ESD-TR-67-516

(Final Report)

SPEECH COMMUNICATION INDEX METER: FURTHER APPLICATIONS AND IMPROVEMENTS

Michael H. L. Hecker
Gottfried von Bismarck
Carl E. Williams

26 July 1967

DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

(Prepared under Contract No. AF 19(628)-5874 by Bolt Beranek and Newman Inc., Cambridge, Mass.)

ESD-TR-67-516

(Final Report)

SPEECH COMMUNICATION INDEX METER: FURTHER APPLICATIONS AND IMPROVEMENTS

Michael H. L. Hecker
Gottfried von Bismarck
Carl E. Williams

26 July 1967

DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

(Prepared under Contract No. AF 19(628)-5874 by Bolt Beranek and Newman Inc., Cambridge, Mass.)

Foreword

This document was prepared by Bolt Beranek and Newman Inc., of Cambridge, Massachusetts, and is the final report under Air Force Contract No. AF 19(628)-5874. This contract was sponsored under Project 2808 and was monitored by Mr. Allen C. Busch of the Decision Sciences Laboratory, Electronic Systems Division, L. G. Hanscom Field, Bedford, Mass.

This Technical Report has been reviewed and approved.

James S. Duva
JAMES S. DUVA
Technical Director
Decision Sciences Laboratory

SPEECH COMMUNICATION INDEX METER:
FURTHER APPLICATIONS AND IMPROVEMENTS

Abstract

To study the feasibility of using the Speech Communication Index Meter (SCIM) to evaluate time-varying communication systems, recordings were made of the transmission characteristics of a troposphere-scatter system. At many specific points in time in these recordings, Articulation Indexes were calculated and intelligibility scores were obtained from listeners with the aid of a special test procedure. For most points, the intelligibility score could be reasonably well predicted from the Articulation Index. This finding was interpreted as indicating that SCIM is potentially capable of evaluating time-varying systems.

In a second study, SCIM was used to predict the intelligibility of peak-clipped speech. For a wide range of signal-to-noise ratios and clipping levels, the performance of SCIM was compared with intelligibility test results reported in the literature. After a minor modification, SCIM computed reliable SCI's which could be related to the published intelligibility data. This study demonstrated that SCIM could accurately evaluate communication systems employing peak clipping.

SCIM was also modified to reduce the computation time, and to improve the synchronization between the signal generator and the analyzer. In addition, various other circuit changes were implemented to improve the performance of SCIM.

TABLE OF CONTENTS

LIST OF FIGURES

1. EVALUATION OF TIME-VARYING COMMUNICATION SYSTEMS

1.1 Introduction

The transmission characteristics and level of performance of many types of speech-communication systems vary with time. In high-frequency radio systems, for example, multipath interference is a common occurrence. Disturbances in the ionosphere may change the phase relations among radio signals that reach the receiver over different paths, and these changes produce rapid fading of the audio signal and fluctuations in the background noise. Fading is also encountered with mobile radio equipment, where the path between transmitter and receiver is constantly changing. Another type of time-varying communication system is the troposphere-scatter system. The tropo system provides radio communication over moderate distances that lie between the capabilities of ultrahigh-frequency short-range systems and low-frequency long-range systems.[4] The transmission characteristics of tropo systems have been studied in detail at the National Bureau of Standards.[17]

The performance of time-varying speech-communication systems is usually specified in terms of the length of time that a particular signal-to-noise (S/N) ratio is available. This S/N ratio is determined from a short-term analysis of the received signal, and the time availability is determined from a long-term analysis. Because S/N ratio is a measure which is related to the intelligibility of the received speech, it is a valid indicator of system performance. However, wide-band S/N ratio is insensitive to spectral changes in the received noise which may also affect intelligibility.

A more accurate indicator of performance can be obtained by measuring the S/N ratio in each of several frequency bands which are equally important to speech intelligibility. Such multi-band measurements are required for the calculation of the Articulation Index (AI),[8] from which the intelligibility of communication systems with non-varying characteristics may be estimated. Since the design of the Speech Communication Index Meter (SCIM)[9] is based upon the concept of AI, it would appear that SCIM is potentially capable of evaluating time-varying systems more accurately than is possible with presently available methods.

Before SCIM can be employed in this manner, the relation between the Speech Communication Index (SCI) and speech intelligibility must be established for time-varying systems. The computed SCI's must be somehow compared with how well speakers and listeners can communicate over the system. Such a comparison is complicated by two factors: (1) The SCI depends strongly on the time interval over which the "speech" and noise are integrated, and (2) no standard test procedures are available for measuring speech intelligibility. In this study, a test procedure was developed and used to investigate how the characteristics of a time-varying system should be measured by a device like SCIM.

1.2 Approach

The usual procedure for measuring speech intelligibility is as follows: One or more trained speakers read several 50-item word lists to a group of listeners over the communication system to be evaluated. Depending on the type of test, the listeners are required either to write down each test word heard or to respond to each item by selecting the test word from a number of given alternatives. The percentage of words that are correctly identified constitutes the test score. Because some words in a given list are inherently more vulnerable to distortion or interference (and therefore more difficult to transmit over a system) than other words, the word lists are fairly well balanced with respect to the distribution of particular phonemes and overall difficulty. The system parameters are assumed to be constant during the presentation of the word lists.

If a time-varying system were evaluated in the manner just described, the intelligibility of a given test word would depend not only on the relative difficulty of the word but also on the system characteristics at the time the word is transmitted. The overall test score obtained would depend on how the system characteristics happen to be synchronized with the order of the test items, and would therefore be difficult to interpret.

To resolve this problem, a new test procedure was developed: The transmission characteristics of a typical tropo system were recorded on magnetic tape. Three 3-minute Program Tapes were prepared from this recording. Twenty-five 3-minute Speech Tapes, each containing a sequence of 60 test words, were also prepared. The words on the Speech Tapes were spaced exactly 3 seconds apart. A given Program Tape was reproduced 25 times, each time together with a different Speech Tape, to provide 25 words for the calculation of each of 60 intelligibility scores. Thus, every 3 seconds an intelligibility score was obtained which was based on listener responses to 25 words. Even lists of only 25 words, often called half lists, offer some degree of balance with respect to word difficulty.

Each Program Tape was also reproduced through 1/3-octave band filters to provide spectral information from which AI's could be computed at the same 60 points in time where intelligibility scores were obtained. These AI's, which represent scores that might be computed by a device like SCIM, were then converted to equivalent intelligibility scores. These latter scores were compared with the scores obtained from listeners to determine how closely the intelligibility of the tropo system could be predicted from AI's.

Various arguments can be advanced as to why the predicted scores might differ from the obtained scores. One argument is based on the fact that a test word "samples" the characteristics of the tropo system for the entire duration of the word, whereas the AI is computed at one particular point in time. Results which support this argument will probably also provide useful information regarding the optimum duration of SCIM measurements. Another argument involves a phenomenon known as temporal masking.[2,12,13] Temporal masking has been demonstrated in experiments such as those illustrated in Fig. 1-1.

A probe tone occurring shortly before or after an intense burst of white noise will be masked by the noise burst even though the two signals are presented sequentially to the listener. The level of masking depends primarily on the intensity of the noise burst and on the interval between the two signals. Temporal masking appears to be negligible for intervals greater than 100 milliseconds.

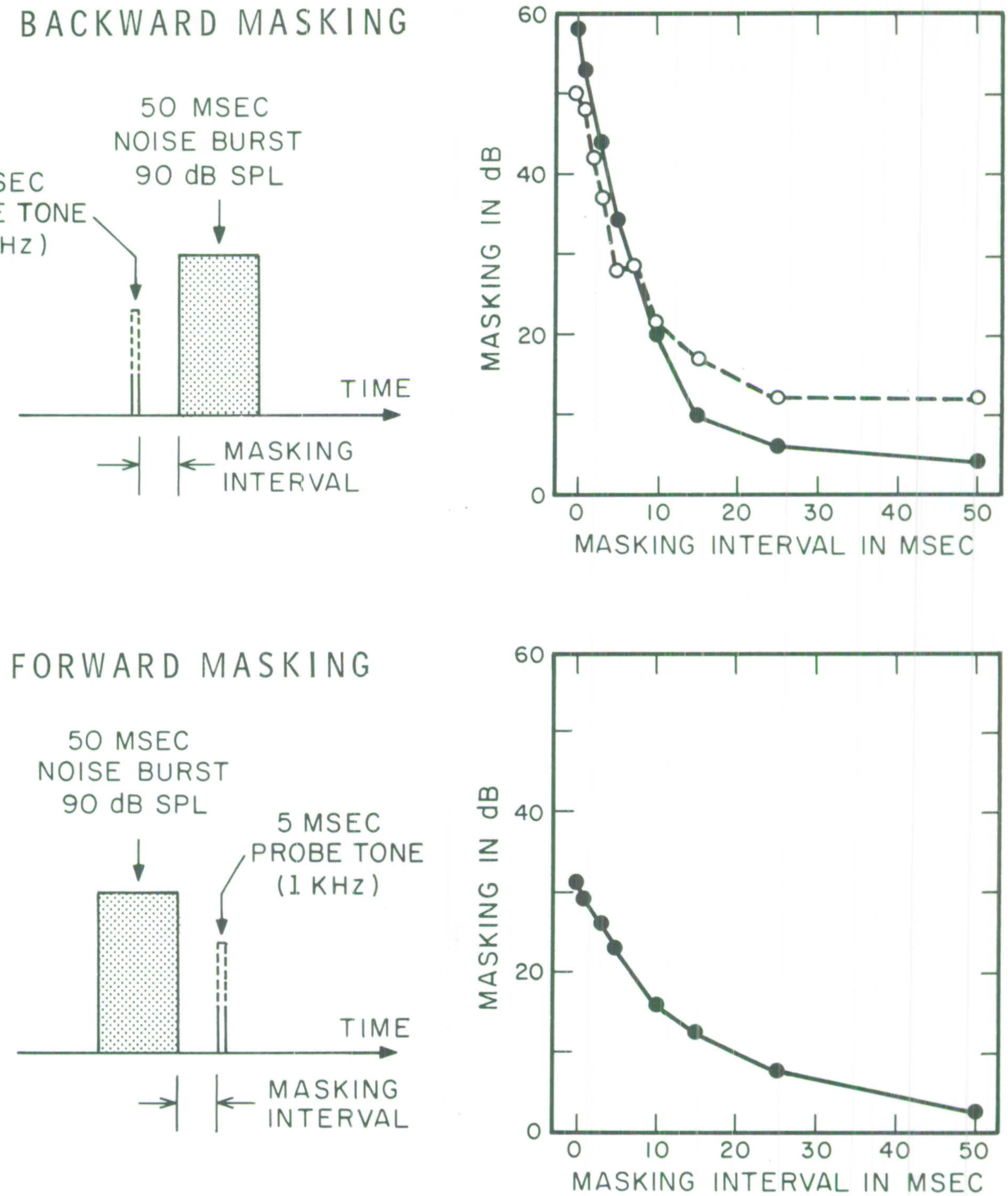

FIG. 1-1 TEMPORAL MASKING OF PROBE TONE OCCURRING BEFORE OR AFTER BURST OF WHITE NOISE. DATA FROM ELLIOTT 1962 (●——●) AND PICKETT 1959 (○--○).

How temporal masking could influence the results of the present study is outlined in the hypothetical situation shown in Fig. 1-2. A number of test words, assumed to be equally difficult, are transmitted over a time-varying system at approximately the same level. These words are received at different levels, together with widely fluctuating background noise. At the location of each test word, the speech and noise spectra are determined and the AI is calculated. The AI values are converted to predicted test scores. Test scores are also obtained from listeners.

Suppose that Test Words A, B, and C are received at the same level. Because Words B and C occur during an interval characterized by a low noise level, their AI values and predicted scores are much higher than the AI value and predicted score for Word A. The obtained scores for these three words may, however, be very similar. Word B, occurring immediately after the cessation of a high noise level, may be subjected to forward masking. Similarly, Word C may be subjected to backward masking.

Considering the critical conditions that must exist for temporal masking to occur, the effect is not likely to be as pronounced as is illustrated in Fig. 1-2. Nevertheless, if some results can be interpreted in terms of this argument, then a correction for temporal masking should be included in the design of SCIM.

1.3 Preparation of Program Tapes

A block diagram of the instrumentation used to record the Program Tapes is shown in Fig. 1-3(a). An experimental Air Force tropo system was used, which consisted of a transmitter and a receiver separated by 200 miles. Although the system could

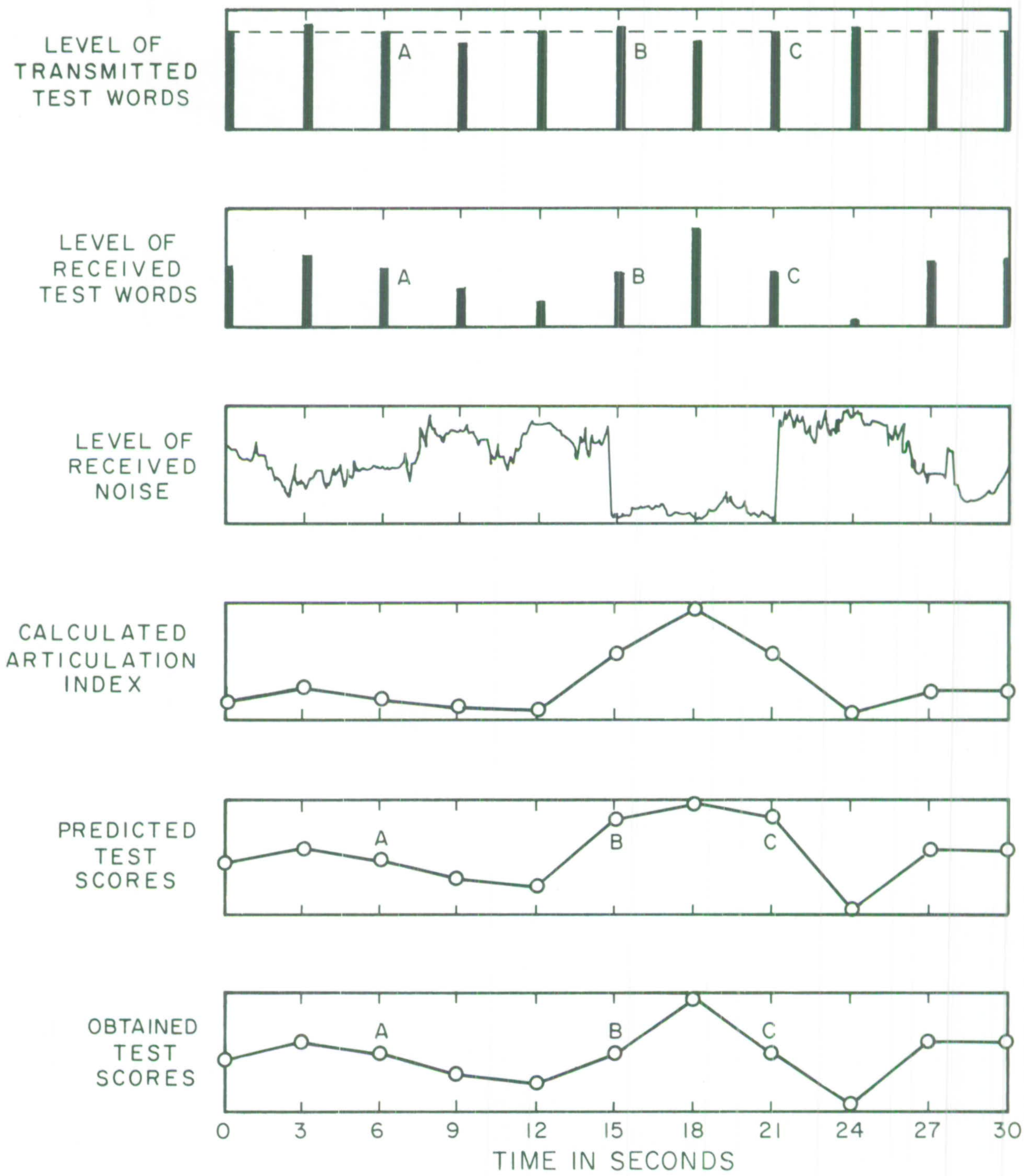

FIG. 1-2 HYPOTHETICAL CHARACTERISTICS OF TIME-VARYING COMMUNICATIONS SYSTEM. INTELLIGIBILITY OF TEST WORDS B AND C IS LOWER THAN PREDICTED FROM ARTICULATION INDEX BECAUSE OF TEMPORAL MASKING.

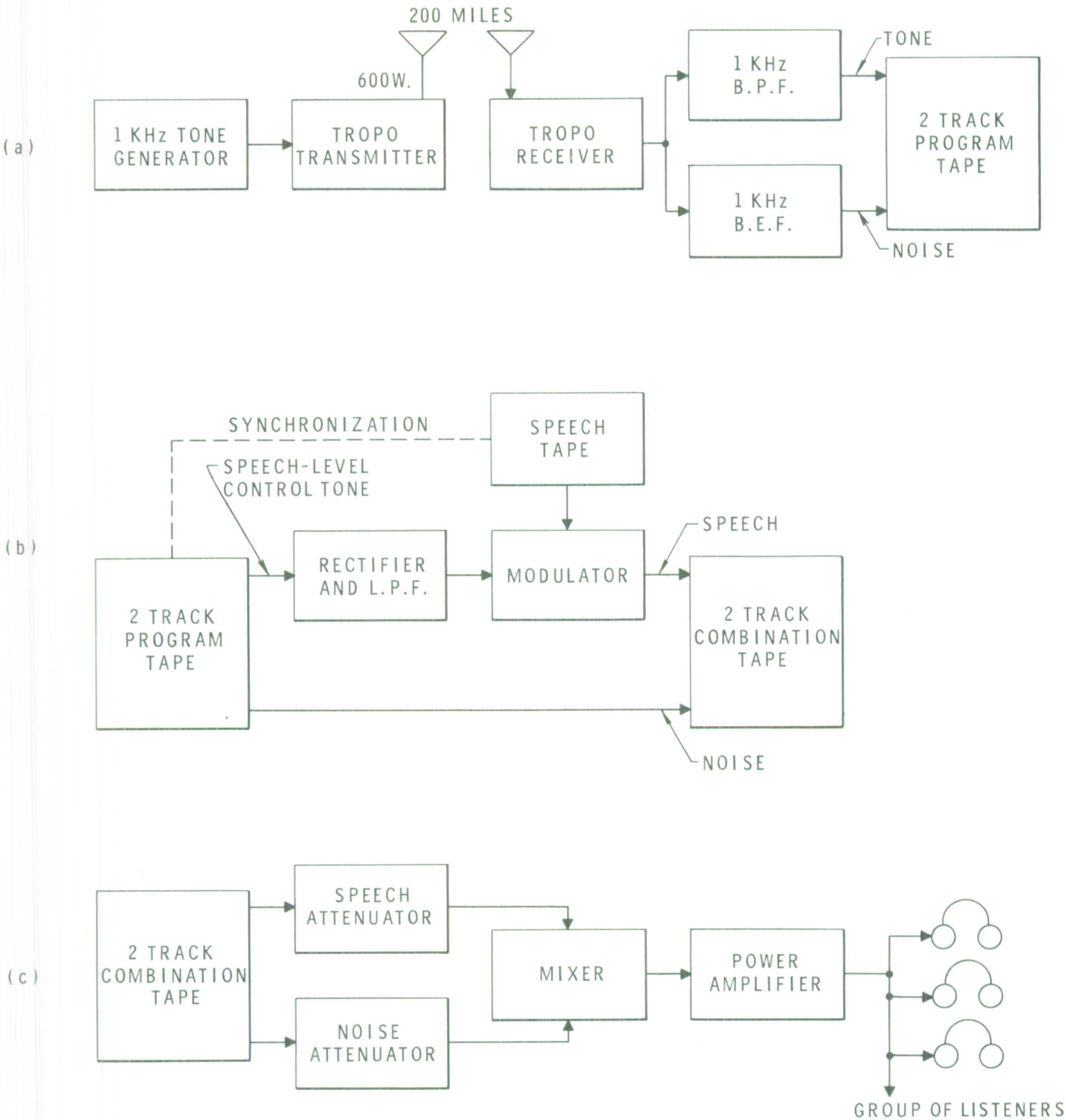

FIG. 1-3 BLOCK DIAGRAMS OF INSTRUMENTATION USED TO (a) RECORD PROGRAM TAPE, (b) PREPARE COMBINATION TAPE, AND (c) REPRODUCE COMBINATION TAPE.

provide several modes of diversity operation, these were not employed. A constant-level 1 KHz tone was applied to one of the voice channels at the transmitter to represent the speech signal ordinarily transmitted over the system. The output of the receiver was processed by two filters. One filter was a sharp-skirted bandpass filter centered on the tone, and the other filter was a complementary sharp-skirted band-elimination filter. The output of the bandpass filter served to describe how the level of a received speech signal might vary as a function of time. The output of the band-elimination filter described how the level and spectrum of the received noise varied. The two filter outputs were recorded simultaneously with a two-channel tape recorder.

The instrumentation just described is based on the assumption that the spectrum of a received speech signal remains constant when the level of the signal changes. The validity of this assumption was determined in an earlier experiment with another tropo system. Nearly all tropo systems employ frequency modulation and transmit a wide-band channel from which a number of voice channels are derived with multiplex equipment. A constant-level 1 KHz tone was applied simultaneously to several voice channels at the transmitter, and the levels of the received tones were recorded with a multi-channel chart recorder. It was noted that the recorded levels tended to vary in unison during intervals of fading. Since the tones correspond to different frequencies in the wide-band signal, it was concluded that the spectrum of the wide-band signal is not appreciably altered when level changes occur.

Various transmitter power settings were employed in recording the output of the tropo receiver. After listening to the recordings obtained and observing graphic-level tracings of the time-varying noise, a 30-minute recording made at a power setting of 600 watts was selected. This recording was reduced to three 3-minute sections to furnish the three Program Tapes which were used in this study. The editing was accomplished so that different types of interference patterns were included and so that the test words (on the Speech Tapes) would occur in a variety of positions with respect to the interference. A short burst of white noise was inserted at the beginning of each Program Tape to facilitate accurate synchronization with the Speech Tapes.

1.4 Preparation of Speech Tapes

It was considered desirable to obtain as many intelligibility scores as possible for each Program Tape without extending the duration of the tape beyond 3 minutes. The duration of each Program Tape determines the length of a listening test, and past experience in intelligibility testing has indicated that listeners may become fatigued when a test lasts longer than 3-5 minutes. Furthermore, it was anticipated that the synchronization between Program Tapes and Speech Tapes could be maintained more accurately for 3-minute tapes than for tapes of longer duration.

On the basis of these considerations, the Modified Rhyme Test (MRT)[6] was selected. Since MRT words can be readily presented and marked by listeners at the rate of one word every 3 seconds, 60 intelligibility scores could be obtained for each 3-minute Program Tape. The speech material of the MRT consists of six 50-word lists which are approximately equally difficult. Each list is composed of 25 words in which the initial consonant is tested and 25 words in which the final consonant is tested.

For the present application, a rearrangement of this material appeared to be necessary. If the standard 50-word lists were to be used, 50 Speech Tapes would be required. The effort of preparing and administrating this many Speech Tapes seemed unwarranted. Also, considering that 60 intelligibility scores were to be obtained for each Program Tape, it seemed desirable to use more than six basic lists. It was therefore decided to construct twelve 25-word lists. In an effort to construct lists of equal difficulty, two sets of data obtained with the MRT were examined.

Data gathered during the development and evaluation of the MRT[7] consisted of scores for each of the 50 words in each list, averaged over 2 speakers, 18 listeners, and 6 S/N ratios. The 25 ensembles with initial variable consonants were divided into two groups (of 12 and 13 ensembles) that would have the same mean score. The 25 ensembles with final variable consonants were similarly divided into two groups. Mean scores were then computed for the resulting 12- and 13-item lists. In order to achieve similar scores across the 12 related lists (variable consonant in same position), some words were permuted within their respective ensembles. Twelve 25-item lists with similar mean scores were thus obtained.

Other available data[21] consisted of scores for each word in each list, averaged over 1 speaker, 13 listeners, and 2 S/N ratios. Using these data, the twelve 25-item lists were again examined for equal difficulty. The original grouping of ensembles was changed slightly, and more words were permuted within ensembles, to achieve similar scores across the 12 related lists. At the same time, care was taken to maintain similar scores across related lists for the first set of data. While an attempt was made to balance the lists with respect to the occurrence of particular phonemes, the requirement for equal difficulty was given priority.

The revised 25-item lists, referred to as Half Lists A-L, are given in Appendix 1. The relative difficulty of these lists was determined experimentally in another study.[22] The scores that were obtained for each list at each of 5 S/N ratios are given in Appendix 2. Although some lists appeared to be slightly more difficult than others, the balance that was achieved was considered adequate.

Five randomizations of the 12 half lists were prepared to provide a half list at each of 60 points in time. The first word in each of the 60 randomized half lists was identified as a word to be included in Speech Tape 1. In the same manner, the remaining words in each half list were marked for inclusion in Speech Tapes 2-25. Each Speech Tape would thus consist of 60 words, but these words do not constitute any particular half list. The relation between the contents of a given Program Tape and the test words of the 25 Speech Tapes is illustrated in Fig. 1-4.

The 25 Speech Tapes were then recorded by an adult male speaker with considerable experience in reading test materials. The recordings were made with high-quality equipment in a sound-treated studio. A flashing light signal was used during

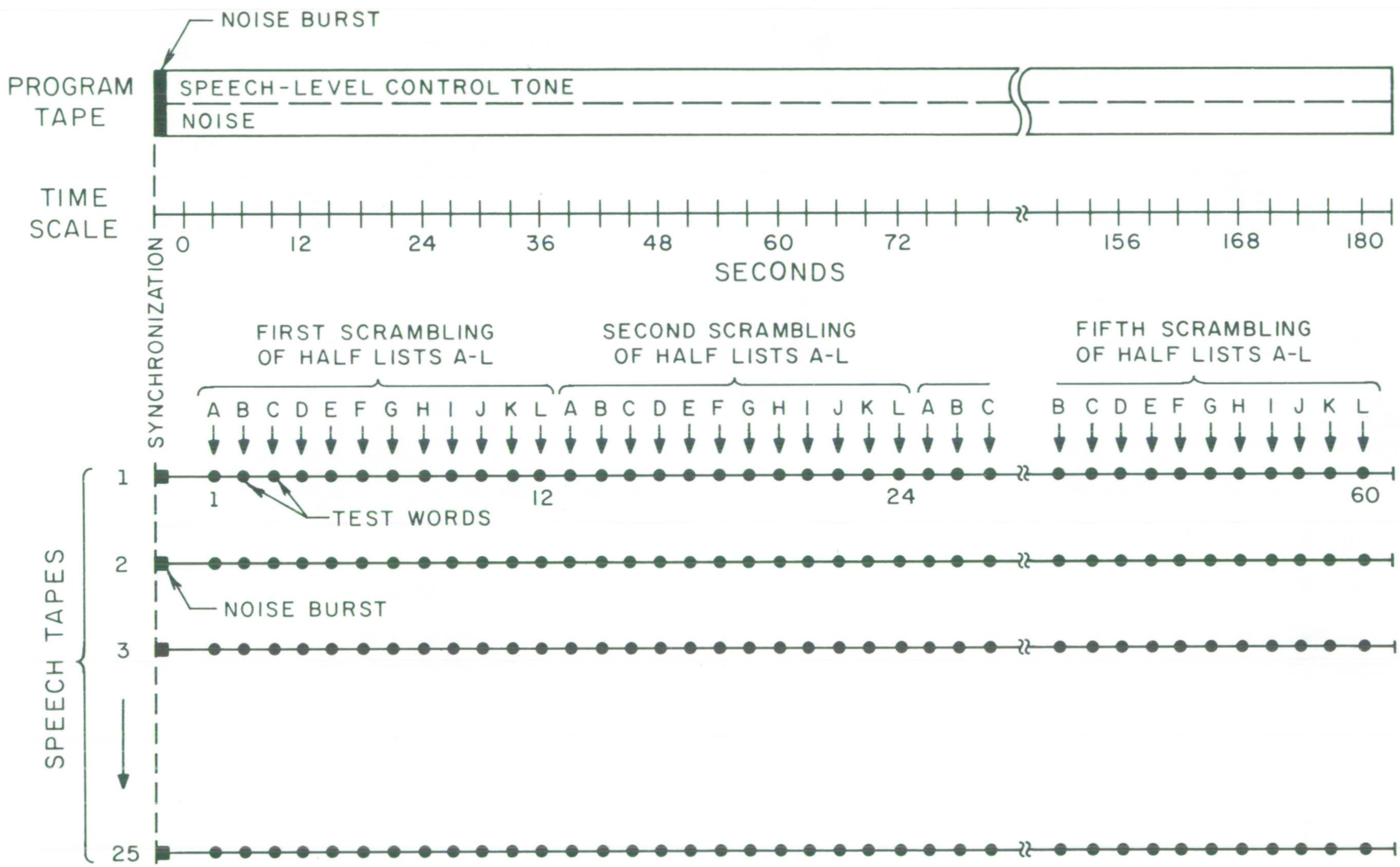

FIG. 1-4 DIAGRAM SHOWING HOW A PROGRAM TAPE WAS COMBINED WITH 25 SPEECH TAPES TO PROVIDE A 25-WORD HALF LIST AT EACH OF 60 POINTS IN TIME OF PROGRAM TAPE.

the recording to help the speaker maintain the required 3-second interval between words. This light signal was controlled from a calibrated pulse generator. Although the words were not embedded in a carrier phrase and no item numbers were recorded, the speaker tried to read all words for a given Speech Tape with the same vocal effort. When the speaker made a mistake, the entire Speech Tape in which the mistake occurred was recorded again.

To check the accuracy of the timing of the words, graphic-level tracings were made of each Speech Tape. These tracings were aligned with a time scale that indicated the exact positions where the words should occur. Corrections were calculated for those words which occurred more than 100 milliseconds too early or too late, and these corrections were then used to edit the Speech Tapes. A short burst of white noise was inserted at the beginning of each Speech Tape to facilitate accurate synchronization with the Program Tapes.

1.5 Preparation and Administration of Combination Tapes

To provide synchronism between a Program Tape and a Speech Tape, two tape recorders were coupled electrically so that the tapes could be started simultaneously. Visual markers on both sets of tapes, positioned just before the noise bursts, were used to accomplish the initial alignment. During the subsequent reproduction of the two tapes, a progressive deterioration of synchronization was observed in some cases. This problem could be related to a dimensional instability of the tape material (acetate base), and was resolved by using a special variable-frequency power source to operate the capstan motor of one of the tape recorders. Because of this requirement, it was considered necessary to combine the Program Tapes and Speech Tapes prior to their presentation to listeners.

A block diagram of the instrumentation used to prepare the Combination Tapes is shown in Fig. 1-3(b). The 1 KHz tone reproduced from the Program Tape was rectified and low-pass filtered, and the resulting d-c signal was applied to a modulator. This modulator was used to vary the level of the speech signal reproduced from the Speech Tape. The output of the modulator and the time-varying noise reproduced from the Program Tape were recorded with a two-channel tape recorder. This arrangement allowed independent settings of the speech and noise levels during the administration of the Combination Tapes.

Since 3 Program Tapes had to be combined with 25 Speech Tapes, a total of 75 Combination Tapes were prepared. These Combination Tapes were ordered according to the test design given in Appendix 3 and administered to a group of listeners during five test sessions. The listeners were provided with response forms of the type shown in Appendix 4. A block diagram of the instrumentation used for administering the Combination Tapes is shown in Fig. 1-3(c). The speech attenuator was adjusted so that the maximum sound pressure level of the speech (corresponding to the highest level of the 1 KHz tone on the Program Tapes) was 71 dB re 0.0002 microbar under the earphones. Similarly, the noise attenuator was adjusted so that the maximum sound pressure level of the noise was 78 dB re 0.0002 microbar. The results obtained from 10 listeners are tabulated in Appendix 5.

For each Program Tape and Speech Tape, Appendix 5 shows the number of errors that were made by 10 listeners at each point in time (referred to as a word location). Also shown for each Program Tape are the mean and standard deviation of the errors made at each word location on all Speech Tapes. The means are also expressed as intelligibility scores in percent, corrected for chance.* Large standard deviations (over 3.20) at particular word locations suggest that rapid changes in the level of the noise occurring at those times may have disturbed the normal balance in difficulty between words with initial consonants and words with final consonants under test. If, for example, the noise level increased rapidly during the presentation of the words constituting a half list, the initial consonants would be masked less than the final consonants. As a result, words with initial consonants under test would be more intelligible than words with final consonants under test. The intelligibility scores shown for such word locations are therefore less reliable.

1.6 Calculation of Articulation Index

In order to calculate an AI at each word location in each Program Tape, it was first necessary to subject the Program Tapes to a detailed spectral analysis. The instrumentation used to obtain the required spectral data is represented in Fig. 1-5. A graphic-level tracing was prepared of the 1 KHz tone to provide information regarding the speech level at each word location.

---

* All obtained intelligibility scores were corrected for chance with the formula:

$$\text{Corrected Score} = \text{Items Correct} - (\text{Items Incorrect})/5$$

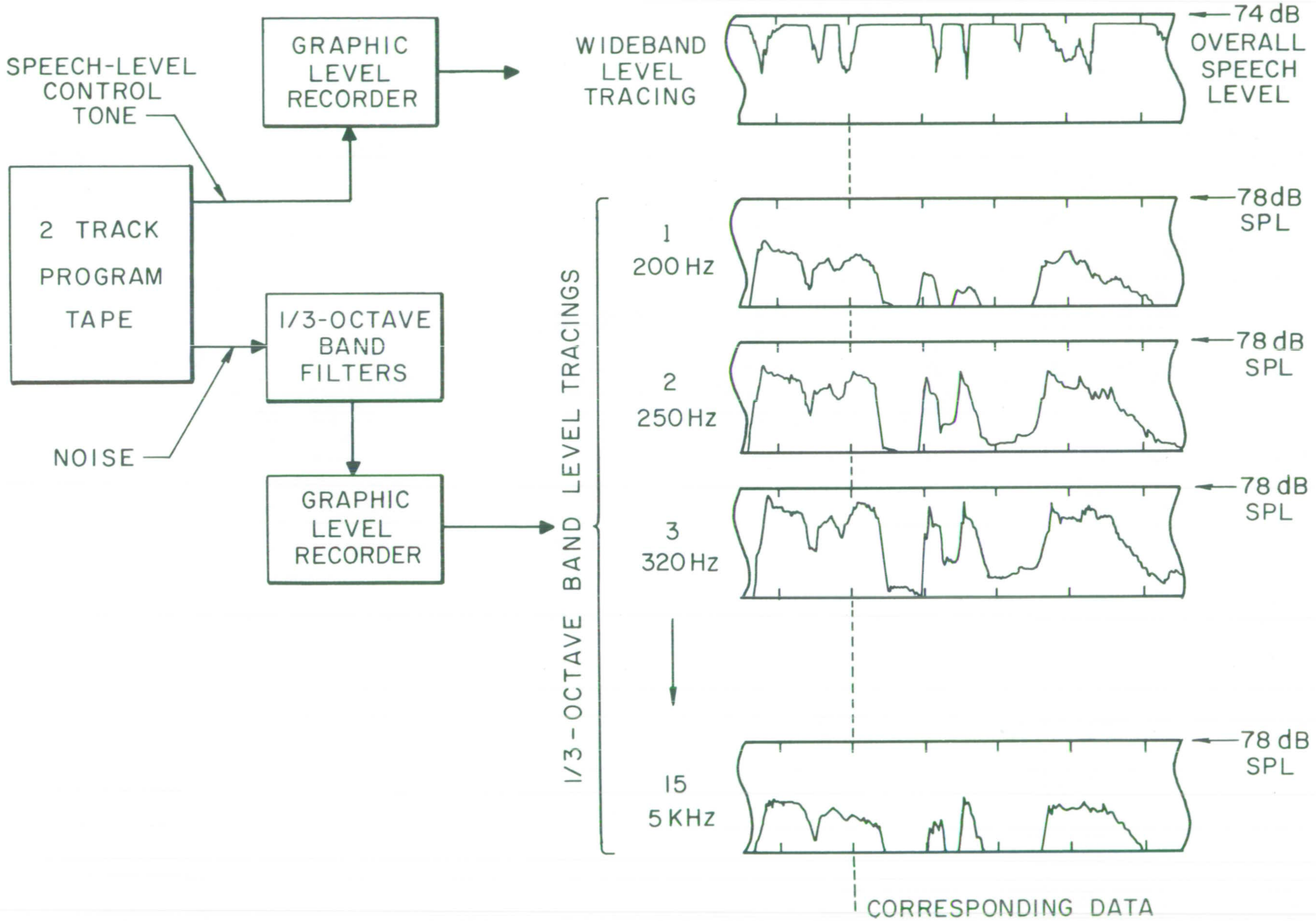

FIG. 1-5 BLOCK DIAGRAM OF INSTRUMENTATION USED TO MEASURE OVERALL SPEECH LEVEL AND NOISE SPECTRUM FOR CALCULATION OF ARTICULATION INDEX.

This tracing was calibrated in terms of the maximum level at which the speech had been presented to the listeners. The time-varying noise was reproduced sequentially through fifteen 1/3-octave band filters. The center frequencies of these filters ranged from 200 Hz to 5 KHz. A graphic-level tracing was prepared from the output of each filter; these tracings were calibrated in terms of the maximum overall level at which the noise had been presented.

By means of a time scale common to all tracings, the speech level and the noise spectrum at each word location could now be uniquely determined. The tracings were read at each word location, and from this information the AI's were calculated. An example of such a calculation is given in Fig. 1-6. For this particular word location (I:52), the tracing of the 1 KHz tone was read as 69 dB. This is 5 dB below the overall level which corresponds to a theoretical reference speech spectrum.* The "measured" speech spectrum is therefore shown 5 dB below the reference speech spectrum. The measured noise spectrum is determined directly from the 15 graphic-level tracings of the noise. For this example, the AI was calculated as shown in Appendix 6.

The AI's calculated in this manner for all word locations in each of the three Program Tapes are tabulated in Appendix 7. The calculations were carried out with the aid of a digital computer. Using the relation between AI and MRT scores shown

---

* As required by the procedure for calculating AI, the reference speech spectrum includes a 12 dB correction for speech peaks.

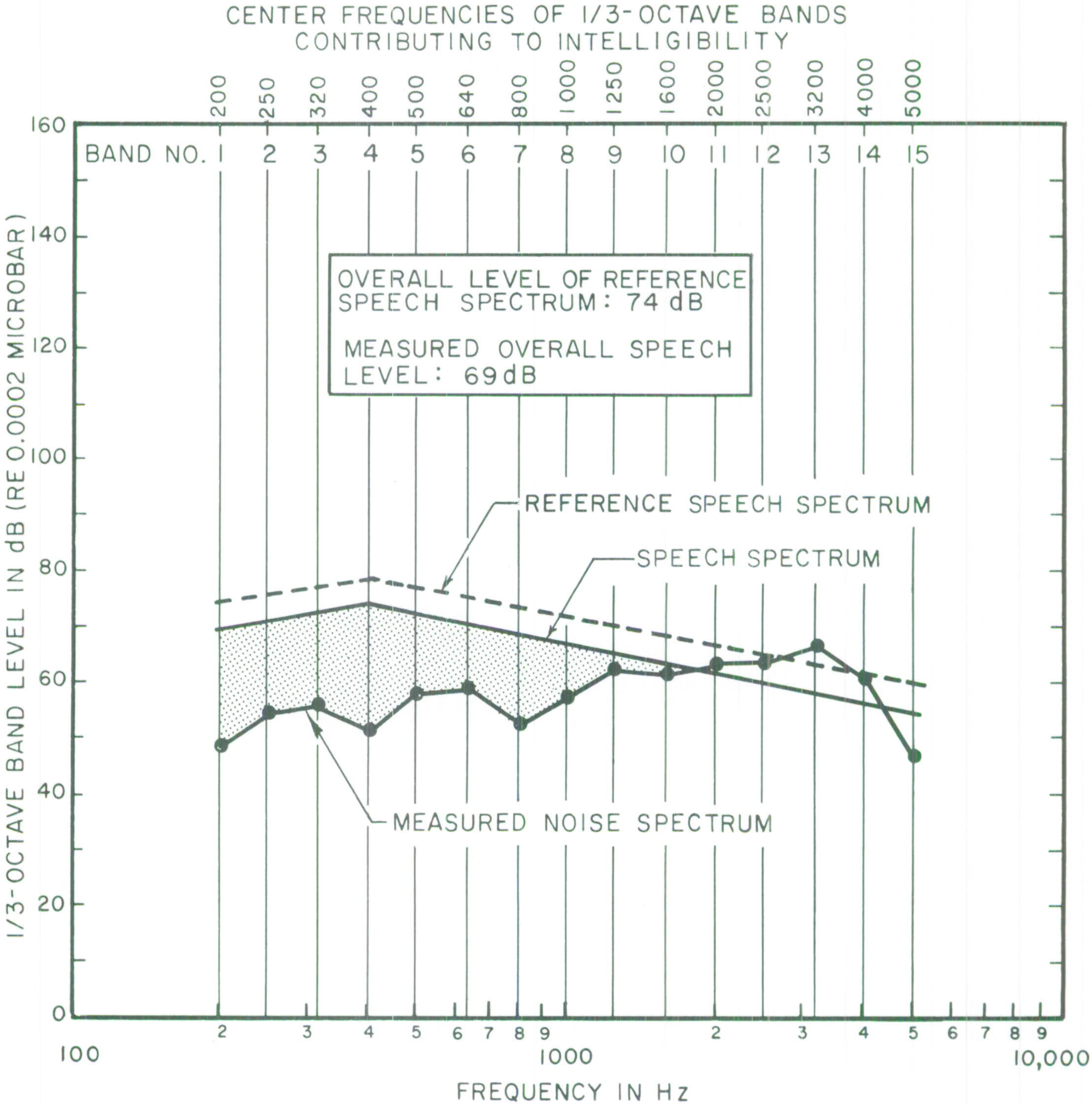

FIG.1-6 CALCULATION OF ARTICULATION INDEX FOR WORD LOCATION 52 OF PROGRAM TAPE I. SHADED AREA INDICATES CONTRIBUTION TO INTELLIGIBILITY.

in Fig. 1-7 (solid curve), the calculated AI's were then converted to equivalent MRT scores. The data shown in Fig. 1-7 were obtained in a related study[22] which deserves some comment.

The related study was concerned with the masking of speech by aircraft noise. Recordings of various aircraft flyovers were assembled into three 3-minute Program Tapes which were similar to the Program Tapes prepared in the present study. Both studies employed the same 25 Speech Tapes and the identical test procedure for obtaining intelligibility scores. In the aircraft study, AI's were also computed at the 60 word locations of each Program Tape. One important difference between the two studies was the rate at which the noise varied. Whereas the overall noise level in the present study could change by as much as 25 dB within an interval of 400 milliseconds (average duration of test words), the overall level of the aircraft noise never changed by more than 5 dB during the same interval. Another difference was the fact that in the aircraft study the Speech Tapes were presented to listeners at a uniform level. Considering these differences, it was assumed that a more reliable relation between AI and MRT scores could be obtained from the aircraft study than from the present study.

As indicated in Fig. 1-7, the aircraft study provided two relations; one was obtained with time-varying noise (aircraft flyovers), and the other with steady-state noise. For a given AI, the time-varying noise masked the test words to a lesser extent than the steady-state noise. This difference in masking may be attributable to such factors as listener attention, adaptation, and fatigue. The relation obtained with time-varying noise was considered more appropriate for the present study than the other relation. The use of this relation, however, excluded

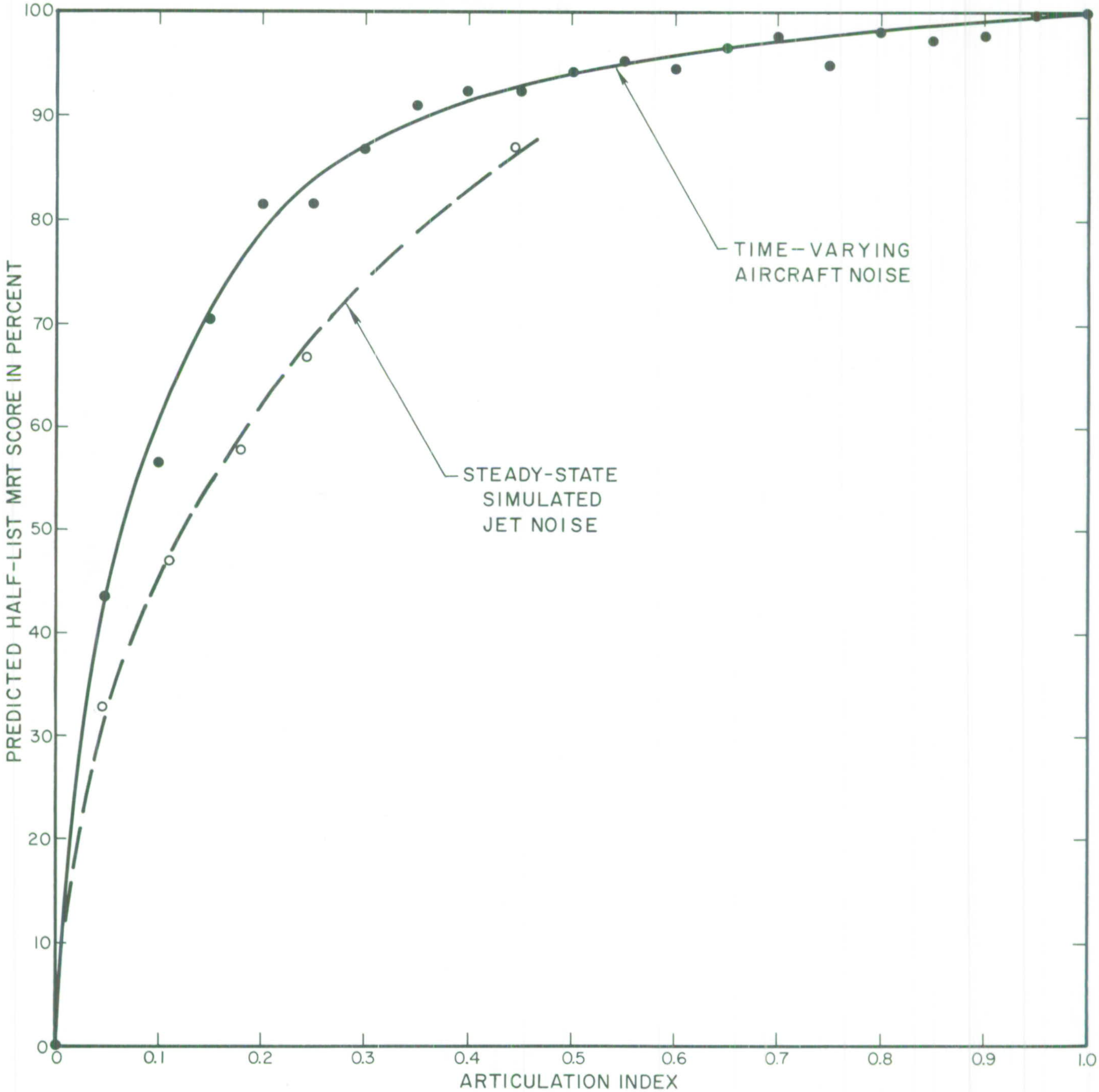

FIG. 1-7 RELATION BETWEEN ARTICULATION INDEX AND HALF-LIST MODIFIED RHYME TEST SCORE (CORRECTED FOR CHANCE) FOR TIME-VARYING AND STEADY-STATE AIRCRAFT NOISE. DATA FROM WILLIAMS ET AL. 1967.

an investigation of the above mentioned factors which might be relevant to the design of SCIM.

1.7 Comparison Between Obtained and Predicted Scores

For each word location of each Program Tape, the intelligibility score obtained from the listeners was compared with the score predicted from the AI calculation. Plots of the relation between these obtained and predicted scores are shown for Program Tapes I, II, and III in Figs. 1-8, 1-9, and 1-10, respectively. Each point in these plots represents a word location; the obtained scores for word locations denoted by solid points are considered more reliable than the obtained scores for word locations denoted by open points. (See Appendix 5.) Word locations where the obtained scores were predicted correctly can be identified by points that lie on an imaginary diagonal line. The horizontal displacement of points from such a line is a measure of how accurately the obtained scores were predicted.

To facilitate an analysis of the incorrect predictions, it was necessary to define a region, centered about the diagonal line, over which the obtained scores could be considered correctly predicted. This region, shown in Figs. 1-8, 1-9, and 1-10 by two dashed lines, was adopted from the aircraft study. When the data obtained in the aircraft study were arranged in the same manner as the present data, the chosen region enclosed 156 of the 180 word locations (87 percent of all data). Assuming that the distribution of the aircraft data was only moderately skewed, the region corresponded to approximately 1.5 standard deviations, measured in one direction from the mean.

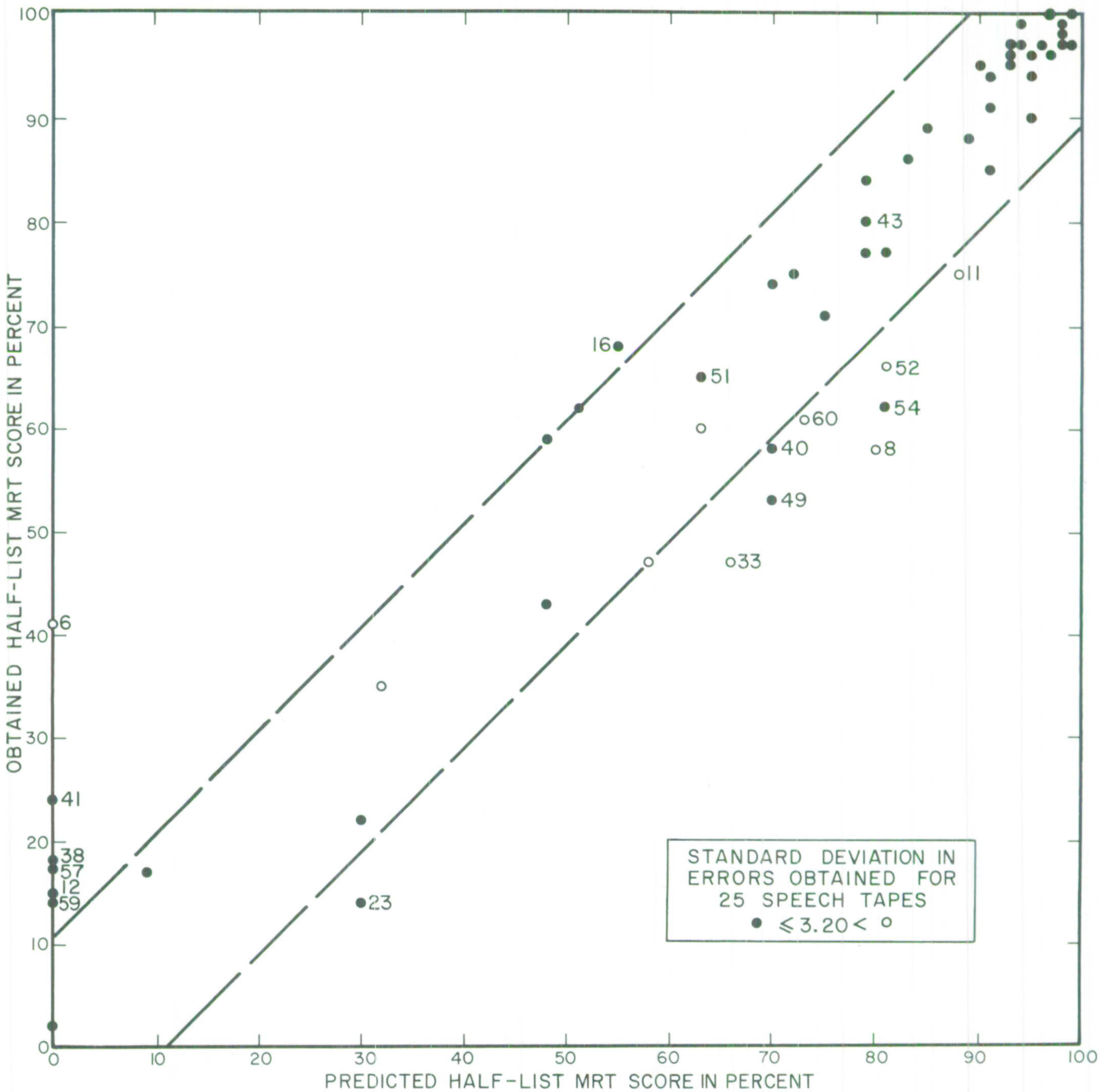

FIG. 1-8 PLOT OF PREDICTED VERSUS OBTAINED HALF-LIST MRT SCORES (CORRECTED FOR CHANCE) FOR 60 POINTS IN TIME OF PROGRAM TAPE I. DIAGONAL LINES ARE EXPLAINED IN TEXT.

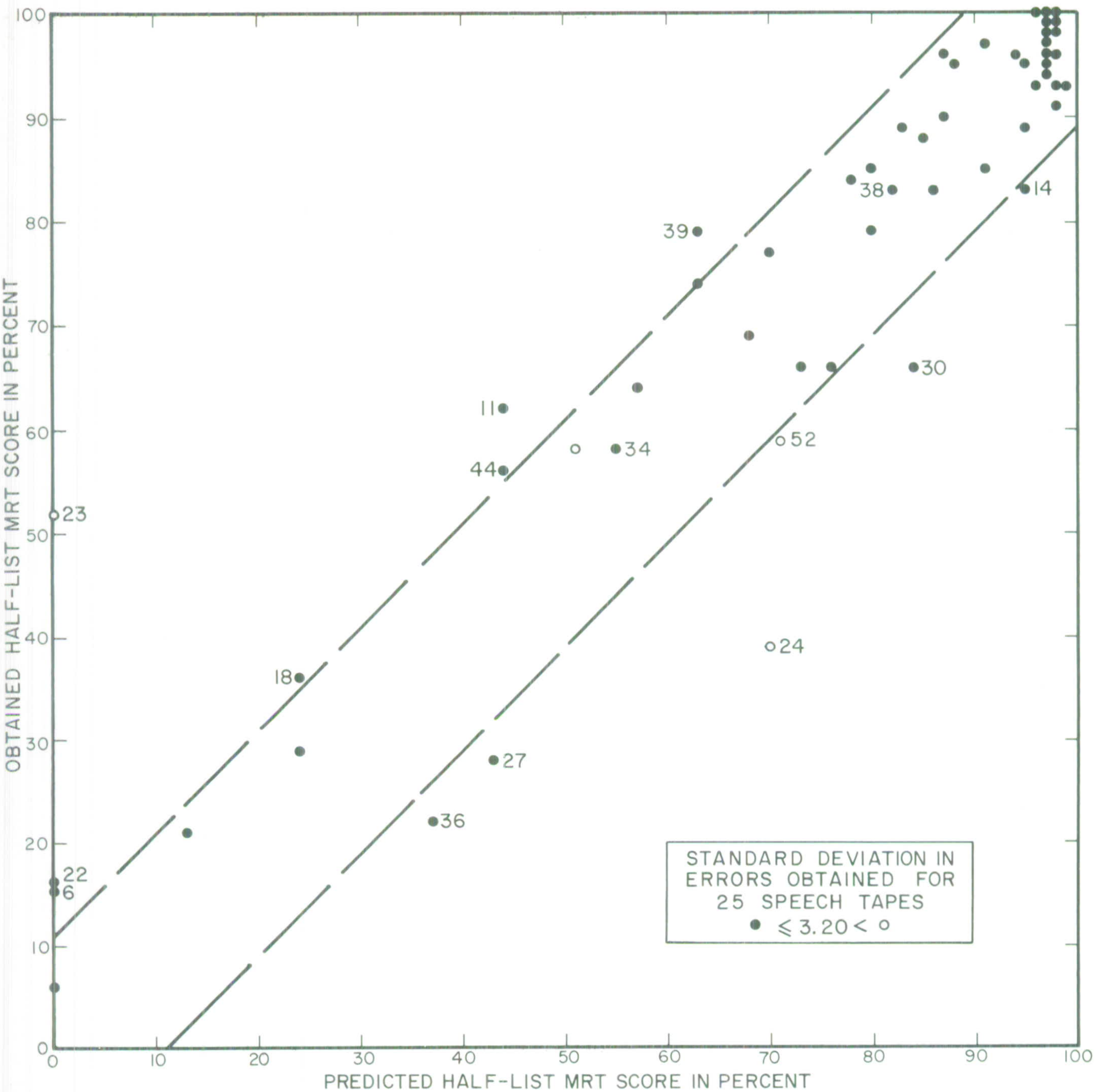

FIG. 1-9 PLOT OF PREDICTED VERSUS OBTAINED HALF-LIST MRT SCORES (CORRECTED FOR CHANCE) FOR 60 POINTS IN TIME OF PROGRAM TAPE II. DIAGONAL LINES ARE EXPLAINED IN TEXT.

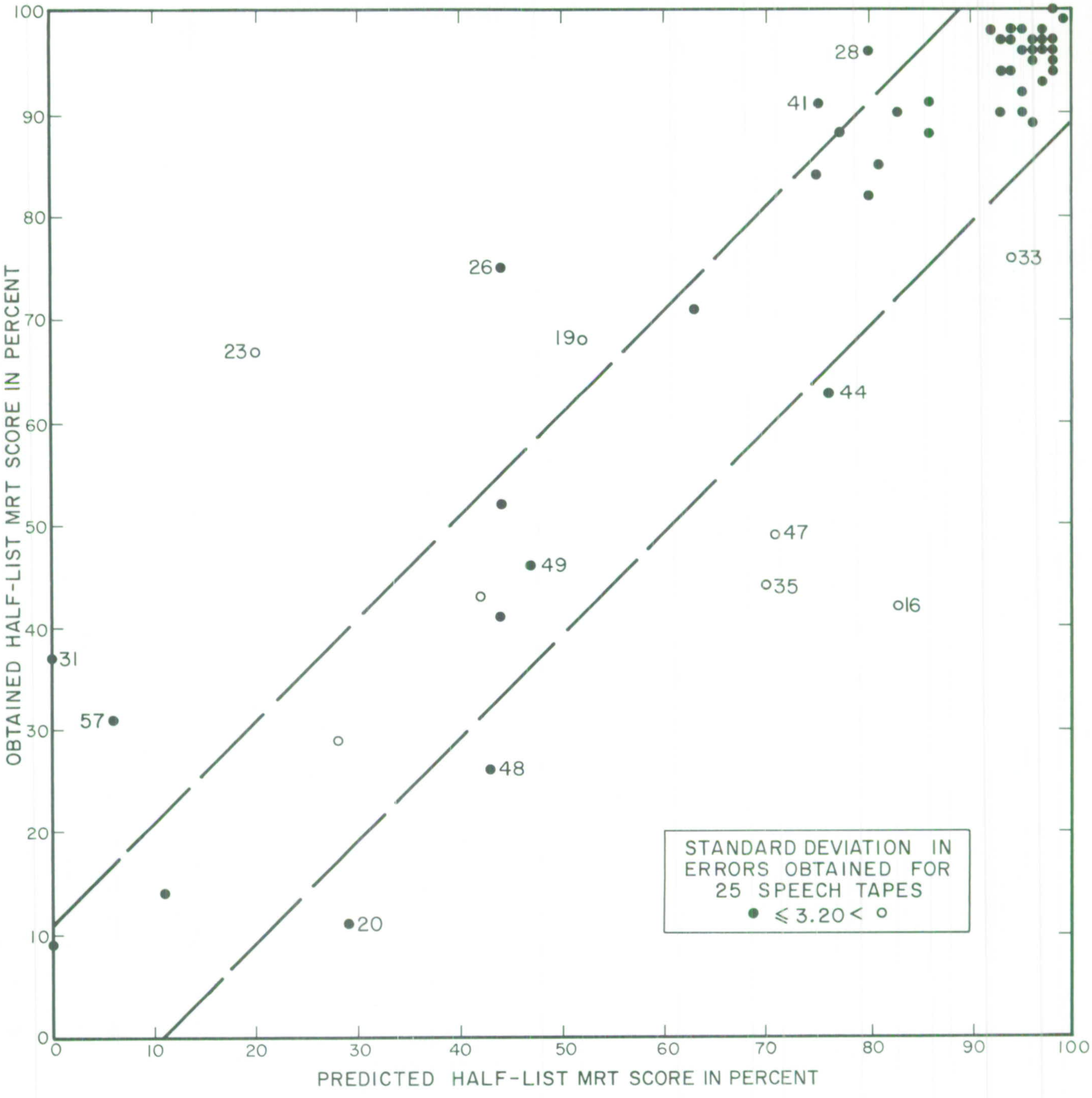

FIG. 1-10 PLOT OF PREDICTED VERSUS OBTAINED HALF-LIST MRT SCORES (CORRECTED FOR CHANCE) FOR 60 POINTS IN TIME OF PROGRAM TAPE III. DIAGONAL LINES ARE EXPLAINED IN TEXT.

In an effort to determine why the obtained scores were predicted incorrectly at certain word locations, the graphic-level tracings of the Program Tapes were carefully examined. This examination revealed that at the word locations of interest either the speech level, the noise level, or both levels varied considerably. As a consequence, different portions of a given word were subjected to different amounts of interference. Considering the nature of the MRT, those portions containing the initial and final consonants were probably more important in the determination of the obtained score than any other portions. The AI calculation, however, from which the predicted score was determined, only sampled the interference at one point in time. This point corresponded to the central portion of the word which contains the vowel.

All incorrectly predicted intelligibility scores could be explained in terms of this general finding: At the boundaries of the average word duration (approximately 400 milliseconds) where the initial and final consonants occurred, the interference conditions were appreciably different from the conditions that prevailed in the center of the duration, where the AI was calculated. The word locations of interest were ordered into two groups according to whether the predicted test score was too low or too high. Within each group, the word locations were further divided according to the probable cause of the incorrect prediction. These groupings are given in Appendix 8. Five arbitrarily selected examples from each group are also illustrated in Figs. 1-11 and 1-12.

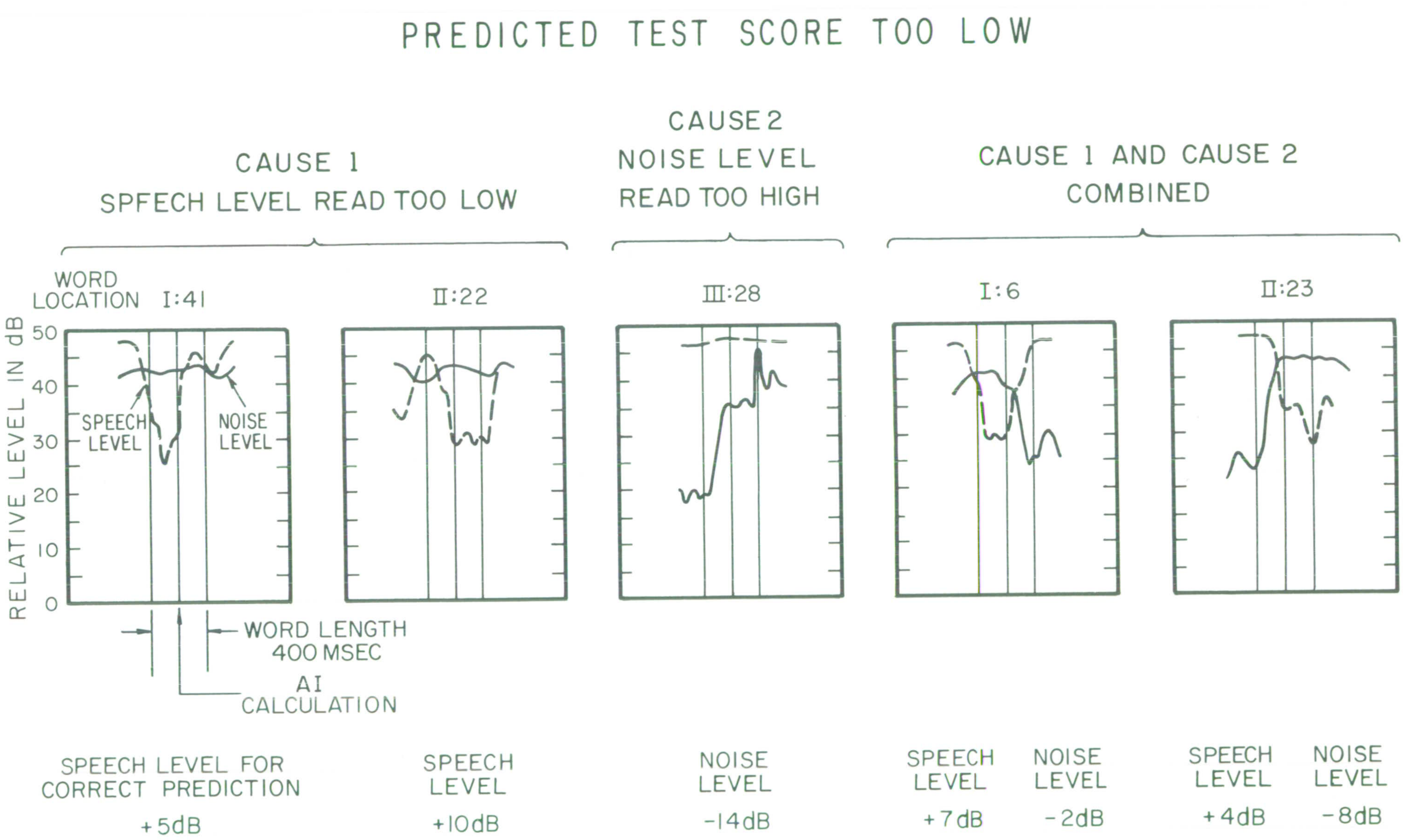

FIG. 1-11 SPEECH AND OVERALL NOISE LEVEL TRACINGS FOR SOME WORD LOCATIONS WHERE THE PREDICTED TEST SCORE WAS TOO LOW.

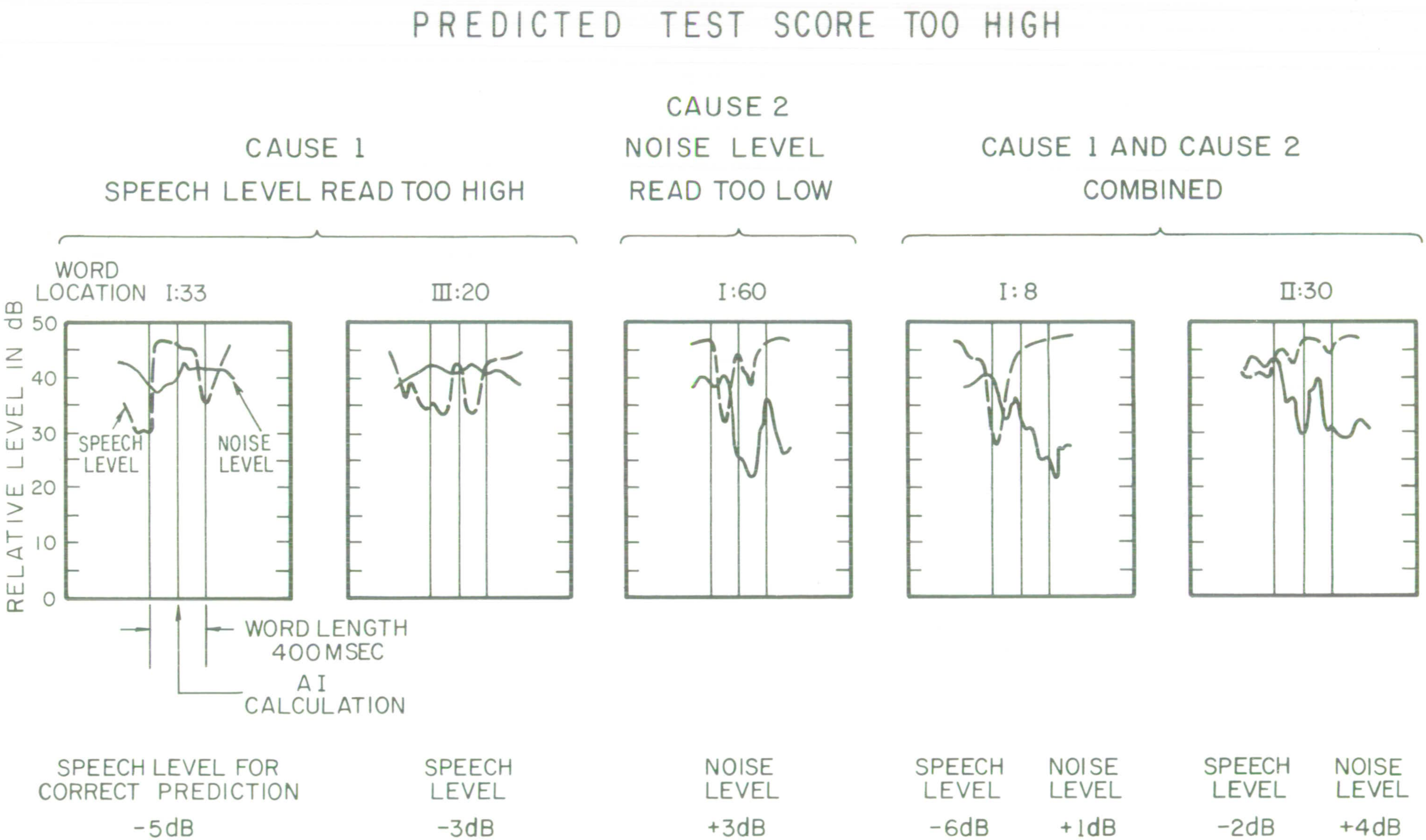

FIG. 1-12 SPEECH AND OVERALL NOISE LEVEL TRACINGS FOR SOME WORD LOCATIONS WHERE THE PREDICTED TEST SCORE WAS TOO HIGH.

Consider the level tracings for Word Location I:41, shown in Fig. 1-11. The temporal course of the speech level suggests that the initial and final consonants were barely attenuated. This was not taken into account by the AI calculation. Because the speech level used in the AI calculation was read too low on the tracing, the predicted test score was too low. To determine the amount by which the speech level was read too low, the AI calculation was repeated with various higher speech levels until the result approximated the AI value that corresponded to the obtained score (Fig. 1-7). For this example, the correction factor was +5 dB, as is indicated in the lower part of Fig. 1-11.

The level tracings for Word Location III:28 illustrate the case where the noise level was read too high. Although the position of the final consonant was masked more than was reflected in the AI calculation, the position of the initial consonant was masked much less. As a result, the predicted test score was again too low. The corrected AI calculation indicated that the noise level was read 14 dB too high. The level tracings for this and several other word locations showed an interesting trend: The speech and noise levels used in the corrected AI calculations tended to agree more with the levels existing at the position of the initial consonant than with those existing at the position of the final consonant. No satisfactory explanation could be found for this trend.

Now consider the level tracings for Word Location I:8, shown in Fig. 1-12. The speech level used in the AI calculation did not take into account the extreme attenuation at the position of the initial consonant and was therefore read too high.

It is also possible that the noise level was read too low, since the noise level used in the AI calculation was lower than the noise level which existed at the position of the initial consonant. While both of these factors undoubtedly contributed to a predicted test score that was too high, it was difficult to determine which factor was the dominant one. From an inspection of the traces, and from the knowledge that a total level change of 7 dB was required, it was determined that the speech level was primarily responsible for the incorrect prediction.

Five word locations for which the predicted score was nearly the same as the obtained score were randomly selected from the three Program Tapes. The level tracings for these word locations are shown in Fig. 1-13. The speech and noise levels appear to be generally more uniform over the word durations than the levels shown in Figs. 1-11 and 1-12. With the possible exception of the noise-level tracing for Word Location II:38 and the speech-level tracing for Word Location I:51, the tracings do not exhibit any features which might have led to incorrect predictions.

1.8 Conclusions

Within the criterion of accuracy employed in this study, approximately 76 percent of the intelligibility scores obtained from listeners could be correctly predicted from the computed AI's. All incorrect predictions could be attributed to the fact that, whereas an intelligibility score was influenced by the system characteristics for about 400 milliseconds, the AI was essentially computed for a single point in time. Contrary to theoretical expectations, the incorrect predictions could

PREDICTED TEST SCORE CORRECT

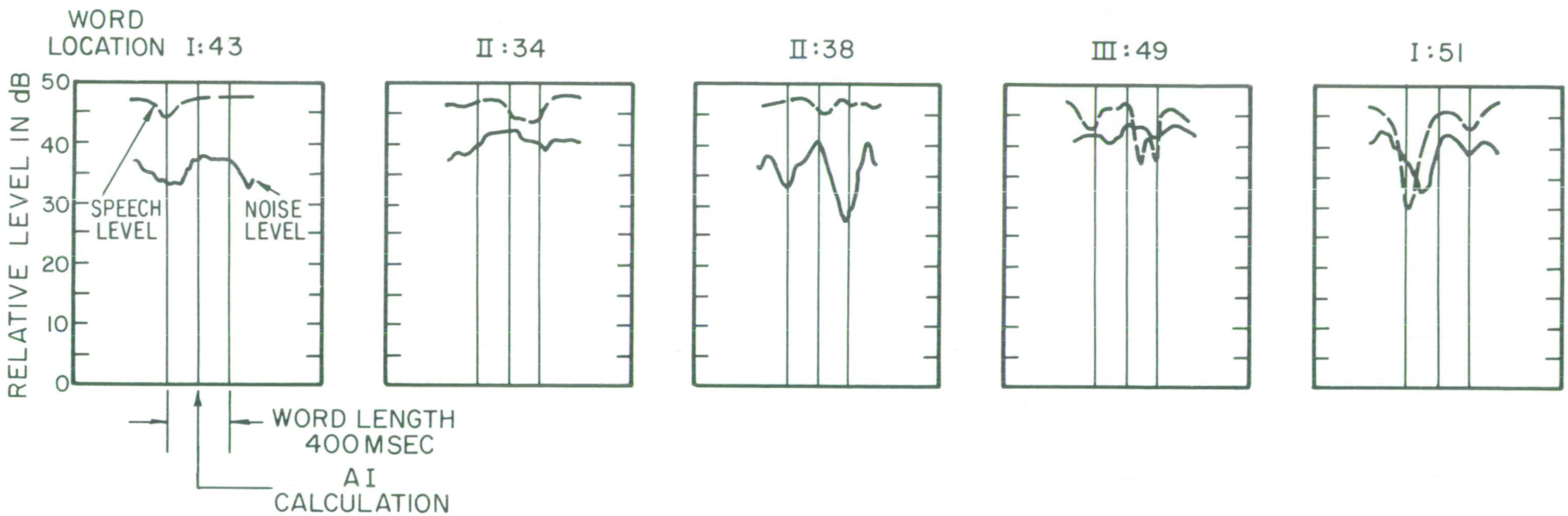

FIG. 1-13 SPEECH AND OVERALL NOISE LEVEL TRACINGS FOR SOME WORD LOCATIONS WHERE THE PREDICTED TEST SCORE WAS CORRECT.

not be related to temporal masking. The results of this study suggest that fewer incorrect predictions would have been made if the speech and noise level tracings had both been averaged over a 400 millisecond interval at each word location. Perhaps the number of incorrect predictions could have been reduced even further if, in obtaining average speech and noise levels for each AI calculation, the initial portion of the interval had been given more weight than the final portion.

The computed AI's were intended to represent scores which might be obtained with a device like SCIM. Despite their general validity, the AI's are clearly not representative in the sense that they are based on instantaneous measurements. SCIM employs circuits which integrate the system characteristics over an extended period of time before an SCI is computed. If SCIM were used to process the Program Tapes prepared in this study, and if the integration time and sampling rate of SCIM were properly adjusted, then the resulting SCI's should accurately predict the obtained intelligibility scores. It is concluded, therefore, that SCIM is potentially suitable for the automatic evaluation of time-varying communication systems.

Although it is difficult to make specific recommendations for the possible modification of SCIM, the results of this study suggest that the time intervals over which SCIM presently integrates the received "speech" and noise signals are too long. The "speech" is integrated over 1 second and the noise over 2 seconds. For the evaluation of time-varying systems, an integration time of approximately 500 milliseconds for both "speech" and noise would appear to be more appropriate. This interval corresponds roughly to the duration of a syllable.

Since individual SCI's are difficult to interpret, it may be desirable to have SCIM compute a large number of SCI's before an average score is displayed. The average score could then be used to estimate the probability that a syllable will be correctly identified over the system, regardless of when the syllable is transmitted. It is clear that such scores must be eventually related to how well longer message units (e.g. phrases and sentences) can be understood. The ultimate measure of system performance is not the intelligibility of a series of independent test words, but the ease and reliability with which system users can transmit and receive operational messages.

2. EVALUATION OF COMMUNICATION SYSTEMS EMPLOYING PEAK CLIPPING

2.1 Introduction and Approach

In many radio-communication systems, the speech signal is subjected to peak clipping at the transmitter. Prior to the modulation stage, the extreme peaks of the speech signal are intentionally removed so that a given level of transmitter power can be more efficiently utilized. Since peak clipping and subsequent restoration of the maximum amplitude increases the long-term RMS power of a speech signal,[20] systems employing even moderate levels of peak clipping tend to transmit more intelligible speech than systems that do not employ peak clipping. In the evaluation of such systems with a device like SCIM, this effect must be taken into account. The computed SCI's should accurately predict the intelligibility scores that might be obtained with the aid of speakers and listeners.

The results of an earlier study[9] indicated that SCIM was unable to predict the intelligibility of peak-clipped speech. In the present study, this problem was investigated in some detail to determine whether SCIM could be modified to compute more reliable SCI's. Since peak clipping changes the amplitude distribution of a speech signal, it was possible that the difficulty encountered with SCIM was inherent in the concept of the AI, on which the design of SCIM is based. Certain assumptions made in the calculation of an AI are violated in the case of peak-clipped speech. If the problem could be explained in terms of this argument, then a simple equipment modification would not be a solution. On the other hand, considering that

SCIM really measures RMS rather than peak S/N ratios, and that only moderate levels of peak clipping are involved, a relatively minor change in SCIM could lead to the desired performance.

Under certain operating conditions of a communication system, the speech signal may be subjected to unintentional peak clipping. This may happen because of a faulty component, an equipment mismatch, or other factors which are difficult to control and correct in a field situation. The clipping may not be symmetrical for positive and negative amplitudes, and the level of clipping may change as a function of time. If SCIM were to be used to evaluate such a system, it would be necessary to provide a device which could detect the type and level of clipping that might be encountered. In view of the difficulties associated with the design of such a device, it was decided that the present study would not be concerned with unintentional peak clipping.

A survey of the literature was made to determine the intelligibility of speech for different amounts of peak clipping and for various S/N ratios. SCIM measurements were then obtained for these same conditions. Using a compilation of the most applicable and reliable data found in the literature as a model, the performance of SCIM was critically evaluated. After the primary cause of the incorrect SCI's was identified and eliminated, the performance of SCIM was found to be adequate. A conversion chart was prepared for relating SCI's to intelligibility scores when the level of peak clipping is known.

2.2 Effect of Peak Clipping on Speech Intelligibility

A block diagram of a typical radio-communication system employing peak clipping is shown in Fig. 2-1(a). After the speech signal is subjected to peak clipping, the maximum amplitude of the signal is restored to its original value. At the receiving antenna, the transmitted radio signal is mixed with radio-frequency noise. Such a system may be simulated in the laboratory by means of the instrumentation indicated in Fig. 2-1(b). All of the system components shown in Fig. 2-1(a) have their counterparts in Fig. 2-1(b), except that the noise source is more clearly defined in the laboratory simulation. Instrumentation such as is shown in Fig. 2-1(b) has been used extensively for the purpose of measuring the effects of peak clipping on speech intelligibility.

Because different studies have used different speech materials, different kinds of noise, different methods of measuring S/N ratio, and even different definitions of clipping level, it was difficult to integrate the results of the studies reported in the literature. Consider, for example, the presumedly simple relation between S/N ratio and Harvard PB-word intelligibility[1] for white masking noise. Here the speech material and the kind of noise are clearly defined, and no clipping is involved. However, as is evident from Fig. 2-2, the results obtained in different studies vary considerably. In the study reported by Stuckey,[18] the listeners may have been over-exposed to particular randomizations of the word lists used. They were also permitted to adjust their individual listening levels. Whether either of these factors was responsible for the unusually high test scores is not known.

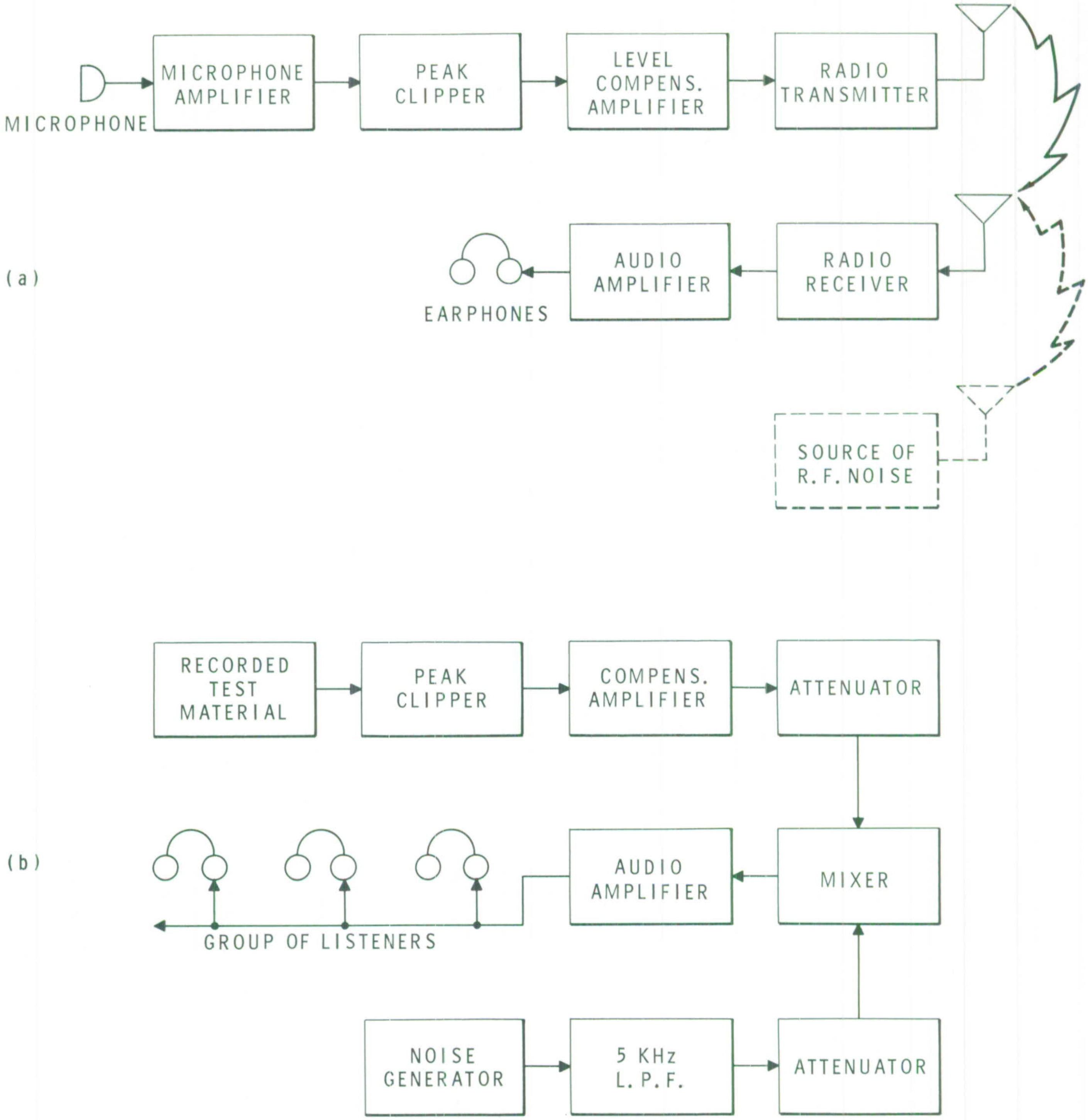

FIG. 2-1 BLOCK DIAGRAMS OF (a) TYPICAL RADIO COMMUNICATION SYSTEM EMPLOYING PEAK CLIPPING, AND (b) INSTRUMENTATION USED TO SIMULATE AND EVALUATE SUCH A SYSTEM.

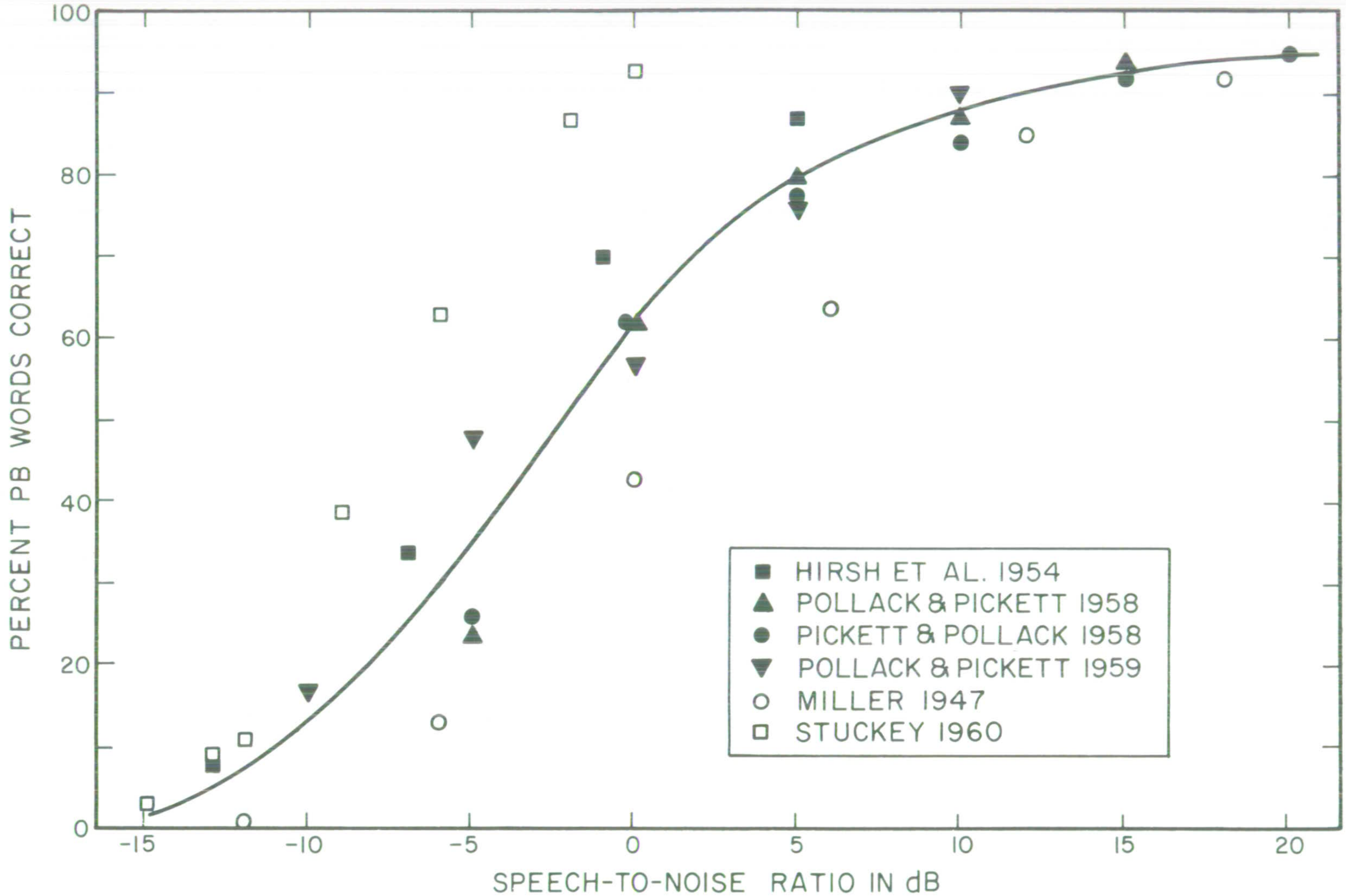

FIG. 2-2 RELATION BETWEEN SPEECH-TO-NOISE RATIO AND PB-WORD INTELLIGIBILITY FOR WHITE MASKING NOISE.

The vocabulary in the study published by Hirsh[5] was limited to only 200 words. Since the full vocabulary of the Harvard PB-Word Test consists of 1000 words, this might explain why the test scores were somewhat higher than those reported in other studies. The oldest study considered, published in a classic paper by Miller,[11] produced scores that are 10 percentage points lower at some S/N ratios than the scores obtained by Pollack and Pickett in three separate studies.[14,15,16] The curve drawn through the data points in Fig. 2-2 therefore only approximates the relation between S/N ratio and speech intelligibility.

When peak clipping is introduced in a given experiment, the prevailing S/N ratio is most meaningfully specified in terms of measurements made when the clipping level is zero. This convention preserves the usual definition of S/N ratio as the amplitude relation between the RMS speech level and the RMS noise level, expressed in dB. The clipping level is usually defined in terms of the level that is exceeded 1 percent of the time in an unprocessed speech signal sampled 8 times per second. When only those few speech peaks with amplitudes larger than this level are clipped, the clipping level is referred to as 0 dB. Clipping levels greater than about 40 dB reduce the speech signal to a sequence of square pulses (infinitely clipped speech).

After peak clipping, the loss in amplitude of the speech signal is usually compensated by amplification before the signal is mixed with noise. Two types of compensation are commonly employed; level compensation and power compensation. In level compensation, the original peak level of the speech signal is restored. Hence the gain of the compensating amplifier, expressed in dB, is exactly equal to the clipping level.

In power compensation, the original RMS power of the speech signal is restored. For power compensation the gain of the compensating amplifier is always less than the clipping level. These two forms of compensation are, of course, related. With the aid of the relation derived by Wathen-Dunn and Lipke,[20] it is possible to convert data obtained with one form of compensation to data that might have been obtained with the other form of compensation.

The results of three studies reported in the literature[3,10,15] are shown in Figs. 2-3 and 2-4. For a given S/N ratio, peak clipping and subsequent level compensation generally increase intelligibility (Fig. 2-3), whereas peak clipping and subsequent power compensation have no appreciable effect on intelligibility (Fig. 2-4). Because level compensation is typically employed in radio-communication systems, the data given in Figs. 2-2 and 2-3 were combined and averaged to obtain a "psychoacoustic model" against which the performance of SCIM could be compared. The data constituting this model are given in Fig. 2-5.

2.3 Performance of SCIM

The results obtained with SCIM are shown in Fig. 2-6. Each point in this figure represents the mean of 3 display readings. In comparing these results with the corresponding data from the literature (Figs. 2-3 and 2-4), it is obvious that the performance of SCIM was grossly inadequate. This finding confirmed the earlier reports that SCIM appeared to be unable to predict the intelligibility of peak-clipped speech. In obtaining the results shown in Fig. 2-6, it was noted that the higher-frequency filters in the SCIM analyzer barely contributed to the total SCI when

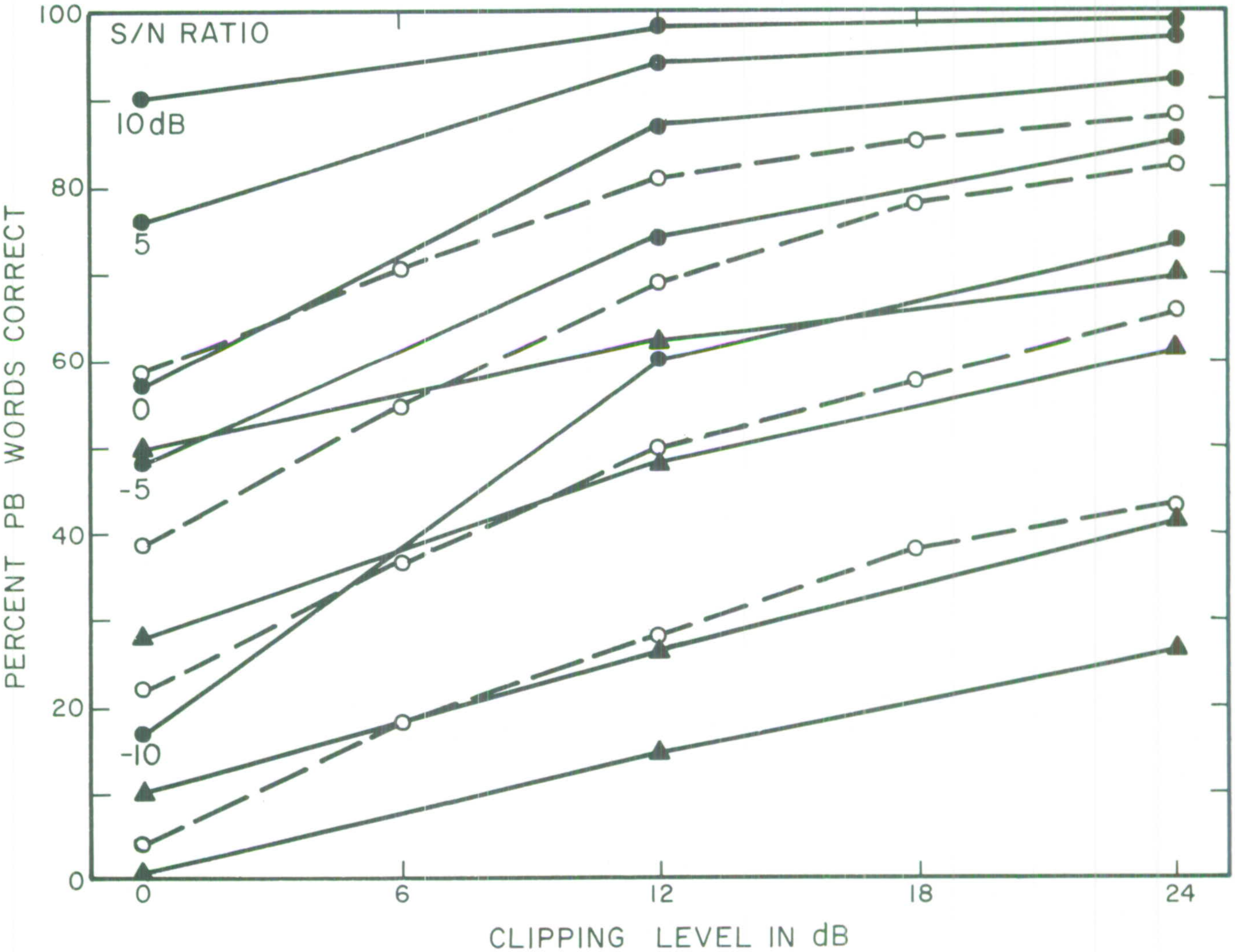

FIG. 2-3 PB-WORD INTELLIGIBILITY AS A FUNCTION OF CLIPPING LEVEL FOR PEAK CLIPPING WITH LEVEL COMPENSATION. THE PARAMETER IS SPEECH-TO-NOISE RATIO. DATA FROM POLLACK AND PICKETT 1959 (●—●), KRYTER ET AL. 1947 (○--○), AND EWING AND HUDDY 1966 (▲—▲).

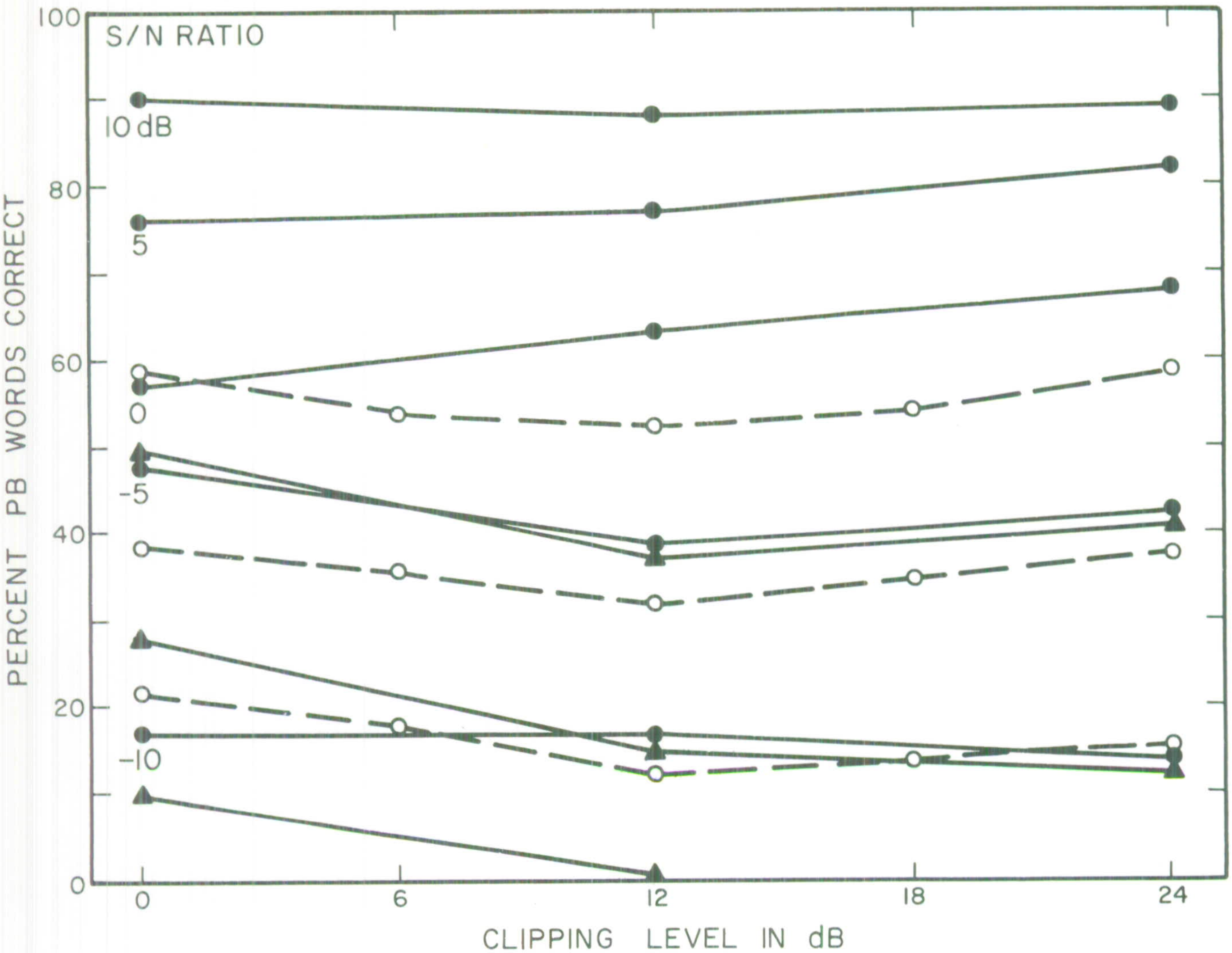

FIG. 2-4 PB-WORD INTELLIGIBILITY AS A FUNCTION OF CLIPPING LEVEL FOR PEAK CLIPPING WITH POWER COMPENSATION. THE PARAMETER IS SPEECH-TO-NOISE RATIO. DATA FROM POLLACK AND PICKETT 1959 (●——●), KRYTER ET AL. 1947 (○– – –○), AND EWING AND HUDDY 1966 (▲——▲).

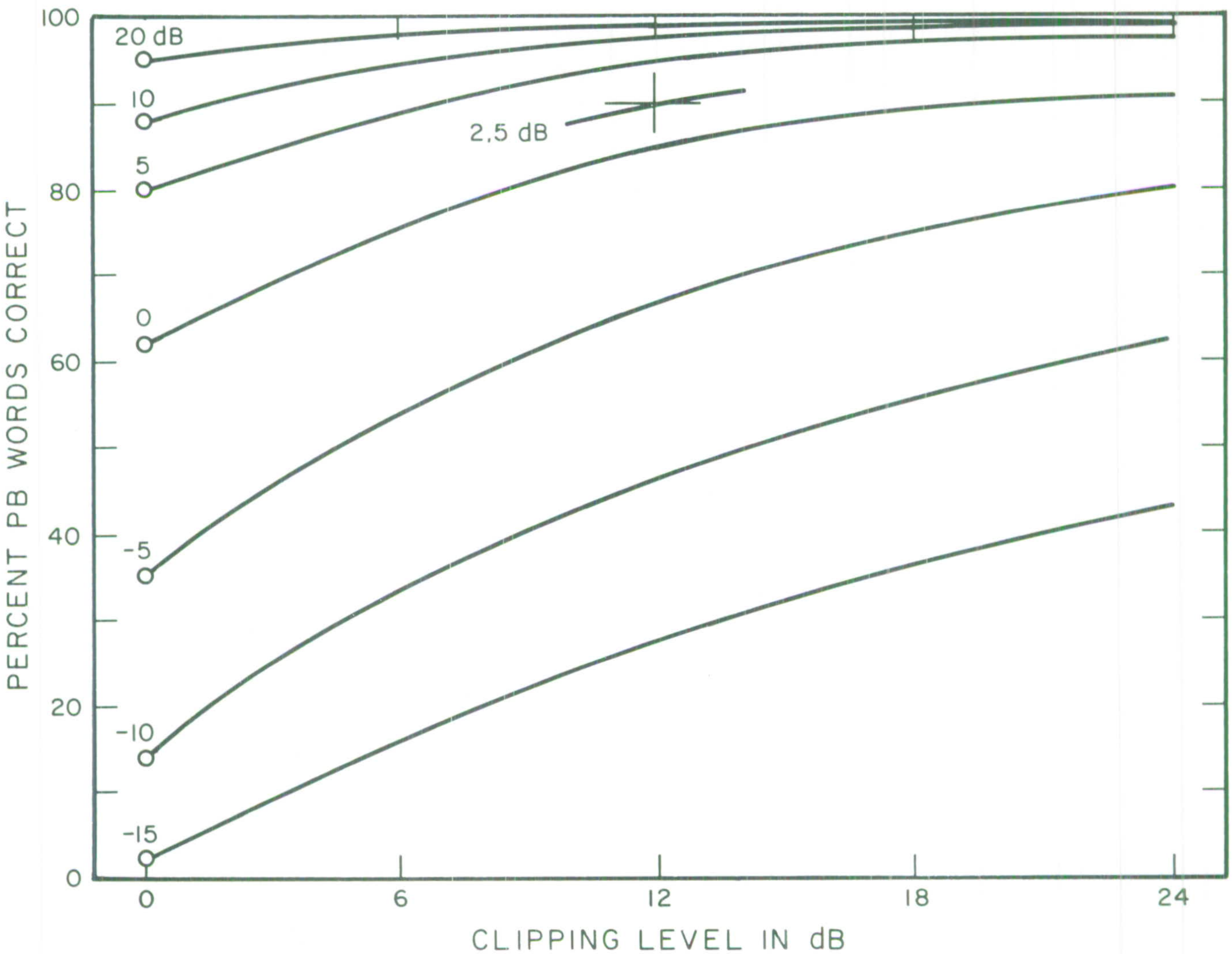

FIG. 2-5 PB-WORD INTELLIGIBILITY AS A FUNCTION OF CLIPPING LEVEL FOR PEAK CLIPPING WITH LEVEL COMPENSATION. THE PARAMETER IS SPEECH-TO-NOISE RATIO. DATA ESTIMATED AND AVERAGED FROM FIGS. 2-2 AND 2-3.

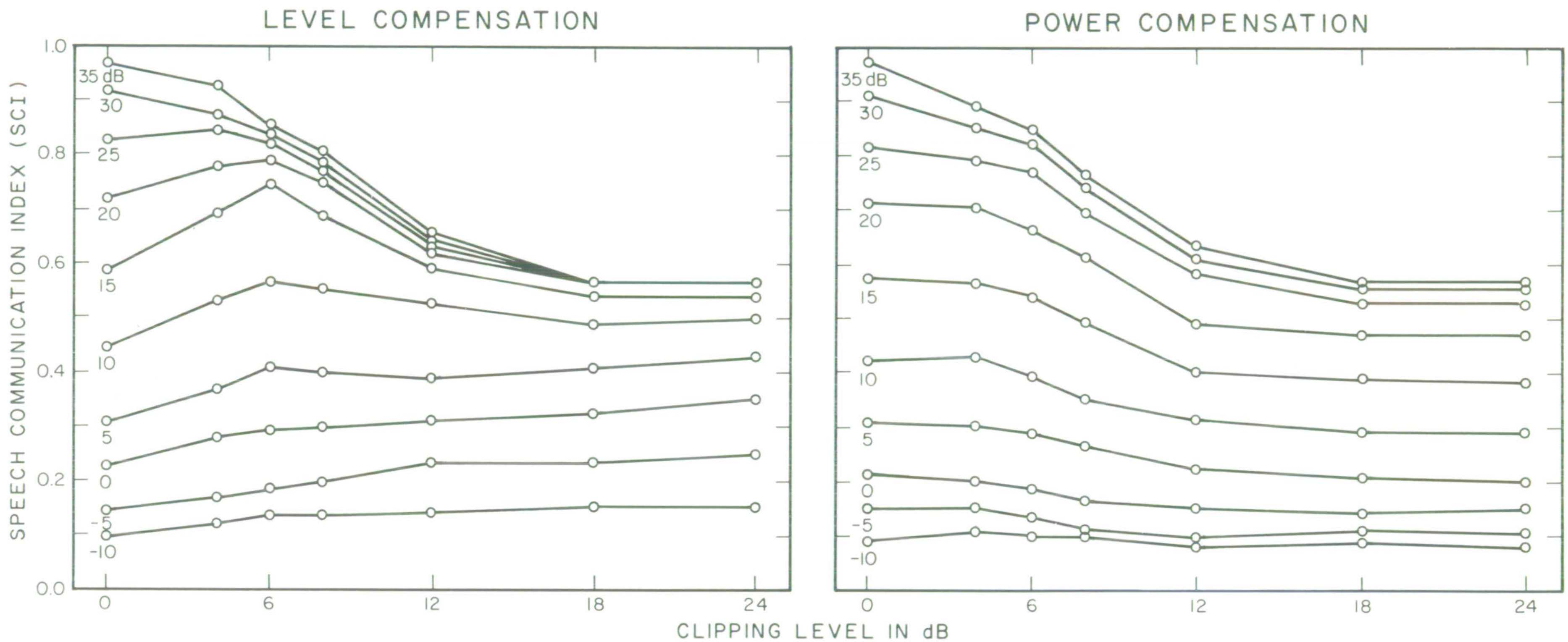

FIG. 2-6 SPEECH COMMUNICATION INDEX AS A FUNCTION OF CLIPPING LEVEL FOR PEAK CLIPPING WITH LEVEL COMPENSATION AND POWER COMPENSATION. THE PARAMETER IS SIGNAL-TO-NOISE RATIO. DATA OBTAINED WITH AGC TONES ENGAGED.

high clipping levels were employed. Despite a very favorable overall S/N ratio, the S/N ratios measured by SCIM in some of the high-frequency bands were consistently low.

The source of this problem was traced to the tones which are normally transmitted by the SCIM signal generator to disable Automatic Gain Control (AGC) circuits in the system under evaluation. During the interval in which the SCIM analyzer measures noise in the upper portion of the spectrum, a 400 Hz tone is received in the lower portion of the spectrum. While this tone does not normally interfere with the measurements, amplitude clipping will introduce strong harmonics of the tone into the upper portion of the spectrum. These harmonics serve to reduce the S/N ratios measured in the high-frequency bands.

With the AGC-tone oscillators removed from the SCIM signal generator, the performance of SCIM was measured again. The new results obtained for peak clipping with level compensation and peak clipping with power compensation are shown in Figs. 2-7 and 2-8, respectively. Since these results compared fairly well with the published intelligibility data (Figs. 2-3 and 2-4), the performance of SCIM was considered adequate. Further work consisted of relating a given SCI and a given clipping level to a unique intelligibility score.

2.4 Preparation of Conversion Chart

When SCIM is used to evaluate a typical communication system employing peak clipping and level compensation, the only available information consists of the SCI and the clipping level. In most situations, the overall S/N ratio (defined for a clipping

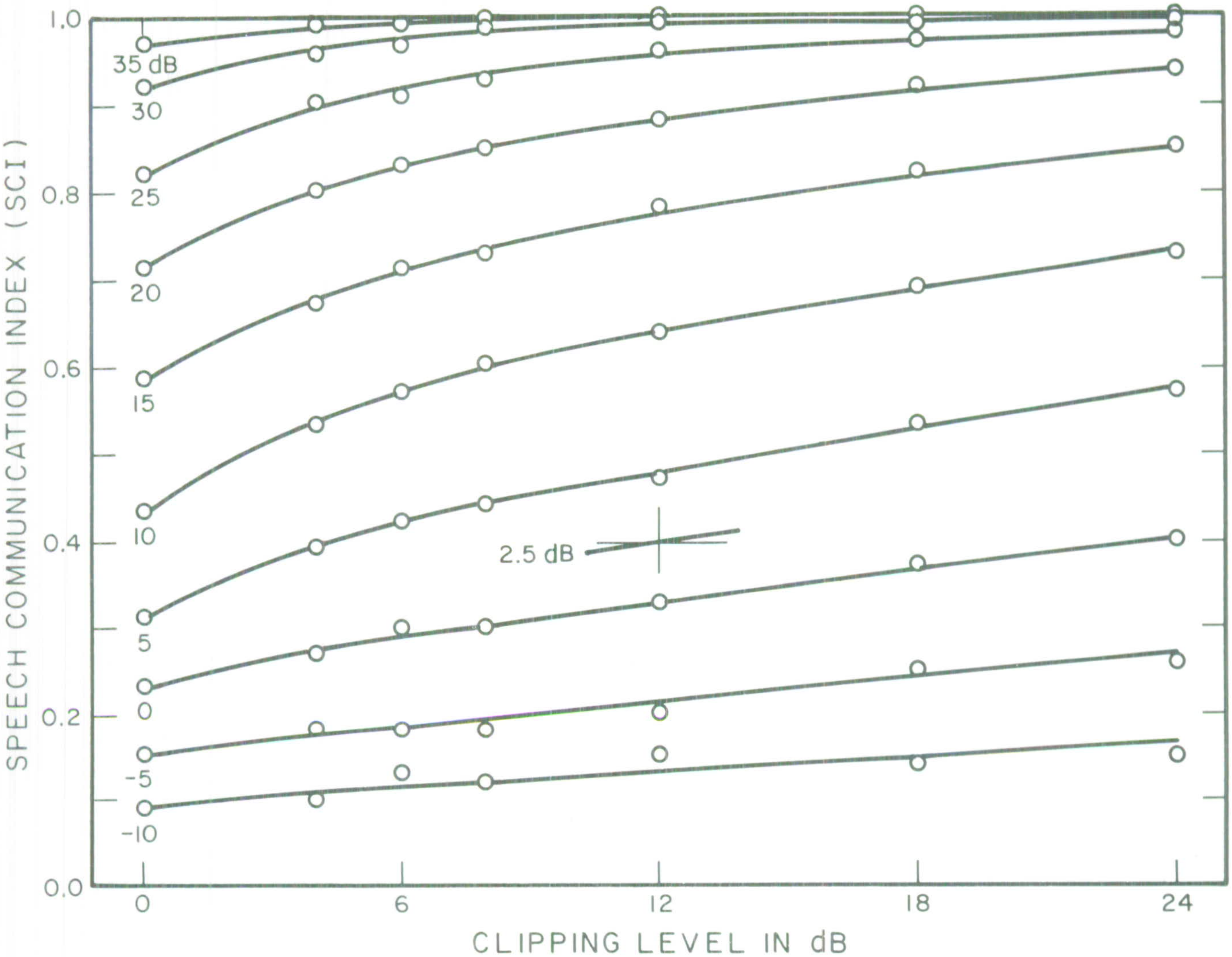

FIG. 2-7 SPEECH COMMUNICATION INDEX AS A FUNCTION OF CLIPPING LEVEL FOR PEAK CLIPPING WITH LEVEL COMPENSATION. THE PARAMETER IS SIGNAL-TO-NOISE RATIO. DATA OBTAINED WITHOUT AGC TONES.

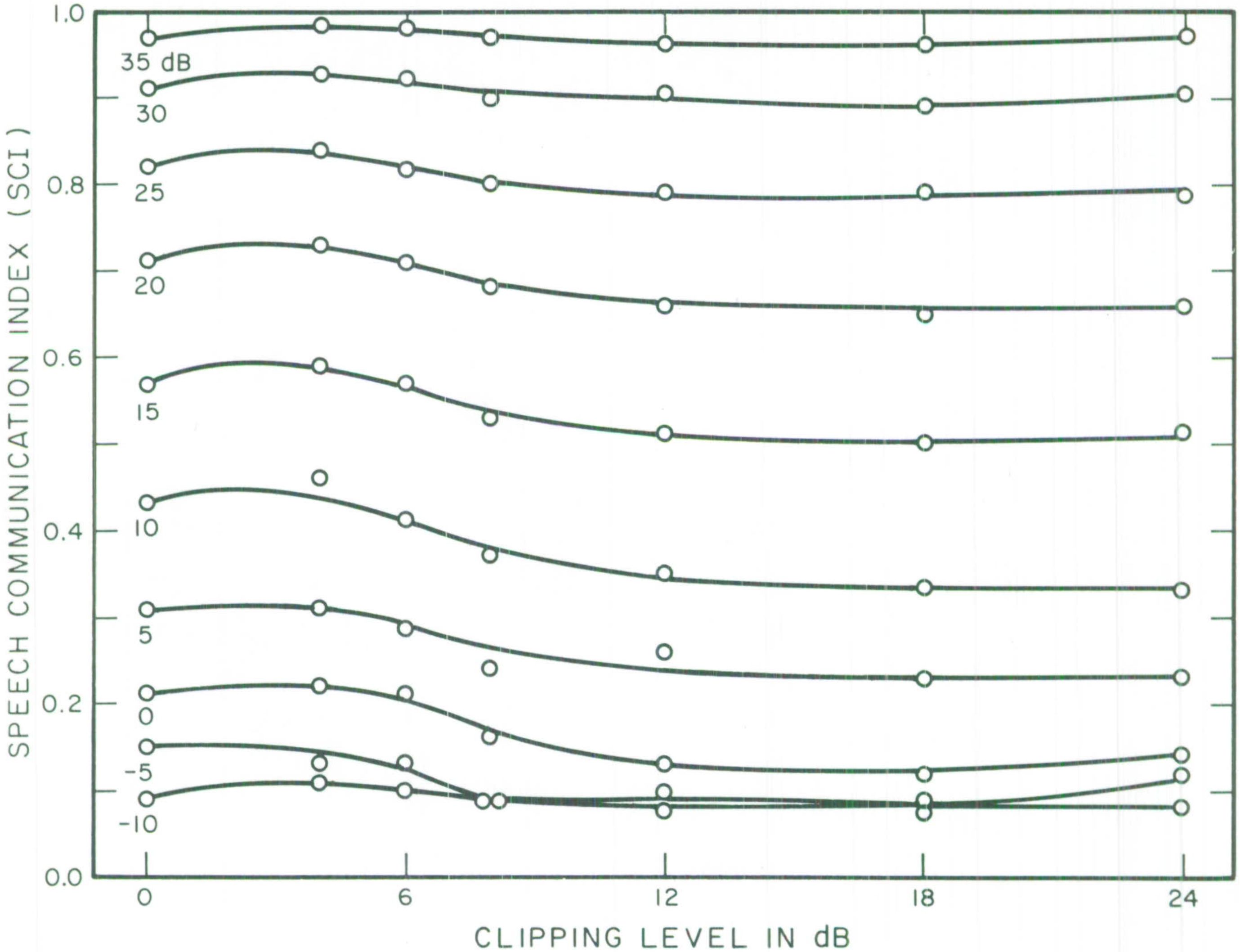

FIG. 2-8 SPEECH COMMUNICATION INDEX AS A FUNCTION OF CLIPPING LEVEL FOR PEAK CLIPPING WITH POWER COMPENSATION. THE PARAMETER IS SIGNAL-TO-NOISE RATIO. DATA OBTAINED WITHOUT AGC TONES.

level of 0 dB) is not known. To be able to convert the available information into an intelligibility score that might be obtained in a formal listening test, it was necessary to derive the relation between SCI's and Harvard PB-word intelligibility for various clipping levels.

Using the data shown in Fig. 2-7, S/N ratios were estimated for 21 SCI's (covering the range from 0.00 to 1.00 in steps of 0.05) at each clipping level. These S/N ratios were then translated into intelligibility scores, using the data shown in Fig. 2-5. For example, the S/N ratio corresponding to an SCI of 0.40 and a clipping level of 12 dB is approximately 2.5 dB. (See Fig. 2-7.) This S/N ratio, at the same clipping level, corresponds to an intelligibility score of about 90 percent. (See Fig. 2-5.) The relation between SCI's and intelligibility scores thus derived is shown in Fig. 2-9. The data from which this figure was prepared are given in Appendix 9.

2.5 Conclusions

The principal conclusion drawn from the results of this study was that SCIM is potentially capable of predicting the intelligibility of peak-clipped speech. The problem encountered earlier was not due to a fundamental shortcoming of the AI concept, but to a minor defect in the design of SCIM which can be easily remedied. In order to retain the advantages of employing AGC tones during the assessment of certain types of communication systems, it was necessary to modify both the SCIM signal generator and the SCIM analyzer. These modifications are described in the following section.

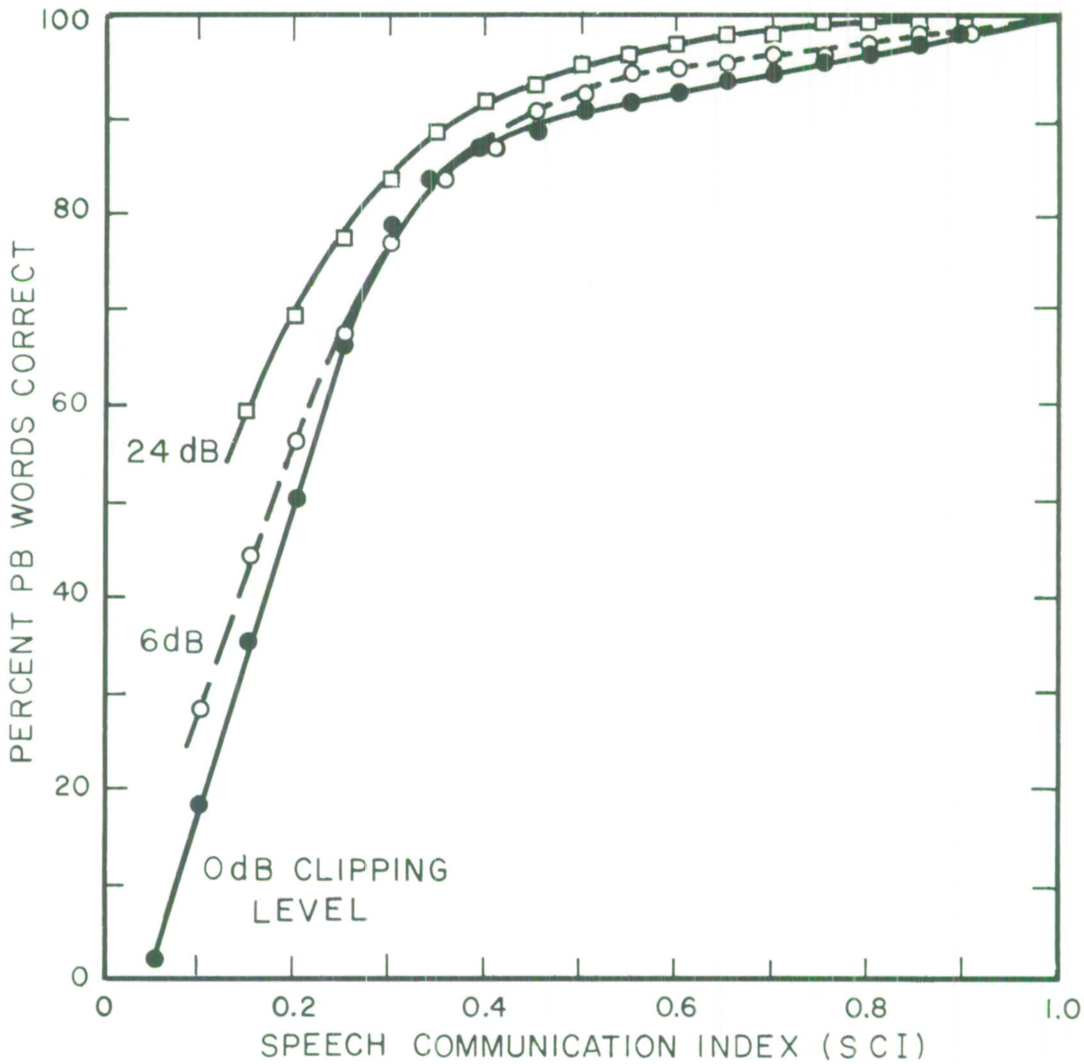

FIG. 2-9 RELATION BETWEEN SPEECH COMMUNICATION INDEX AND PB-WORD INTELLIGIBILITY FOR PEAK CLIPPING WITH LEVEL COMPENSATION. THE PARAMETER IS CLIPPING LEVEL.

3. EQUIPMENT MODIFICATIONS

3.1 Introduction

During the course of this study, SCIM was modified and improved in several respects. Some of these modifications were made in direct compliance with the requirements stated in the contract. These included the changes necessary for the evaluation of systems employing peak clipping, and the reduction in computation time. Other modifications and improvements were implemented because they appeared generally desirable and could be easily accomplished. These included replacing some components by improved versions, and minor changes that would facilitate equipment calibration.

A second addendum to the Operation/Maintenance Manual for SCIM was prepared.[19] This addendum contains technical details and schematic diagrams for all equipment modifications that were made. General descriptions of some of the more important modifications are given below.

3.2 Modifications for the Evaluation of Systems Employing Peak Clipping

Examination of the measuring procedure inherent to SCIM revealed that the clipped 400 Hz AGC tone, which is present when noise measurements are being made in the high-frequency bands, has harmonics of considerable amplitude in the frequency region of these bands. The presence of this tone, therefore, seriously interfered with the measurements made by SCIM. In view of this problem, it was decided to modify SCIM as follows: (1) A 3-KHz AGC tone would be substituted for the 400-Hz AGC tone so that the harmonics of the AGC tone would fall outside of the last band employed by SCIM. (2) A sharp-skirted 3-KHz notch filter would be designed to provide sufficient attenuation of the AGC tone so that the presence of the tone does not appreciably affect the performance of SCIM. The center frequency of the notch filter would be manually variable over a limited range to accommodate possible frequency translation, as might be encountered with single-sideband radio systems.

The SCIM signal generator was modified to provide a 3-KHz AGC tone during both time intervals in which the analyzer performs noise measurements. For calibration purposes, the 3-KHz tone can also be transmitted independent of a test-signal sequence.* A 3-KHz notch filter was constructed and installed into the SCIM analyzer. During calibration, the transmitted 3-KHz tone is "tuned out" at the notch filter. A special circuit and associated panel meter have been installed in the SCIM analyzer to facilitate this procedure.

3.3 Reduction of Computation Time and Improved Synchronization

A timing diagram of the present test-signal sequence and analyzer timing sequence is shown in Fig. 3-1. In the previous

---

* By means of two switches, the former arrangement of AGC tones is still available.

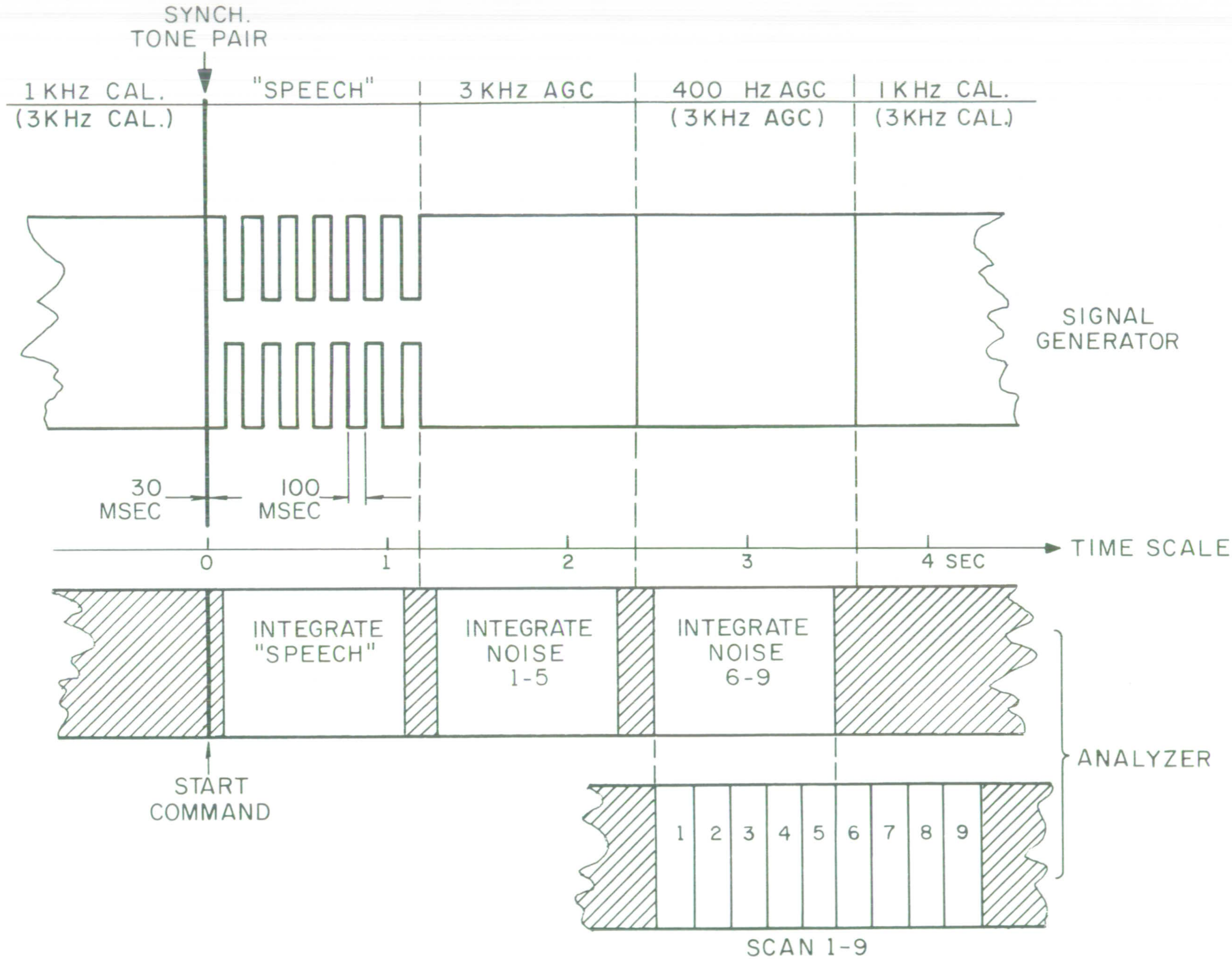

FIG. 3-1 TIMING SEQUENCE OF SIGNAL GENERATOR AND ANALYZER.

analyzer timing sequence, each change in mode was mediated by a new tone-pair synchronization signal from the generator. In the present scheme, only one tone pair is transmitted, which triggers an analyzer timing sequence that is independent of the test-signal sequence. To allow for slight inaccuracies in the synchronization between the test-signal sequence and the analyzer timing sequence, the transmission times of the "speech" signal and the AGC tones were increased from 1.0 to 1.2 seconds, while the integration times remained fixed at 1 second. Thus, approximately 200 milliseconds were provided between integration intervals to insure that the correct signals are integrated, even if the analyzer timing sequence should be slightly shifted in time relative to the test-signal sequence.

The time intervals necessary for sampling the integrator outputs were reduced from 1.0 to 0.2 seconds. The measurement accuracy was found to be unaffected by this change. This modification reduced the computation time from 10 to 2 seconds, and the total run time from 12 to 4.4 seconds.

A disadvantage of the old timing scheme was that if any one of the tone pairs was missed, or if a spurious tone pair was received, the analysis was negated and the analyzer had to be manually reset. This occurred quite often in the evaluation of time-varying communication systems. When SCIM was equipped with a data logger (printer and clock), the printed data were often unreliable. Another disadvantage was that the tone pairs were integrated along with the "speech" and noise signals. Although this error was minor, it could not be completely eliminated.

Since the present synchronization scheme requires only one initial tone pair, the probability that a successfully started analysis run will be disturbed has been reduced considerably. The amplitude of the tone pair was increased above its former value (this was permissable because it is no longer integrated) to reduce the chances of a missed start command. In the new scheme, the synchronization detection system is disabled until an analysis run is complete, so that a spurious tone pair occurring after the reception of the initial tone pair will not be interpreted as another start command.

This new scheme allows for the complete automation of measurements if SCIM is used with a data logger. If a start command is missed, the complete run would be missed. If a spurious tone pair occurs, which can only be detected during a missed run, the analyzer would reset itself and be ready for the next valid tone pair. It is doubtful that the latter condition would occur, because if the start command were too weak to be detected, a spurious tone pair within the next four seconds would probably also be too weak to be detected.

3.4 Other Modifications

Although the dynamic range of the SCIM analyzer was adequate for most noise spectra that are likely to be encountered, a signal with appreciable high-frequency energy could overload the detectors. This potential problem was eliminated by increasing the dynamic range of the detectors, and by resetting the VU meter to allow better utilization of the increased range. This modification is illustrated in Fig. 3-2. The shaded areas in this figure represent the filter input ranges for linear detector output, before the modification. The total column lengths represent the input ranges of the present system. The improvement in linearity is shown for Filter No. 4.

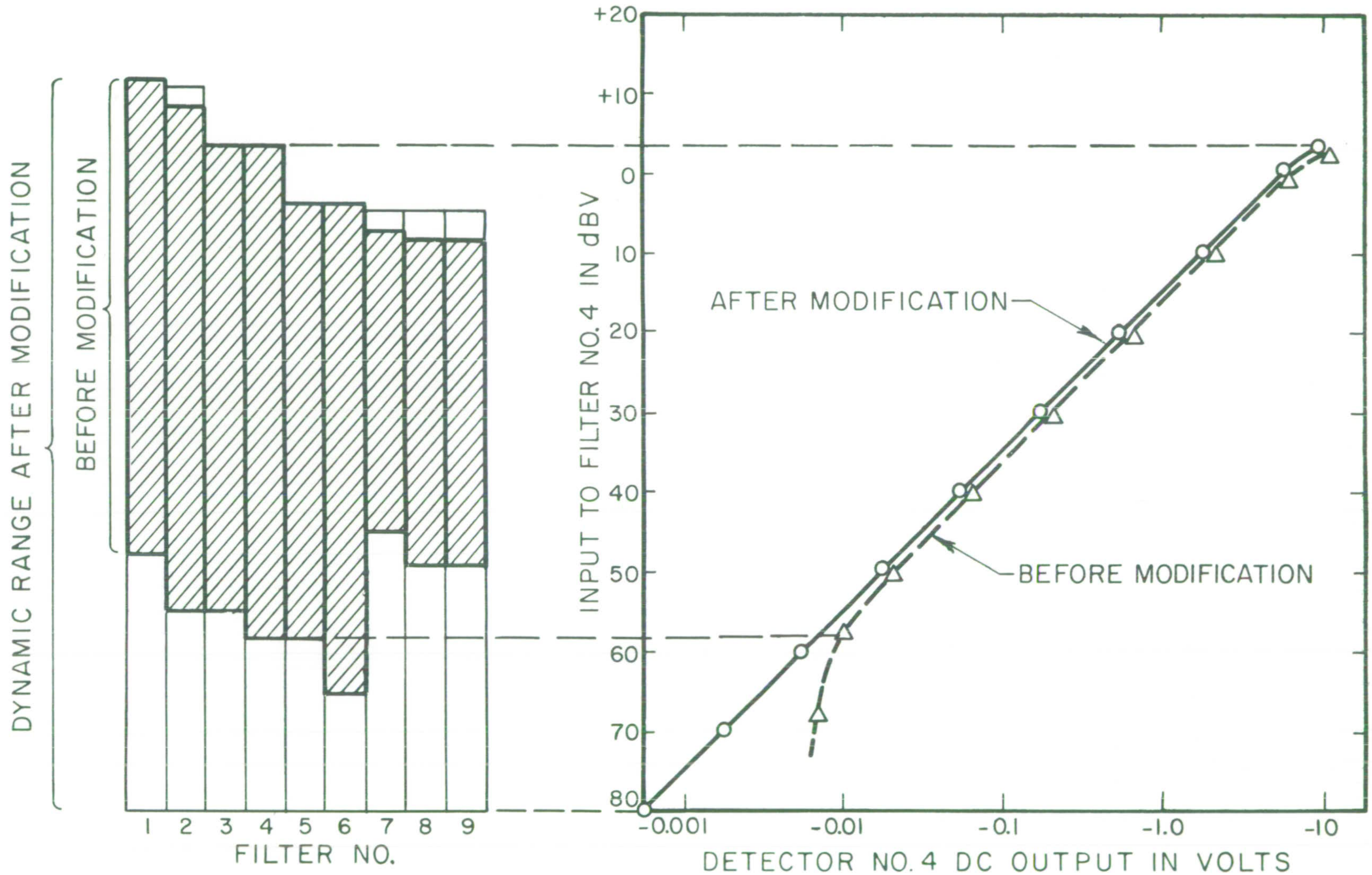

FIG. 3-2 ILLUSTRATION OF IMPROVEMENT IN DYNAMIC RANGE OF FILTERS IN ANALYZER. DYNAMIC RANGE IS REGION OVER WHICH FILTER INPUT AND DETECTOR OUTPUT ARE RELATED LINEARLY.

Previously, the back-to-back calibration was accomplished with the aid of a random-noise source. During the calibration procedure, as much as ± 0.10 SCI of scatter has been encountered at low SCI settings. A new calibration module, consisting of 9 oscillators whose outputs are summed, has been designed and fabricated. Using this module, the maximum scatter was found to be ± 0.01 SCI at all SCI settings. The new module thus permits a faster and more accurate calibration of SCIM.

4. REFERENCES

1. J. P. Egan, "Articulation Testing Methods." Laryngoscope 58, 955-991 (1948).

2. L. L. Elliott, "Backward Masking: Monotic and Dichotic Conditions." J. Acoust. Soc. Am. 34, 1108-1115 (1962).

3. G. D. Ewing and N. W. Huddy, Jr., "RF Clipping and Filtering to Improve the Intelligibility of Speech in Noise." IEEE Trans. Audio and Electroacoustics, Vol. AU-14, No. 4 (Dec. 1966).

4. F. A. Gunther, "Tropospheric Scatter Communications." IEEE Spectrum 3, No. 9, 79-100 (Sept. 1966).

5. I. J. Hirsh, E. G. Reynolds, and M. Joseph, "Intelligibility of Different Speech Materials." J. Acoust. Soc. Am. 26, 530-538 (1954).

6. A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, "Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set." J. Acoust. Soc. Am. 37, 158-166 (1965). Also Rept. ESD-TDR-64-507.

7. A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, "Psychoacoustic Speech Tests: A Modified Rhyme Test." Rept. ESD-TDR-63-403, Decision Sciences Lab., Electronic Systems Div., Air Force Systems Command (June 1963).

8. K. D. Kryter, "Methods for the Calculation and Use of the Articulation Index." J. Acoust. Soc. Am. 34, 1689-1697 (1962). Also Rept. ESD-TDR-62-35.

9. K. D. Kryter and J. H. Ball, "SCIM - A Meter for Measuring the Performance of Speech Communication Systems." Rept. ESD-TDR-64-674, Decision Sciences Lab., Electronic Systems Div., Air Force Systems Command (Oct. 1964).

10. K. D. Kryter, J. C. R. Licklider, and S. S. Stevens, "Premodulation Clipping in AM Voice Communication." J. Acoust. Soc. Am. 19, 125-131 (1947).

11. G. A. Miller, "The Masking of Speech." Psychol. Bull. 44, No. 2, 105-129 (1947).

12. E. Osman and D. H. Raab, "Temporal Masking of Clicks by Noise Bursts." J. Acoust. Soc. Am. 35, 1939-1941 (1963).

13. J. M. Pickett, "Backward Masking." J. Acoust. Soc. Am. 31, 1613-1615 (1959).

14. J. M. Pickett and I. Pollack, "Prediction of Speech Intelligibility at High Noise Levels." J. Acoust. Soc. Am. 30, 955-963 (1958).

15. I. Pollack and J. M. Pickett, "Intelligibility of Peak-Clipped Speech at High Noise Levels." J. Acoust. Soc. Am. 31, 14-16 (1959).

16. I. Pollack and J. M. Pickett, "Masking of Speech by Noise at High Sound Levels." J. Acoust. Soc. Am. 30, 127-130 (1958).

17. P. L. Rice, A. G. Longley, K. A. Norton, and A. P. Barsis, "Transmission Loss Predictions for Tropospheric Communication Circuits." Tech. Note 101, Vols. 1 and 2, National Bureau of Standards (1965).

18. C. W. Stuckey, "Development of a Method to Evaluate Learning in Articulation Testing." Rept. RADC-TN-60-198, Rome Air Development Center, Air Research and Development Command (Sept. 1960).

19. G. von Bismarck, "Addendum to Operation/Maintenance Manual for Speech Communication Index Meter." Rept. 1215B, Bolt Beranek and Newman Inc., Cambridge, Mass. (Apr. 1967).

20. W. Wathen-Dunn and D. W. Lipke, "On the Power Gained by Clipping Speech in the Audio Band." J. Acoust. Soc. Am. 30, 36-40 (1958).

21. C. E. Williams, M. H. L. Hecker, K. N. Stevens, and B. Woods, "Intelligibility Test Methods and Procedures for the Evaluation of Speech Communication Systems." Rept. ESD-TR-66-677, Decision Sciences Lab., Electronic Systems Div., Air Force Systems Command (Dec. 1966).

22. C. E. Williams, K. N. Stevens, M. H. L. Hecker, and K. S. Pearsons, "The Speech Interference Effects of Aircraft Noise." Final report prepared under Contract FA66WA-1566 for the Federal Aviation Administration, Washington, D. C. by Bolt Beranek and Newman Inc., Cambridge, Mass. (1967).

5. APPENDICES

Appendix 1

Word contents of twelve 25-item half lists of the Modified Rhyme Test.

Initial Consonant Variable

No.	Half List A	B	C	D	E	F
1	sold	fold	gold	told	hold	cold
2	dig	wig	pig	rig	big	fig
3	book	took	look	cook	hook	shook
4	gale	male	tale	bale	sale	pale
5	will	hill	kill	till	fill	bill
6	fame	same	came	name	game	tame
7	rang	fang	hang	bang	sang	gang
8	sip	rip	tip	hip	dip	lip
9	feat	meat	heat	seat	beat	neat
10	hot	got	not	pot	lot	tot
11	dust	just	rust	must	gust	bust
12	top	cop	pop	hop	shop	mop

Final Consonant Variable

No.	A	B	C	D	E	F
13	dig	dip	dim	did	dill	din
14	duck	dun	dung	dub	dug	dud
15	fizz	fig	fin	fill	fib	fit
16	came	cape	cane	case	cave	cake
17	kick	king	kid	kit	kin	kill
18	late	lay	lake	lace	lane	lame
19	pale	pane	pace	pay	page	pave
20	pig	pick	pip	pin	pill	pit
21	pun	puff	pup	puck	pus	pub
22	rave	rake	race	rate	raze	ray
23	seep	seen	seethe	seem	seed	seek
24	sub	sum	sud	sun	sup	sung
25	teach	tear	tease	teal	team	teak

Appendix 1, cont.

Initial Consonant Variable

No.	Half List					
	G	H	I	J	K	L
1	bed	shed	red	led	fed	wed
2	kick	lick	sick	tick	wick	pick
3	hark	dark	mark	lark	park	bark
4	peel	reel	feel	heel	keel	eel
5	foil	oil	coil	boil	soil	toil
6	ten	pen	men	hen	then	den
7	pin	sin	tin	win	din	fin
8	sun	nun	gun	fun	run	bun
9	tent	bent	dent	went	rent	sent
10	raw	paw	law	jaw	thaw	saw
11	kit	bit	fit	sit	wit	hit
12	nest	vest	west	test	best	rest
13	may	way	say	gay	day	pay

Final Consonant Variable

14	ban	bath	back	bass	bat	bad
15	bean	beach	beat	beam	bead	beak
16	bun	bus	but	buff	buck	bug
17	heath	heap	heal	hear	heat	heave
18	cut	cub	cuff	cup	cud	cuss
19	map	mat	math	man	mass	mad
20	pass	pack	pat	pad	path	pan
21	peace	peas	peak	peal	peat	peach
22	sake	sale	save	sane	safe	same
23	sad	sag	sass	sat	sap	sack
24	sit	sing	sin	sick	sip	sill
25	tab	tan	tam	tang	tack	tap

Appendix 2

Half-list MRT scores (corrected for chance) obtained from 8 listeners in an experiment employing steady-state simulated aircraft noise. Data from Williams et al. 1967.

Half List	S/N Ratio					
	+6dB	0dB	-2dB	-4dB	-6dB	Mean
A	84.4%	57.4%	61.6%	50.8%	34.6%	64.8%
B	88.6	55.0	50.8	40.6	37.0	62.0
C	86.2	70.6	58.6	34.6	26.8	62.8
D	90.4	71.8	60.4	56.2	43.0	70.3
E	86.2	73.6	65.2	49.6	41.2	69.3
F	84.4	62.8	53.8	44.2	29.8	62.5
G	85.0	68.8	54.4	44.2	33.4	64.3
H	93.4	75.4	64.6	50.2	25.0	68.1
I	81.4	62.2	67.0	49.0	28.6	64.7
J	89.2	73.6	53.2	46.0	35.2	66.2
K	86.2	65.8	56.2	43.6	28.0	63.3
L	88.0	67.6	49.6	51.4	33.4	65.0

Appendix 3

Order in which 75 Combination Tapes were administered to listeners during five test sessions. Each Combination Tape is identified according to the Speech Tape (1-25) and Program Tape (I-III) from which it was prepared.

Test Session				
1	2	3	4	5
1-I	16-III	2-I	15-III	6-II
2-II	17-II	13-II	4-I	25-III
3-III	18-I	1-III	24-II	15-I
4-II	20-II	11-II	12-I	9-III
5-III	19-I	18-III	21-II	17-I
6-I	21-III	22-I	11-III	12-II
7-III	22-II	7-II	16-II	5-I
9-II	23-III	25-I	3-I	14-II
8-I	24-I	10-III	20-III	4-III
10-II	25-II	23-II	1-II	21-I
11-I	8-III	19-III	13-III	18-II
12-III	14-I	20-I	7-I	22-III
13-I	5-II	3-II	2-III	19-II
14-III	9-I	17-III	23-I	10-I
15-II	6-III	16-I	8-II	24-III

Appendix 4 Typical response form.

NAME________________________ DATE________________ TEST NO.____________

No.			
1	bust rust	just gust	dust must
2	cake came	cape case	cave cane
3	fill fig	fizz fit	fib fin
4	dug dung	duck dud	dub dun
5	pop shop	top cop	hop mop
6	dim dip	dill dig	did din
7	fed red	shed bed	wed led
8	tack tam	tab tap	tan tang
9	sick sing	sit sill	sip sin
10	sick kick	tick pick	wick lick
11	sap sag	sad sack	sass sat
12	park dark	hark bark	lark mark
13	bale male	gale pale	sale tale
14	fill kill	will bill	hill till
15	rate rake	rave ray	raze race
16	game same	fame name	tame came
17	puck pun	pus pub	pup puff
18	pin pit	pig pick	pill pip
19	pin tin	sin fin	din win
20	man math	map mad	mass mat
21	run nun	sun gun	bun fun
22	rent bent	tent went	sent dent
23	jaw paw	thaw saw	raw law
24	cuff cuss	cut cud	cup cub
25	kid king	kin kill	kit kick
26	teal tease	teach teak	team tear
27	sud sum	sung sub	sup sun
28	dig rig	pig wig	big fig
29	seed seen	seep seek	seem seethe
30	hook look	shook book	cook took
31	keel feel	reel eel	peel heel
32	oil coil	toil foil	soil boil
33	save sale	sake same	safe sane
34	men hen	then den	ten pen
35	peak peach	peat peace	peal peas
36	path pack	pass pat	pad pan
37	hang bang	sang gang	rang fang
38	pane pale	pace pave	page pay
39	dip rip	sip tip	lip hip
40	heat meat	feat neat	beat seat
41	not hot	lot tot	pot got
42	lace lame	late lay	lane lake
43	hear heath	heap heave	heat heal
44	wit fit	kit hit	bit sit
45	but bus	bun bug	buff buck
46	beat beam	bean beak	bead beach
47	back bath	ban bad	bat bass
48	west nest	rest best	test vest
49	dig din	did dim	dip dill
50	hold gold	sold cold	fold told
51	teak team	teal teach	tear tease
52	sum sub	sup sun	sud sung
53	rig wig	dig fig	big pig
54	seep seek	seed seethe	seen seem
55	dark park	bark hark	mark lark
56	peel keel	heel reel	feel eel
57	coil soil	boil oil	toil foil
58	sake safe	sane save	sale same
59	den then	hen men	pen ten
60	peas peat	peal peach	peak peace

Appendix 5

Errors made by 10 listeners at each word location of Program Tape I.

SPEECH TAPE	WORD LOCATION																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1				2		8	6	2		2	1	6						2	8	
2				5	1	2	2			1		10			1	6				
3				6		6	1	5		2	1	8		2				1		
4				9		2	3					10			2	1			4	
5				1		9	1	6	6		6	1				2	1		7	
6	1			2	2	9	1	9	1	1	10	5				2				
7		4								1	8	2			7					
8				9		10		10	1			10				6			1	
9				2		2	3	6		1		8	6		5	4				
10				1		9		2		6		10			2	2	1		3	
11			1	3		9		2				9			1	2			3	
12				3		7		4		5	2	9							3	
13						6	3	7		6		7		1		5	2		1	
14				6	1					3	1	4			1	8	3		7	
15						6						6				6			7	
16						3	5	2		5	1	8			1	2				
17	1							6				8		3		1				
18						6				5		10	1						1	
19				7		8				2	1	7				2		1		
20				9		1	1	1			1	7				2				
21	4			5		7		2		7	2	6				8			2	
22		1		7	1	1		8			9	6	3		1					
23						5	1			3		2			1	1				
24				1		6	1	5				9			1	6		1		
25				6		1	4	10		2	9	10		1						3
Mn Ɛ	0.24	0.20	0.04	3.36	0.20	4.92	1.28	3.48	0.32	2.08	2.08	7.12	0.40	0.28	0.92	2.64	0.28	0.20	1.88	0.12
SD	0.83	0.82	0.20	3.23	0.50	3.38	1.74	3.45	1.22	2.29	3.34	2.68	1.32	0.74	1.68	2.66	0.74	0.50	2.68	0.60
SCORE	97.0	97.5	99.5	59.5	97.6	41.0	84.7	58.2	96.0	74.9	75.0	14.6	95.0	96.5	88.9	68.4	96.5	96.5	77.4	98.5

Appendix 5, cont.

Errors made by 10 listeners at each word location of Program Tape I.

SPEECH TAPE	WORD LOCATION																			
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1		5	7	8		1		7				3	3		1		4	8	4	7
2			8	6				7	1			5	8		4		7	7		8
3		2	7	8			5	1		2		6					8	9		7
4			9	9			10	4	1			7	6					5		6
5		8	10	10			1	2	10			8	1				6	9		
6		2	10	8	2			7	7			7	4	5			10	8		4
7	2	6	8	9				2	2	8		8	3	5		1	8	10		4
8		10	10	9			4	8	6				6	1			5	6		4
9		5	8	7					1		1	5	3		3		7	8		10
10		1	8	8			5		6		6	7	1			1	9	9		8
11		4	6	10			4	9	5			6					9	5		6
12			1	10					4			5	6	1			9	7		4
13	5	5	8	10			4		3			4	7	5			8			
14		1	6	8			4	3				4	5				3	6		2
15			7	8			8	8	2					2	2		8	7	1	1
16	1	1	9	10			7	2	8			8	6				9	9		1
17		6	5	8			7	10	1	3	2	5	9	4		3	7	6		
18			5	8	1			1			4	6	8				1	10		4
19	2		5	8		4	9	10	3		3	2	9	6	5		8	8		3
20			4	9			3	7	7		4	2	3				8	1		
21			6	9			4	3	2		3	3			1		10	8		6
22			6	7	6		5	6			1	9	8					10		
23		4	7	9		3	2	6	4				5				7	7		1
24			9	6				8			1	4			1		9	6		2
25		1	10	10			3		7			6	9		2		3	3		
Mn Ɛ	0.40	2.44	7.16	8.48	0.36	0.32	3.40	4.44	3.20	0.52	1.00	4.80	4.40	1.16	0.76	0.20	6.52	6.88	0.20	3.52
SD	1.12	2.93	2.17	1.19	1.25	0.99	3.11	3.51	3.01	1.71	1.68	2.60	3.21	2.03	1.39	0.65	3.02	2.59	0.82	3.06
SCORE	95.1	70.6	14.3	2.3	95.5	96.0	59.2	46.6	61.6	93.5	88.0	42.8	47.3	86.1	90.6	97.5	21.8	17.9	97.5	57.7

Appendix 5, cont.

Errors made by 10 listeners at each word location of Program Tape I.

SPEECH TAPE	WORD LOCATION																			
	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	6					7		10	6		3			8	6		8		10	
2								9	7	2	3	7		4	2		7		6	
3	8	6				1		9			2			8	9		9		7	3
4	8		2			2		9	6		1			1			7		4	5
5	8	3	5			1		5	1		8	10		7	2		7		6	10
6	8	5	2					9	8		7			7	10		8		7	5
7	8		5			4	2	6	4			4			8	2	8		6	
8	10		1		1	3		4	5		6			5	9		8	1	7	7
9	7	1	6					9	10		2	7	6		8		5		9	2
10	10	5	3			3		9						4	5	1	9		10	4
11	7	1						8	5		5			1	8		8		7	1
12	2		1		1										1		3		7	
13	6	5			2			1	3	2				1	9		10		6	9
14			4			1			1		3		4		5		8		9	2
15	3		2		3	2			7			8		3	10	8	6		8	5
16	6	5						7	6		3	7	1	6	8	1	8		7	3
17	8	3			5	2		2			5		1	2	6				7	6
18	7	9	2					9	2			9	1	6	8		5		10	2
19	9		3	1		5		3	6		9			5	9		6		5	8
20	6	6				8		8	3	1		4	3	1	9		6	1	7	1
21	9					8		3				1	8		10	1	7		7	
22	1							9			1	3	8	2	8		10		6	1
23	5	3	5	1					2		4	1		1	6		5		7	
24	6					1			9	2	2		2	5	10		6		7	8
25	10	3			9	1		7	6		10	9		3	7		7		9	
Mn Ɛ	6.32	2.20	1.64	0.08	0.84	1.96	0.08	5.44	3.88	0.28	2.96	2.80	1.36	3.20	6.92	0.52	6.84	0.08	7.24	3.28
SD	2.98	2.66	2.00	0.28	2.08	2.56	0.40	3.73	3.17	0.68	3.08	3.65	2.50	2.74	2.94	1.64	2.17	0.28	1.54	3.22
SCORE	24.4	73.6	80.4	99.5	89.8	76.5	99.0	34.6	52.6	96.5	64.5	66.3	83.8	62.1	16.9	94.1	17.9	99.0	13.8	60.6

Appendix 5, cont.

Errors made by 10 listeners at each word location of Program Tape II.

SPEECH TAPE	WORD LOCATION																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1			2		8	5	6		1	1	4		2					7		
2			3	5	2	6	2				9	1	7	2			4	6		
3			1	2		8	5				4			7				2	1	6
4		3	3	1		5	1				7		9	1			1	9	3	3
5	1		3	5		8							3							5
6			5	1	1	9					3		8	1				5		
7		2				7	2		1		1	1	9	2	2		2	6		
8			2	3		4	5		2		2	1	8	3				5		3
9	4		3			10	1				1	1	2		6	1		3		1
10			3			10					2		2				1	2	1	
11	1		6			10	1				5	1	4	3				9	3	4
12				1		9		4			2		2					2		
13						10	4			4	4		9	7			6	6	4	
14			1	3	4		4				7		2			1	5	2	6	2
15		1			2	9	2	1			8		9	1		7		6	9	3
16			2			9	6	2					8	1				7	1	
17			3		2	1	2	1			3		9	3				4		3
18		1			1	9					4		10					6	1	1
19			8		4	8						2	6					9		
20		1		4		2	2	4			4		5					7	1	3
21	8		6	2	2	7	8				3	1	6	1				9		
22			5	1	1	9	2						8	1				6		
23			9			7	6	1	5	1		7	7					4		
24						8	9				4		6				2	10	6	5
25				4	3	6	6			1	1		7	2	1			2		6
Mn E	0.56	0.32	2.60	1.28	1.20	7.04	2.96	0.52	0.36	0.28	3.12	0.60	5.92	1.40	0.36	0.36	0.84	5.36	1.44	1.80
SD	1.76	0.75	2.63	1.72	1.91	2.84	2.73	1.16	1.08	0.84	2.60	1.44	2.96	1.98	1.25	1.41	1.70	2.74	2.42	2.12
SCORE	93.0	96.0	68.8	84.7	85.1	15.5	64.3	93.6	95.6	95.5	61.6	92.6	29.1	83.2	95.6	95.5	89.9	35.7	82.8	78.5

Appendix 5, cont.

Errors made by 10 listeners at each word location of Program Tape II.

SPEECH TAPE	WORD LOCATION																			
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1		6	6	9									8		7			1	7	
2		10	2							6			7		1	9		5	1	2
3		8		5			8			3			6	4	2	8	4	2		
4		6					9			8			8	8		6		1	1	2
5		8		10			2		2	10			10		1	7		3	1	
6		8	9	8					3	5			7	8	6	4			1	
7	1	6		10	1		8			7	1		9	6	1	10		7	1	
8		8	10	2			4		4				9	3	8	5				
9		6		9		1	10			8			7		4	5			1	5
10		9		3	1		4				2		9	5		10	1	1	6	
11		6		2			3						10	4	2	3		2		
12		10	6	10			8			1			8	4	6	8				
13	2	6		5			6						7	8	1	5			2	
14		8	7	1		2	5	1					9	1		8	1	3	1	
15	1	10	9	9			7			1			9	10	1	7			5	
16		9		3			9		3	1			7	3	8	10	4		1	
17		3	9	8			9			5			8		8	3			5	
18		2		2			8			3			8			9		1	3	
19	2	6	6	5			9		4	4	1		7	6	1	9				
20		5		4			8				6	1	8	4	5	1				
21		8	5	1			3	9		2			10	8	9	9		2		1
22		9	4	7	7		8						8	4		7		8		
23		8	1	6		3	8		2	5			8	2		6			7	
24		3	9	6			7			2	1		7			8				
25		7	9	1			7						8		1	7				
Mn Ɛ	0.24	7.00	3.68	5.04	0.36	0.24	6.00	0.40	0.72	2.84	0.44	0.04	8.08	3.52	2.88	6.56	0.40	1.44	1.72	0.40
SD	0.60	2.18	3.92	3.42	1.41	0.72	3.12	1.80	1.36	3.10	1.26	0.20	1.07	3.20	3.17	2.75	1.12	2.24	2.34	1.12
SCORE	97.0	15.9	51.9	38.6	95.5	97.0	28.3	95.0	91.2	66.0	94.6	99.5	5.9	57.7	65.5	22.1	93.1	82.9	79.3	95.2

Appendix 5, cont.

Errors made by 10 listeners at each word location of Program Tape II.

SPEECH TAPE	WORD LOCATION																			
	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	9			1				5						2			1			
2	8			7				1		2	4	4			1				3	
3	8			5				5				5			2				1	
4	6		2	2							7		1	5				1		
5	7	5		6				7			9	10		5	3				3	
6	7				1	1		4			4	1	7	8					6	
7	9		2	1			2	1				4		1			2	1	3	
8	9							6				1	5	6	5			8	2	
9	7			10	1			3			2	6	3		1				2	
10				2				2				1						1	1	
11	6			4				10			2		2		3					
12	6			6				1						1	2					
13	7			2						4	8		6	1						
14	1		1	3							7		4							
15	9			3	3			2		1		2			2			1		
16	1			3				2	1		5	6	1						1	
17	8							4			8	7	4	1						
18	6			9				8	1		6	9		9	1					
19	8			7							1		1							
20	5			4				4		4		10	7	4	8			7	1	
21	7							1			5	8	4		4			4		
22	7			4							6	3	9	1	1					
23	8		4	7							3	5		3	1					
24	6			5						2	10	2		2						
25	9			1				5				1							1	
Mn Ɛ	6.56	0.20	0.36	3.68	0.20	0.04	0.08	2.84	0.08	0.52	3.48	3.40	2.16	1.96	1.36	0.00	0.12	0.92	0.96	0.00
SD	2.50	1.00	0.95	2.90	0.65	0.20	0.40	2.85	0.28	1.19	3.39	3.42	2.79	2.68	1.98	0.00	0.23	2.16	1.49	0.00
SCORE	21.4	97.5	95.5	55.8	97.5	99.5	99.0	65.8	99.0	93.7	58.3	59.2	74.2	76.5	83.6	100.0	88.5	88.8	88.3	100.0

Appendix 5, cont.

Errors made by 10 listeners at each word location of Program Tape III.

SPEECH TAPE	WORD LOCATION																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1			3				2			3	2					3		9	10	10
2		1		6								2				10	1			3
3														1				4		9
4		4							1									3	6	5
5	9	1		1			1		1				1			5				6
6	1								4									8	1	9
7		2													3	6	1	1		10
8		1										2		1		10		3	1	8
9	4	1				1						1	4		2	8		3	6	5
10															1	7		2	2	10
11			1									2				8		7		9
12								4								4		3	7	8
13						7				9							1	3	4	10
14							1				1					10	4	4	6	8
15		8						1			1	1				6	5	4	10	10
16							2						1		1	6		9	5	6
17	5										1			2		9		3		2
18	6								8			1	1			10		1	1	10
19			2						7			2	3			2		6	1	6
20		2						3										1		5
21	7					1	1			2		2	1			2		5	3	8
22					1	1						3	1					3		9
23	1					3				2		9			1	6	1	8		5
24	2															4		8	4	8
25	3						2									6		3		8
Mn Ɛ	1.52	0.80	0.24	0.28	0.04	0.52	0.36	0.32	0.84	0.64	0.20	1.00	0.48	0.16	0.32	4.88	0.52	4.04	2.68	7.48
SD	2.62	1.78	0.72	1.20	0.20	1.50	0.70	0.99	2.17	1.91	0.50	1.91	1.00	0.47	0.75	3.63	1.26	2.76	3.24	2.33
SCORE	81.9	90.3	97.0	96.5	99.5	93.5	95.6	96.0	89.7	92.0	97.5	87.8	94.2	98.0	96.0	41.7	93.6	51.6	67.9	10.9

Appendix 5, cont.

Errors made by 10 listeners at each word location of Program Tape III.

SPEECH TAPE	WORD LOCATION																			
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	7	3	1		7	1					3	5	2		7				4	
2	9					3					9	9	10		9					
3	10				1		2			1	1	3	10		2	1	5		1	3
4	10				3	4	1				2	1	1		10				2	1
5			1		10	2	7			1	8	9			7					2
6	10	2	9		2	6					5	5	6		10					
7	5	2	1		8		5			9	9	9		2	2					
8	1		8		8				3		1	1			8					
9			4		6	2					7	4	1		10					4
10	1				4						4	7			6					
11	7			8	6						5	7			3				1	1
12	9			1	4	5	1				9		9							1
13	8	3			10		5				6	1								2
14	1	2	1		8	4		3			8	6		1	1					
15	10		8		4	6	3				7	2			5	1			1	
16	4	1				1		1	1		8	6			3		7			
17	8				9	1	1			3	3	4	1			4				6
18	6				8	1	2				5	7	6		2				1	
19	6		6			6	1		1		7	8	4	2	8					
20	8				6						9	6								
21	8	1	9	1	6	5	2	3			6	7	1	1	6	2				2
22	7		1	1	2		1	1			5	8			5			1		
23	2		5		1	5	2			6	1	3								
24	5		9		6						4	6			6		1			
25	7		6												7					
Mn ε	5.96	0.56	2.76	0.44	4.76	2.08	1.32	0.32	0.20	0.80	5.28	4.96	2.04	0.24	4.68	0.32	0.52	0.04	0.40	0.88
SD	3.36	1.00	3.53	1.61	3.33	2.31	1.89	0.85	0.65	2.16	2.85	2.91	3.37	0.60	3.52	0.90	1.69	0.20	0.91	1.54
SCORE	28.8	93.1	66.8	94.5	42.8	75.2	84.2	96.0	97.5	90.2	36.8	40.6	75.5	97.0	44.0	96.0	93.5	99.5	95.1	89.4

Appendix 5, cont.

Errors made by 10 listeners at each word location of Program Tape III.

SPEECH TAPE	WORD LOCATION																			
	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	1			1	1		5	6	8						6	5	9			
2				4			9	2	3	1					5	8	2			
3				1			1	8	1	1	1				10	8				
4				3			1	8	6						5	9	6			
5	1	6		9	1			7	4	3	1				2	10	7			
6	1	1			2		7	8	1		2			2	10	10	8			
7	4		3	5			10	9	3	2					10	8	6			1
8	5		1				4	8	7					1	8	8	8	3		
9	1			6			8	10	2			1	1		2	10	8			
10				3	2		5	8		1					10	5	9			
11					2			8	6						6	7	5			
12	1		1	1			5	5	5						9	6	2			
13				2	3		10	5	7						6	7	10	1		
14							7		4	9	4				7	10	8	4		1
15				1	4		1	2	7	9					10	5	6			
16		1		2	1		3	8	8	6	2	1			8	3	1	7		
17					5		3	4	2	3	5				4	8				
18				6			4	5	2	1				3	10	10	2			
19				6			6	6	1	1					9	10	7			
20				8		7		8	7	7	1	1	1		7	6	7	2		
21	2						3	7	1	9	2	1	3		8	9	6	2		
22	1			1	1		3	6	6	2	1		1		10	8	9	3		
23			2	9				10	10						8	6	6	3		
24	1			8			1		5	6				1	3	7	3			
25	1	1			9		10	8	8						9	9	9			
Mn ε	0.76	0.36	0.28	3.04	1.24	0.28	4.24	6.24	4.56	2.44	0.76	0.16	0.24	0.28	7.28	7.68	5.76	1.00	0.00	0.08
SD	1.27	1.22	0.74	3.16	2.13	1.40	3.39	2.80	2.83	3.20	1.33	0.37	0.66	0.74	2.61	1.95	3.06	1.78	0.00	0.40
SCORE	90.8	95.5	96.5	63.4	85.2	96.5	49.2	25.5	45.7	70.7	90.7	98.0	97.0	96.5	14.2	9.1	30.9	88.1	100.0	99.0

Appendix 6

Calculation of Articulation Index for Word Location 52 of Program Tape I. (See also Fig. 1-6.)

Band No.	Speech Level	Noise Level	Difference	Weight	Contribution to AI
1	70	49	21	0.0003	0.0063
2	71	54	17	0.0007	0.0119
3	72	56	16	0.0010	0.0160
4	74	51	23	0.0016	0.0368
5	72	58	14	0.0017	0.0238
6	70	59	11	0.0017	0.0187
7	68	53	15	0.0027	0.0405
8	67	57	10	0.0030	0.0300
9	65	62	3	0.0033	0.0099
10	63	61	2	0.0037	0.0074
11-14	-	-	0	-	0.0000
15	54	46	8	0.0017	0.0136
				AI →	0.2149

Appendix 7

Tabulation of Articulation Indexes for 60 word locations of each of 3 Program Tapes.

Word Location	Program Tape I	II	III	Word Location	Program Tape I	II	III
1	0.928	0.735	0.207	31	0.337	0.723	0.002
2	0.767	0.656	0.239	32	0.059	0.642	0.052
3	0.839	0.131	0.498	33	0.121	0.000	0.498
4	0.107	0.211	0.630	34	0.235	0.081	0.453
5	0.813	0.377	0.761	35	0.387	0.184	0.138
6	0.000	0.004	0.498	36	0.742	0.038	0.747
7	0.368	0.088	0.544	37	0.027	0.838	0.777
8	0.205	0.717	0.580	38	0.000	0.231	0.740
9	0.649	0.495	0.458	39	0.744	0.108	0.620
10	0.147	0.783	0.569	40	0.144	0.309	0.617
11	0.321	0.052	0.477	41	0.002	0.012	0.170
12	0.000	0.618	0.281	42	0.137	0.749	0.720
13	0.444	0.021	0.436	43	0.203	0.300	0.753
14	0.593	0.538	0.412	44	0.647	0.051	0.181
15	0.258	0.703	0.593	45	0.548	0.690	0.223
16	0.081	0.760	0.242	46	0.199	0.712	0.664
17	0.494	0.285	0.492	47	0.746	0.674	0.144
18	0.434	0.020	0.054	48	0.029	0.162	0.048
19	0.221	0.275	0.072	49	0.143	0.788	0.056
20	0.506	0.205	0.025	50	0.751	0.661	0.113
21	0.357	0.394	0.024	51	0.110	0.066	0.281
22	0.171	0.000	0.718	52	0.215	0.160	0.715
23	0.028	0.000	0.017	53	0.195	0.106	0.715
24	0.000	0.143	0.785	54	0.221	0.137	0.630
25	0.542	0.735	0.045	55	0.006	0.188	0.008
26	0.434	0.674	0.051	56	0.378	0.727	0.000
27	0.059	0.049	0.169	57	0.000	0.536	0.004
28	0.089	0.567	0.212	58	0.753	0.241	0.187
29	0.067	0.735	0.532	59	0.003	0.257	0.756
30	0.534	0.254	0.558	60	0.162	0.801	0.826

Appendix 8

A. Word locations where predicted test scores were too low, tabulated according to probable cause of error. Underlined word locations are illustrated in Fig. 1-11.

Cause 1: Speech level read too low
I-12, I-16, I-38, I-41, I-57, I-59,
II-6, II-18, II-22,
III-57.

Cause 2: Noise level read too high
II-39,
III-28.

Cause 1 and Cause 2 combined
I-6,
II-11, II-23, II-44,
III-19, III-23, III-26, III-31, III-41.

B. Word locations where predicted test scores were too high, tabulated according to probable cause of error. Underlined word locations are illustrated in Fig. 1-12.

Cause 1: Speech level read too high
I-11, I-23, I-33, I-40, I-52,
II-24, II-27,
III-20, III-35.

Cause 2: Noise level read too low
I-60,
II-36,
III-33, III-44, III-47.

Cause 1 and Cause 2 combined
I-8, I-49, I-54,
II-14, II-30, II-52,
III-16, III-48.

Appendix 9

Harvard PB-Word scores (in percent) corresponding to various Speech Communication Indexes for peak clipping with level compensation.

SCI	Clipping Level				
	0 dB	6 dB	12 dB	18 dB	24 dB
1.00	100	100	100	100	100
0.95	99	99	99	99	99
0.90	98	98	99	99	99
0.85	97	98	98	99	99
0.80	96	97	98	98	99
0.75	95	96	98	98	99
0.70	94	96	97	98	98
0.65	93	95	97	98	98
0.60	92	95	96	97	97
0.55	91	94	96	96	96
0.50	90	92	95	95	95
0.45	88	90	94	93	93
0.40	86	86	90	91	91
0.35	83	83	86	87	88
0.30	78	77	82	81	83
0.25	66	67	73	75	77
0.20	50	56	63	66	69
0.15	35	44	51	55	59
0.10	18	28	39	-	-
0.05	2	-	-	-	-
0.00	-	-	-	-	-

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Bolt Beranek and Newman Inc., Cambridge, Mass. 02138	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b GROUP N/A

3. REPORT TITLE

SPEECH COMMUNICATION INDEX METER:
FURTHER APPLICATIONS AND IMPROVEMENTS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Final Report

5. AUTHOR(S) *(Last name, first name, initial)*

Hecker, Michael H.L., von Bismarck, Gottfried, and Williams, Carl E.

6. REPORT DATE 26 July 1967	7a. TOTAL NO. OF PAGES 85	7b. NO. OF REFS 22

8a. CONTRACT OR GRANT NO. AF 19(628)-5874 b. PROJECT NO. 2808 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ESD-TR-67-516 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* BBN Report No. 1481

10. AVAILABILITY/LIMITATION NOTICES

THIS DOCUMENT HAS BEEN APPROVED FOR PUBLIC RELEASE AND SALE; ITS DISTRIBUTION IS UNLIMITED

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Decision Sciences Laboratory, Electronic Systems Division, Air Force Systems Command, USAF L. G. Hanscom Field, Bedford, Mass. 01730

13. ABSTRACT

To study the feasibility of using the Speech Communication Index Meter (SCIM) to evaluate time-varying communication systems, recordings were made of the transmission characteristics of a troposphere-scatter system. At many specific points in time in these recordings, Articulation Indexes were calculated and intelligibility scores were obtained from listeners with the aid of a special test procedure. For most points, the intelligibility score could be reasonably well predicted from the Articulation Index. This finding was interpreted as indicating that SCIM is potentially capable of evaluating time-varying systems.

In a second study, SCIM was used to predict the intelligibility of peak-clipped speech. For a wide range of signal-to-noise ratios and clipping levels, the performance of SCIM was compared with intelligibility test results reported in the literature. After a minor modification, SCIM computed reliable SCI's which could be related to the published intelligibility data. This study demonstrated that SCIM could accurately evaluate communication systems employing peak clipping.

SCIM was also modified to reduce the computation time, and to improve the synchronization between the signal generator and the analyzer. In addition, various other circuit changes were implemented to improve the performance of SCIM.

DD FORM 1 JAN 64 1473

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Clipped Speech						
Evaluation of Communication Systems						
Psychoacoustics						
Speech Communication						
Speech Intelligibility						
Tropo Scatter Communications						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
______________________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
______________________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
______________________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.