 \tag{4.21}$$

$$T1E2 = 0.82$$
$$T2E2 = 2.46 \quad \text{if } E2N < 2.46 \tag{4.22}$$

$$T1E2 = 0.82 + 0.82(E2N - 2.46)$$
$$T2E2 = 2.46 + 0.82(E2N - 2.46) \quad \text{if } E2N > 2.46 \tag{4.23}$$

With the adjustment of the thresholds based on the character of the noise, the two metrics as defined in Equations 4.14 and 4.15 were modified as shown next.

$$\text{METRIC1} \begin{cases} = 1, & \text{if } E1 < T1E1 \\ = 0, & \text{if } E1 > T2E1 \\ = \dfrac{T2E1 - E1}{T2E1 - T1E1}, & \text{otherwise} \end{cases} \tag{4.24}$$

$$\text{METRIC2} \begin{cases} = 1, & \text{if } E2 < T1E2 \\ = 0, & \text{if } E2 > T2E2 \\ = \dfrac{T2E2 - E2}{T2E2 - T1E2}, & \text{otherwise} \end{cases} \tag{4.25}$$

Modified thresholds were established for each segment of speech to be analyzed and METRIC1 and METRIC2 calculated. METRIC1 and METRIC2 were averaged over all segments of the word being analyzed to produce $\overline{\text{METRIC1}}$ and $\overline{\text{METRIC2}}$. $\overline{\text{METRIC1}}$ and $\overline{\text{METRIC2}}$ are then averaged to produce DM3.

$$\overline{\text{METRIC1}} = \frac{1}{NS} \sum_{k=1}^{NS} \text{METRIC1}\ (k) \tag{4.26}$$

$$\overline{\text{METRIC2}} = \frac{1}{NS} \sum_{k=1}^{NS} \text{METRIC2}\ (k) \tag{4.27}$$

$$\text{DM3} = \frac{\overline{\text{METRIC1}} + \overline{\text{METRIC2}}}{2} \tag{4.28}$$

where

NS is the number of 128-point speech segments contained in the word being analyzed.

Distance Measure 4

Distance Measure 4, DM4, is identical to DM3 except that $\overline{\text{METRIC1}}$ and $\overline{\text{METRIC2}}$ are a weighted average based on the distribution of signal power in the word being analyzed. The average power, $\overline{RO}$, in the word being analyzed is calculated

by computing the average sample squared value of the undistorted word. Values obtained for METRIC1 and METRIC2 when the average signal power in the segment of speech being analyzed was greater than $\overline{RO}/2$ were averaged to produce $\overline{M1H}$ and $\overline{M2H}$, respectively. Conversely, values for METRIC1 and METRIC2 for segments with an average signal power less than $\overline{RO}/2$ were averaged to yield $\overline{M1L}$ and $\overline{M2L}$. DM4 was then defined by

$$DM4 = \frac{\overline{M1} + \overline{M2}}{2} \tag{4.29}$$

where

$$\overline{M1} = \frac{\overline{M1H} + \overline{M1L}}{2}$$

$$\overline{M2} = \frac{\overline{M2H} + \overline{M2L}}{2}$$

In general, only about one-third of the speech segments will have an average power less than $\overline{RO}/2$. Therefore, the weighted average used to calculate DM4 has the effect of emphasizing the intelligibility of low power segments more than high power segments.

V. Results

As mentioned in Chapter II, the data base used in this study had been subjectively scored by a listener group of ten people. Each listener was given a score sheet containing each pair of rhyming words from the master lists. As the listener heard each word of the data (noise corrupted) list, he marked which word of the master pair he perceived was said. A subjective score was developed for each data list by calculating the percent of right answers on each listener's score sheet. The average subjective score for the listener group was used as the subjective measure of speech intelligibility. The results of the subjective scoring for each word list are shown in Table II.

TABLE II

Subjective Scores

Jamming Level	1	2	3	4	5	6	7
Subjective Score	90	92	93	84	84	86	79

The subjective measure for each word list was the standard by which the effectiveness of the LPC based measures and the Articulation Index were judged.

Articulation Index

The Articulation Index (AI) was calculated using the one-third octave band method (Ref 2:11-15). This method differs

slightly from the standard 20-band method described in Chapter I, but can be more easily calculated in an automated manner using existing audio test equipment. Table III illustrates the measurements and calculations involved in the one-third octave method.

A Bruel and Kjaer Digital Frequency Analyzer Type 2131 was used to measure the signal power in each of the one-third octave bands, and a Sony Model TC-850 tape recorder was used for analog data input to the analyzer. A Hewlett-Packard 9845A mini-computer was used to control the frequency analyzer and to calculate the Articulation Index. The steps involved in computing the AI are detailed below.

Step 1. A baseline analog tape was played into the frequency analyzer. The analyzer computed the average power in each of the one-third octave bands over a period of 128 seconds.

Step 2. At the end of 128 seconds, the mini-computer sampled the average power figures for each of the frequency bands and stored the results (Column 2 of Table III).

Step 3. A noise corrupted data tape was played through the frequency analyzer and the average power calculated in each band over a 128 second interval.

Step 4. The mini-computer sampled the average power figures for the noisy tape and stored the results (Column 3 of Table III).

Step 5. The average noise power in each of the bands was calculated by subtracting the baseline signal power from the signal plus noise power (Column 4 of Table III).

Step 6. The signal-to-noise ratio (SNR) was calculated for

TABLE III

One-Third Octave Band Method for Computing Articulation Index

COLUMN 1 One-Third Octave Band (Hz)	COLUMN 2 Average Speech Power (Watts)	COLUMN 3 Average Speech Plus Noise Power (Watts)	COLUMN 4 Average Noise Power (Watts)	COLUMN 5 Average SNR (dB)	Column 6 Adjusted SNR (dB)	COLUMN 7 Weight	COLUMN 7 Value
180-224						.0004	
224-280						.0010	
280-355						.0010	
355-450						.0014	
450-560						.0014	
560-710						.0020	
710-900						.0020	
900-1120						.0024	
1120-1400						.0030	
1400-1800						.0037	
1800-2240						.0038	
2240-2800						.0034	
2800-3550						.0034	
3550-4500						.0024	
4500-5600						.0020	
						ARTICULATION INDEX _____	

each frequency band (Column 5 of Table III).

Step 7. The SNR value for each band was adjusted according to Equation 5.1.

$$\begin{aligned} &\text{SNR} = 30, \quad \text{if SNR} > 30 \\ &\text{SNR} = 0, \quad \text{if SNR} < 0 \\ &\text{Otherwise no adjustment is made} \end{aligned} \tag{5.1}$$

Step 8. The adjusted SNR for each band was multiplied by the weighting factor shown in Column 6 and the values for each band summed to produce the Articulation Index.

TABLE IV

Articulation Index

Jamming Level	1	2	3	4	5	6	7
Articulation Index	.49	.50	.52	.44	.37	.37	.36

The Articulation Index was calculated for each of the noise corrupted tapes. A scatter plot comparing the AI to the subjective scores is shown in Figure 6. The correlation between the AI and the subjective scores was computed using Equation 5.2.

$$C = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{[\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2]^{1/2}} \tag{5.2}$$

where

X_i denotes the ith value of the Articulation Index

$\bar{X}_i$ is the mean value of the Articulation Index

Y_i denotes the ith subjective score

$\bar{Y}_i$ is the mean value of the subjective scores

LPC Measures

Computer programs were developed for computing each of the LPC based distance measures. A Xerox Sigma 7 computer was used for exercising each of the measures on the data base. Figures 7 through 10 are scatter plots comparing each of the measures to the subjective scores. As with the Articulation Index, the correlation between the LPC measures and the subjective scores is shown.

TABLE V

LPC Distance Measures

Jamming Level	1	2	3	4	5	6	7
DM1	.47	.09	.17		.07	.05	.05
DM2	.27	.10	.16	.10	.11	.09	.10
DM3	.96	.89	.93	.62	.88	.85	.94
DM4	.96	.89	.94	.63	.88	.87	.95

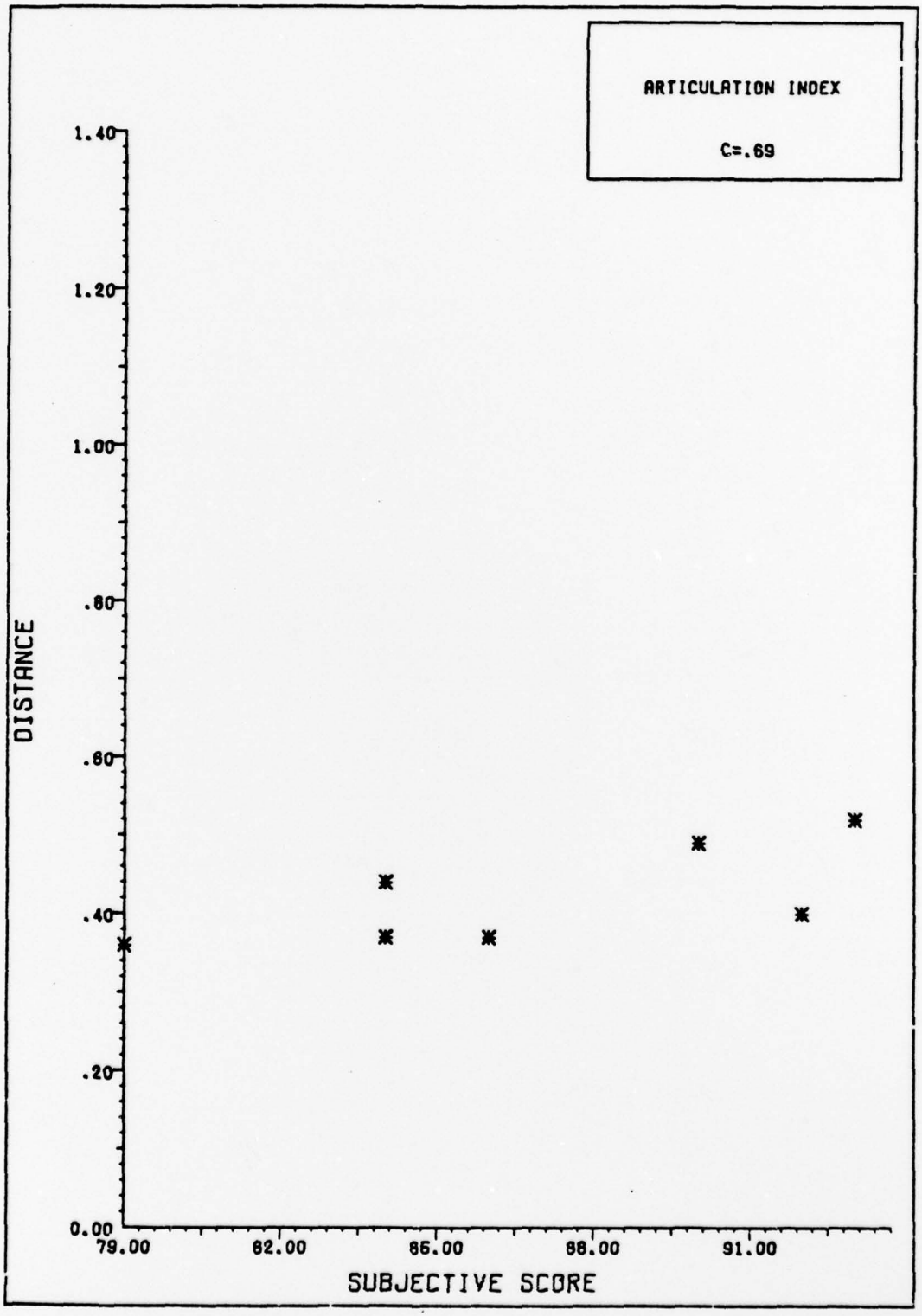


Fig 6. Articulation Index

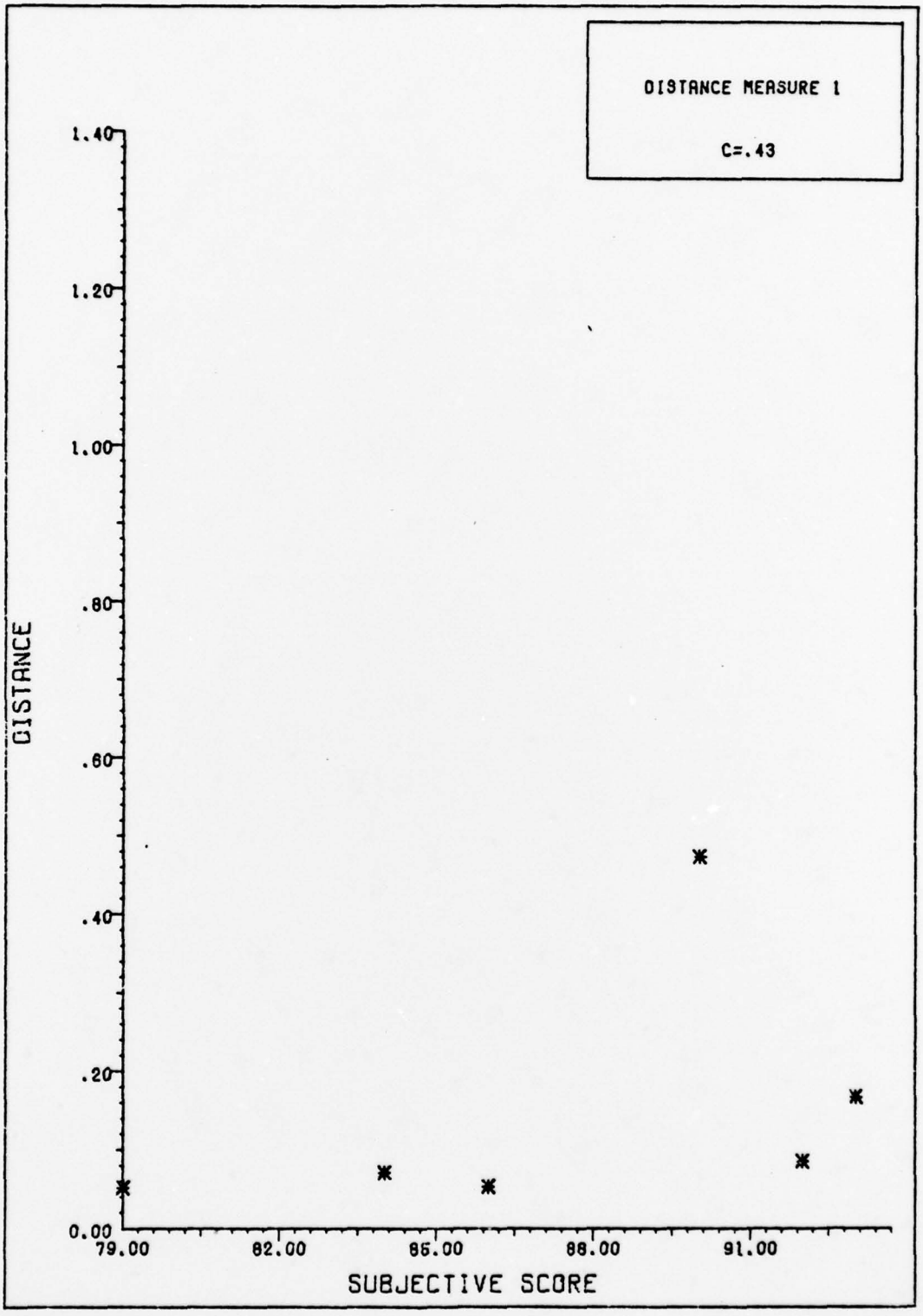


Fig 7. Distance Measure 1

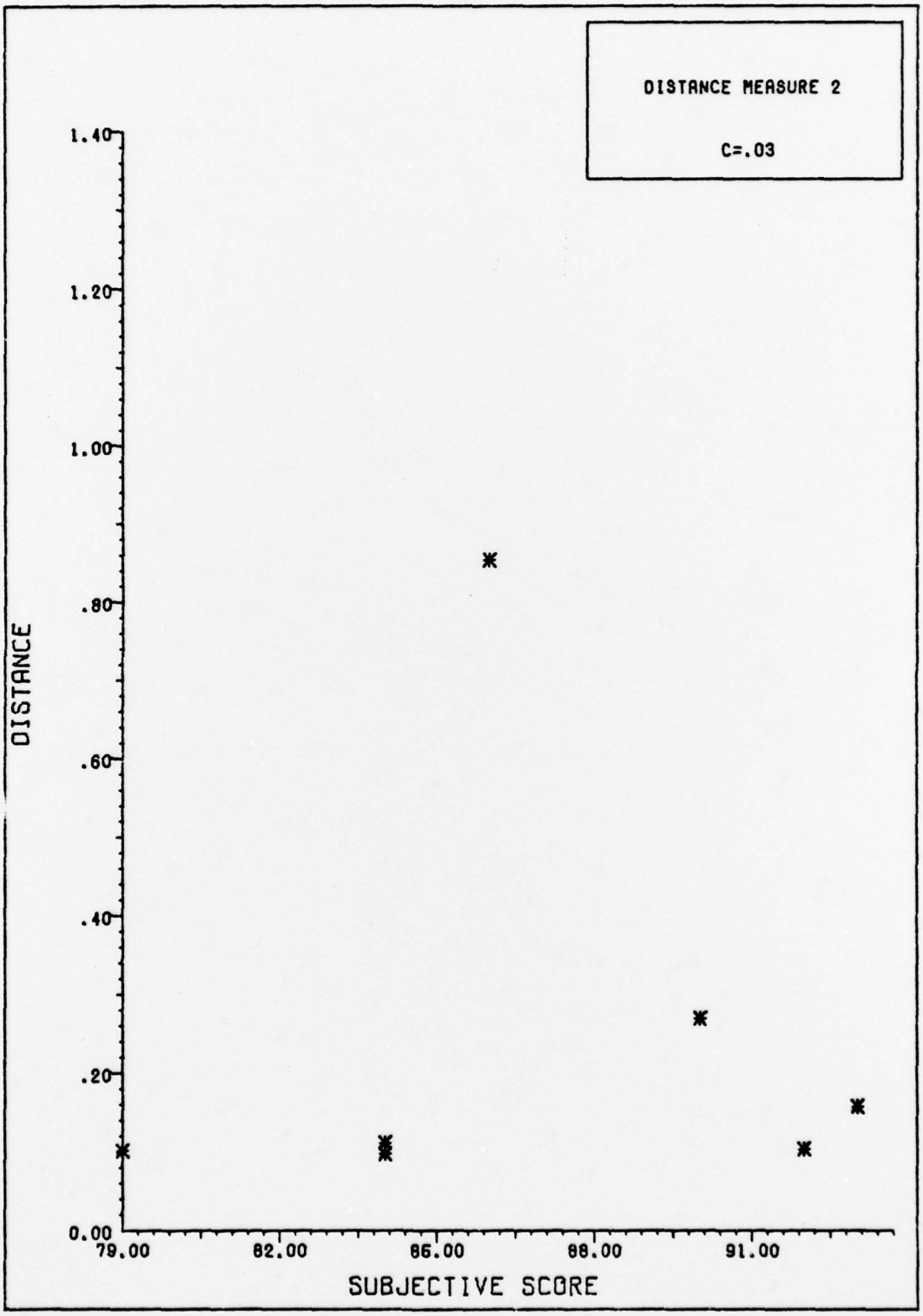


Fig 8. Distance Measure 2

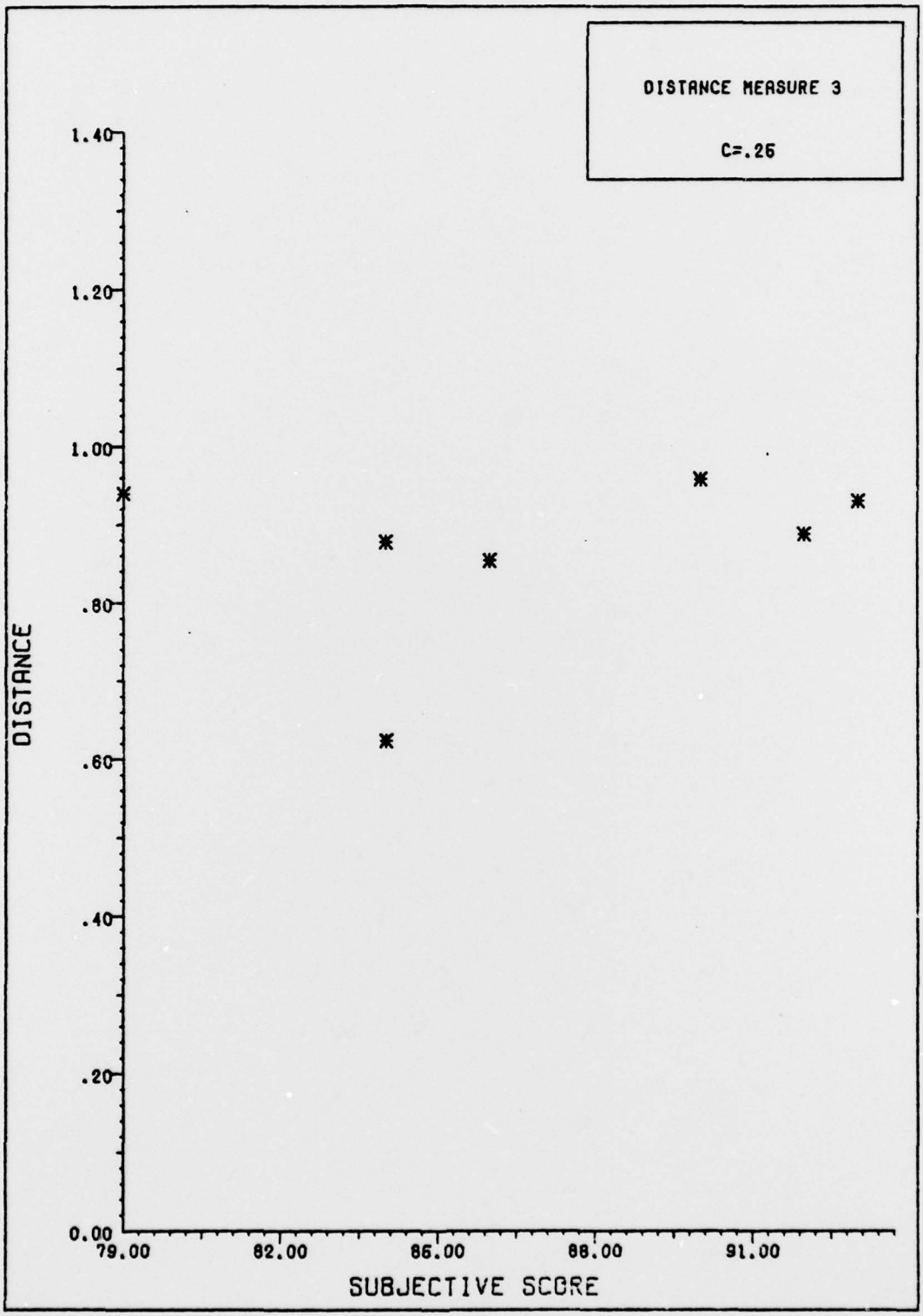


Fig 9. Distance Measure 3

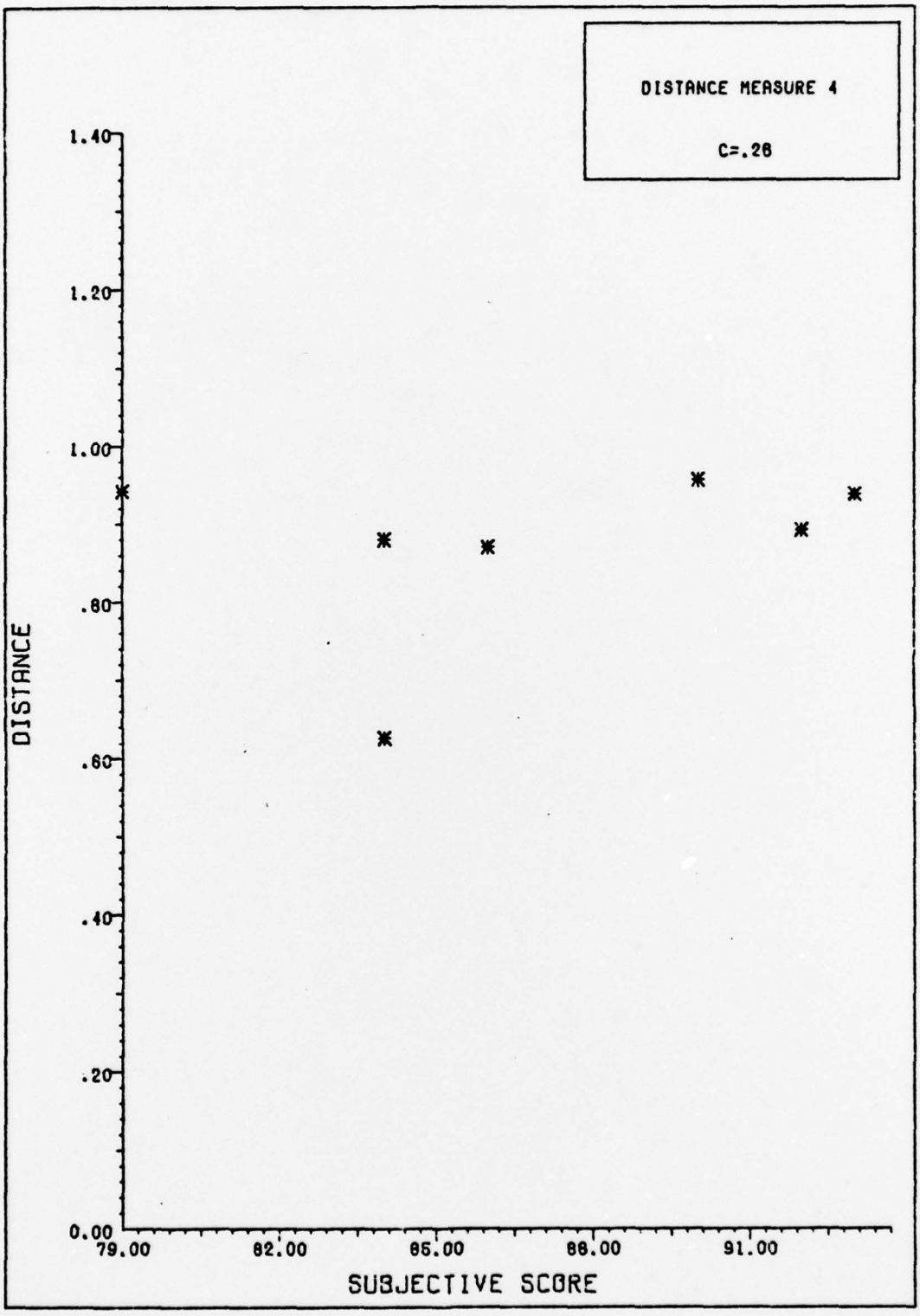


Fig 10. Distance Measure 4

VI. Conclusions and Recommendations

Conclusions

No LPC based distance measure produced a correlation with the subjective scores that can be considered significant for a valid distance measure. As mentioned in Chapter II, two reasons may explain the failure of the LPC based measures.

First, inadequate tape alignment procedures may have caused the noise corrupted word samples to have been shifted relative to the corresponding baseline word samples. However, as stated before, the intent of this research was to evaluate the use of LPC based measures under conditions similar to those under which the Articulation Index is presently used. The use of timing marks on the analog data tapes can be considered a reasonable and realistic way of providing tape synchronization under normal data collection conditions. The failure of LPC measures solely because of tape alignment does not invalidate the findings of this report, but rather reinforces the claim that LPC measures require synchronization in excess of what can be realistically provided under field use outside the laboratory. Hartman reported that software implementation of his LPC based measures resulted in 70% of the computer time and 85% of the manpower requirements being directly attributable to the data alignment procedures. Unless LPC based measures can be proven to be considerably more effective than the Articulation Index, the extensive overhead associated with data

alignment makes the LPC measures impractical for field use.

The second reason the LPC measures may have failed is that LPC measures may be incapable of predicting speech intelligibility in the presence of the distortion types used here. Computation of the Articulation Index did provide some insight into the differences between the baseline speech signal spectrum and the noise corrupted data that may indicate the types of distortion present. The digital frequency analyzer indicated that in the frequency range 0-355 Hz, the baseline speech signal contained more power than the baseline signal plus noise as output from the receiver. Two possible explanations may account for this inconsistency in power spectrums. First, spread spectrum communications involve spreading a relatively narrowband signal (4 kHz) into a relatively wideband signal (30 mHz). The spreading of the signal, in addition to any non-linear processing within the transmitter/receiver pair, could cause a frequency translation in the speech signal. Second, the use of high pass filters, such as in pre-emphasis, would have the effect of decreasing the power in the lower frequencies of the transmitted speech signal. Since the baseline speech signal represents speech before it enters the transmitter, it is conceivable that the low frequency spectrum of the baseline signal would be greater than the received version of the baseline signal plus noise. Hartman showed that like the AI, LPC intelligibility measures compare the frequency spectrum of two signals (Ref 8:29). This indicates that if there is frequency distortion present

in the signal to be measured for intelligibility, the accuracy of LPC measures is in doubt. It is interesting to note, however, that the AI apparently predicts intelligibility in the presence of frequency distortion of the kind evident in this experiment.

The Articulation Index proved to be significantly more efficient to compute in terms of manhours and equipment as compared to the LPC based measures. The inefficiency of LPC measures is primarily due to the synchronization requirements and the present lack of commercially available LPC hardware. Additionally, the AI provided much more accurate results. Since the LPC based measures, like the AI, are based on the comparison of the frequency spectrum of two signals, a performance advantage of LPC measures over the AI is doubtful. Interestingly, both studies of LPC measures mentioned in this report, References 4 and 8, failed to compare LPC measures with the Articulation Index. Unless a clear performance advantage of LPC based measures over the AI can be proved, the continuation of research measuring speech intelligibility using LPC measures is questionable. Once the superiority of LPC measures is proven, work must be done on developing an efficient LPC based system which can function with limited overhead and under field conditions such as real-time testing of voice communications channels.

Recommendations

As mentioned in the introduction to Chapter I, a real need exists within the military for an efficient and effective

automated speech intelligibility measure. The most successful automated measure presently used by the military is the Articulation Index. Research into the use of LPC for intelligibility measures must first examine the sensitivity of LPC calculations to data alignment. If adequate alignment techniques can be developed and proven useable under field conditions, the further development of LPC measured is warranted.

The most pressing need identified during the conduct of this research is the need for an extensive data base of testable speech received over an actual communications channel in the presence of common channel distortions. To date, the majority of data bases are extremely small and involve distortions simulated in the laboratory or by computer, rather than real word distortions. Once such a data base can be created, an extremely useful and valuable tool will be present for judging the relative merits of automated speech intelligibility measures.

Bibliography

1. Ablett, Charles B. Measuring Speech Intelligibility During the EW/CAS Test. Final Technical Report. Contract Number F6000-77-90133 (Task 2). Kalman Sciences Corporation, Colorado Springs, Colo., 10 October 1977.

2. Acoustical Society of America. American National Standard: Methods for the Calculation of the Articulation Index. ANSI S3.5-1969. New York: American National Standards Institute, Inc., 1970.

3. Atal, B.S., and S.L. Hanauer. "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave." The Journal of the Acoustical Society of America, 50:637-55 (August 1971).

4. Barnwell, T.P., and A.M. Bush. Speech Quality Measurement. Final Report, E21-655-77-TB-1, Defense Communications Agency. Georgia Institute of Technology, December 1977.

5. Bauer, John E. Evaluation of Two Sampled Data Communications Systems. Thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December 1977.

6. Beeson, Wayne R. An Algorithm for Determining Speech Intelligibility. Thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December 1977.

7. Flanagan, J.L. Speech Analysis, Synthesis, and Perception. New York: Academic Press, Inc., 1965.

8. Gamauf, K.J., and W.J. Hartman. Objective Measurement of Voice Channel Intelligibility. Final Report, FAA-RD-77-153. Institute for Telecommunications Sciences, Department of Commerce, Boulder, Colo., October 1977.

9. Hartman, W.J., and S.F. Bell. Voice Channel Objective Evaluation Using Linear Predictive Coding. Final Report, FAA-RD-75-189. Institute for Telecommunications Sciences, Dept. of Commerce, Boulder, Colo., October 1977.

10. Makhoul, J. "Linear Prediction: A Tutorial Review," Proceedings of the IEEE, 63:561-580 (April 1975).

11. Markel, John D., and A.H. Gray. Linear Prediction of Speech. New York: Springer-Verlag, 1976.

12. Sambur, M.R., and N.S. Jayant. "LPC Analysis/Synthesis Inputs Containing Quantizing Noise or Additive White Noise," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol 24, No. 6. December 1976.

13. Wiener, Norbert. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. New York: John Wiley and Sons, Inc., 1949.

Vita

Donald M. Ottinger, Jr. was born on 14 July 1951 in Thorntown, Indiana. He graduated from high school in Thorntown, Indiana in 1969 and attended the U.S. Air Force Academy from which he received the degree of Bachelor of Science in Engineering Mechanics and a regular commission in the USAF. He completed the USAF communications officer course in April 1974 and then served as a communications systems officer in the 728 Tactical Control Squadron and 507 Tactical Air Control Wing, Shaw AFB, S.C. He entered the School of Engineering, AFIT, in December 1977.

Permanent Address: 508 S. Market St.
Thorntown, Indiana
46071

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/GE/EE/78-35	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Objective Measure of Speech Intelligibility Using Linear Predictive Coding		5. TYPE OF REPORT & PERIOD COVERED MS Thesis 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Donald M. Ottinger, Jr., Capt. USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology AFIT/EN Wright-Patterson AFB, Ohio 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS HQ AFCS/OA Scott AFB IL 62225		12. REPORT DATE December 1978 13. NUMBER OF PAGES 55
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

Approved for public release IAW AFR 190-17
J.P. Hipp, Major, USAF
Director of Information 1-23-79

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Speech
Speech Intelligibility
Linear Predictive Coding

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Four distince measures of speech intelligibility based on linear predictive coding (LPC) are developed and evaluated. The data base used for evaluating the measures consisted of lists of 58 words from Diagnostic Rhyme Test IV. The lists were transmitted over a spread spectrum radio communications channel and subjected to 7 different levels of non-white, non-Gaussian jamming noise. The lists were all scored subjectively

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

for intelligibility by a trained listener panel. The subjective scores were used to judge the effectiveness of the four distance measures.

The Articulation Index was also calculated for each of the word lists and compared to the LPC measures as to effectiveness and efficiency in measuring speech intelligibility. The Articulation Index was significantly more effective than the LPC measures. The best LPC measure provided 42% correlation with the subjective scores. The Articulation Index provided 69% correlation. The overhead associated with data tape alignment and parameter computation makes LPC measures extremely inefficient as compared to the Articulation Index.