17618-H174-R0-00

PROJECT TECHNICAL REPORT

Task 707

NEW CRITERIA DEVELOPMENT

NAS 9-8166 29 June 1971

Prepared for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

Prepared by

Communications and Sensor Systems Department
Electronics Systems Laboratory

TRW
SYSTEMS GROUP

17618-H174-R0-00

PROJECT TECHNICAL REPORT

Task 707

NEW CRITERIA DEVELOPMENT

NAS 9-8166

29 June 1971

Prepared for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

Prepared by

Communications and Sensor Systems Department
Electronics Systems Laboratory

TRW
SYSTEMS GROUP

CR-115079

17618-H174-R0-00

PROJECT TECHNICAL REPORT

Task 707

NEW CRITERIA DEVELOPMENT

NAS 9-8166

29 June 1971

Prepared for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

Prepared by

J. C. Miller

Approved by: W B Warren
W. B. Warren, Task Manager

Approved by: John DeVillier
John DeVillier, Manager
Communications and Sensor Systems
Department

TRW
SYSTEMS GROUP

CONTENTS

CONTENTS (Continued)

ILLUSTRATIONS

TABLES

1. INTRODUCTION

The purpose of this technical memorandum is to describe and evaluate several methods of specifying performance criteria for voice, video, and digital data systems. The selection of a specific criteria for use in evaluating a given system is highly dependent on the detailed nature of the system under test. Consequently, it has not been possible in the work described in this memorandum to make a definitive selection of criteria for use in each of the three general system categories considered. However, it has been possible to determine, for each category some general characteristics which are common to most systems considered. These characteristics were used to arrive at particular choices for test criteria in each system category. The method employed is to propose several candidate evaluation systems in each category, and make an evaluation of each system based on a number of parameters. These parameters include precision of test data, conciseness of results, required data reduction, pertinence to actual system performance, and ease of simulation. Each of these parameters is assigned a relative weighting value. In order to evaluate the candidate systems, the factors affecting each of the performance parameters of each of the candidate systems are listed, and a numerical rating value is assigned to each. By this procedure an overall score for each of the candidate systems is achieved which will provide the basis for recommendations concerning these methods of specifying performance criteria. Before these criteria are applied in the evaluation of a specific system, the numerical ratings assigned in the trade-off matrix should be re-examined to determine if the more detailed information available is sufficient to warrant the selection of a different test criteria.

2. METHODS OF SPECIFYING PERFORMANCE CRITERIA

Judgement of performance of a communication system depends upon a subjective evaluation in most cases (what is the message, how good is the picture, what is the data). Even though it is sometimes possible to express these evaluations in numbers (word intelligibility for voice systems, picture quality rating for video) it is usually simpler and easier to measure physical quantities and to attempt to determine a calibration relationship between these quantities and subjective ratings. Since the amplitude and frequency distribution of both the desired signal and noise introduced by the system under consideration have obvious bearing of the performance of the system, the idea of a signal-to-noise ratio is inherent in most criteria, either explicitly or implicitly. Because of the non-linear response of the human sensory organs to amplitude-frequency and amplitude-time relationships, and to the redundancy inherent in many systems, methods other than average signal-to-noise ratios are often employed. Methods listed in this memo include:

- Methods based on frequency considerations
  - Weighted signal concept (articulation indices for voice, equal importance bands for video)
  - Weighted noise concept (video SNR)
  - Mean squared difference of input and output spectra (voice and video)
- Methods based on amplitude-time considerations
  - Analog and digital methods of computing speech SNR
  - Cross-correlation techniques - (Voice and video)

Methods based on statistical counting of errors (digital data)

Bit-by-bit counting to determine BER

Extrapolation of pseudo-error rates

3. CANDIDATE CRITERIA SYSTEMS

Section 3 contains a brief description of the systems considered as means of specifying performance criteria for voice, video, and television. Most of them represent well established principles. An attempt has been made to suggest possible means of mechanization with the intention of simplifying test procedures and data reduction. At least one method, method B for television, represents an unproven principle - that video signals can be separated into frequency bands having equal subjective importance to the viewer, in a manner similar to the principle involved in determining the articulation index of a voice system.

The discussion of each candidate system includes a description of the method involved in mechanizing the test and computing the criteria, a simplified block diagram of the process, and a listing of the factors considered in assigning a numerical rating for each system for each of the five parameters used in Section 4 to compare the candidate systems.

Two of the candidate systems, Methods C and F for determining speech SNR for voice systems have been developed and tested under NASA MSC direction by the Philco-Ford corporation. (Reference 3 and 5).

3.1 Voice Criteria - Method A Determination of Articulation Index of Output

The use of the articulation index (AI) as a means of relating articulation testing to more easily measured physical quantities such as signal to noise ratios has been the subject of interest, research, and testing for a long period of time (Reference 1). Since it is based on the spectral distribution of both speech signal and noise, it tends to be a more accurate method of describing voice quality than an average signal-to-noise value taken over the entire voice band (Reference 1).

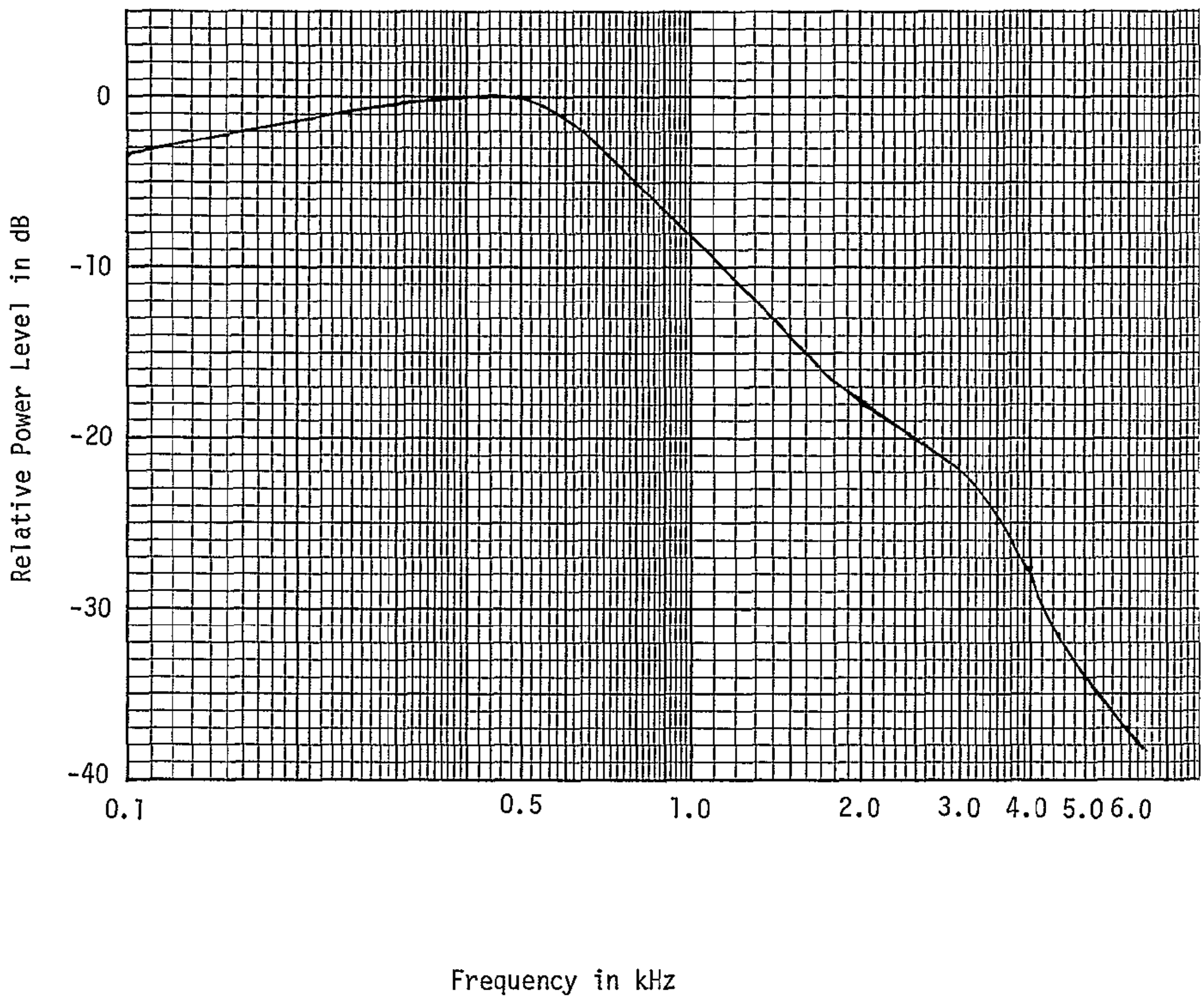
The concept of AI is based on experimental evidence which has shown that normal speech signals can be divided into a number of frequency bands such that the speech information in each band has an equal effect on the overall intelligibility (Reference 1). A list of these "equal importance" frequency bands is given in Table I. It has also been established that numbers representing the articulation values or each band can be considered as probabilities (Reference 10). Thus, the overall articulation can be represented as the product of the individual articulation values of each band. This fact allows us to define AI (in Step I of Method A given below) as the sum of the weighted signal-to-noise ratios of each band (w_i), since they are expressed as decibels.

Since the composition of the "equal importance" frequency bands was determined using real speech, any mechanization of the AI method using a substitute for real speech should employ weighting factors experimentally determined to be representative of real speech. Such a curve, shown in Figure 1, should be used in Method B, an alternate means of determining AI using discrete input frequencies.

Table 1

Equal Importance Bands for Normal Speech

Articulation Band Number	Frequency Range in Hz	Bandwidth in Hz
1	200 - 330	130
2	330 - 430	100
3	430 - 560	130
4	560 - 700	140
5	700 - 840	140
6	840 - 1000	160
7	1000 - 1150	150
8	1150 - 1310	160
9	1310 - 1480	170
10	1480 - 1660	180
11	1660 - 1830	170
12	1830 - 2020	190
13	2020 - 2240	220
14	2240 - 2500	260
15	2500 - 2820	320
16	2820 - 3200	380
17	3200 - 3650	450
18	3650 - 4250	600
19	4250 - 5050	800
20	5050 - 6100	1050



Normal Voice Spectrum

Figure 1

Method A described below is a method of mechanizing the testing to determine the AI of a voice communication system. A block diagram is shown in Figure 2.

3.1.1 Description of Method

A. The input tape is composed of a standard phonetically balanced word list scored for word intelligibility.

B. Modulation, RF source and RF link simulation are determined by the system under test. If the test is an end-to-end test of a complete link, the modulator and RF source would be replaced with components of the system under test.

C. The input tape signal is fed into a variable equalizer to match the magnitude, $|A(\omega)|$, and phase, $\phi(\omega)$, of the system under test.

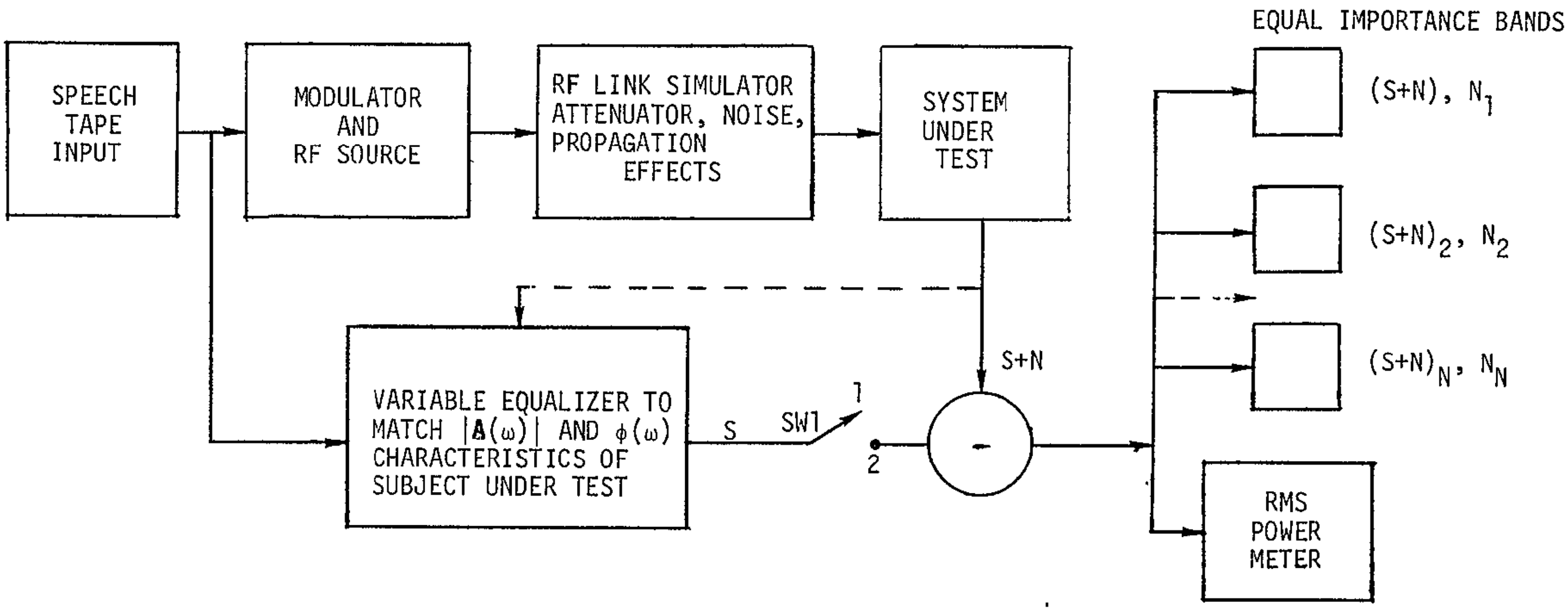
D. The output signal S is fed through switch S_1 to a difference network, to which the output S+N from the system under test is also applied.

E. In switch position 1, only the output S+N from the system under test is fed to the difference network, resulting in an output S+N.

F. In switch position 2, both the matched output S and the output S+N are fed to the difference network resulting in N, the system noise, as an output.

G. The output of the difference network is applied to an RMS meter to read the S+N and N values in order to compute an overall S/N ratio.

H. The output is also sent in parallel to 20 bandpass filters corresponding to the equal importance bands of the articulation index. The S/N ratio of each band is then computed, and the weighting functions w_i determined for each band as indicated below.



VOICE CRITERIA - METHOD A

ARTICULATION INDEX

Figure 2

The weighting functions w_i applied to each equal importance frequency band depend on the observation that if the SNR in a particular band is less than -12 dB, the effect of that band on the overall articulation index (AI) can be disregarded, and also that the effect of increasing the SNR in a particular band beyond +18 dB has no addition affect. For each band the values of SNR between -12 dB and +18 dB have a linear effect on AI, resulting in the values of w_i given below. Use of this method results in a value of AI which never exceeds 1 (corresponding to a word intelligibility score of 100%).

$$w_i = 0 \quad \text{if } (S/N)_i < -12 \text{ dB}$$

$$w_i = \frac{(S/N)_i + 12}{30} \quad \text{if } -12 \text{ dB} \leq (S/N)_i \leq 18 \text{ dB} \tag{1}$$

$$w_i = 1 \quad \text{if } (S/N)_i > 18 \text{ dB}$$

I. The articulation index (AI) is then computed

$$AI = \frac{1}{N} \sum_{i=1}^{N} w_i \tag{2}$$

where N = number of articulation bands in the pass band of the system under test. N = 20 for the band pass of the normal speech ranging from 200 to 6,100 Hz.

J. The AI is converted to word intelligibility (WI) by means of an empirical relationship determined by subjective testing. Such a relationship is controlled by the word list used in preparing the input tape. An example of this type of relationship is shown in Figure 3.

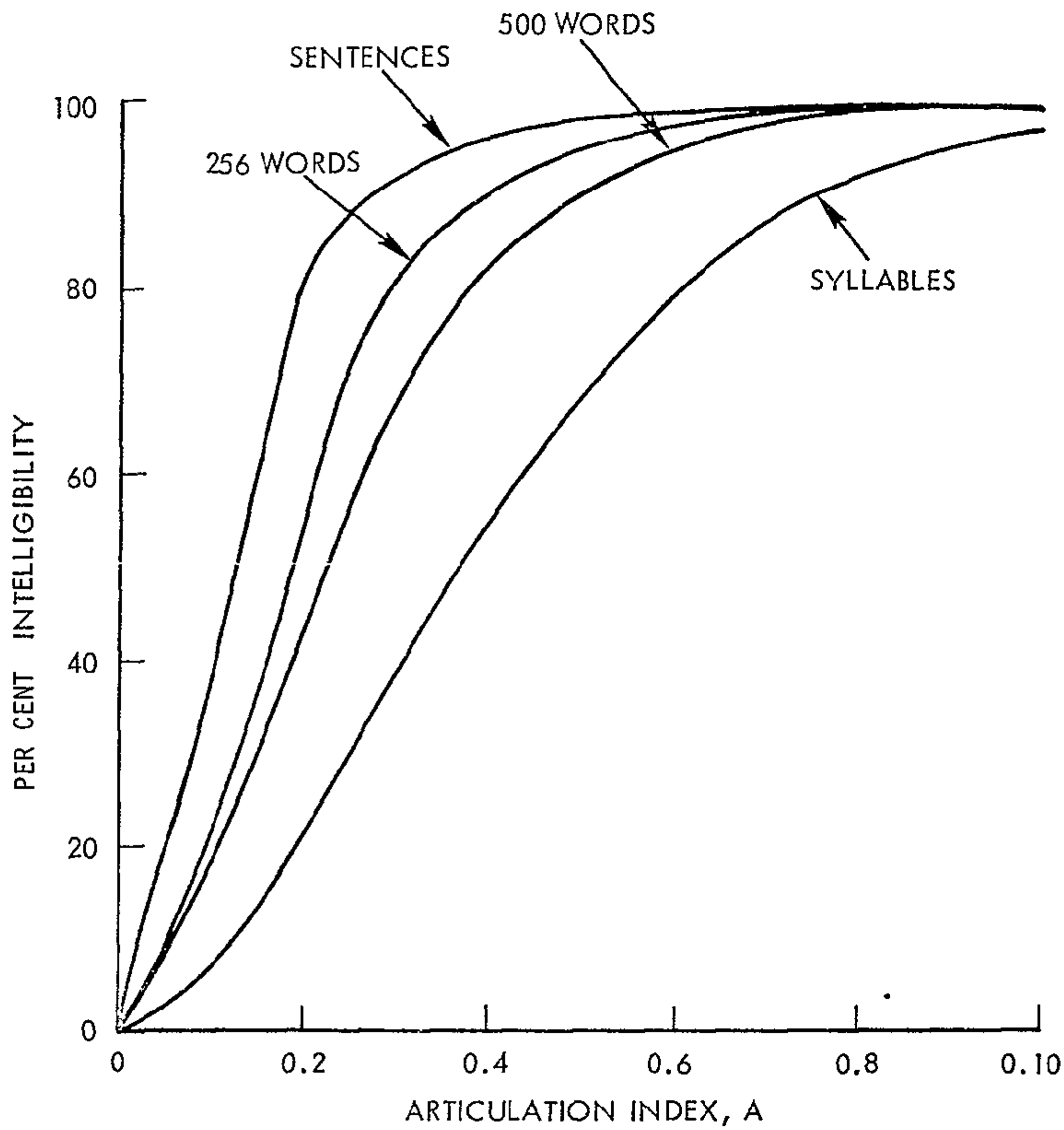


Figure 3. Intelligibility Versus Articulation Index

K. The overall S/N ratio computed in Step G could also be converted to WI by an appropriately developed empirical relationship as shown in Figure 4.

3.1.2 Discussion of Comparison Parameters

Parameter One - Precision of Test Data

The following factors affecting precision of data were considered in determining the rating for Parameter One:

A. Precision of equal importance band filters

B. Precision of measurement of individual SNR's in equal importance bands

C. Precision of amplitude and phase comparisons in variable equalizer

D. Precision of difference circuit (S from S+N)

E. Precision of RMS meters in individual S+N circuits

F. Precision of calculations of AI and WI

G. Precision of relationship of WI versus SNR

H. Precision of data reduction

Parameter Two - Conciseness of Results

This method results in a single criteria value, the articulation index (AI), which is a weighted average of SNR's calculated for a number of "equal importance" frequency bands within the speech pass band. The AI value must be converted to percent word intelligibility (WI) by means of empirical relationships previously determined by performing standard WI tests on audio outputs and comparing the results to the AI.

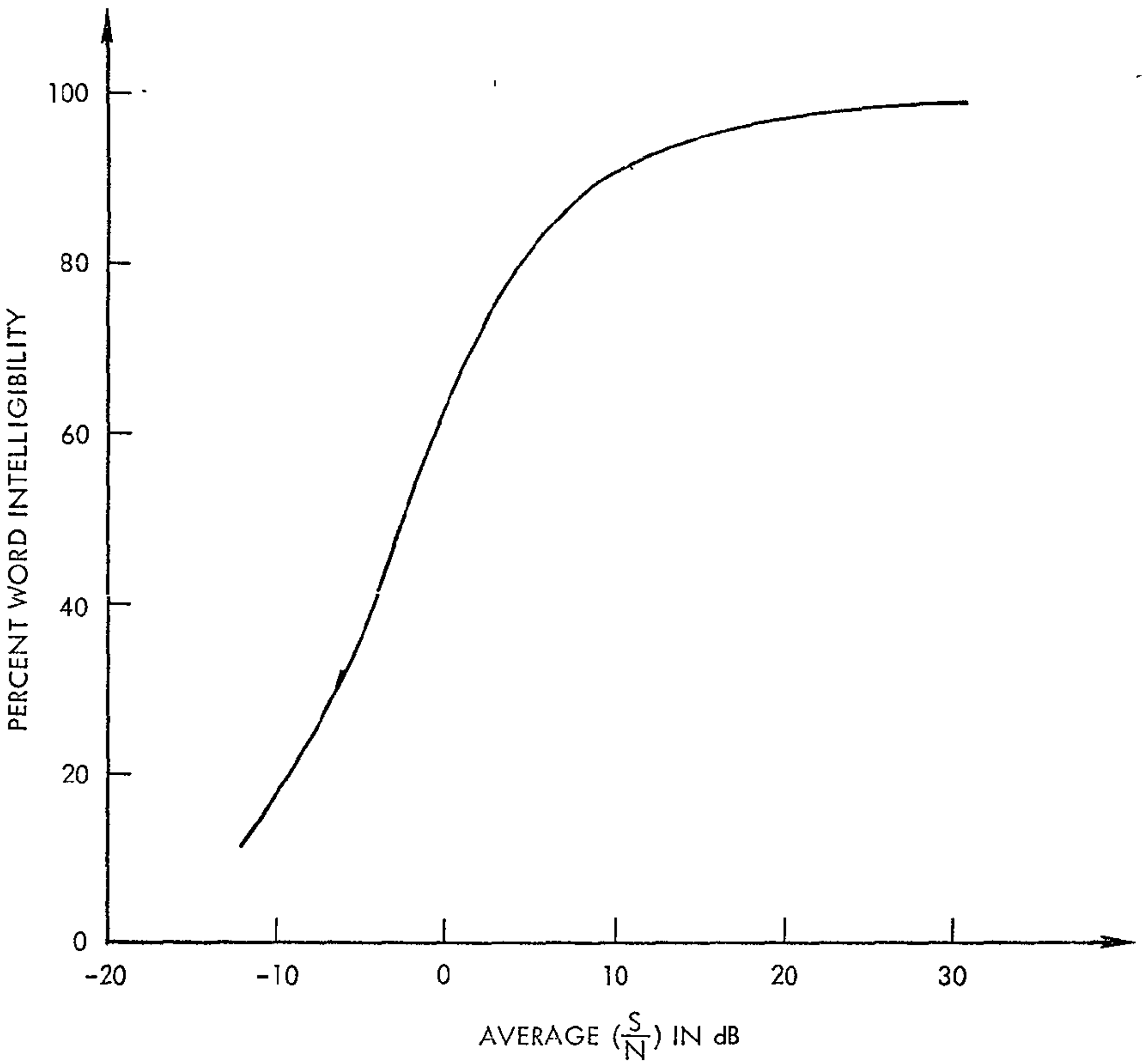


Figure 4. Word Intelligibility as a Function of Average S/N

Parameter Three - Data Reduction Required

Manual recording and calculation of the S+N/N ratios for up to 20 equal importance bands for each test calculation of AI are required. Conversion to WI is made by reference to calibration curves.

Parameter Four - Relationship to Actual System Performance

Methods based on AI should show excellent relationship to actual system performance based on a long history of experimentation and use. Areas of concern include the accuracy of the variable equalizer in matching $|A(\omega)|$ and $\phi(\omega)$ of input and output.

Parameter Five - Ease of Mechanization

Method A would require hardware development of the variable equalizer used to match the amplitude and phase characteristics of the subject under test. No computer use would be required.

3.2 Voice Criteria - Method B Determination of Articulation Index Using Discrete Input Frequencies

Method B is a variation of the method for mechanizing the articulation index testing in which the input voice tape is replaced by a series of discrete frequencies centered in each of the "equal importance" frequency bands. Such a system should be easier to implement since the requirement for an equalizer to match the input and output would not exist. This method is suggested by references to "Test Tone SNR" contained in a draft document of a CCIR study group (Reference 2). A block diagram is shown in Figure 5.

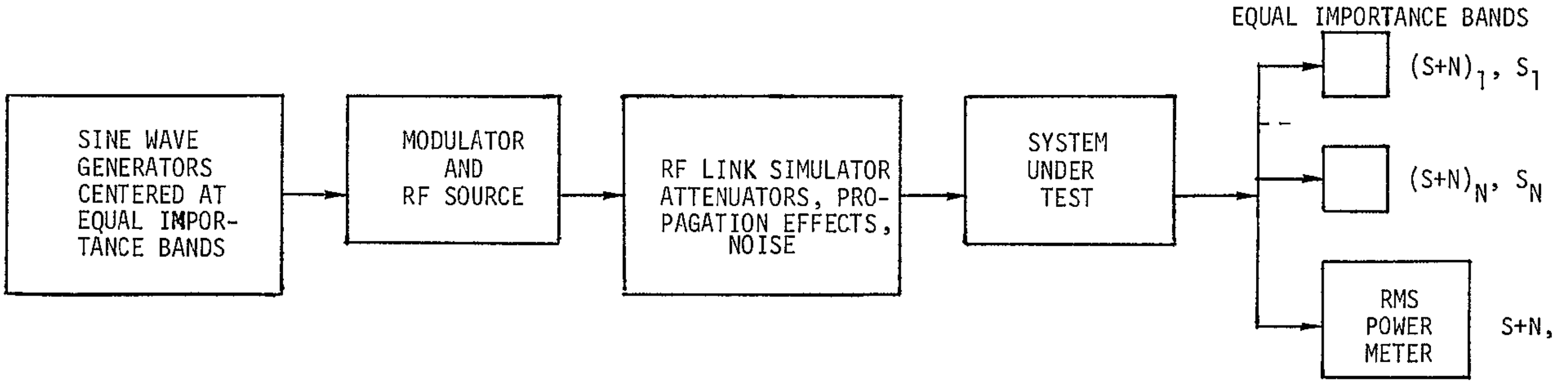
3.2.1 Description of Method

A. The input to the equipment under test consists of a series of single frequencies centered in each of the "equal importance" frequency bands. The amplitudes of the signal frequecies would be adjusted relative to each other in order to conform with a standard speech frequency distribution.

B. Modulation, RF source and RF link simulation are determined by the system under test. If the test is an end-to-end test of a complete link, the modulator and RF source would be replaced with components of the system under test.

C. A wave analyzer is used to measure the signal power of each of the center frequencies of the equal importance bands. Since the filter characteristics of the analyzer can be very sharp, only a small amount of broad-band system noise will be measured, resulting in measuring only signal power.

D. Signal plus noise (S+N) for each of the equal importance bands is measured using filters with the proper response for each band.



VOICE CRITERIA - METHOD B

ARTICULATION INDEX

Figure 5

E. The S/N ratio of each band is then computed from the signal and signal plus noise measured in steps C and D, and weighting functions derived:

$$W_i = 0 \text{ if } (S/N)_i < -12 \text{ dB}$$

$$w_i = \frac{(S/N)_i + 12}{30} \text{ if } -12 \text{ dB} \leq (S/N)_i \leq 18 \text{ dB} \tag{3}$$

$$w_i = 1 \text{ if } (S/N)_i > 18 \text{ dB}$$

F. The articulation index (AI) is then computed

$$AI = \frac{1}{N} \sum_{i=1}^{N} w_i \tag{4}$$

where N = the number of equal importance articulation bands in the pass band of the system under test.

N = 20 for the band pass of normal speech ranging from 200 Hz to 6100 Hz.

G. The AI can then be converted to word intelligibility (WI) by means of an empirical relationship. Since the WI versus AI characteristic depends upon the length of the word list used, determination of WI would require subjective testing of the system using standard word lists and techniques such as those developed by the U.S. Standards Association. The WI would be determined using the same average SNR as was used in determining the AI. A relationship between the test signal AI and WI would thus be determined for one value of average SNR.

H. Determination of test signal AI for several different average SNR's could also be determined, and conversion into WI could be accomplished using methods indicated in Step G, resulting in a calibration curve.

3.2.2 Discussion of Comparison Parameters

Parameter One - Precision of Test Data

The following factors affecting precision of test data were considered in determining the rating of Parameter One:

1. Precision of test sine wave generators used as input.
2. Precision of equal importance band filters.
3. Precision of RMS meters used to measure S+N and S.
4. Precision of calculations of AI and WI.
5. Precision of relationship of WI versus SNR.
6. Precision of data reduction.

Parameter Two - Conciseness of Results

Since the difference between Method A and Method B is in the method of obtaining the individual SNR's for each equal importance band, the same considerations for Parameter Two of Method A are applicable to Method B.

Parameter Three - Data Reduction Required

Manual recording and calculation or S+N/N ratios are required for up to 20 equal importnace bands for each test calculation of AI. Conversion to WI is also required.

Parameter Four - Relationship to Actual System Performance

Results of this method should approach those of Method A. Unknown is the effect of using the single line spectra centered in the equal importance bands instead of using a typical voice response over each band.

Parameter Five - Ease of Mechanization

Method B would require little hardware development and no computer use.

3.3 Voice Criteria - Method C Development of Speech-to-Noise Ratio Using Analog Measurements

The unique feature of Method C is the method of identifying speech in the presence of noise, based on the identification of the speech plus noise and noise only segments of an analog record. These analog records are used to separate and identify the corresponding binary coded decimal (BCD) print-outs of a digital voltmeter used to integrate the instantaneous squared wave-form of the voice input. The BCD values of speech plus noise and noise only are then averaged to produce a speech to noise ratio. This method was devloped under NASA MSC direction by the Philco Ford Corporation (Reference 3).

3.3.1 Description of Method

A. This method is based on the ability of recognizing vowel sounds which represent approximately 90% of speech power spectrum.

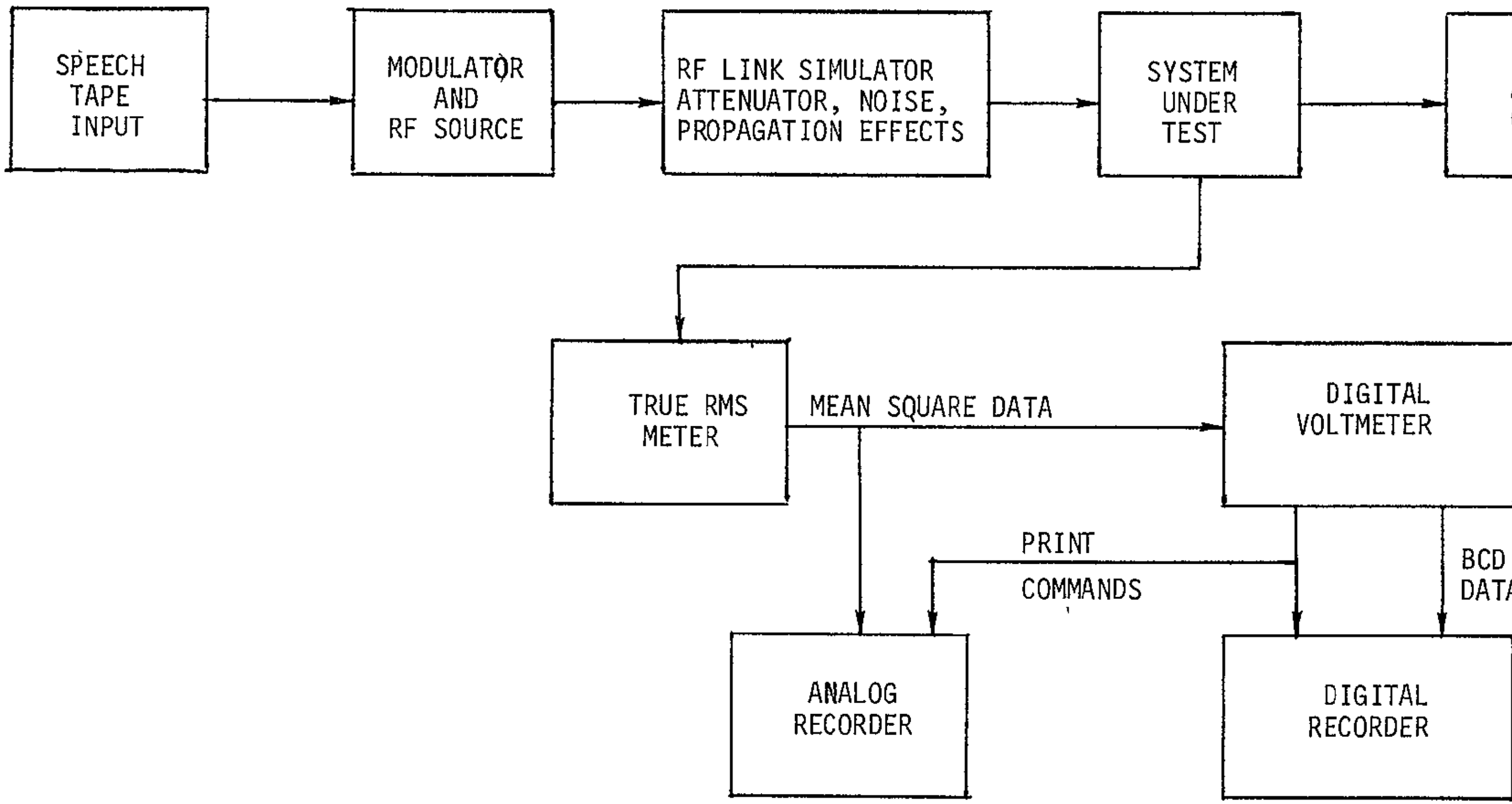
B. Standard test tape records scored for word intelligibility will be used as input.

C. Modulation, RF source and RF link simulation are determined by the system under test. If the test is an end-to-end test of a complete link, the modulator and RF source would be replaced with components of the system under test.

D. The output is converted to an instantaneous squared waveform by the true RMS meter; and recorded on a paper recorder.

E. Also, a time averaged output over a pre-selected time interval is provided by an integrating digital voltmeter.

F. The digital voltmeter converts the average power in each sample time to BCD and prints out the decimal values. At the same time "print commands" from the DVM are recorded on the paper tape along with the squared analog output. An example of the combined analog and digital



VOICE CRITERIA - METHOD C

SPEECH SNR - ANALOG METHOD OF IDENTIFICATION

Figure 6

printer output is shown in Figure 6.

G. The analog record is used to time correlate the digital printout of S+N and N. (To distinguish noise samples from speech + noise samples)

H. Speech SNR is then calculated by speech SNR = $10 \log \frac{P1}{P2}$

P1 = average value of speech + noise

P2 = average value of noise

I. To determine word intelligibility (WI) versus SNR relationship, several tapes of different quality will be used as input to the system under test. The output take will be scored by a trained listening team for WI, and compared to the SNR's.

3.3.2 Discussion of Comparison Parameters

Parameter One - Precision of Test Data

Factors Affecting Precision of Data

1. Precision of true RMS meter
2. Precision of BCD output of digital voltmeter
3. Precision of relating the time of the BCD "Print Commands" to analog record of vowel sounds
4. Precision of calculation of SNR
5. Precision of WI versus SNR relationship
6. Precision of data reduction

Parameter Two - Conciseness of Results

This method results in an average of speech SNR's taken over a number of time increments. This average SNR must be converted to percent WI by means of empirical relationships determined by standard word intelligibility tests.

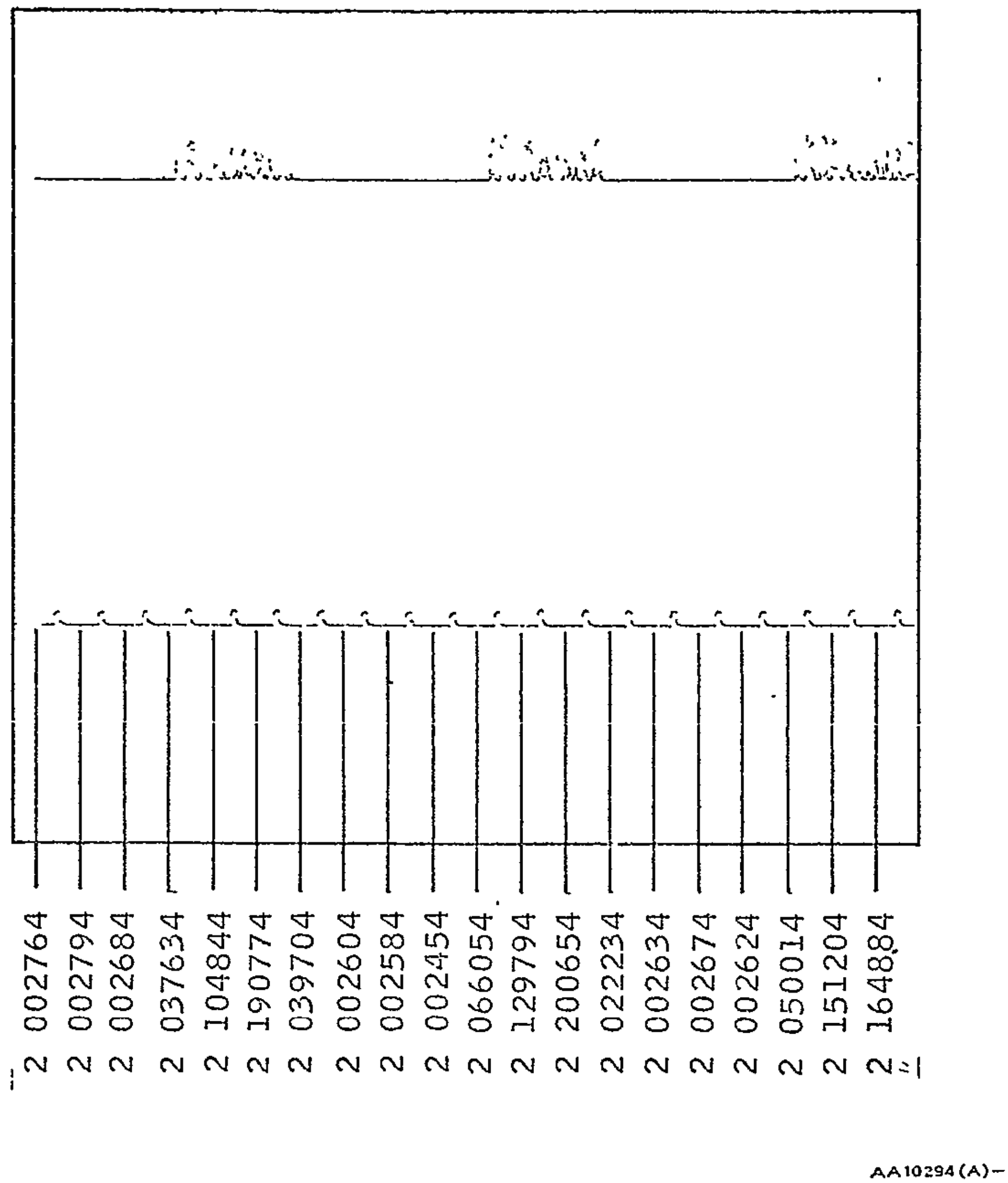


Figure 7. Example of Analog and Digital Printer Output (Reference 3)

Parameter Three - Data Reduction Required

This method requires manual determination of S+N and N intervals by comparison of DVM "print commands" with analog data, averaging of the BCD values for each of the intervals, determination of overall average of S+N and for N, and calculation of speech SNR. Correlation of SPNR with WI would require calibration curves produced by subjective testing.

Parameter Four - Relationship to Actual System Performance

Use of input speech tapes which have been scored for WI and for which test data on WI-SPNR (speech power to noise ratio) relationships have already been developed should produce results which simulate actual system performance very closely for noise distributed over the voice band. Noise concentrated in the high end of the speech band would tend to produce lower WI scores than should be expected in actual practice.

Parameter Five - Ease of Mechanization Voice Systems

Method C would require no hardware development but considerable manual data reduction.

3.4 Voice Criteria - Method D Cross Correlation of Input and Output

One method of assessing overall system performance is to measure how well the system input and output functions are correlated (Reference 4). More specifically, the cross correlation function, $R_{xy}(\tau)$, defined by

$$R_{xy}(\tau) = \int_{0}^{T} X(t + \tau)\, y(t)\, dt \tag{5}$$

where x(t) is the system input and y(t) is the system output, can be used as a measure of system performance. The value of $R_{xy}(\tau)$ varies with the delay, τ, reaching its maximum value when the value of τ approximates the delay through the system under test. Since the presence of a fixed delay in a transmission system generally causes no degradation of the transmitted information, only the maximum value of $R_{xy}(\tau)$ is needed to rate system performance. If the cross correlation function defined by Equation (5) is normalized by dividing by the geometric mean of the mean square values of the input and output signals, the maximum value of the normalized correlation function is confined to the range $-1 \leq \rho_{xy}(\tau) \leq 1$. The quantity, R, needed to specify performance is given by

$$R = |\rho_{xy}(\tau)|_{max} = \frac{|R_{xy}(\tau)|_{max}}{\sqrt{R_{xx}(0)\, R_{yy}(0)}} \tag{6}$$

and is confined to the range $0 \leq R \leq 1$.

The quantities $R_{xx}(0)$ and $R_{yy}(0)$ are the autocorrelation functions of the input and output signals evaluated at $\tau = 0$, and are equal to the mean square values of input and output signals respectively. If the system under test provides distortionless transmission, the output, $y(t)$, is simply a delayed version of the input, $x(t)$, i.e.,

$$y(t) = Kx(t + \tau) \quad .$$

Then from Equation (2)

$$R = \frac{\left| K \int_0^T x(t+\tau)\, x(t+\tau)\, dt \right|_{max}}{\sqrt{\left[\int_0^T x(t)\, x(t)\, dt\right]\left[K^2 \int_0^T x(t)\, x(t)\, dt\right]}}$$

$$= \frac{K \int_0^T [x(t + \tau)]^2 dt}{K \int_0^T [x(t)]^2 dt} = 1 \quad . \tag{7}$$

If, on the other hand, the system output, $y(t)$, were pure noise when a deterministic $x(t)$ was used as a system input, then

$$y(t) = Kn(t),$$

and

$$R = \frac{K \left| \int_0^T x(t + \tau)\, n(t)\, d\tau \right|_{max}}{K \sqrt{R_{xx}(0) \quad R_{nn}(0)}}$$

but

$$\int_0^T x(t + \tau)\, n(t)\, d\tau \rightarrow 0 \quad , \text{ for large } T$$

Consequently, $R \rightarrow 0$.

The block diagram of Figure (8) outlines one method of implementing the normalized correlation coefficient measurement for a receiving system for voice transmission.

3.4.1 Description of Method

A. The input tape is composed of a phonetically balanced word list scored for percent word intelligibility.

B. Modulation, RF source and RF link simulation are determined by the system under test.

C. The input tape signal, x(t), and the system output signal, y(t), are fed to the correlation coefficient computer if the test is being made in real time. Otherwise, the two signals, x(t) and y(t), are recorded for off-line processing at a later time.

D. The value of the performance parameter, R, is obtained from the computation of the normalized cross correlation function.

E. This value of R is used to determine a corresponding value of the Articulation Index, AI, from a calibration curve which has been previously determined by experiment.

F. The value of AI is converted to percent Word Intelligibility by means of the standard calibration curve already determined by subjective testing for the particular type and length of word list used on the input voice tape.

3.4.2 Discussion of Comparison Parameters

Parameter One - Precision of Test Data

The precision of the test data obtained by the cross correlation method hinges primarily on two factors. They are first,

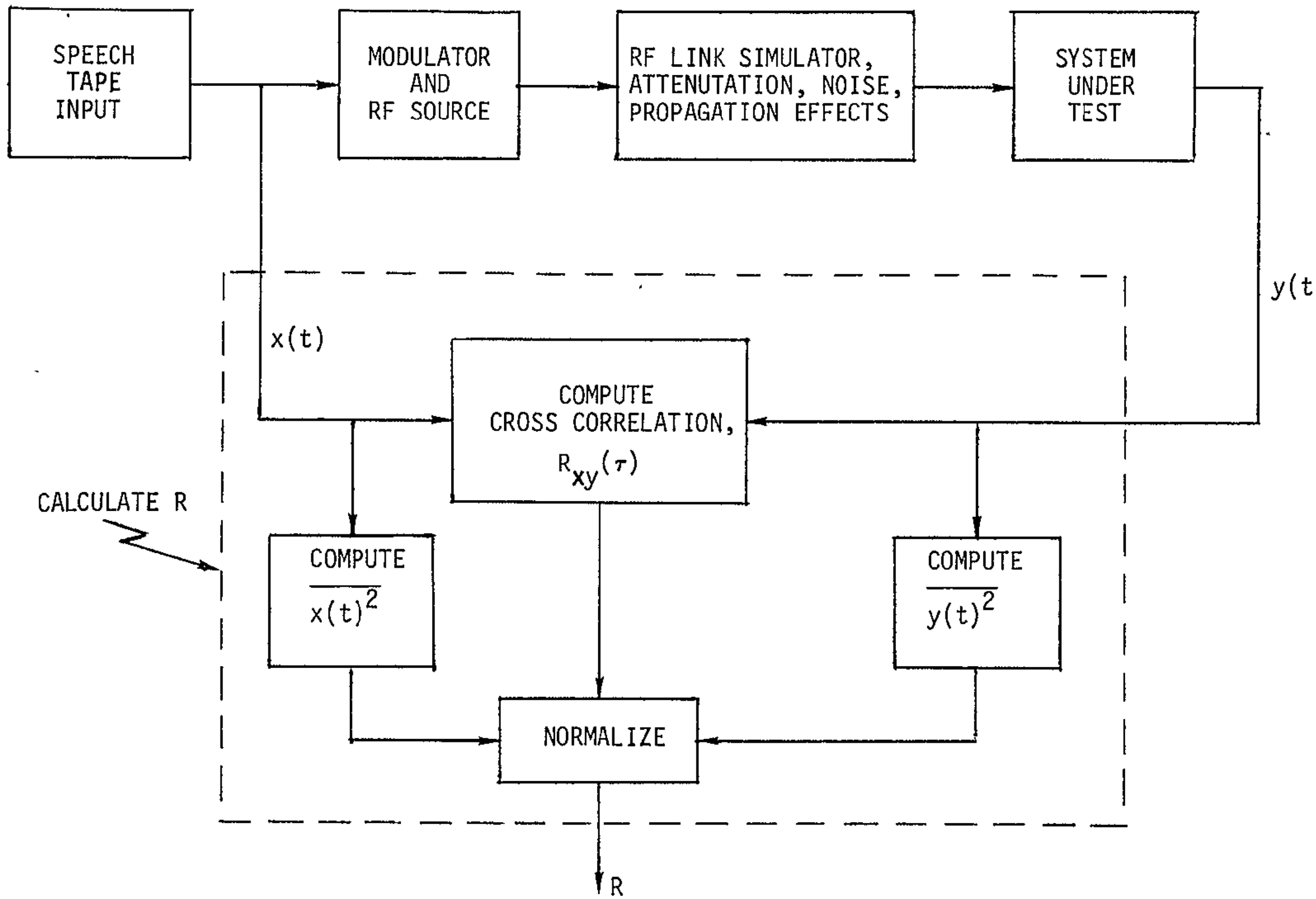


Figure 8. Measurement of the Performance Parameter R

the length of the integration time used in evaluating the cross correlation integral and second, the precision with which the experimental relationship between AI and R can be determined and repeated.

The length of the integration time determines the degree to which the effects of system noise perturb the computed value of R. As long as this time interval is long with respect to the coherence time of the system noise, these perturbations will be small. Since the coherence time is of the order of the reciprocal of the system bandwidth, B, the integration time, T, should meet the criterion

$$T \gg \frac{1}{B} .$$

The accuracy with which the connection between AI and R can be established is difficult to assess without having sufficient experimental data on which to base an accuracy estimate.

Parameter Two - Conciseness of Results

The provision of a single parameter, R, to describe the system performance under a fixed set of conditions is highly concise.

Parameter Three - Data Reduction Required

If the output data is considered to be the parameter R, then the only data reduction required is the translation of this output R data into the corresponding AI or percent word intelligibility data.

Parameter Four - Relationship to Actual System Performance

The lack of experimental data on the use of this technique makes it difficult to determine whether or not the parameter R is one-to-one related to actual system performance as determined by subjective listener testing.

Parameter Five - Ease of Mechanization

The current availability of commercial equipment with the capability to compute correlation functions indicates that the measurement of the parameter, R, could be readily mechanized. The use of tape recording with subsequent computer processing of the recorded input and output signals offers an alternate method of obtaining data on R.

3.5 Voice Criteria - Method E Difference of Power Spectra Between Input and Output

This method is based on the fact that the essential intelligence of speech signals is contained in the short term running power spectrum. The criterion used therefore is a measure of the mean squared error between the input and output power spectra of the system under test. The input and output tapes are converted to digital form, accumulated over a specified time period, subjected to a Fourier transform routine, and compared in a difference circuit. The output error is then squared and accumulated at the end of each test word. The average of the accumulated errors is the evaluation criterion.

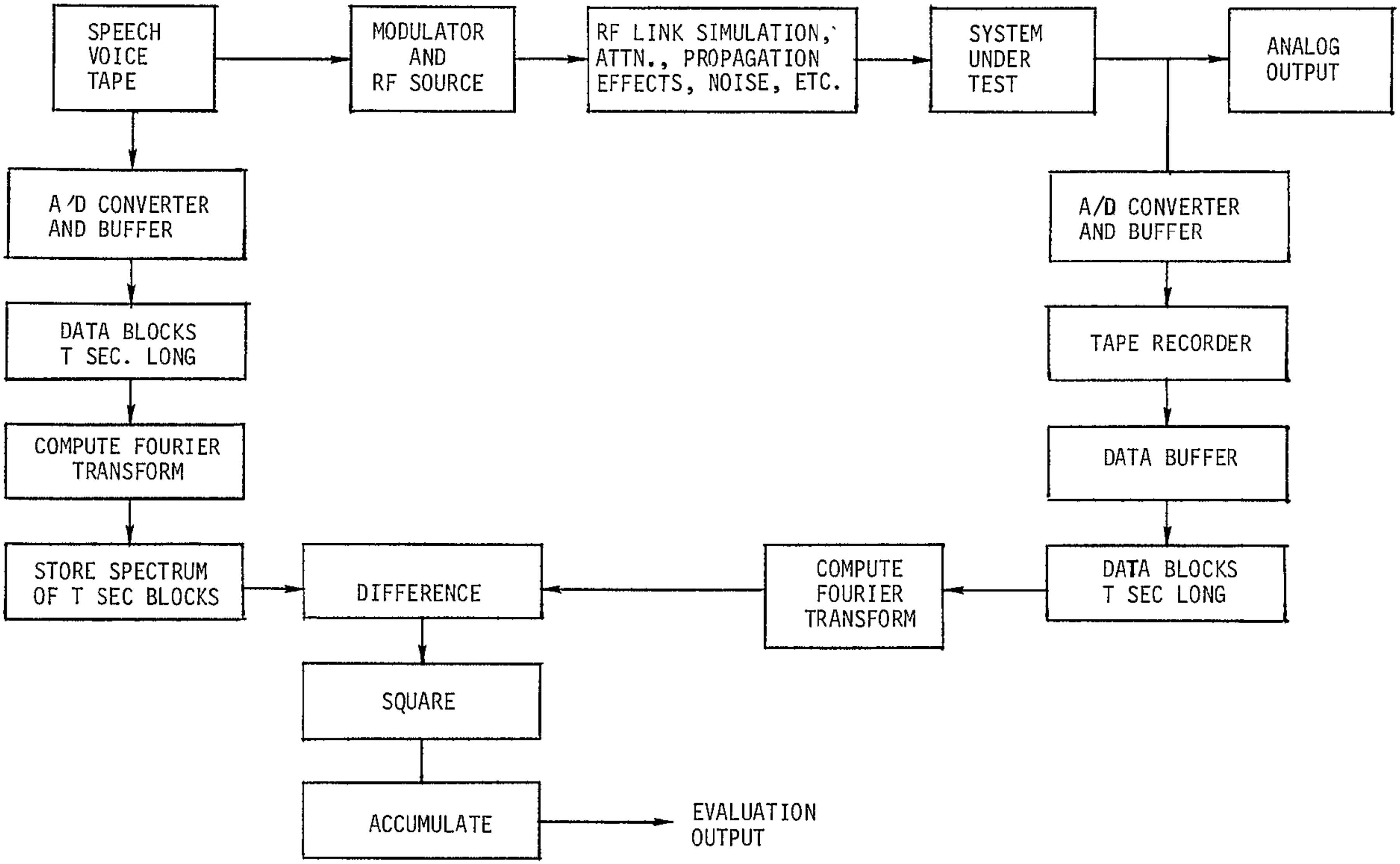
3.5.1 Description of Method

A. Standard voice test tapes scored for word intelligibility are used as the input.

B. Modulation, RF source and RF link simulation are determined by the system under test. If the test is an end-to-end test of a complete link, the modulator and RF source would be replaced with components of the system under test.

C. The tape recording of the output of the system under test is converted into digital form by an analog-to-digital converter, stored in a computer and examined in short time blocks; subjected to a Fourier transform routine to compute the power spectra of each block, and compared to the power spectra of the input tape processed by a similar routine.

D. The difference between the input and output spectra is squared and accumulated for each test word. The average difference is the criteria for voice quality.



VOICE CRITERIA - METHOD E

MEAN SQUARED ERROR

E. To relate to word intelligibility, the output tape would also be scored by an experimental test team. Several different tapes would be used as inputs, and the output word intelligibility for each would be plotted as a function of the differences of the power spectra, providing a calibration curve.

3.5.2 Discussion of Comparison Parameters

Parameter One - Precision of Test Data

Factors Affecting Precision

1. Precision of A-D conversions
2. Precision of Fourier transformer routine in computing power spectra of input and output
3. Precision of difference and squaring routing of input and output power spectra
4. Precision of WI versus power spectra relationship

Parameter Two - Conciseness of Results

This method results in an average of squared differences between input and output power spectra. Since no known relationship between this criteria value and percent WI exists, this relationship would have to be determined by testing using evaluation teams to score the output tapes.

Parameter Three - Data Reduction Required

The input and output digital data are stored in computer and examined in short blocks. The blocks are processed by a Fourier transform routine, the difference of the input and output power spectra thus produced is squared and accumulated as a measure of voice quality. Intially, analog output tapes must be scored for

WI by subjective testing to provide calibration of the quality rating. Further signal conditioning will probably be necessary to provide means for compensating for system time delay and level shifts through the system.

Parameter Four - Relationship to Actual System Performance

Assuming that sufficient quantizing levels are employed in the A/D process to insure that quantizing noise is small with respect to the system noise, this method should produce acceptable results. Since the spectral content of the speech and noise are considered in the process, the problem concerning noise concentrated in the high end of the speech band should be minimized.

Parameter Five - Ease of Mechanization

Method E would possibly require some hardware development, and considerable computer programming and use.

3.6 Voice Criteria - Method F Digital Method of Determining Speech SNR

Method F is based on the fact that vowel sounds are much longer and stronger than consonants. It has been estimated that vowels contribute over 90 percent of the total power spectrum of speech. Use can then be made of this fact to determine speech to noise ratios if monosyllabic test words (or words spoken so slowly that the syllables can be separated) are used in preparing input speech tables.

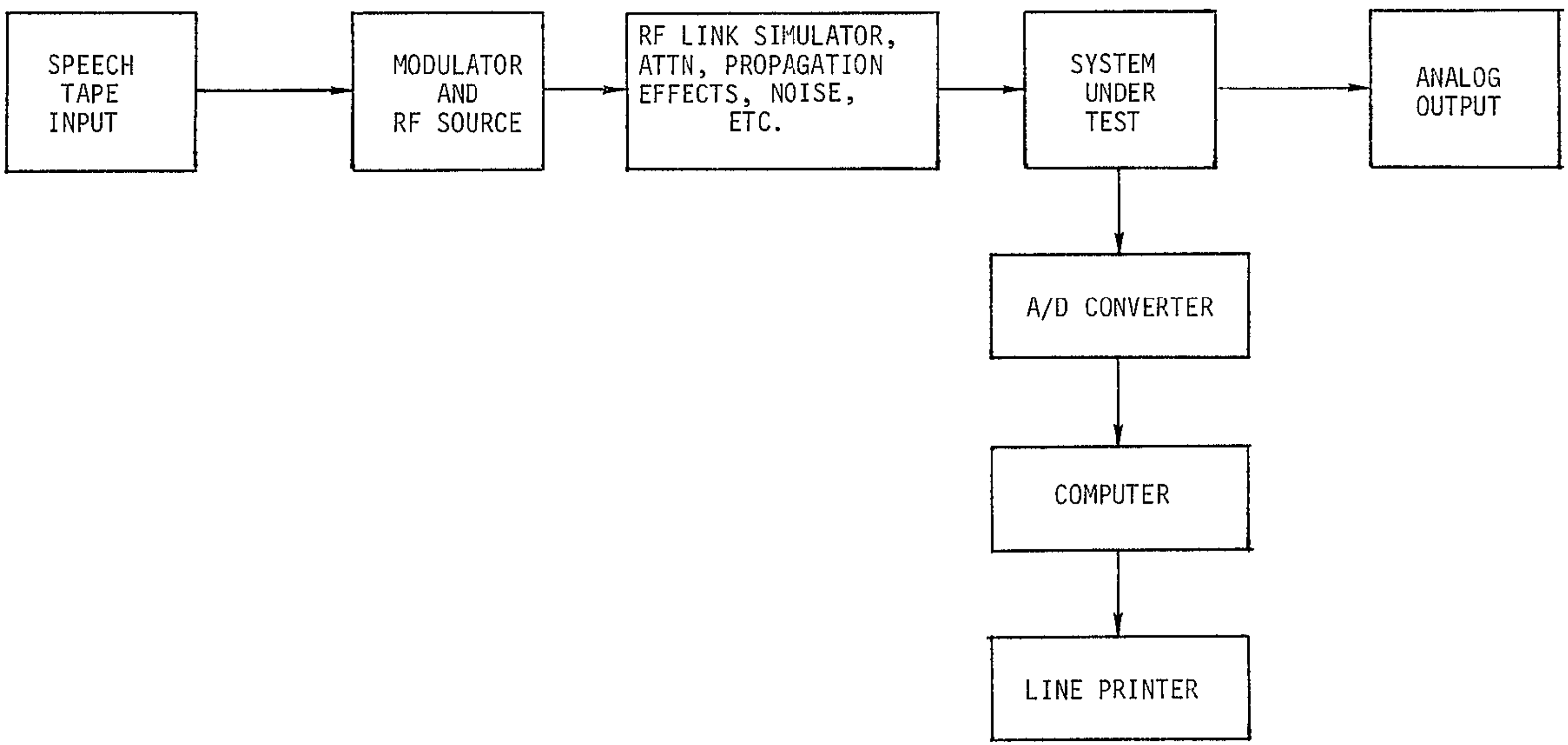
This method converts the output of the system under test to digital form and uses a computer routine to separate speech plus voice from the noise that occurs between words or syllables. To effect this separation the following assumptions are made:

1. The average power of a vowel plus noise waveform will not deviate more than 1 dB throughout the duration of the word or syllable (100 msec minimum to 200 msec maximum).
2. The average power of the in-between-syllable noise (or in-between-word noise) will not deviate more than 1 dB for approximately the same time interval as that of a vowel sound.

This method was developed under NASA MSC direction by the Philco Ford Corporation (Reference 5). A block diagram is shown in Figure 10.

3.6.1 Description of Method

A. Standard voice test tapes scored for WI are used as the input.



VOICE CRITERIA - METHOD F

DIGITAL DETERMINATION OF SPEECH SNR

Figure 10

B. Modulation, RF source and RF link simulation are determined by the system under test. If the test is an end-to-end test of a complete link, the modulator and RF source would be replaced with components of the system under test.

C. After analog to digital conversion, the output is applied to a computer with a program designed to mechanize solution of the problem and provide the output on a line printer.

D. The computer program is based on:

1. Use of 20 msec as the measurement interval, with the average power in at least three consecutive intervals being compared to be within 1 dB, (each 20 msec interval contains 400 samples at sample rate of 20K).
2. Three or more consecutive intervals arè averaged and placed in storage until 50 consecutive intervals have been accumulated.
3. The logic assumes that the smallest value of average power for three or more consecutive intervals (which agree within 1 dB) represents noise and this value is taken as a reference level.
4. All other values of (three or more consecutive intervals) are compared to this reference level.
5. If a value compares within 1 dB of the reference value, it is considered noise for the purpose of computation.
6. If a value is 3 dB or greater than the reference value, it is considered S+N for the purpose of computation.
7. Values between 1 dB and 3 dB of the reference value are ignored.
8. The mean is calculated for all values of N and for all values of S+N, thus providing the basis for speech signal-to-noise ratio (SPNR) computation for each are second interval.

$$SNR = \frac{(S+N) - N}{N} \quad .$$

Thus, the minimum SPNR is 0 dB and occurs when S+N = 2N.

E. The value of SPNR calculated above are converted to word intelligibility by producing a system output analog tape for scoring by trained observers. Several input tapes of different quality would be used to produce a SPNR versus WI calibration.

3.6.2 Discussion of Comparison Parameters

Parameter One - Precision of Test Data

Factors Affecting Precision of Data

1. Validity of assumptions concerning uniformity of average power for duration of one word, and for uniformity of power of the "in between syllable" and "in between word" time periods.
2. Precision of analog to digital conversions.
3. Precision of computer program in averaging, sorting, comparing and calculating speech SNR.
4. Precision of WI versus speech SNR relationship.
5. Precision of data

Parameter Two - Conciseness of Results

This method results in a weighted average of speech SNR's determined for a number of time increments. The weighting is a result of a rather arbitrary differentiation of speech and noise levels. To convert to WI would require test results based on tapes scored for WI.

Parameter Three - Data Reduction Required

After A to D conversion, output tape is processed by a computer program which measures S+N and N by defining as noise the smallest value of consecutive samples whose amplitudes agree within 1 dB, and as signal + noise the values of consecutive samples whose average value is 3 dB above this base line. The computer routine computes the mean SNR. This SNR must be then converted to WI by comparison to known WI versus SNR-WI relationships.

Parameter Four - Relationship to Actual System Performance

Good relation to actual system performance has been demonstrated in lab tests. Some problem areas that became apparent in testing are: sudden shifts in noise levels, and fast continuous speech which can cause speech SNR (SPNR) errors. Optimization of test parameters and computer programs should minimize these problems.

Parameter Five - Ease of Mechanization

Method F would require some hardware development, considerable computer use and program development.

3.7 Voice Criteria - Method G Bit-by-Bit Comparison of Input and Output Tapes

This method is simple in concept in that it merely takes the input and output of the system under test, converts them into digital form and compares them bit-by-bit to determine the bit-error-rate. BER versus word intelligibility calibration would have to be made by subjective testing of the analog output of the system under test.

3.7.1 Description of Method

A. As a system input, standard voice test tapes of scored for word intelligibility (WI) will be used.

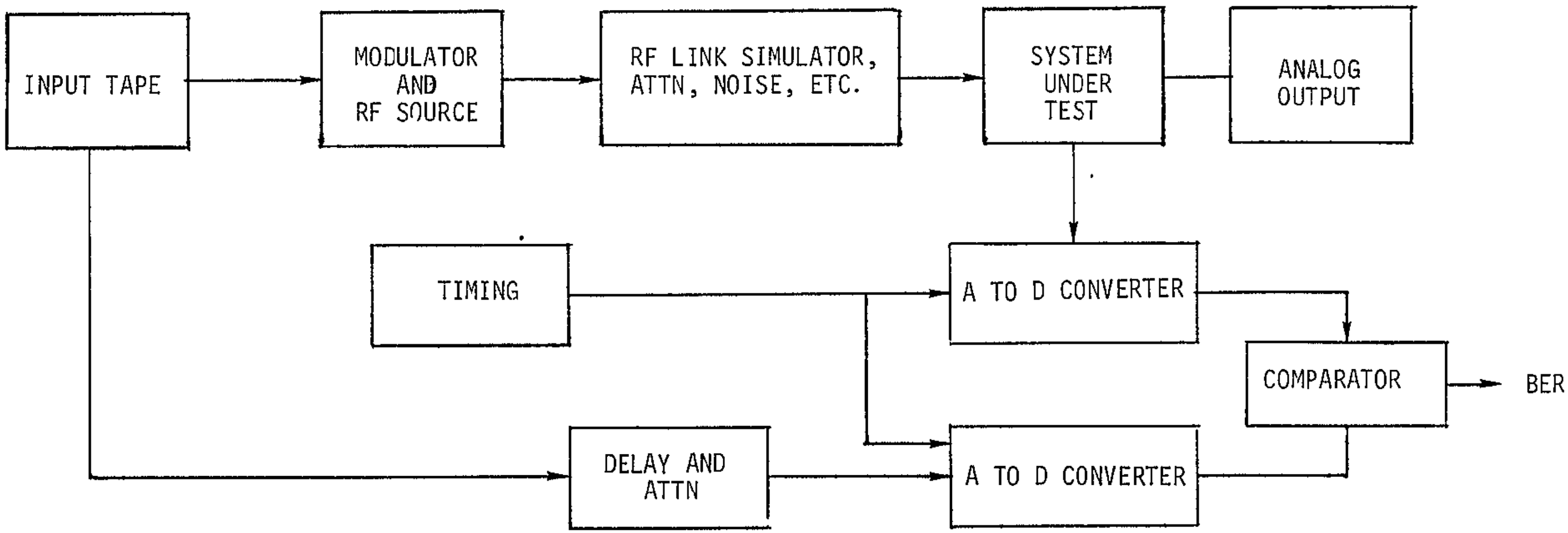
B. Modulation, RF source and RF link simulation are determined by the system under test. If test is end-to-end, the modulator and RF source would be replaced with components of the system under test.

C. The output of system under test is converted into digital form by an A/D converter; in addition, an analog tape output is provided for comparison.

D. The input is also fed through a delay calibrated to match the system delay to an identical A/D converter.

E. The output of the two A/D converters are compared on a bit-by-bit basis and a bit error rate (BER) is calculated.

F. To establish a calibration of BER versus WI, the output analog tape would be scored by a trained observer team. Input tapes of different qualities would be used to produce a BER-WI calibration curve.



VOICE CRITERIA - METHOD G

BIT-BY-BIT COMPARISON

Figure 11

3.7.2 Discussion of Comparison Parameters

Parameter One - Precision of Test Data

The following factors affecting precision of data were considered in determining the rating for Parameter One:

1. Precision of input tape (amount of "Jitter").
2. Precision of A/D converters.
3. Precision of delay circuit.
4. Precision of comparator circuits.
5. Precision of data reduction.

Parameter Two - Conciseness of Results

This method results in bit-error-rate (BER) directly. BER relation to word intelligibility (WI) would have to be established by WI testing of analog output and comparison with digital BER.

Parameter Three - Data Reduction Required

Bit-by-bit comparison of input and output tapes would require little data reduction since BER is computed directly. The bit error rate would have to be converted to WI by appropriate relationship and determine by WI testing.

Parameter Four - Relationship to Actual System Performance

Bit-by-bit comparison should yield accurate bit error rates (BER), assuming that problems such as tape jitter are solved. Use of disc recording should help to minimize this problem. The BER-WI relationship would require subjective testing to provide calibration.

Parameter Five - Ease of Mechanization

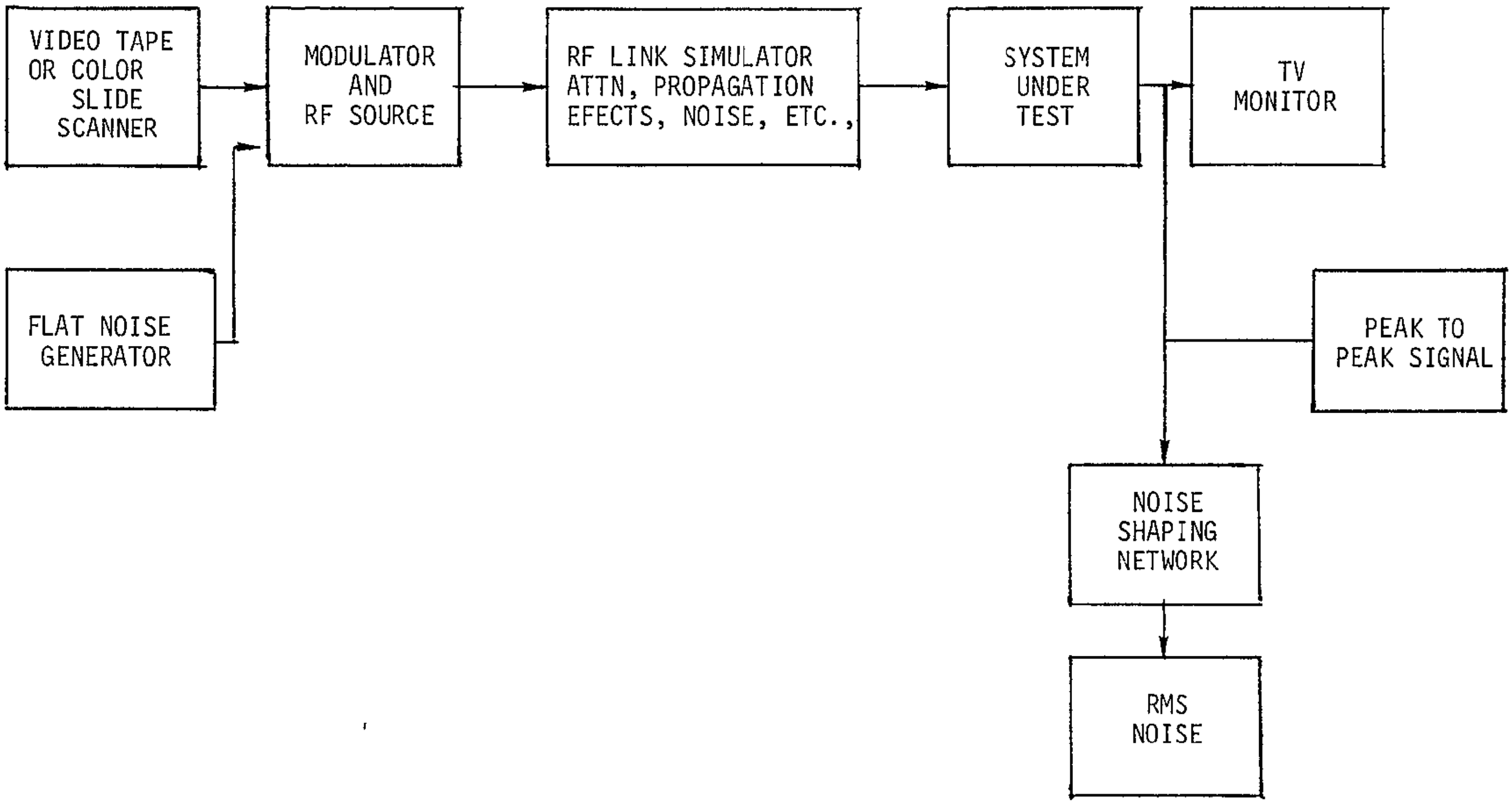
The principal problem in the bit-by-bit comparison technique is that of timing and synchronization, which is complicated by tape recorder jitter in the input. Disc recordings or digital test word generators would help to reduce this problem.

3.8 Video Criteria - Method A Video Signal to Noise Ratio Measurement Using a Weighted Noise Concept

Use of a noise weighting scheme in determining picture signal-to-noise ratios (SNR) is based on the fact that noise in the lower end of the video spectrum has a greater effect on picture quality than noise at the upper end of the spectrum. Methods using noise weighting have been investigated by several groups of researchers such as the International Radio Consultative Committee (CCIR), the Electronic Industries Association (EIA), Bell Telephone Laboratories, the Television Allocation Study Organization (TASO), and the United States Standards Association. This method is described in more detail in TRW Document No. 17618-H123-R0-00, the project technical report covering phase one of Task 707 (Reference 1). A graph of weighted SNR versus two picture quality rating scales is shown in Figure 13. Since the noise weighting curve is an experimentally determined relationship describing relative video picture degradation as a function of noise frequency, use of the noise weighting curve should be applicable to systems which have non-flat video noise spectrums as well as to those which have white noise. For example, the noise weighting function should be applicable to frequency modulated television systems which result in parabolic noise.

3.8.1 Description of Method

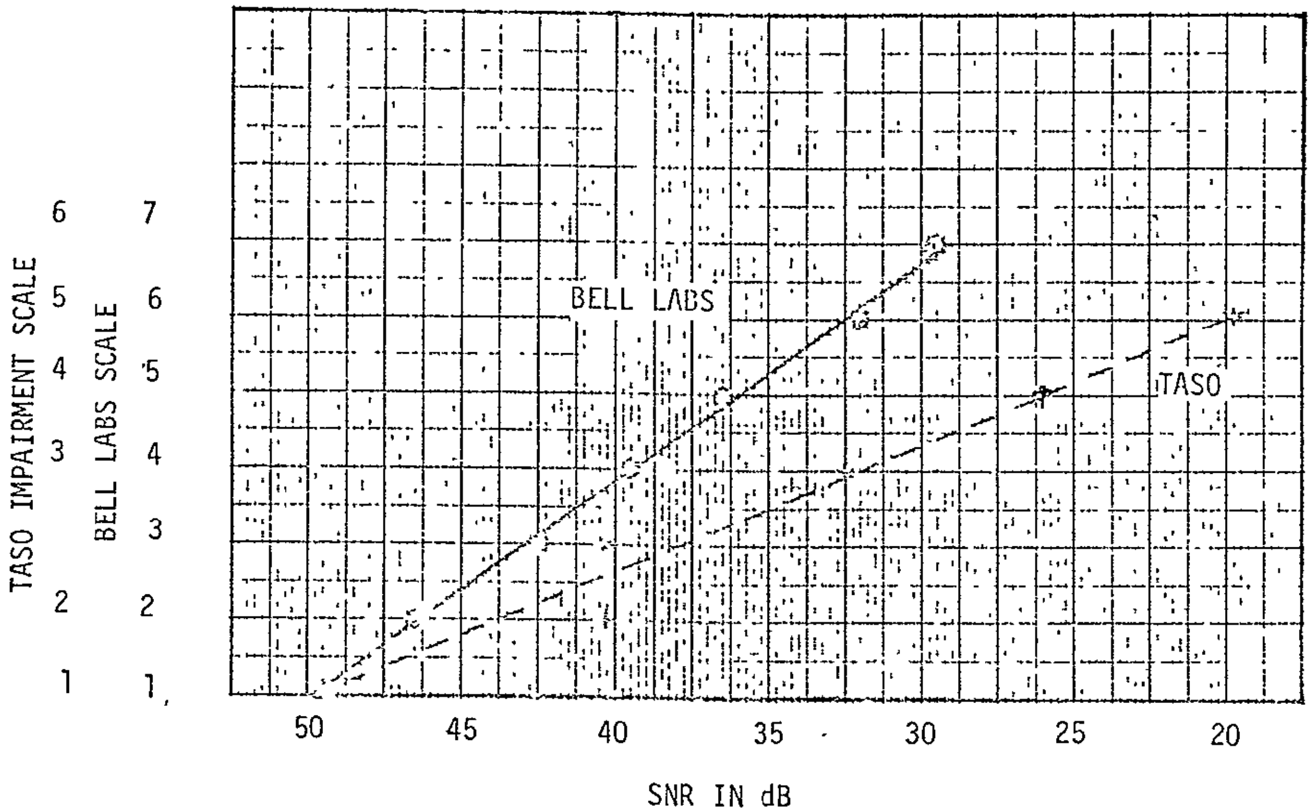
A. A video tape or a color slide scanner and selected color slides will be used to provide video input to the system under test. A noise generator capable of providing a flat noise spectrum over the video bandwidth may be used to inject noise into the system either at the input or as part of the link simulator.



VIDEO CRITERIA - METHOD A

PICTURE SNR USING WEIGHTED NOISE

Figure 12



Quality Scales in Terms of Impairment

Bell Labs	TASO (Impairment Comments)
1. Not Perceptible	1. Excellent
2. Just Perceptible Impairment	2. Perceptible Interference
3. Definitely Perceptible Slight Impairment	
4. Impairment but not Objectionable	3. Interference not Objectionable
5. Somewhat Objectionable	4. Somewhat Objectionable Interference
6. Definitely Objectionable	5. Definitely Objectionable Interference
7. Extremely Objectionable	6. Unusable

Figure 13

B. Modulation, RF source and RF link simulation are determined by the system under test. If the test is an end-to-end test of a complete link, the modulator and RF source would be replaced with components of the system under test.

C. A noise shaping network at the output will provide noise weighting in accordance with the noise weighting curve provided by the United States Standards Association. This curve is shown in Figure 14. The weighted noise will be measured by a true RMS voltmeter when no video signal is provided to the input of the system under test.

D. With the video signal applied, the white to blank video signal will be measured.

E. The weighted picture SNR will be calculated:

$$S/N = \left(\frac{\text{blank to white video voltage}}{\text{weighted RMS voltage of video noise}}\right)^2 \tag{8}$$

F. At the same time, an observer team will make an assessment of the quality of the output picture on the TV monitor, using a standard 5 or 6 point Bell Laboratories, CCIR or TASO scale. Use could be made of the curves shown in Figure 13, but improvements made in TV equipment since the time the curves were taken make it important to repeat this step.

G. Several combinations of video signal level input and noise input (either at the input or as part of the link simulator) will be measured and scored by the observer team, resulting in a calibration curve of picture SNR versus picture quality.

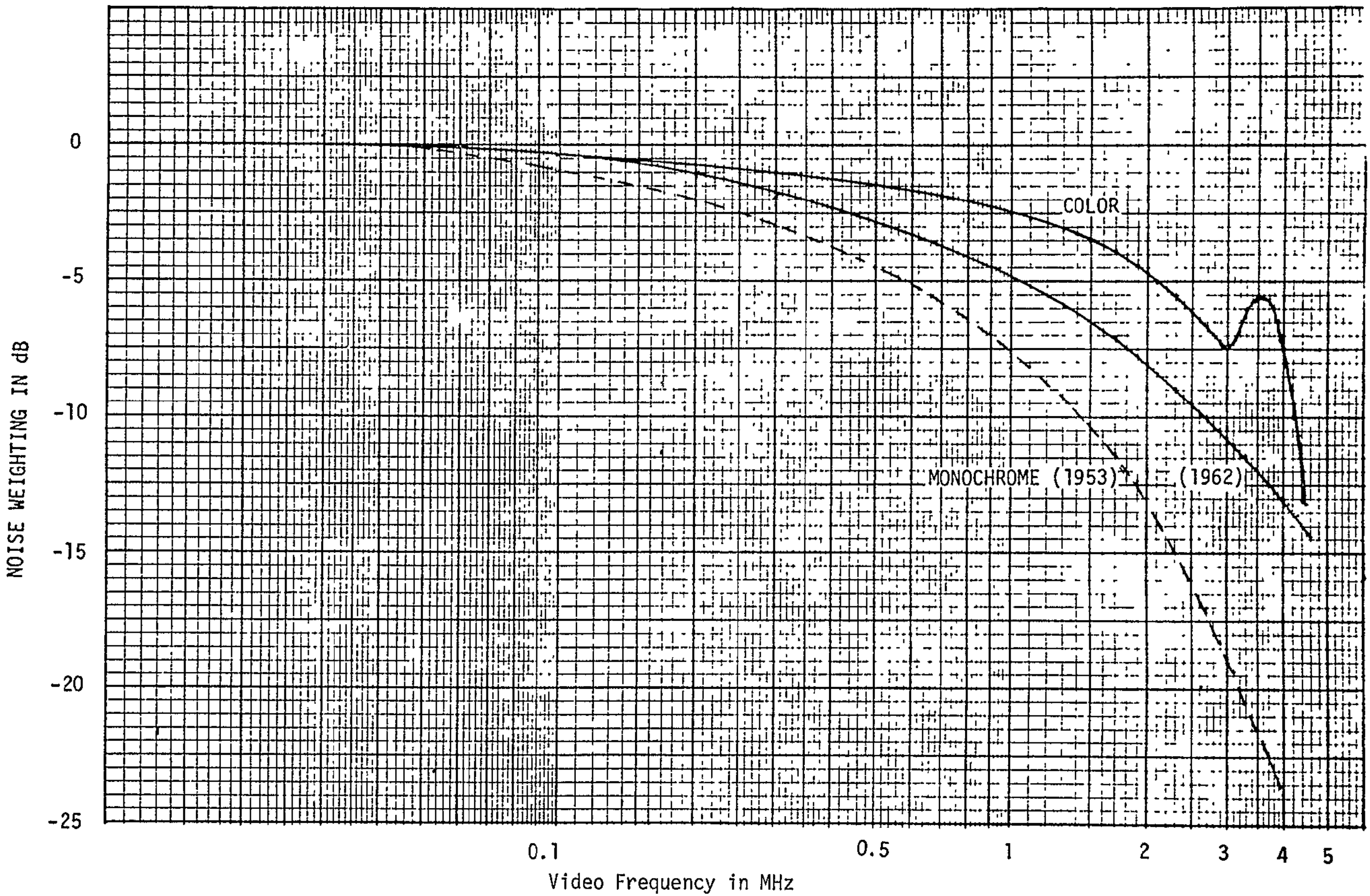


Figure 14. Noise Weighting Curves

3.8.2 Discussion of Comparison Parameters

Parameter One - Precision of Data

The following factors affecting precision of data were considered in determining the rating for parameter one:

1. Precision of peak measurement of blank to white signal.
2. Precision of measuring RMS noise.
3. Precision of filters for weighting noise.
4. Effectiveness of filters matching subjective effects of noise.
5. Precision of data reduction.

Parameter Two - Conciseness of Results

This method results in a measurement of overall SNR which is converted in picture quality rating index by use of previously determined empirical curve.

Parameter Three - Data Reduction Required

This method would require only one manual recording of signal and noise and calculation of SNR. The picture quality would be manually determined from a previously determined relationship.

Parameter Four - Relation to Actual System Performance

The weighted noise concept used in measuring SNR and the quality rating judgement concept used in this method has received considerable attention in the past, and seems to be a reasonable approach to the problem. The noise weighting factors should be checked with up-to-date hardware.

Parameter Five - Ease of Mechanization

Once the noise weighting factor and SNR-picture quality relationship have been established, the method is reasonably simple and straight forward. The picture SNR is determined by measuring the peak blank-to-white signal level and the RMS noise in the video band. The picture quality is then determined by use of the known relationship of SNR versus picture quality.

3.9 Video Criteria - Method B Cross Correlation of Input and Output

3.9.1 Description of Method

This method is essentially the same as that described for Method D, Voice Criteria, with the substitution of a standard video tape or the output of a slide scanner used in place of the voice tape as input to the system under test. The analog-to-digital conversion of the input x(t) and output y(t) to the correlation computer would operate at a higher data rate because of the video bandwidth, but the principles and method would be the same as those described for the voice system.

3.9.2 Discussion of Comparison Parameters

Parameter One - Precision of Test Data

As discussed in Section 3.4.2, one of the principal factors affecting precision in the correlation method is the length of the integration time used in evaluating the cross correlation integral. Since the coherence time of the system noise is of the order of the reciprocal of the system bandwidth, and it is desirable that the integration be long with respect to the coherence time of the system noise, this factor should be more easily realized in the video system since the bandwidth is larger. However, this advantage is offset by the fact that more samples per second are required, and the difficulty in synchronizing the analog to digital mechanization of the input and output to the correlation computer.

Parameter Two - Conciseness of Results

The single parameter R results from determining an average of values of R computed over fixed time intervals.

Parameter Three - Data Reduction Required

Assuming that the calculation of the parameter R is completely mechanized, the data reduction required would consist of determining the corresponding value of picture quality from calibration curves previously established by subjective testing.

Parameter Four - Relationship to Actual System Performance

The relationship of this method to actual system performance will have to be established from experimental test results. Since the subjective effects of picture quality are more difficult to assess than those of voice systems, this relationship will probably be more difficult to establish.

Parameter Five - Ease of Mechanization.

The higher data rates required to mechanize the correlation coefficient computer would increase the size and cost of the computer as compared to that needed to mechanize the system for voice.

3.10 Video Criteria - Method C Equal Importance Frequency Bands

3.10.1 Description of Method

This method is hypothetical and is based on the fact that in the weighted noise concept of measuring video signal-to-noise ratios, bands of noise centered at different frequencies in the video band cause equal subjective interference effects when the applied through a noise weighting network (reference 7). The frequency composition of a typical series of noise bands is shown in Figure 15. When the noise amplitudes in these bands were weighted in accordance with the standard of the U.S. Standards Association, equal SNR's caused equal subjective effects as judged by a panel of observers. Additionally, a more or less linear relationship of a picture quality rating scale and the weighted noise level in dB appears to exist. Thus it is postulated that it might be possible to divide the video frequency bands into a number of "equal importance" frequency bands which could be used as a basis for establishing video criteria in a manner similar to that used in calculating the articulation index of voice systems.

The method used is such a system would be similar to that described in Method A, Voice Criteria.

3.10.2 Discussion of Comparison Parameters

Parameter One - Precision of Data

The following factors affecting precision of data were considered in determining the rating for parameter one:

1. Precision of equal importance band filters
2. Precision of measuremnt of individual SNR's in equal importance bands.
3. Precision of amplitude and phase comparisons in variable equalizer
4. Precision of difference circuit (S from S+N)

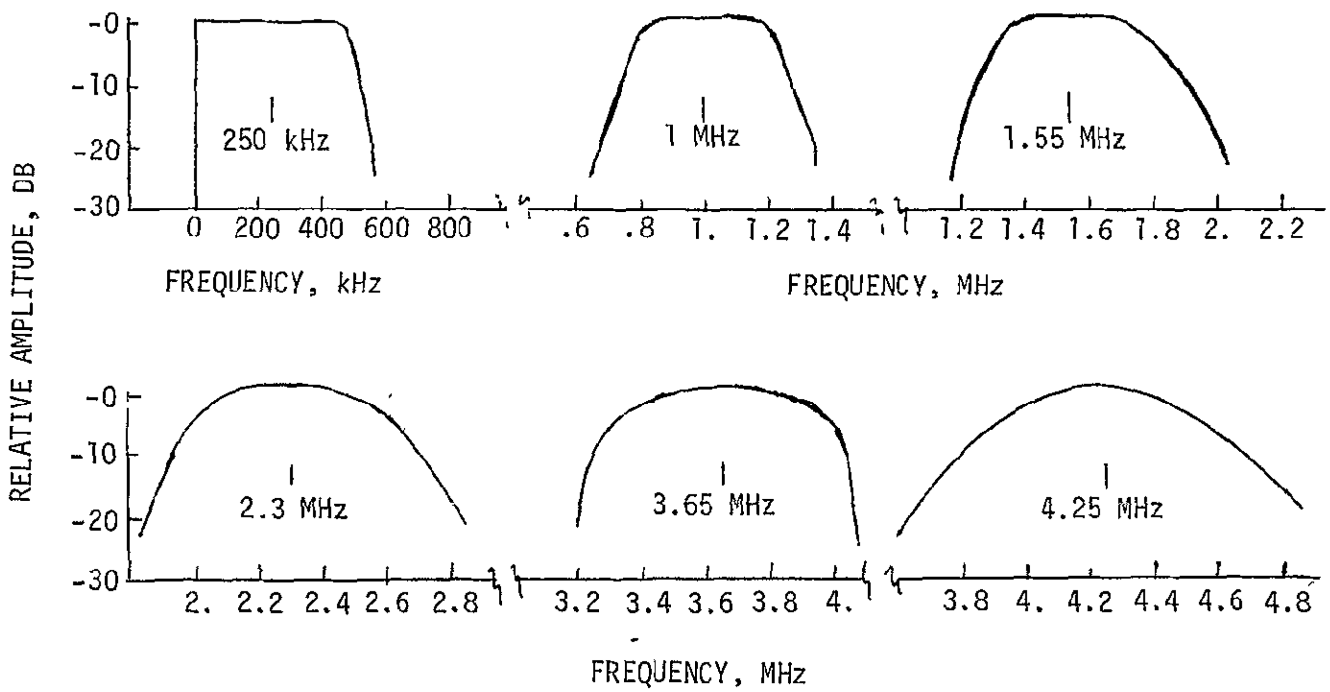


Figure 15. Frequency Composition of 500-kHz Bands of Random Interference Centered at Various Video Frequencies

5. Precision of RMS meters in individual S+N circuits.
6. Precision of calculation of picture quality index.
7. Precision of relationship of picture quality index and SNR.
8. Precision of data reduction.

Parameter Two - Conciseness of Results

This method results in a single criteria value, the picture index, which is a weighted average of SNR's calculated for a number of "equal importance" frequency bands in the TV pass video band.

Parameter Three - Data Reduction Required

Manual recording and caluclation of signal to noise ratios for a number of frequency bands are required by this method. Calculation of the picture index from the weighted average of the individual SNR's is also required. Calibration of picture index would require subjective testing.

Parameter Four - Relation to Actual System Performance

The idea of determining "equal importance" frequency bands for video has not been proven. However, the process used in determining the noise weighting factors used in Method A involved a similar idea - the magnitude of noise in different bands was adjusted to give equal impairment to the picture.

Parameter Five - Ease of Mechanization

The test mechanization would be complicated by the number of band pass filters necessary to separate the "equal importance" signal bands. Manual computation of the "picture index" based on the weighted average of the band pass SNR's would be required.

3.11 Video Criteria - Method D Mean Squared Error of Input and Output Spectra

3.11.1 Description of Method

This method would be essentially the same as that described for Method E - Digital Voice Criteria, with the substitution of a standard video tape or the output of a slide scanner in place of the voice tape as input to the system under test. The A to D converters would require a larger number of quanitizing levels (128 levels, or 7 bits has been suggested as adequate for picture information encoding, Reference 8) and higher data rates, but the technique would be the same. The analog video output would be scored for quality by an observer team to provide a calibration relationship to the difference of the mean squared errors.

3.11.2 Discussion of Comparison Paramters

Parameter One - Precision of Data

The following factors affecting precision of data were considered in determining the rating for Parameter One

1. Precision of A-D conversions.
2. Precision of Fourier transform routine in computing power spectra of input and output.
3. Precision of difference and squaring routine of input and output power spectra.
4. Precision of relationship of picture quality to power spectra.
5. Precision of data reduction

Parameter Two - Conciseness of Results

This method results in an average of squared difference between input and output power spectra. Since no known relationship between this criteria and picture quality rating index is known. this relationship would have to be established by testing using an evaluation team to score the output tapes.

Parameter Three - Data Reduction Required

The data reduction requirements for Method D are similar to those for the same type of criteria for voice systems, except that larger, faster computers would be necessary because of the higher data rate required for video.

Parameter Four - Relation to Actual System Performance

This method should produce acceptable results if the number of quantizing levels is high enough to keep the quantizing noise low with respect to the system noise.

Parameter Five - Ease of Mechanization

The analog-to-digital equipment required would be complicated by the high data rate required to reduce quantization error. Considerable software development would be required to implement the Fourier transform, squaring and computation of the mean square error of the spectra. Assuming that the test configuration had been fully developed, this method would be fairly simple to mechanize.

3.12 Video Criteria - Method E Bit-by-Bit Comparison of Input and Output

3.12.1 Description of Method

This method would be similar to Voice Criteria - Method G. A standard video tape or the output of a slide scanner would provide the input. The A to D converters and digital comparators would be required to operate at a much higher data rate. The analog video output would be scored for quality by an observer team in order to provide a calibration relationship of picture quality versus BER.

3.12.2 Discussion of Comparison Parameters

Parameter One - Precision of Data

The following factors affecting precision of data were considered in determining the ratings for Parameter One:

1. Precision of input tape (amount of jitter).
2. Precision of A-D conversions.
3. Precision of redundancy removal-coding and decoding (if required).
4. Precision of digital comparison circuits.
5. Precision of BER versus picture quality.
6. Precision of data reduction.

Parameter Two - Conciseness of Results

Method E would result in an error rate determined by bit-by-bit comparison of input and output digital tapes. The error rate versus picture quality relationship would have to be established by subjective testing of picture quality.

Parameter Three - Data Reduction Required

Bit-by-bit comparison of input and output tapes would require fast computers with a large storage. In addition, the BER would have to be converted to picture quality rating by use of appropriate relationship determined by subjective testing.

Parameter Four - Relation to Actual System Performance

Accurate BER should result from this method assuming that synchronizing and timing problems can be solved. These problems are more acute than those experienced by bit-by-bit methods for voice because of the higher data rate required for video systems. The BER versus picture quality rating would have to be determined by subjective testing, assuming that such a relationship, hopefully monotones, exists.

Parameter Five - Ease of Mechanization

The timing and synchronizing problems expressed in the bit-by-bit comparison method for voice systems would be intensified because of the higher data rates required. If redundancy removal decoding were required as part of the test process, the complexity of mechanization would obviously be increased.

3.13 Digital Data Systems Criteria - Method A Bit-by-Bit Comparison

Bit-by-bit comparison of digital tapes of the input and output of a system under test is conceptually one of the most simple methods of determining bit error rate. Timing and synchronization of the input and output can be a problem. For very low error rates, the counting time may be appreciable.

3.13.1 Description of Method

A. The input to the system under test is digital test tape of known message content.

B. Modulation, RF source and RF link simulation are determined by the system under test. If the test is an end-to-end test of a complete link, the modulator and RF source would be replaced with components of the system under test.

C. The output of the system under test is fed to a comparator where it is compared to the input bit stream, after the input bit stream has been corrected for system delay.

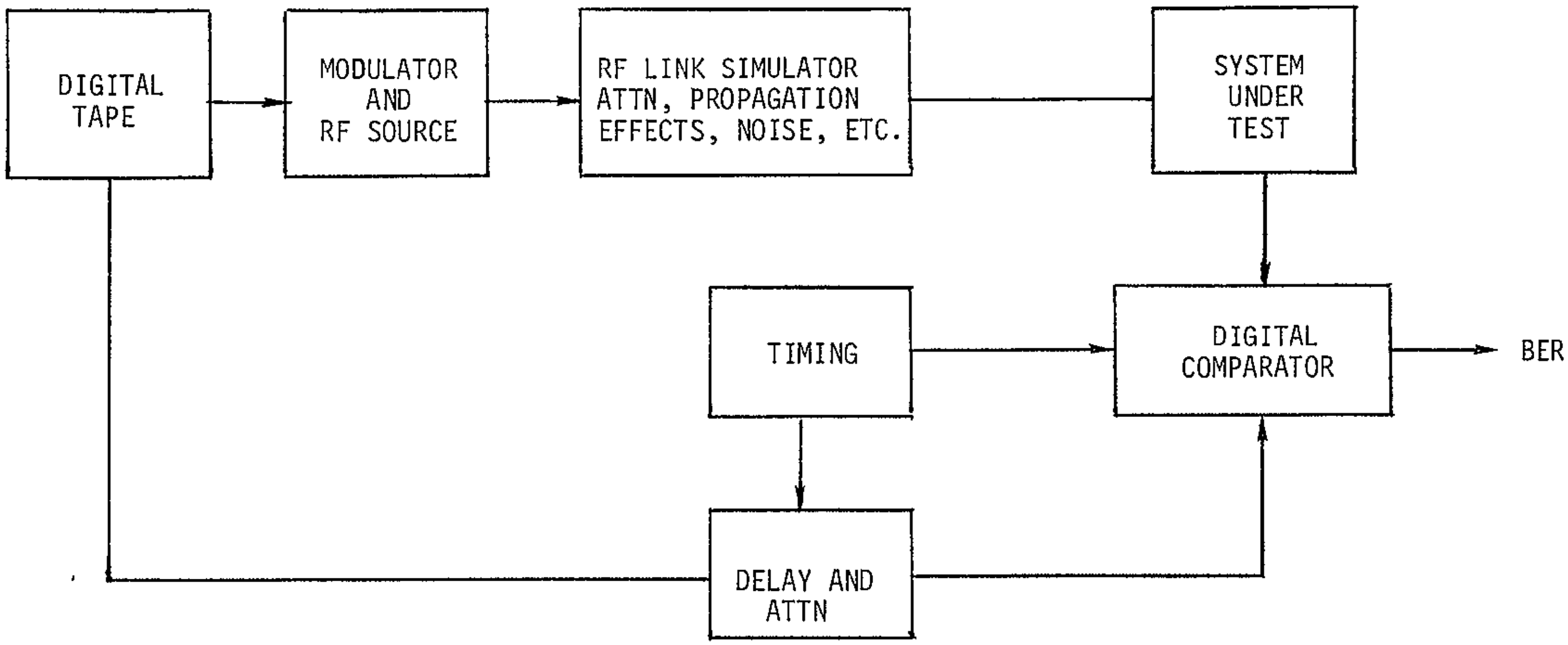
D. The comparator produces a BER directly by counting the errors in a specified length of time. A problem with this method is that if the BER is very low, an unacceptably long time may be required to count enough errors to give a reliable estimate of the actual error rate.

3.13.2 Discussion of Comparison Parameters

Parameter One - Precision of Data

The following factors affecting precision of data were considered in determining the rating for Parameter One:

1. Precision of input tape (amount of jitter).



DIGITAL DATA CRITERIA - METHOD A

BER BY BIT-BY-BIT COMPARISON

Figure 16

2. Precision of delay circuit.

3. Precision of comparator circuits.

4. Precision of data reduction.

Parameter Two - Conciseness of Results

This method results in bit error rate (BER) directly.

Parameter Three - Data Reduction Required

Bit-by-bit comparison would require little data reduction since the BER is computed directly. If the error rate is very low, an unacceptably long time may be required to estimate the actual error rate.

Parameter Four - Relationship to Actual System Performance

Bit-by-bit should yield accurate bit error rates (BER) assuming that problems such as tape jitter are solved.

Parameter Five - Ease of Mechanization

The principal problem in the bit-by-bit comparison technique is that of timing and synchronization, which is complicated by tape recorder jitter in the input. Disc recording or digital test word generators would help reduce this problem.

3.14 Digital Data Criteria - Method B Pseudo-Error Extrapolation

In an effort to overcome the problem of long counting time which may occur in conventional bit-by-bit comparison of input and output data streams when the error rate is very low, the technique of computing pseudo error rates which are much larger than the actual error rate, has been developed (Reference 9). This method creates large pseudo error rates by biasing modified mark-space decision circuits in favor of the incorrect decision, and is characterized by the following features:

1) The pseudo error rates are generated by use of modified decision thresholds in the "mark" and "space" channels.
2) A method of estimating the pseudo error rates corresponding to two or more modified decision thresholds.
3) Two or more estimated pseudo error rates based on different decision thresholds are used to generate a function of pseudo error rates versus a parameter representing the modified decision thresholds.
4) This function is extrapolated to a point where the decision threshold parameter corresponds to that of the actual decision threshold. Thus, at this point the estimated pseudo error rate equals the estimated actual error rate.

The principle of operation of a device designed to produce a pseudo-error rate (P_p) is based on the following observation: "for a given type of modulation and given form of probability distribution of the noise and fading processes, it is possible to define a threshold parameter K such that the logarithm of the pseudo-error rate P_p is a linear function of K for a wide range of values of P_p. The linear portion of this curve, when extended to K=0, coincides with the logarithm of the actual error rate. Thus, by measuring P_p for two values of the parameter K and linearly extrapolating through these two points to the value K=0, one obtains an estimate of the logarithm of the actual error rate." (Reference 9).

Figure 18 shows the graph (A) of the logarithm of the actual receiver error rate P_e as function of signal to noise ratio (R) for some propagation criteria, plotted with curve (B) of the logarithm of the pseudo error rate P_p versus the parameter K for a particular value of received signal to noise ratio, R_0. The point where the linear extrapolation of curve (B) intersects the R_0 ordinate is also the point when the curve (A) intersects the R_0 ordinate.

Since this method of computing pseudo-errors depends on modified thresholds in the "mark" and "space" channels, access to these points prior to the threshold detector must be made available in the system under test.

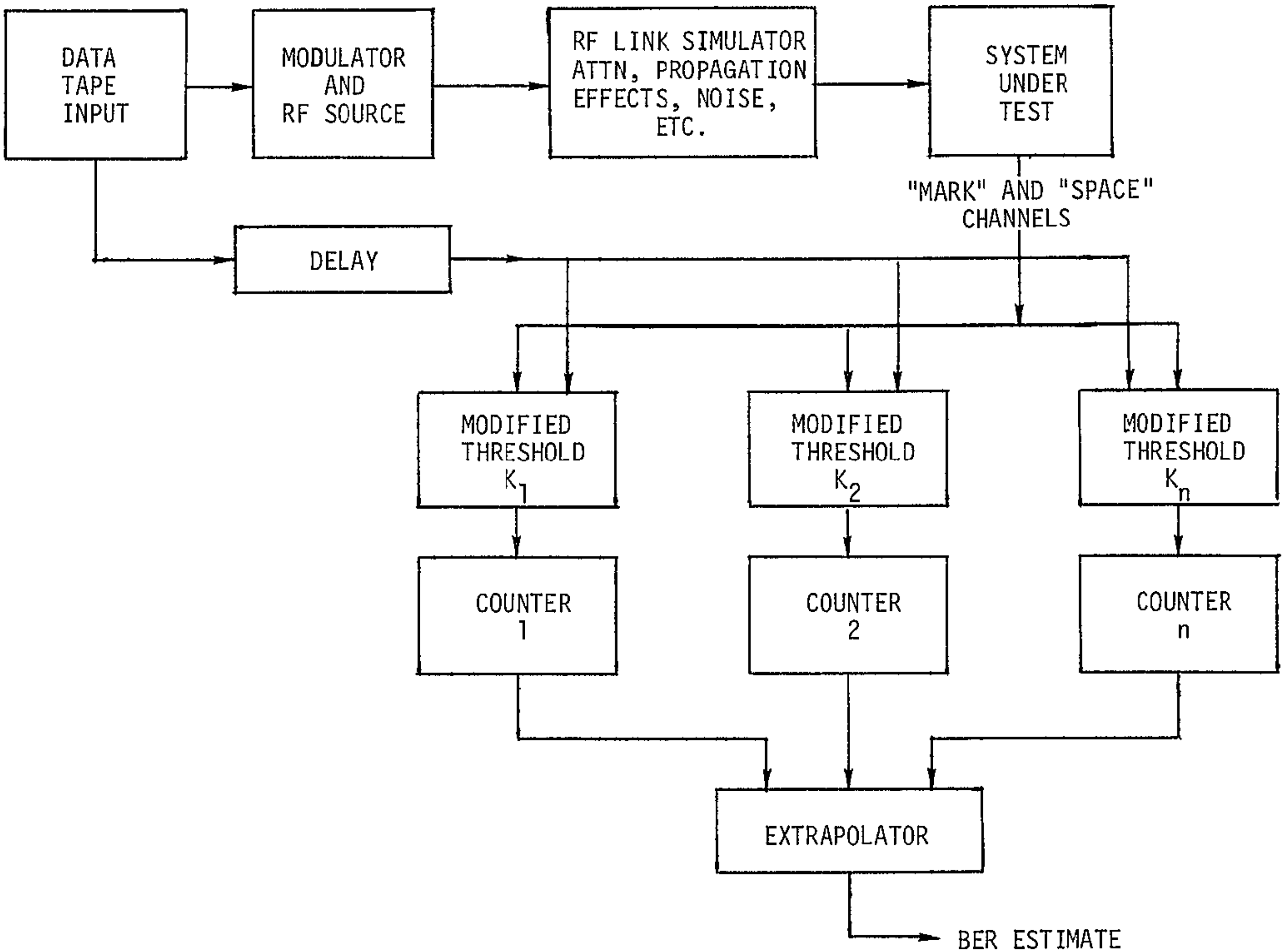
3.14.1 Description of Method

A. The input to the system under test is a digital test tape of known message content.

B. Modulation, RF source and RF link simulation are determined by the system under test. If the test is an end-to-end test of a complete link, the modulator and RF source would be replaced with components of the system under test.

C. The "mark" and "space" outputs (prior to the threshold detection) of the system under test are fed to a series of modified decision circuits where they are compared with the input data stream in such a manner that the pseudo error rates generated are larger than the actual error rate of the output.

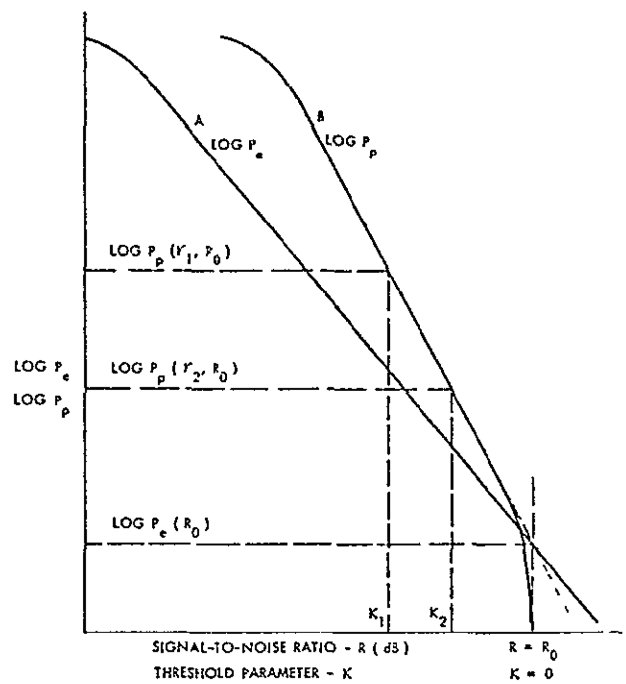
D. The pseudo error rates are counted and fed to an extrapolator where a linear extrapolation of the pseudo errors versus their respective



DIGITAL DATA CRITERIA - METHOD B

PSEUDO-ERROR RATE EXTRAPOLATION

Figure 17



Plot of Pseudo Error Rate (P_p) versus Threshold Parameter K compared to plot of Actual Error Rate (P_e) versus SNR.

Figure 18

threshold parameters K, K_2, K_n is made to extend to the point K=0 (the point at which the modified threshold is equal to the actual threshold in the system under test). At this point, the estimated pseudo error rate is equal to the actual estimated error rate.

E. Since the pseudo error rates are larger than the actual error rates, the time to count the estimated error is much shorter than that required by a bit-by-bit counting process.

3.14.2 Discusion of Comparison Parameter

Parameter One - Precision of Test Data

The following factors were considered in determining the rating for Parameter One:

1. Precision of input tape (amount of jitter).
2. Precision of the delay circuit.
3. Discrepancy between the actual noise statistics and those assumed in determining the value of the threshold parameters.
4. Linearity of the pseudo error rate relationship with the threshold parameters.
5. Precision of data reduction.

Parameter Two - Conciseness of Results

The output BER is determined as result of the extrapolation of pseudo error rate.

Parameter Three - Data Reduction Required

This method requires the calculation of the extrapolation of the pseudo error rate curve to determine the estimated true error rate.

This involves solving for the logarithm of the pseudo error rate (P_p) for n numbers of modified thresholds, in order to construct the P_p versus modified threshold parameter (K) curve and then extending it to the point K=0.

Parameter Four - Relationship to Actual System Performance

Tests have shown this method to be accurate within a factor of 3 at an actual error rate of 10^{-6}. (Reference 9).

Parameter Five - Ease of Mechanization

This method requires a number of modified threshold decision circuits, comparators and a computer program to mechanize the pseudo error curve calculation and extrapolation.

4. COMPARISON OF CANDIDATE SYSTEMS

4.1 Rating of Comparison Parameters

In order to determine the relative value of each of the candidate criteria systems, some scheme of numerical rating must be used. The method used here is to assign a rating number for each comparison parameter for each of the candidate systems. This number ranges from 1 to 5, with number 1 representing the best. However, the numbers are not exclusive. That is, if for any particular parameter, such as Precision of Test Data, it is felt that two of the candidate systems result in about the same precision of data, each is assigned the same numerical rating. These rating numbers, when multiplied by the weighting values discussed in Section 4.2 yield a value for each parameter for each of the candidate systems. These parameter values, when added for each candidate system, give an indication of the overall rating for each system, with the system resulting in the lowest total value being considered the best.

Each parameter rating was assigned after consideration of the factors listed in Section 3 for each of the candidate systems. One of the basic difficulties in the rating scheme is the definition of the comparison parameters. For the purpose of this report the parameter "Precision of Test Data" is used more or less synonymously with accuracy. In assigning values to Parameter One, Precision of Test Data, each factor listed was considered for accuracy, and an "average" accuracy of each system was determined. Those systems with the best "average" accuracy were assigned a rating of 1, etc.

The parameter "Conciseness of Results" presents something of a problem of interpretation. Since all of the candidate systems provide a number value output for a given input, it could be said that they were equally concise.

The values arrived at in this report, however, are based on the relative amount and complexity of the computations required to achieve the final value for each of the candidate systems.

The parameters "Data Reduction Required" and "Ease of Simulation" are interdependent and are nearly redundant in concept. As used herein, "Data Reduction" is taken to indicate not only the total amount of data computation involved in a given system, but also the amount of computation required by the operator after the process has been completed. It can be seen, then, that "Data Reduction Required" and "Ease of Simulation" are somewhat reciprocal - a system with a low rating score for required data reduction could be expected to have a relatively high rating score for ease of simulation.

"Relation to Actual System Performance" is the most difficult of comparison parameters to rate for some of the candidate systems. Most of the systems require calibration by subjective testing to establish a relationship between the quantity derived as a result of the test and the desired result - percent word intelligibility or picture quality. Since some of the candidate systems have not been mechanized, the question of whether there is a monotonic relationship between the quantity derived from the test and a subjective evaluation can only be surmised.

4.2 Weighting of Comparison Parameters

After first determining the parameters to be used in evaluating performance criteria, a second question to be considered is that of the relative weighting of each of the parameters. Do all of them affect the overall value of performance equally, or are some of them more important than others? The answer to this question depends upon those who are using the criteria to evaluate the performance of a system.

In determining a weighting system, the precision of the test data appears an obvious choice as the most, or one of the most, important parameters and was thus assigned a value of 1. Relationship to actual system performance seems almost as important as precision of data and was also assigned a value of 1. Ease of simulation and the amount of data reduction required were determined to be of about equal importance, but were judged to be of less critical nature, and were assigned values of 2. The last parameter, "Conciseness of Results" was assigned a value of 3, not so much because it was felt to be of less importance than the others, but because the very nature of the candidate criteria systems is such that the results tend to represent averages of measured and computed values. In addition, the results of criteria systems for voice and video quality must be related to subjective evaluation.

This arbitrary numerical weighting, which is listed below, has been used in Table 2, which lists numerical values of the parameters for each of the candidate systems, as well as the total rating for each system:

Parameter	Weighting
Precision of Data	1
Ease of Simulation	2
Data Reduction Required	2
Relation to Actual System Performance	1
Conciseness of Results	3

Another weighting system, or none at all (assuming all the parameters to have equal importance) could be used, depending upon the needs of the users of the performance criteria. The intention here is to indicate how such a weighting scheme, when combined with a rating for each parameter assigned to each of the candidate systems can assist in determining the overall ranking of a particular candidate system. Table 3 lists the unweighted values, assuming that each of the parameters has equal weight.

Criteria Evaluation Weighted Parameters

Voice Systems

		Precision of Data	Conciseness of Results	Data Reduction Required	Relation to Actual System	Ease of Simulation	Total Values	Overall Ranking
A	AI-Equal Importance Bands	1	3	4	1	2	11	1
B	AI-Discrete Frequency	1	3	4	5	2	15	3
C	Speech SNR-Analog	2	6	6	2	6	22	5
D	Cross Correlation	2	6	6	5	8	27	7
E	Mean Squared Error	2	9	2	3	8	24	6
F	Speech SNR-Digital	2	6	2	3	8	21	4
G	Bit-by-Bit Comparison	1	3	2	4	4	14	2

Video Systems

		Precision of Data	Conciseness of Results	Data Reduction Required	Relation to Actual System	Ease of Simulation	Total Values	Overall Ranking
A	Picture SNR	1	3	2	1	2	9	1
B	Equal Importance Bands	3	6	6	5	6	26	5
C	Cross Correlation	2	6	6	5	6	25	4
D	Mean Squared Error	3	9	6	3	4	25	3
E	Bit-by-Bit Comparison	2	6	6	3	6	23	2

Digital Data Systems

		Precision of Data	Conciseness of Results	Data Reduction Required	Relation to Actual System	Ease of Simulation	Total Values	Overall Ranking
A	Bit-by-Bit Comparison	1	3	2	1	2	9	1
B.	Pseudo Error Extrapolation	2	3	2	3	4	14	2

TABLE 2

Criteria Evaluation Unweighted Parameters

Voice Systems

		Precision of Data	Conciseness of Results	Data Reduction Required	Relation to Actual System	Ease of Simulation	Total Values	Overall Ranking
A	AI-Equal Importance Bands	1	1	2	1	1	6	1
B	AI-Discrete Frequency	1	1	2	5	1	10	3
C	Speech SNR-Analog	2	2	3	2	3	12	4
D	Cross Correlation	2	2	3	5	4	16	6
E	Mean Squared Error	2	3	1	3	4	13	5
F	Speech SNR-Digital	2	2	1	3	4	12	4
G	Bit-by-Bit Comparison	1	1	1	4	2	9	2

Video Systems

		Precision of Data	Conciseness of Results	Data Reduction Required	Relation to Actual System	Ease of Simulation	Total Values	Overall Ranking
A	Picture SNR	1	1	1	1	1	5	1
B	Equal Importance Bands	3	2	3	5	3	16	5
C	Cross Correlation	2	2	3	5	3	15	4
D	Mean Squared Error	3	3	3	3	2	14	3
E	Bit-by-Bit Comparison	2	2	3	3	3	13	2

Digital Data Systems

		Precision of Data	Conciseness of Results	Data Reduction Required	Relation to Actual System	Ease of Simulation	Total Values	Overall Ranking
A	Bit-by-Bit Comparison	1	1	1	1	1	5	1
B	Pseudo Error Extrapolation	2	1	1	3	2	9	2

TABLE 3

5. SUMMARY AND CONCLUSIONS

5.1 General Considerations

Some of the difficulties pertaining to the assignment of numerical ranking of the different criteria systems which have been pointed out in this report are:

1. The difficulty associated with defining the parameters.
2. Use of a weighting system for the comparison parameters; and the weighting values assigned to each parameter if used.
3. Rating each system for each parameter.
4. The uncertainties associated with untried methods.

One problem not treated in the report is that of differentiating between criteria and methods used to test a system to meet that criteria. The approach taken in this report is to consider criteria systems or methods. This leads to some duplication as regards criteria - for instance, the two methods discribed to achieve voice-articulation index. Since the two methods result in different ratings, it is felt that this approach is of value.

An additional factor not considered in the final selection of a performance criteria is the ease with which the criteria can be used by the systems or equipment designer in developing the design of the system. For example, the performance of a portion of a system might be specified in terms of a criterion which is accurately related to system performance, but which is very difficult for the designer to compute. Although this computational difficulty is not an overriding consideration in criteria selection, it could be considered in assigning a weight to the trade-off parameter "Ease of Simulation."

5.2 Voice Systems

Based on both the unweighted and weighted parameters, the articulation index method using the equal importance frequency bands appears to be the best method for specifying voice performance. This ranking may result because the method is one which has been used most in the past with proven results. The alternate method B is downgraded primarily because of its unproven relation to actual system performance.

References to Tables 2 and 3 reveals the composition of the ratings of the other systems. It is of interest to note that the weighted and unweighted overall ranks are quite similar. The important feature of the Tables is that is assists users with different requirements to determine which of the systems would be more suited to his needs. For instance, if the amount of data reduction required were not of prime importance to a particular user, he could downgrade or ignore this particular parameter and re-compute the total for each system, thus arriving at a rating suited to his requirements.

5.3 Video Systems

The picture SNR method using a standard noise weighting is the clear choice based on both the weighted and unweighted parameter systems. Again, the fact that this scheme has had considerable proven experience undoubtedly affected the results.

5.4 Data Systems

There is really only one choice for the digital systems criteria - bit error rate based on comparison of input and output. Method B is actually a sub-method, and under the circumstances of very low BER conditions could be the number one choice.

5.5 Recommendations

It is recommended that continued study and possibly hardware testing be made for at least two criteria systems in each category of voice, video, and digital data. The criteria system with the highest rank should obviously be considered for further investigation and mechanization. The choice of the other method in each category should not necessarily be restricted to the second ranked system, but could depend on the desires and needs of the user.

6. REFERENCES

1. Project Technical Report, "Survey of Performance Criteria for Voice, Television, and Digital Data Transmission Systems", TRW Document No. 17618-H123-RO-00, 30 March 1971.

2. Draft Report "Technical Characteristics of Systems Providing Communication and/or Radio Determination Using Satellite Techniques for Aircraft and/or Ships," Document No. USSG IV/W-1142, 1 August 1970.

3. Technical Report, "Development of a Speech-to-Noise Ratio Measurement Utilizing Analog Techniques," Philco-Ford Document PHO-TN248, 6 September 1968.

4. Peter T. Roth, "Effective Measurements Using Digital Signal Analysis," IEEE Spectrum, April 1971, pp 62-70.

5. Technical Report "Development of a Speech-to-Noise Ratio Measurement Utilizing Digital Techniques," Philco-Ford Document PHO-TN228, 12 June 1968.

6. EIA Standard RS-250-A, "Electrical Performance Standards for Television Relay Facilities," Electronics Industries Association, February 1967.

7. J. M. Barstow, H. N. Christopher, "The Measurement of Random Video Interference to Monochrome and Color Television Pictures," Proc. AIEE, pp. 313-320, November 1962.

8. J. O. Weston, "Transmission of Television by Pulse Code Modulation," Electrical Communication, Vol. 42, No. 2, 1967.

9. D. J. Gooding, "Performance Monitor Techniques for Digital Receivers Based on Extrapolation of Error Rate," IEEE Transactions on Communication Technology, Vol. COM-16, No. 3, June, 1968.

10. Harvey Fletcher, "Speech and Hearing in Communication," pp. 278-302, D. Van Nostrand Company, Inc, 1953.