NTIA Report 85-188

Voiceband Quality-of-Service Issues in the Post Divestiture Environment

James A. Hoffmeyer

U.S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

Rodney L. Joyce, Acting Assistant Secretary
for Communications and Information

December 1985

PREFACE

The author gratefully acknowledges the contributions of Ms. Evelyn M. Gray whose assistance in performing the literature search for this report was invaluable. She not only acquired more than 600 documents, but also developed short abstracts of each document. These efforts saved the author many hours of work in the research necessary for writing this report.

The author is also indebted to Mr. Neal B. Seitz and Mr. Robert F. Linfield for their many valuable suggestions.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF FIGURES

LIST OF TABLES

VOICEBAND QUALITY-OF-SERVICE ISSUES IN THE POST DIVESTITURE ENVIRONMENT

James A. Hoffmeyer*

This report discusses quality-of-service (QOS) issues for telephone networks. Deregulation and divestiture have fostered increased competition in the United States in the telephone equipment and service industries. There are many economic, policy, and technical issues that remain to be solved as the result of the plethora of equipment and services now available. This report addresses the technical problems associated with the interconnection of equipment from many vendors. In order to maintain a satisfactory quality of service to the end user, performance standards must be developed, approved, and implemented. The work of IEEE, CCITT, and ANSI-accredited standards groups responsible for telephone QOS is reviewed. The problem of interconnecting different national networks in the international community is seen to be analogous, in part, to the problem of interconnecting the numerous public and private networks within the United States. Although progress has been made by both national and international telephone QOS standards groups, unsolved issues remain. Principal among these are the development of objective measures of voice quality, the mapping of these objective measures into five levels of quality, enhancement of IEEE and CCITT telephony QOS standards (including the development of standards for the transmission of data on voiceband networks), and the development of QOS standards for Integrated Services Digital Networks. These are discussed in this report along with recommendations for new programs that would contribute to their resolution.

Key words: competition; divestiture; objective quality evaluation; quality of service; subjective quality evaluation; telephone systems standards

1. INTRODUCTION

The divestiture of the Bell Operating Companies (BOC's) and the emergence of a competitive environment for interexchange services are having profound effects on the telecommunications industry. These may be seen as the culmination of deregulation initiatives and technology advances set in motion over a decade ago with the Federal Communications Commission's landmark Carterphone and Specialized Common Carrier decisions in 1968 and 1971, respectively. The Carterphone decision permitted the connection of non-telephone-company-provided equipment to the public telephone system, thereby creating the interconnect industry, which today provides to telephone companies and their customers more than $2 billion annually in

*The author is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80303.

customer-premises equipment (CPE), such as private branch exchanges (PBX's), key-sets, and telephones. The Specialized Common Carrier decision made possible the growing variety of competitive long-distance services now available. The Modified Final Judgment (MFJ) in the American Telephone and Telegraph Company (AT&T) anti-trust case required AT&T to divest itself of the Bell Operating Companies (BOC's) and gave each entity new incentives to innovate and compete in telecommunications and related information processing markets.

Much of the discussion to date on the effects of the AT&T divestiture has been focused on economic issues (Jensen, 1983). There are technical issues as well, many of which can be lumped together under the category of quality of service. Previously, the vast majority of customer premises equipment, local loops, end-office and toll switches, and interswitch trunks were part of a single integrated system provided by AT&T and its subsidiary operating companies. The telecommunication system of the future will be comprised of equipment and services manufactured and operated by a multitude of vendors--each potentially involved in a typical long-distance call. Clearly, this will make it more difficult to provide uniformly high quality of service (QOS) to the end user. For voice, quality of service includes <u>transmission performance factors</u> such as intelligibility, loudness, noise, echo, speaker recognizability, delay, and naturalness of speech. Quality of service also includes <u>switching performance factors</u> such as the number of digits that must be dialed, access time, blocking probability, and disengagement time. This report emphasizes the transmission performance factors because these are currently of more direct concern to end users.

A typical long-distance call today involves three independent service providers: the access BOC, the interexchange carrier, and the terminating BOC. It will be difficult for the end user to determine which service provider is at fault when poor service is perceived. Minimum performance standards for each segment of an end-to-end circuit will be helpful in ensuring that the high quality of service traditionally provided to, and expected by, end users is maintained.

The quality of service issues arising from the divestiture include the following:

1. What does the term "quality of service" mean to telecommunication users? To telecommunication service providers? Will more comprehensive or more precise quality parameters be needed in the post-divestiture environment, or are existing measures adequate?

2. How will end-to-end quality of service objectives be allocated among component networks in the post divestiture environment? Who will be responsible for overall service planning? For representing service performance to customers? For service monitoring and network optimization? For fault isolation? For billing, and the granting of credit for service interruptions? Will competition among interexchange carriers complicate these processes?

3. Existing international communications involve joint planning and interoperation among many independent service providers. How valid is the international model in representing the U.S. post-divestiture environment? Are there lessons to be learned from international experience? Are the principles used in allocating quality objectives among national networks applicable to the multinetwork environment emerging in the United States?

4. What does the divestiture requirement that BOC's provide "equal exchange access" mean in engineering terms? Will "equal exchange access" require explicit quality-of-service specifications for exchange networks? Will demonstration measurements be needed?

5. The U.S. Congress has recently considered legislation aimed at stabilizing local telephone rates in the post divestiture environment. What are the implications of such legislation on service quality?

The above questions are very difficult to answer. It is not the intent of this report to answer these questions fully, but rather to (1) summarize the current status of performance assessment work pertinent to them, and (2) identify related problem areas that have not been adequately addressed. The report recommends further research into several such areas.

Section 2 of this report provides a more complete definition of the QOS problem. Although the QOS issue is technical in its origin, its resolution has both economic and policy ramifications. Differences in the QOS requirements for voice and voiceband data are briefly discussed.

Section 3 summarizes the status of standards committee efforts to address the question of quality of service for analog voice circuits. A standards committee of the Institute of Electrical and Electronic Engineers (IEEE) has developed a draft standard (P823) on QOS for voice circuits. International Telegraph and Telephone Consultative Committee (CCITT) Study Group XII is developing models for predicting user-perceived transmission quality from objective measurements. Accredited Standards Committee (ASC) Working Groups T1Q1.1 and T1Q1.2 are investigating 4 kHz voice and 4 kHz voiceband data, respectively. These two working groups are developing standards for 4 kHz voice and 4 kHz voiceband data performance.

Section 4 is devoted to a discussion of QOS models that have been formulated outside of the various standards committees. These include transmission performance models developed by AT&T and British Telecom; delay, timing jitter, and speed loss models; and customer behavior models.

Section 5 summarizes the more significant quality issues that have not been addressed (or that have been addressed but not resolved) to date. These include:

- the enhancement of existing IEEE and CCITT telephony QOS standards
- the development of objective measures of speech quality
- the development of procedures for mapping objective measures to five levels of quality
- the development of QOS standards for voiceband data and other services
- the development of QOS standards for Integrated Services Digital Networks
- investigation of the relevancy of the Open Systems Interconnection (OSI) model for voice.

Section 6 provides recommendations for future work in developing meaningful user-oriented voice QOS parameters and measurement techniques. This proposed new work would support CCITT standards efforts as well as contributing to the resolution of domestic QOS issues such as equal exchange access.

The appendixes provide supplemental information on both subjective and objective speech-quality measures. Although substantial research has been devoted to the development of objective speech-quality measures, no single objective measure has been widely accepted. Such a widely accepted measure is urgently needed.

2. QUALITY-OF-SERVICE PROBLEM DEFINITION

Quality of service in this report refers to the end-user's perception of the entire process of call setup, conversation (which may be thought of as the information transfer phase), and call disengagement. As noted earlier, the emphasis in this report is on the conversation phase--specifically, on the intelligibility of received speech signals. Brief discussions of voiceband data (VBD) requirements and QOS specification in the future Integrated Services Digital Network (ISDN) are also provided.

Some uniform terminology will facilitate the QOS problem definition. An <u>end user</u> is an individual or a computer program that produces or ultimately consumes information transferred over a telecommunication system (Nesenbergs et al., 1981; ANSI, 1983). Typical end users are the calling and called parties in a telephone conversation. User-oriented performance parameters are those that describe services provided to end users, as opposed to particular network facilities (Gruber and Le, 1983a). The access phase of the telecommunications process encompasses the activities required to establish a communication path between end users. Once such a path has been established, information transfer phase begins. The transfer phase consists of the flow of information between the end users. It encompasses formatting, transmission, storage, error control, and protocol or media conversion activities performed between start of output and completion of delivery. The transfer may be unidirectional or bidirectional. Completion of the transfer phase marks the beginning of the disengagement phase, which ends when the system is returned to its initial state and is ready to allocate the access path to other users (Nesenbergs et al., 1981; ANSI, 1983).

A recently approved American National Standard (ANSI, 1983) specifies data communication performance in terms of access, transfer, and disengagement parameters. Voice quality parameters may also be classified in this way. The term grade of service as used in this report refers primarily to the access phase--specifically, the probability of a call being blocked or delayed. Quality of service refers to all three phases.

2.1 Relating Network Parameters to User-Oriented QOS

The need for QOS standards and models for analog voice connections has been well described (DiBiaso, 1983; Cavanaugh et al., 1983; Silverthorn, 1983; Jensen, 1983; Palladino and Wilkens, 1983; and Kort, 1983). This need for standards has increased as a result of divestiture and deregulation (Johnson, 1983). As shown in Figure 1, a typical long-distance call involves customer-premises equipment at both ends of the circuit, exchange network facilities at both ends of the circuit, and interexchange network facilities. In the past, all of these facilities were typically provided by one company (AT&T). Now there is a plethora of CPE and interexchange service vendors (Jensen, 1983).

In the access areas, the Regional Bell Operating Companies (RBOC's) dominate, but facilities designed to bypass the local telephone network are becoming increasingly available. As noted by Ebert (1984), the access network is becoming increasingly competitive. Technologies such as privately owned microwave, infrared, local area networks, and satellite terminals are being employed in bypass systems.

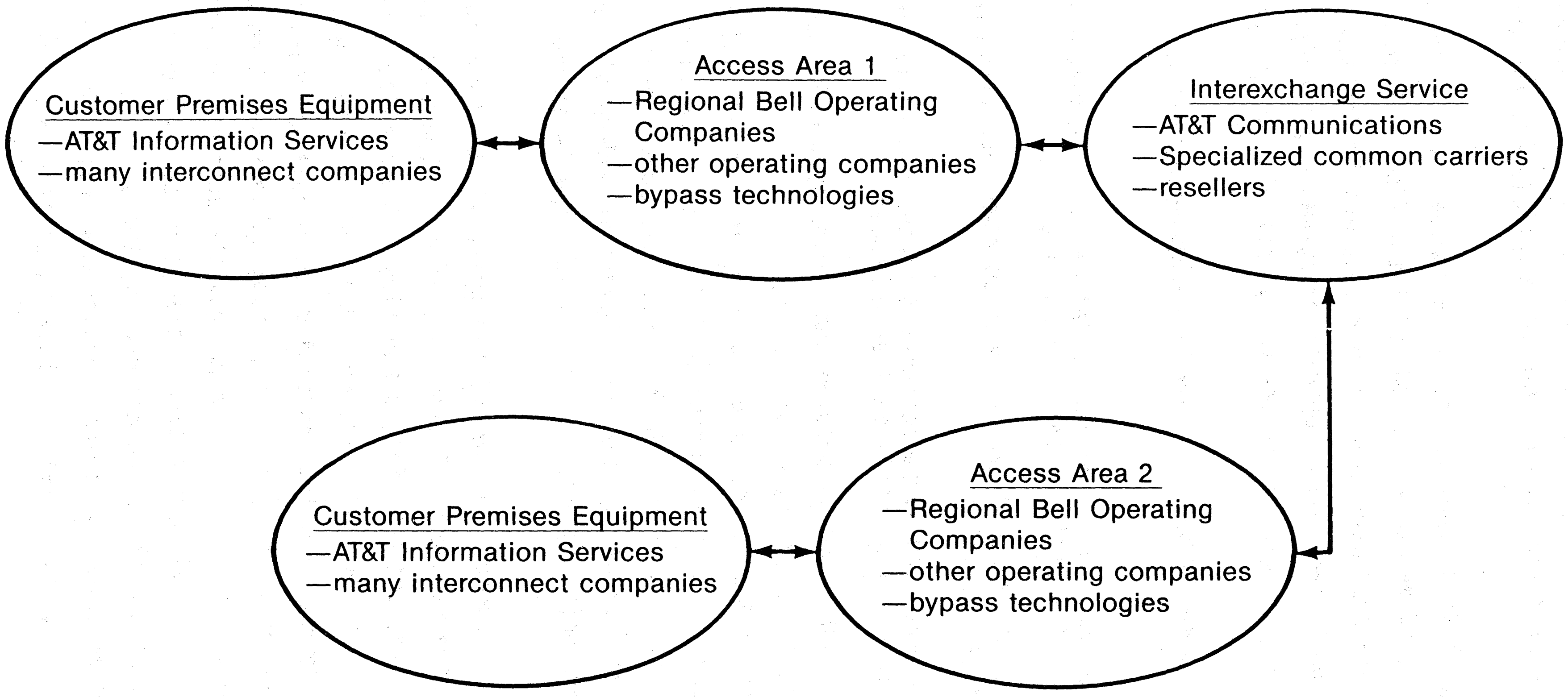

Figure 1. Services and facilities involved in a typical long-distance call.

As a result of the divestiture, AT&T Communications will gradually become one of many interexchange carriers with no special relationship with the Regional Bell Operating Companies.

Standards for quality of service should be independent of the method of service provision and should be user-oriented (DiBiaso, 1983). Both DiBiaso and Silverthorn (1983) note that end-to-end performance may be unsatisfactory even when the individual parts meet their separate performance criteria. Quality of service depends not only on the individual parts, but also on the way the parts interact.

Palladino and Wilkens (1983) list the following needs for loop performance standards:

- no method of measurement exists to determine if customers are obtaining a "satisfactory" level of quality
- terminal equipment manufacturers need to know loop transmission performance capabilities
- criteria for settling jurisdictional disputes regarding trouble (loop vs terminal equipment) must be enhanced.

The latter is a particularly thorny problem that commonly appears when there are multiple equipment suppliers in any single system.

The rapid advance of technology is encouraging new communication facilities and services. New services such as remote call forwarding are allowing multiple passes through the network (DiBiaso, 1983). These cascaded connections are being implemented through the flexibility that is inherent in modern electronic switches. The opportunity for encountering poor quality connections is increased for these multiple-pass calls.

Figure 2 provides a conceptual view of the types of standards and models that are needed. First, there is the model (box a) that maps objective network parameters such as loss, noise, and echo into subjective voice quality measures (end-user's perception of voice quality). Second, there is the model (box b) that maps subjective voice quality measures into objective voice quality measures and vice versa. Third, there is the model (box c) that maps subjective voice quality measures and/or objective voice quality measures into the economic impact. Finally, there is the model (box d) that maps network parameters such as echo, noise, and loss into parameters that are objective measures of the speech signal.

Appendix B discusses objective speech quality measures such as signal-to-noise ratio, spectral distortion, etc. Clearly, these parameters are different from, but

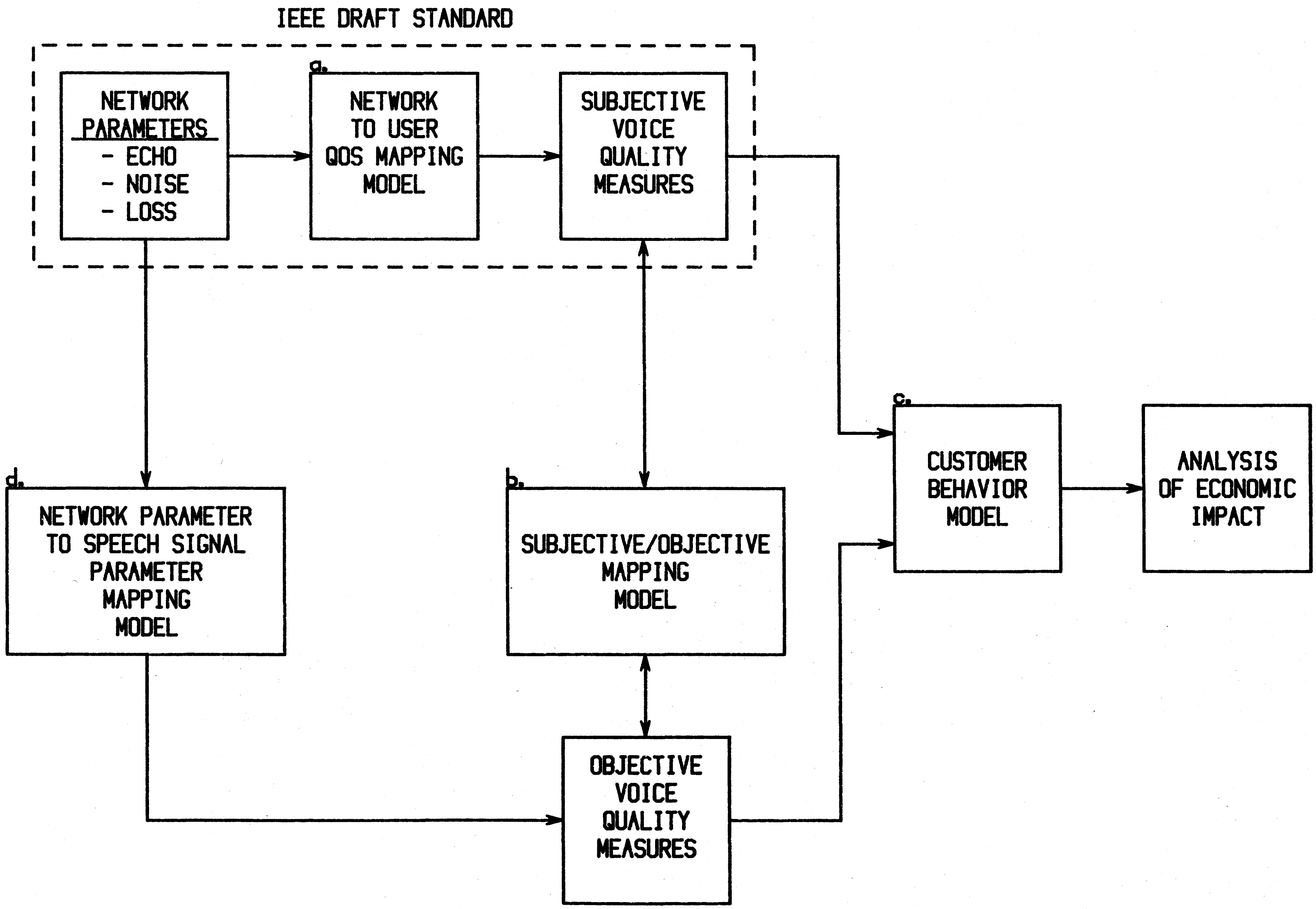

Figure 2. Voice QOS models.

related to, network parameters such as echo, noise, and loss. For an analog system, for example, the signal-to-noise ratio of the speech signal is directly dependent upon loss, which is a network parameter.

Certain network parameters such as echo, noise, and loss can be measured relatively easily. The question of how these measurable parameters can be translated into the end user's perception of the voice quality has not been fully answered. The work of IEEE and CCITT standards committees in the QOS area will be reviewed in Section 3. The IEEE work, as can be seen in Figure 2, does not fully address all of the questions involved. The draft IEEE standard is said to be a first version of a QOS standard. A later revision, to include additional parameters such as delay, can be expected.[1] The draft standard also does not address the questions of objective voice quality measures and how they relate to subjective measures or the question of how the customer behaves when he or she encounters a connection that provides poor service. Does the customer hang up and call the operator if the line is noisy? What is the economic impact of this behavior by the customer? Will the customer eventually cancel the service? From this one can see some of the economic aspects of the quality-of-service question. As can be seen from Figure 2, the IEEE draft standard addresses only a portion of the voiceband QOS issues.

Figure 3 depicts conceptually how a measurable network parameter such as loss, noise, or echo may be mapped into a user's evaluation of voice quality. For a given set of network parameters, a number of listeners are asked to evaluate the quality by selecting one of five subjective descriptions of the service. The listener categorizes the service as being excellent, good, fair, poor, or bad. Numerical scores from 5 to 1 can be assigned to these five categories (excellent is assigned a value of 5). The mean opinion score (MOS) is then determined using the subjective evaluations from several listeners. (See Appendix A for a further discussion of MOS subjective listening tests.)

In a typical laboratory experiment, a single network parameter (such as loss) is varied while other parameters are held constant. For each test condition the listeners evaluate the service using the five categories given above. This series of test results is then plotted. The plots give percent of good or better (GOB) and poor or worse (POW) as a function of the network parameter.

---

[1]Private communication: R. Donald Silverthorn, Bell Northern Research, Ottawa, Canada, April 1984.

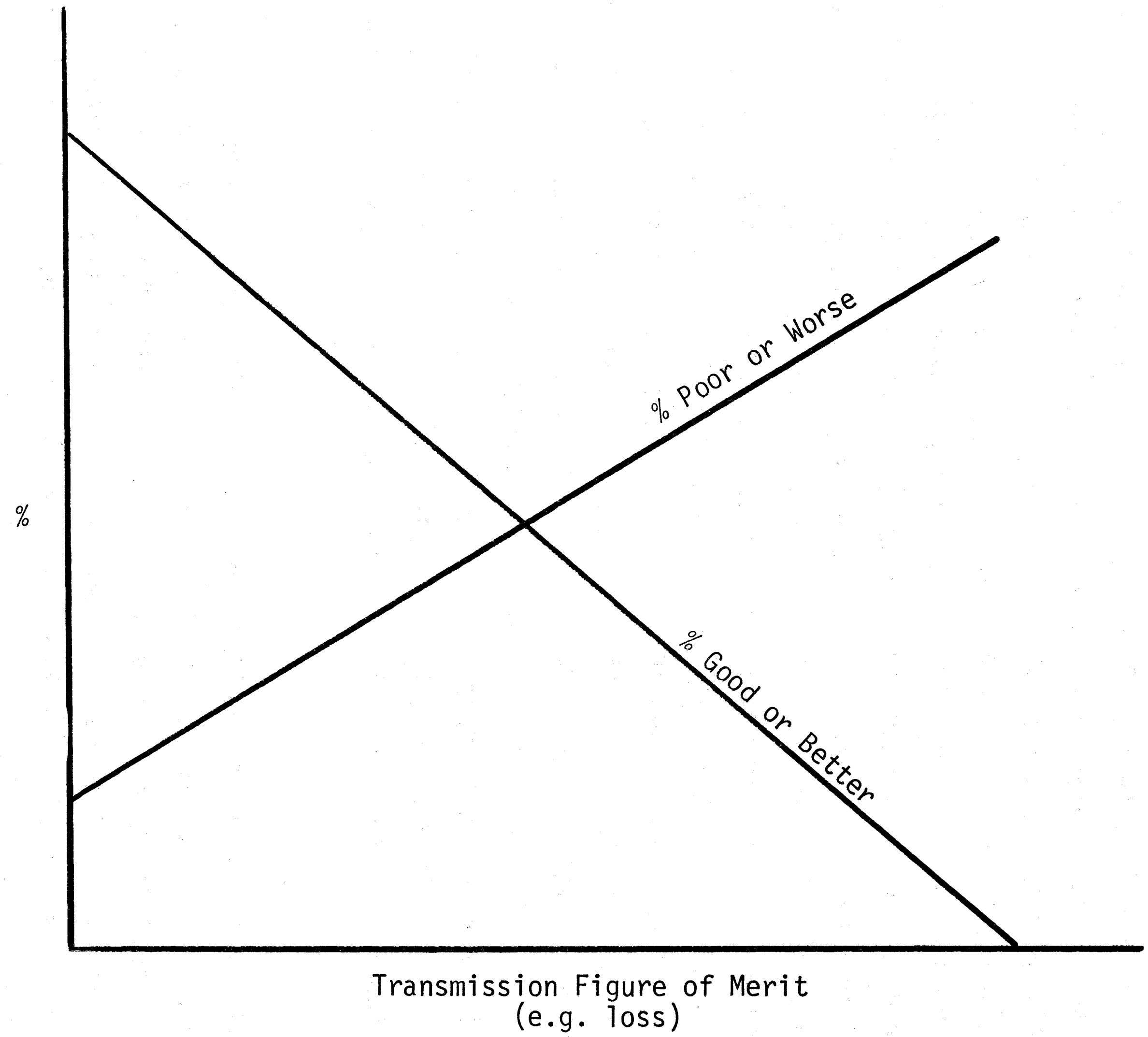

Figure 3. Mapping of transmission figure of merit into end-user's subjective evaluation of voice quality.

One of the problems that must be resolved is that of determining the combined effects of noise, loss, echo, etc., for the several segments (typically from several vendors) of a long-distance circuit. Figure 4 depicts the additive model for computing circuit noise for a three-channel system. Each vendor, in theory at least, would measure parameters for his network and make this information available to the user. As mentioned, the typical long-distance call utilizes the facilities of several service providers. The user could, in theory, obtain network performance measurement information, combine this information as in the additive model for noise shown in Figure 4, and translate this to an expected subjective quality of service through the use of curves such as these shown in Figure 3. One very significant unresolved question in this methodology is that if the end user finds the actual quality of service to be much worse than would be expected from whatever parameter measurement information that has been given, he or she will still have a very difficult time in determining which vendor is at fault. Each vendor may believe that his network is performing as specified and that the problem must lie in one of the other vendor's circuits. This is a quandary one must face whenever products or services from multiple vendors are utilized in a single system.

2.2 Differences in Requirements of Voice and Voiceband Data

As mentioned previously, the emphasis in this report is on voice quality of service. The methodology developed in the proposed IEEE standard (discussed in Section 3) does not apply to service such as data or facsimile (Silverthorn, 1983). There are some fundamental differences between quality-of-service requirements for voice and data. Because of these differences, different interpretations are necessary for the term "user-oriented parameters" when applied to voice users and to data users (Gruber and Le, 1983a, 1983b). For voice users, performance is usually subjectively perceived, while for data users, objective performance parameters are relatively easy to measure.

The emphasis on voice in this report, and in the work of the standards committees, is due to the following:

1) For voice service there will always be an element of analog technology regardless of the state of the network evolution, and

2) it is expected that network evolution will eventually obviate the need for voiceband data by supporting end-to-end data service (Gruber and Le, 1983).

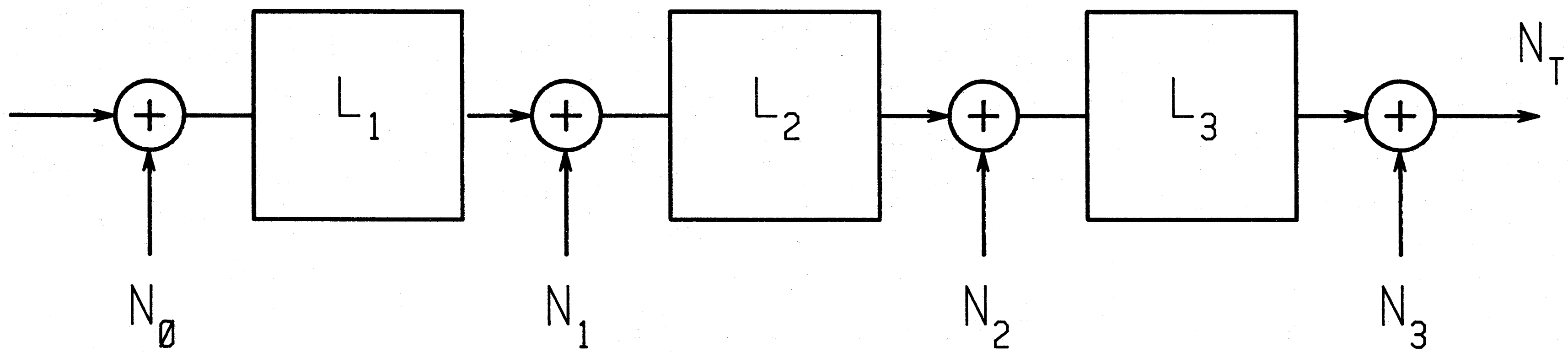

Figure 4. Additive model for computing circuit noise for a three-channel system (after Silverthorne, 1983).

Even when there are parameters common to voice and data services, the acceptable values for these parameters are likely to differ markedly. Consider digitized voice. Acceptable bit error rates for voice are of the order of 10^{-3} or 10^{-4}, while for data the corresponding value may be 10^{-6} or less. Absolute delay and delay jitter are two other parameters whose acceptable values for voice and data are quite different. Typical data service users are much more tolerant of delays in transmission than are voice users. Requirements for transmission delay jitter are typically less stringent for voice than for data. Some interexchange carriers have found that some voice end users are very sensitive to delay, and have begun to phase out the use of satellite channels for carrying voice traffic. There are, of course, differences in service requirements for different types of data. For example, interactive data requires relatively small delays or quick system response but relatively low transmission rates, while bulk data transfer is less stringent regarding delays but may require higher transmission rates.

As can be readily surmised, the differences between voice and data service requirements result in different approaches to the specification of user-oriented performance parameters. Gruber and Le (1983a) briefly discuss approaches for accommodating these differences in the Integrated Services Digital Network (ISDN). The approaches they suggest are:

1) Use the most stringent values in all cases (e.g., use the requirements for data when considering bit error rate, and the requirements for voice when considering delay).

2) Segregate services.

3) Introduce the concept of classes of performance.

4) Require users to enhance performance of their particular service by means of suitable end-to-end protocols and associated terminal equipment (e.g., use of channel coding for error detection and correction for data transmission).

There are disadvantages to each of these approaches. Alternative 1 is excessively expensive, alternative 2 and to a lesser extent alternative 3 tends to be contrary to the integrated services concept, and alternative 4 probably would not satisfactorily resolve all differences in voice and data service requirements. Gruber and Le do not attack these issues, but rather constrain themselves in their paper to the identification of the differences in requirements. We note in passing that the interim ISDN will provide a "C" channel which is analog and which may carry telemetry, packet-switched data, and signaling information on the same line with analog voice (ITS Staff, 1983). This service will be available during a transitional period.

The remainder of this report addresses voice services only.

2.3 Economic and Policy Aspects of the Quality-of-Service Issue

Three primary network transmission impairments that affect voice quality are loss, noise, and echo. The public switched network (PSN) was designed so that echo is not the controlling impairment. When loss in the network is too low, echo control may start to degrade and the received volume may be too high. However, when loss is too high the received volume may be too low. When noise is too high, the received speech may be unintelligible. The PSN design is a compromise of these three parameters (DiBiaso, 1983). Variations in the amount of noise and losses are substantial. This statistical aspect of system performance is intentional, however. It would have been cost prohibitive to have designed the network to provide excellent service quality 100% of the time. Economic considerations have been major factors in the design and development of the network, and they are expected to be even more significant in the future.

Economics is recognized in the Modified Final Judgment (MFJ) in regard to equal exchange access. Appendix B of the MFJ sets forth requirements for the phased-in RBOC provision of equal exchange access. It specifically states in that appendix that:

> ..."a BOC may not be required to provide equal access through a switch if, upon complaint being made to the court, the BOC carries the burden of showing that for particular categories of services such access is not physically feasible except at costs that clearly outweigh potential benefits to users of telecommunications services."

The technical question of equal exchange access and the related question of QOS through the exchange access network is therefore very much influenced by the economics of the situation. An example would be the replacement of a mechanical switch with a modern electronic stored program switch, which is required for equal exchange access.

A full discussion of economic and policy aspects of the quality-of-service issue for telephone networks is outside the scope of this report. The interested reader is referred to McManamon (1984) for a discussion of these topics.

3. STATUS OF STANDARDS ACTIVITIES ON VOICEBAND QOS

This section provides a summary of the voiceband QOS standards that have been and are continuing to be developed by three standards-development organizations. The activities of the Institute of Electrical and Electronic Engineers (IEEE), the International Telegraph and Telephone Consultative Committee (CCITT), and the Exchange Carriers Standards Association (ECSA) T1 Committee will be reviewed. Other standards organizations such as the American National Standards Institute (ANSI), the Electronic Industries Association (EIA), the International Electrotechnical Commission (IEC), and the U.S. Telephone Association (USTA) are also active in standards development that may be related to QOS. Where appropriate, we shall identify areas in the IEEE and CCITT standards that are thought to be incomplete or where additional effort may be required. The T1 working groups are just beginning their standards development activities.

3.1 IEEE Standards Activity on Voiceband QOS

DiBiaso (1983) and Johnson (1983) provide a discussion of the needs for telecommunications system performance standards. Silverthorn (1983) provides an overview of the IEEE Draft Standard P823 entitled "Methodology for Specifying and Evaluating Voiceband Performance Criteria." Palladino and Wilkens (1983) discuss the IEEE Standard P820, which provides standards of performance on voice frequency telephone loops. These two standards are responsive, at least in part, to the requirements as summarized in DiBiaso's paper. Jensen (1983) discusses the application of the IEEE Draft Standard P823. Cavanaugh et al. (1983) describe the transmission rating model that provided the foundation, in part, on which the IEEE Draft Standard P823 was developed.

Telecommunications vendors have developed their own set of parameters and measurement methodologies that have been designed to ensure that their networks provide acceptable service to the end users. However, these methodologies have generally not been reported in sufficient detail to enable general industry and end users to determine and evaluate transmission performance (Silverthorn, 1983). As a result of this deficiency, the IEEE Communications Society Transmission Systems Committee formed, in 1978, a subcommittee on Telecommunications Systems Performance characteristics to assist industry in defining methodologies for the specification and evaluation of telecommunications networks. A working group to develop a standard (P823) for voiceband channel performance criteria was established in February 1980. The objective of the working group is "to prepare a standard to define and specify methodologies, parameters, and performance criteria relevant to the transmission of speech and data through voiceband telecommunications networks" (Silverthorn, 1983). Although the objective included both speech and data, the current draft standard includes only voice. This standard is in the final draft stage.

A second working group on loop performance characteristics was organized in March 1980. This standards activity (IEEE Standard P820) is discussed briefly in Section 3.1.5.

3.1.1 IEEE Draft Standard P823--What is it?

The IEEE Draft Standard P823 defines a specification method for transmission parameters that cause degradations in the transmission of speech. The specific parameters that are currently included are loss, noise, and echo. Noise includes both idle-circuit noise and quantization noise. Echo includes both the talker echo path loss (TEPL) and talker echo path delay (TEPD). The standard recommends that these parameters be specified in statistical terms. The statistical specification may be:

1) a specification of the cumulative distribution function, or

2) a specification of the specific type of distribution function and its moments (mean and variance).

The distance for which the specification applies should be stated. The following paragraphs will provide some detail regarding the specification of loss, noise, and echo for a single link and the method by which the values are added for multiple links.

LOSS

Although losses are detrimental to system performance in terms of the degradation of received signal strength, they are beneficial in that they also reduce the echo path loss and therefore improve echo performance for both speech and data. The channel loss is to be specified at 1004 Hz. The addition method for losses is simply:

$$L_c = \sum_{i=1}^{n} L_i \qquad (1)$$

where L_i = loss for each link (i) expressed in decibels (dB)

L_c = total mean loss for the end-to-end connection not including the end instruments.

The total loudness loss (L_e) in dB is given by:

$$L_e = L_c + \text{TOLR} + \text{ROLR}. \qquad (2)$$

TOLR and ROLR are the transmitting and receiving objective loudness ratings, respectively, and are measures of the efficiency in converting acoustic signals into electrical signals and vice versa.

Idle-circuit noise

Idle-circuit noise is defined as the short-term weighted noise-power during the quiescent state. The weighting is frequency-weighting--either the C-message weighting used in North America or psophometric weighting used in CCITT Recommendations. It is specified in dBrnc0. [The unit dBrnc0 is the noise power measured in dBrnc but referenced to the zero-level transmission level point. The term dBrnc refers to the power level of noise with C-message weighting expressed in dB relative to 1 picowatt reference noise. The reference noise power is -90 dBm. For more details see Bellamy (1982).]

The addition method for a three-link connection (see Figure 4) for noise is given by Silverthorn (1983) as:

$$N_c = \{[(N_1 - L_2) \boxtimes N_2] - L_3\} \boxtimes N_3 \qquad (3)$$

where L_2 and L_3 are losses for links 2 and 3, respectively,

N_1, N_2, and N_3 are the idle-circuit noises for the three links,

Nc is the total noise for the connection, and the symbol $\boxtimes$ means power addition.

The extrapolation of the method to additional links is clear.

Quantization noise

Quantization noise is defined as the distortion that is introduced by the analog-to-digital and digital-to-analog conversions. Since no single objective method for the measurement of quantization noise gives a suitably accurate indication of the effect of a digital coder on the perceived quality of speech, subjective procedures are recommended at this time (Silverthorn, 1983). The procedure recommended is that adopted for study by CCITT Study Group XII. The recommended procedure consists of a series of opinion tests in which speech signals are processed through 1) the digital coder (coder/decoder) under test, and 2) a reference system called the Modulated Noise Reference Unit (MNRU).

The MNRU adds white noise to the speech signal. The ratio of the instantaneous magnitude of the speech signal to the noise signal is designated as "Q". Both signals are unweighted and the ratio "Q" is measured in dB. The coder performance is expressed as the value of Q for which the subjective opinion score is equal to the score for the MNRU.

The addition method for quantization noise is given by Silverthorn (1983):

$$Q_c = -15 \log 10 \sum_{i=1}^{n} 10^{-Q_i/15} \tag{4}$$

where Q_i = Q of the ith link

Q_c = total Q for the end-to-end connection.

The equivalent noise of the complete connection (N_e) is given

$$N_e = [N_c \boxtimes N_q] - \text{ROLR} + 46 \tag{5}$$

where $\boxtimes$ indicates power addition, and

$$N_q = 77.3 - Le - 2.36\ Q_c. \tag{6}$$

Talker echo path loss (TEPL)

The TEPL is defined as the ratio of the incident signal to the reflected signal at the transmission connection interface. The ratio is weighted in frequency and expressed as a single number in dB. Echo return loss (ERL) is defined as the frequency-weighted loss at the impedance discontinuity. Calculation of TEPL is based upon the following equation:

$$\text{TEPL} = 2L + \text{ERL}$$

where L = loss from the near-end of the channel/connection to the point of impedance discontinuity.

Figure 5 depicts the talker echo path loss model for multiple echo paths. The current draft standard recommends that two distributions be provided for TEPL; one for echoes less than 5 ms delay and one for echoes of greater than 5 ms delay.

The echo path loss for multiple echo paths as in Figure 2 is given by:

$$\text{TEPL}_{ci} = 2\ L_k + \text{TEPL}_i \tag{7}$$

where TEPL_{ci} = talker echo path loss from the i_{th} channel referred to the near end of the connection

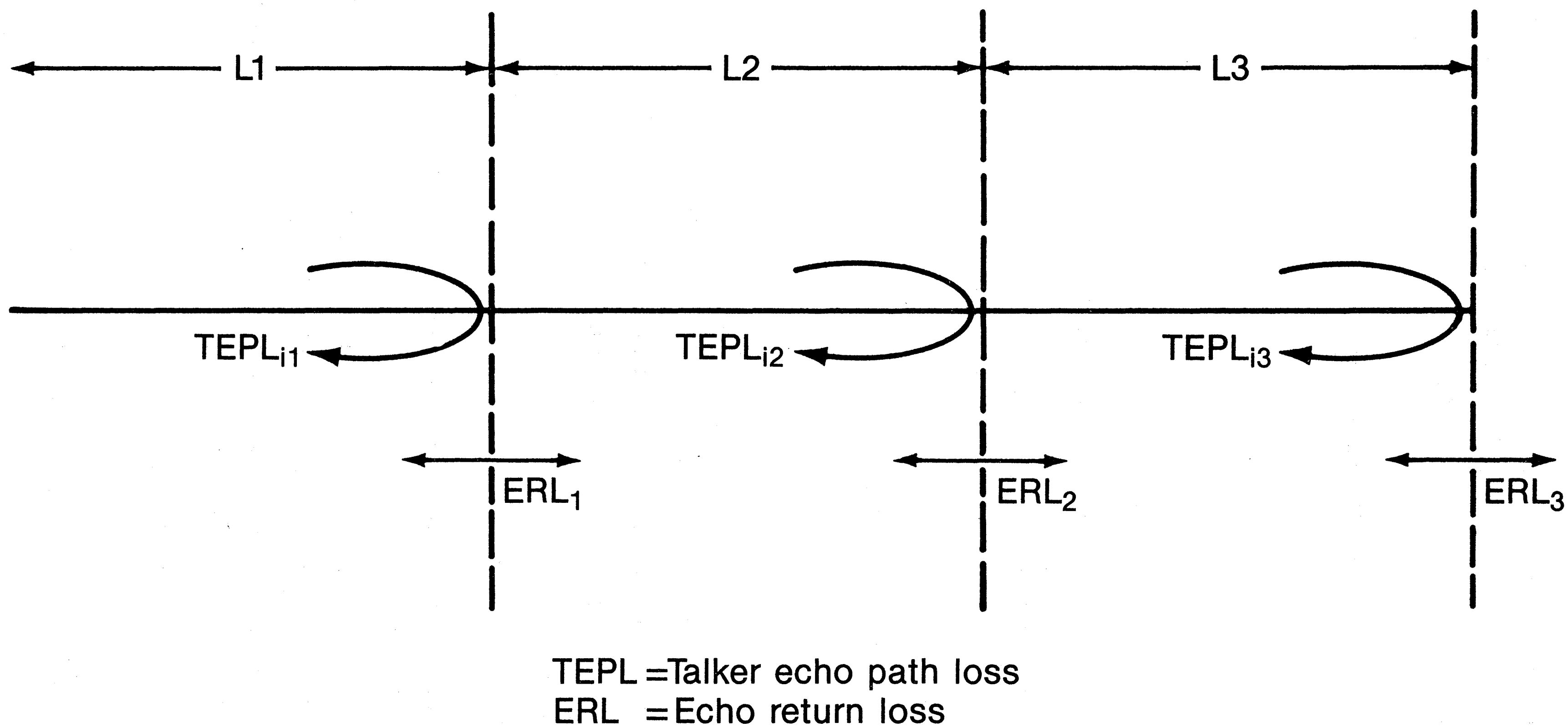

Figure 5. Talker echo path loss model (shown for a three-channel connection) after Silverthorne (1983).

$TEPL_i$ = the echo return loss from the i_{th} channel at the near end terminal of the link

L_k = the sum of 1004 Hz losses for the first nearest through the (i - 1) channels.

The method of addition of the individual channel contributions to the total talker echo path loss for the connection is dependent on the path delays involved. If <u>any</u> difference between echo path delays is less than 5 ms, losses are combined on a power addition basis as follows:

$$TEPL_{ci} = -10 \log_{10} (10^{-TEPL_{ci1}} + 10^{-TEPL_{ci2}}) \tag{8}$$

where $TEPL_{ci1}$ and $TEPL_{ci2}$ are the echo path losses for which the echo path delay difference is less than 5 ms.

The $TEPL_{ci}$ for which the echo path (round trip delay) is greater than 5 ms are to be separately specified. Any $TEPL_{ci}$ which is 10 dB greater than the smallest $TEPL_{ci}$ can be ignored. The $TEPL_{ci}$ can be neglected if the echo path delay of the near-end channel is less than 5 ms.

The talker echo path loudness loss of a complete connection (including end instruments) is given by:

$$TEPL_e = TEPL_c + TOLR + ROLR. \tag{9}$$

Talker echo path delay (TEPD)

The TEPD of a connection is defined as the difference between the time the incident signal is applied to the one end of the connection and the time the delayed replica of the signal is returned to the same terminal. The difference is expressed in ms at that frequency in the voiceband (i.e., 300 - 3300 Hz) at which the delay is the lowest. In the case of multiple-echo paths there will be a delay value for each of the paths (Silverthorn, 1983).

The method of addition for talker echo path delay is found from:

$$D_{ci} = 2 \sum_{i=1}^{n} D_i \tag{10}$$

where D_{ci} = talker echo path delay of the i_{th} link referred to the near end of the connection

D_i = one way delay at 1700 Hz of the i_{th} link.

Subjective opinion models have been developed by telephone network planners (see Cavanaugh et al., 1983 for example) which relate measurable transmission parameters such as noise, echo, and loss to the subjective quality of the speech connection. The IEEE draft standard provides a comparison method for predicting the quality of a speech connection based upon measured circuit parameters.

The standard defines the quality of the connection in terms of four performance classes (A1, A2, A3, and A4), as shown in Figures 6 and 7. Classes A1, A2, and A3 are characterized by excellent, good, and fair quality, respectively. Performance for Class A4 is characterized as severely degraded but usable. Class A4 performance is rarely encountered on the Public Switched Network, but may occur when private and public switched networks are connected (Silverthorn, 1983). This class of connection is not suitable for long social conversations but may have some utility for business transactions of a limited duration. Performance below Class A4 is considered unsuitable.

3.1.2 Application of the IEEE Draft Standard P823

The concept behind the IEEE draft standard is that vendors of transmission services will specify the performance of their services in terms of probability distributions for talker-echo, loss, and noise. The user needs to obtain these distributions from each of the vendors involved in the typical long-distance connections of primary interest. Then through the use of simulation and the addition methodologies summarized in the previous section, the user can evaluate, on a statistical basis, the expected quality of service (Class A1, A2, A3, or A4).

In most situations, the draft standard assumes the use of Monte Carlo simulation to determine the expected performance (Jensen, 1983). This technique makes the 10th and 90th percentile points available for the resultant performance distribution. This information, in addition to the means, is useful in making comparisons between different services. The performance class is determined independently and statistically for both the loss-noise performance (see Figure 6) and talker-echo-delay performance (see Figure 7). The overall performance class for a complete connection is the lower of the two performance classes for loss-noise and talker-echo.

Jensen (1983) provides an example of the application of the IEEE draft standard.

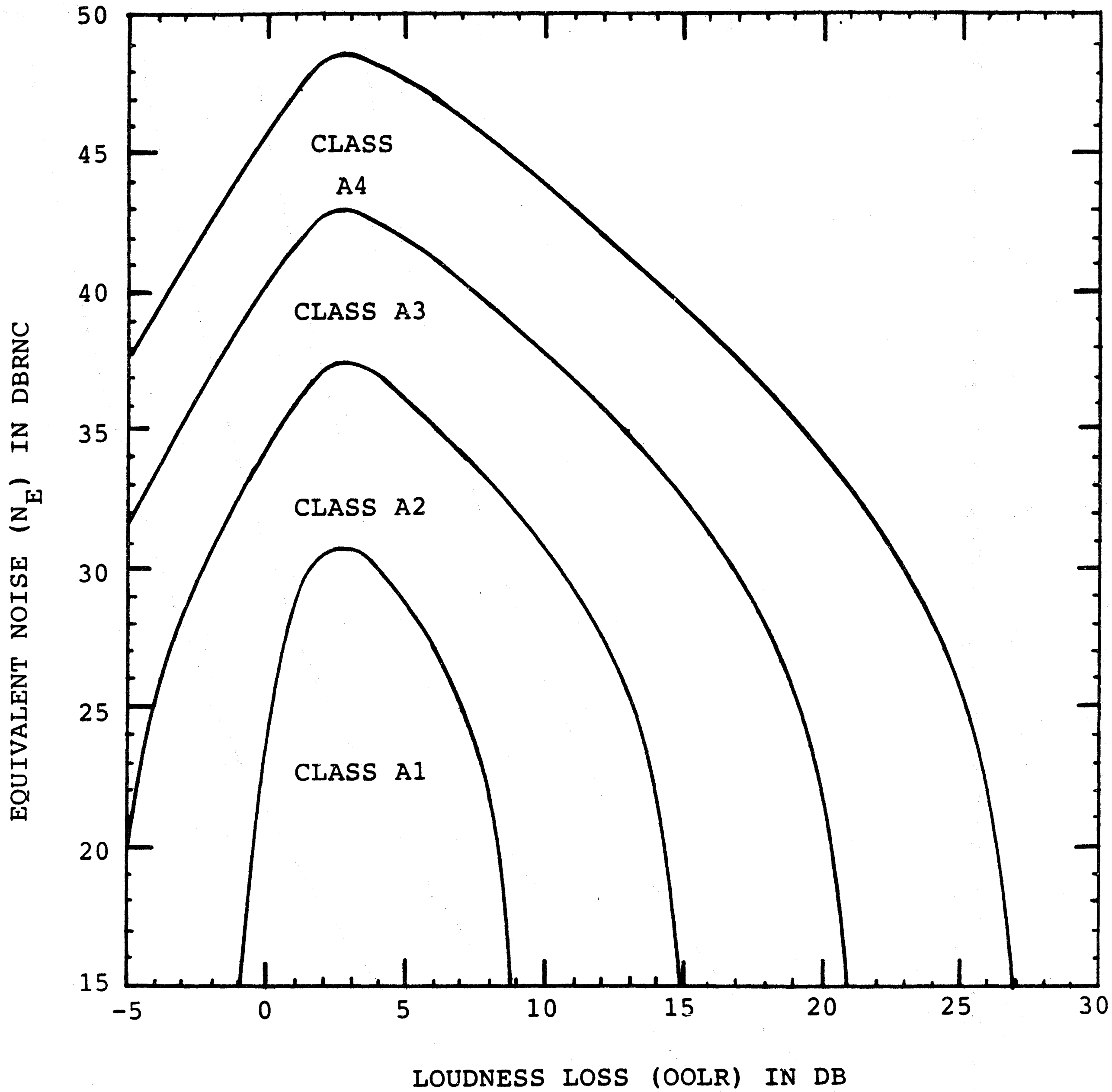

Figure 6. Loss-noise performance classes.

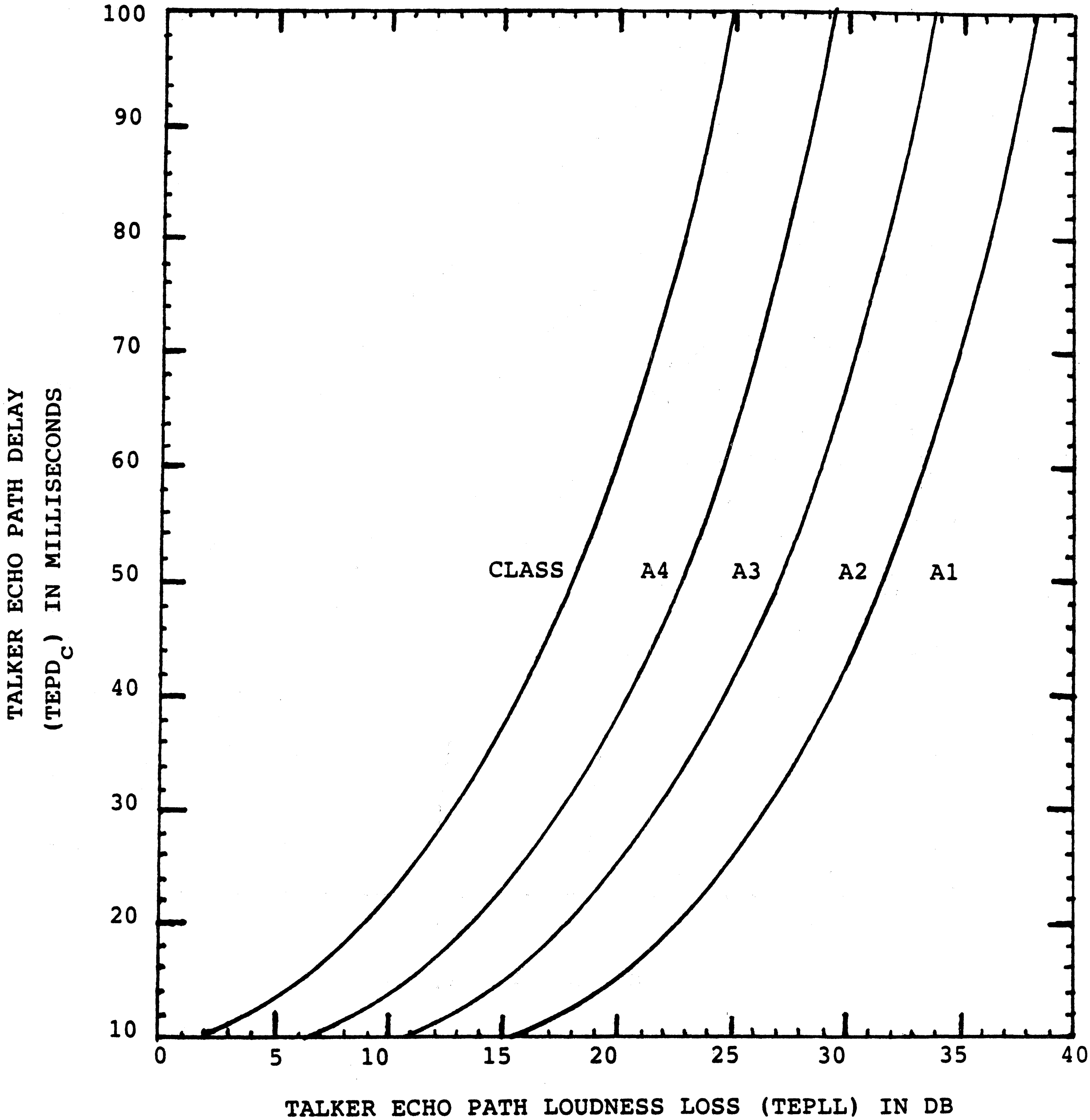

Figure 7a. Talker-echo performance classes for delay less than 100 ms.

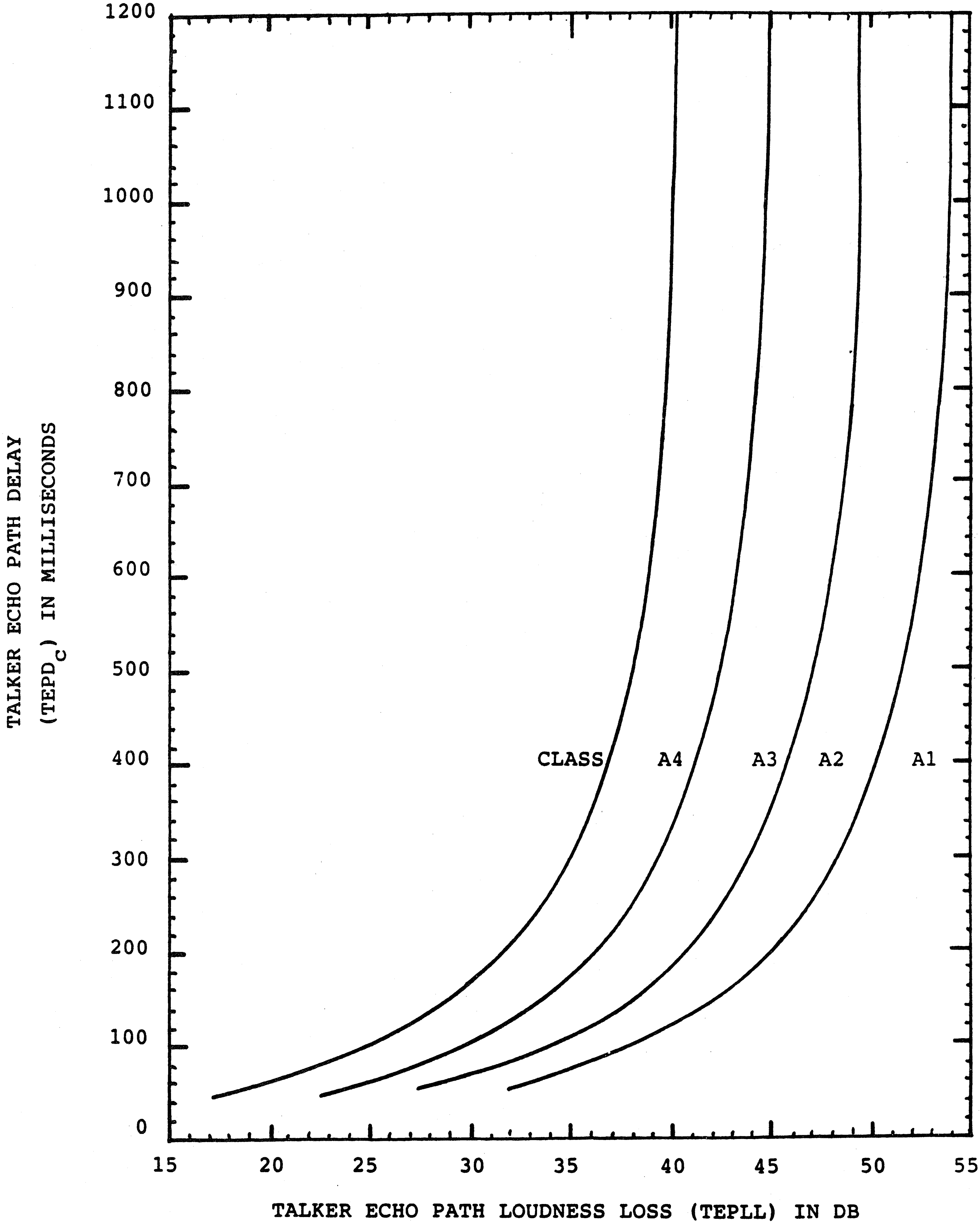

Figure 7b. Talker-echo performance classes for delays greater than 100 ms and less than 1200 ms.

3.1.3 Future Additions to IEEE Draft Standard P823

Much of the IEEE draft standard on specifying and evaluating voiceband channel performance criteria is based upon the transmission rating model developed by the Bell Telephone Laboratories (BTL). The BTL model (Cavanaugh et al., 1976; Hatch and Sullivan, 1976; Cavanaugh, 1980; Cavanaugh et al., 1980; and Cavanaugh et al., 1983) contains all of the parameters (loss, noise, and talker-echo) currently specified in the IEEE draft standard. The BTL model also contains some parameters not in the draft standard. These are: listener echo, bandwidth/attenuation distortion, room noise, sidetone, and echo-control devices. These parameters should be considered for possible inclusion in the next version of the IEEE standard. Silverthorn (1983) indicates that future issues of this standard are expected to deal with other speech impairments such as frequency response, listener echo, speech clipping, etc. Future issues of the standard will include voiceband data as well as voice.

Another parameter that should be considered is delay of the speech signal and its effect on the user's subjective perception of the quality of service. This parameter was considered initially, but a decision was made not to include this parameter in the first version of the standard.[2] Delay is a particular characteristic of satellite voice channels. Although some vendors of interexchange services are in the process of reducing the number of satellite channels in their networks because of users' complaints of degradation in quality due to delay, it can reasonably be expected that satellite channels will continue to be utilized to provide interexchange services. Therefore the delay parameter should be considered for inclusion in future versions of the IEEE standard.

3.1.4 End-User Oriented Issues Associated with IEEE Draft Standard P823

The IEEE draft standard specifies <u>what</u> parameters are to be measured and how to relate these measurable parameters to subjective voice quality. It also specifies the methodology on <u>how</u> parameters are to be measured. For each parameter, there is a section in the standard that specifies the method and units of measurements.

The end user, or his/her agent, has the problem of determining which vendor(s) is at fault if the perceived quality of service is unsatisfactory on a long-distance

---

[2]Private communications with Dr. R. Donald Silverthorn, Bell Northern Research, Ottawa, Canada, April 1984.

call. The user can measure echo, noise, and loss end-to-end on a statistical basis. The user can also obtain distributions of these parameters for each of the segments in the end-to-end connection. Using Monte Carlo simulation and the addition methods previously discussed, the end user can predict the performance class for both loss-noise and talker-echo delay. But what if the estimated performance obtained by combining inputs from the various vendors involved does not match the end-to-end measurements that have been made? What if the predicted performance class does not match the subjective evaluation of a large population of the network users? How does the telecommunications manager determine who is at fault? All of the vendors involved may believe that their portion of the connection is performing properly, and yet the end user is not receiving the specified QOS. This makes it very important that a common measurement methodology be defined and used, and that the end user be given full access to the measurement information.

3.1.5 IEEE Draft Standard P820 (Loop Performance Characteristics)

The purpose of the IEEE Standard P820 is to describe quantitatively the performance characteristics of telephone loops and indicate current acceptable performance criteria (Palladino and Wilkens, 1983). While there has been some general agreement on the requirements that control the performance of voice frequency telephone loops, there has been no formalized performance standard that has been approved by industry. The current IEEE effort under Project P820 is directed toward the creation of a widely accepted standard for loop performance. This effort has been coordinated with the International Electrotechnical Commission (IEC), the Electronics Industries Association (EIA), and the United States Telephone Association (USTA). A working group was created in March 1980 to work on the development of loop performance standards. A final standard will be published in early 1984.

The standard covers the subscriber signaling and an analog voice frequency interface to the local Class 5 switch interface. This standard, like the draft standard P823, is restricted to voice circuits. Insufficient data were available to establish performance limits for other types of loop services such as voiceband data. The standard assumes that the loop has a two-wire analog interface at the customer premises end and either a digital or two-wire analog interface at the central office end of the loop.

The performance characteristics specified are:

loop loss
frequency response characteristics
loop current
loop noise
longitudinal balance.

Palladino and Wilkens (1983) describe the methodology by which these parameters are specified. The standard establishes a range of performance acceptability classifications. They are:

recommended
acceptable
conditionally acceptable
not recommended.

The standard also provides general information on types of loop facilities, interfacing with local switching equipment, interfacing with customer equipment, use of statistical information, and measurement procedures.

3.2 CCITT Activities Related to Quality of Service

The CCITT is currently studying QOS questions similar to those being investigated by the IEEE. This section provides tables of QOS-related questions currently under study, a brief review of QOS-related recommendations published at the conclusion of the last study period (1980), and a summary of some of the contributions submitted in response to these questions.

Maitre and Aoyama (1982) summarize the speech coding activities within the CCITT, both the current status and future trends. Speech coding standardization activity includes questions on the quality of coded speech. Both subjective and objective measures are of interest to those involved in the development of new voice digitizing techniques. The CCITT Study Group XII, for example, is attempting to define methods for evaluating the subjective quality of coded speech. An IEEE group is also investigating this issue (Goodman and Nash, 1982). Objective measures, such as signal-to-noise ratio, which are classically used to evaluate PCM (pulse coded modulation), are not valid for the evaluation of more complex speech coding schemes (Maitre and Aoyama, 1982). Those investigating the end-to-end quality of service issue of concern in this report may benefit from a review of activity in the speech coding arena since both issues require a solution to the question of how quality is to be measured. Appendixes A and B summarize this activity for subjective and objective measures, respectively.

Schweizer (1983), Listanti and Villani (1983), and Decina and de Julio (1982) discuss CCITT activity in the area of voice standards for the integrated services digital network and X.25 packet-switching networks. A special issue of the IEEE Journal on Selected Areas in Communications (SAC-1, No. 6, December 1983) is devoted to packet-switched voice communication. One paper by Gruber and Le (1983a) pertains to the performance requirements of voice and digital data services.

3.2.1 CCITT Questions for Study Related to QOS

Table 1 lists the titles of CCITT Study Groups for the 1985-1988 study period. Study Group XII is concerned with voice quality. Study Groups II, IV, XV, XVI, and XVIII are also studying questions that may be applicable to the voice and voiceband data QOS issue. Tables 2 through 6 list the questions for study for the current (1985-1988) study period and the four previous study periods for Study Group XII, Tables 7 and 8 list the questions investigated by Study Group II, and Table 9 lists current QOS-related questions assigned to various other study groups. Topics of particular interest are underlined in each of the tables. Contribution No. 1 (CCITT, 1981e) provides a detailed description of the SG XII questions. Questions for the 1985-1988 period were obtained from working group papers to be issued as CCITT red books in 1985.

As can be seen from a comparison of Tables 2 through 6, parameters that have an impact on the service quality as perceived by the user have been under study for some time. For example, users' tolerance of echo and propagation time has been under study at least since 1968 (see Question 6 in Tables 2 through 6). The determination of transmission quality by objective measurement (Question 7) has also been under study at least since 1968. However, the nature of the question and the specific issues being investigated have changed somewhat in each study period. For example, early work on Question 7 dealt primarily with the articulation reference equivalent (AEN) and procedures for its measurement in the laboratory (see CCITT 1969 and 1973). Later emphasis on Question 7 has focused on models similar to the IEEE model discussed in the previous section (see CCITT 1981d). Although the CCITT reference apparatus for the determination of transmission performance ratings (such as AEN) is part of the last P Series Recommendations (Section 3 of CCITT, 1981d), we choose to emphasize the modeling aspect of the issue. This will be discussed further in Section 3.2.2.

The focus in this section is on the questions under investigation by SG XII, particularly Question 7. Some of the questions being addressed by SG II that

Table 1. CCITT Study Group Titles for the 1985-1988 Study Period

Study Group Number	Title
I	Definition, Operation, and Quality of Service Aspects of Telegraph, Data Transmission and Telematic Services (Facsimile, Teletex, Videotex, etc.)
II	Operation of Telephone Networks and ISDN
III	General Tariff Principles Including Accounting
IV	Transmission Maintenance of International Lines, Circuits and Chains of Circuits; Maintenance of Automatic and Semi-Automatic Networks
V	Protection Against Dangers and Disturbances of Electromagnetic Origin
VI	Outside Plant
VII	Data Communication Networks
VIII	Terminal Equipment for Telematic Services (Facsimile, Teletex, Videotex, etc.)
IX	Telegraph Networks and Terminal Equipment
X	Languages and Methods for Telecommunications Applications
XI	ISDN and Telephone Network Switching and Signalling
XII	Transmission Performance of Telephone Networks and Terminals
XV	Transmission Systems
XVII	Data Transmission Over the Telephone Network
XVIII	Digital Networks Including ISDN

Table 2. Summary of Questions Allocated to Study Group XII for the Period 1968-1972

Question Number	Short Title	Remarks
1/XII	National System Reference Equivalents in the New Transmission Plan	Of concern to S.G.XVI
2/XII	Assessment of Service Transmission Quality	Of concern to S.G.II and XIII
3/XII	Asymmetry Between the Two Directions of Transmission	Reply to be transmitted to S.G.XVI (Question 5/XVI)
4/XII	Effect of Circuit Noise on Transmission Performance	
5/XII	Specification of Sound Level Meters	
6/XII	Users' Tolerance of Echo and Propagation Time	Reply to be transmitted to S.G.XVI (Question 3/XVI)
7/XII	Determination of Transmission Quality by Objective Measurement	
8/XII	Measuring the Efficiency of a Microphone or a Receiver	
9/XII	Limits Applied in National Trunk and Local Networks	
10/XII	Increase in the Sensitivity of Local Systems	Of concern to S.G.XVI (Question 1/XVI, point 6). Reply to be transmitted to S.G.XVI
11/XII	Limits for Intelligible Crosstalk	
12/XII	Artificial Voices, Mouths and Ears	
13/XII	Nonlinear Distortion of Telephone Apparatus	
14/XII	Extension of the Bandwidth Transmitted	
15/XII	Measuring of Loudness Ratings	
16/XII	Maintenance of Subscriber Sets	Question Africa H

Table 2. (continued)

Question Number	Short Title	Remarks
17/XII	Loudspeaker Telephones	
18/XII	Statistical Study of the Implications of Spanish Phonetics for Telecommunication Systems	Question Latin America 5
19/XII	Impedance Variations in Subscriber Lines and Telephone Sets	Of concern to S.G. XVI (Question 1/XVI, points 3 and 4)
20/XII	Synthetic Speech and Frequency Compression Systems	
21/XII	Transmission Performance of Pulse Code Modulation Systems	Linked with Question 2/D
22/XII	Revision of the Manual on Local Telephone Networks	

Table 3. Summary of Questions Allocated to Study Group XII for the Period 1973-1976

Question Number	Short Title	Remarks
1/XII	National System Reference Equivalents in the New Transmission Plan	
2/XII	Assessment of Service Transmission Quality	
3/XII	Reference Equivalents of Operators' Headsets	
4/XII	Effect of Circuit Noise on Transmission Performance	
5/XII	Hourly Noise Clause	Question 8/XVI
6/XII	Users' Tolerance of Echo and Propagation Time	
7/XII	Determination of Transmission Quality by Objective Measurement	
8/XII	Measuring the Efficiency of a Microphone or a Receiver	
9/XII	Sidetone	
10/XII	Increase in the Sensitivity of Local Systems	
11/XII	Limits for Intelligible Crosstalk	
12/XII	Artificial Voices, Mouths and Ears	
13/XII	Nonlinear Distortion of Telephone Apparatus	
14/XII	Effect of Attenuation Distortion	
15/XII	Measuring of Loudness Ratings	
16/XII	Impedance Variations in Subscriber Lines and Telephone Sets	
17/XII	Loudspeaker Telephones	
18/XII	Transmission Performance of Pulse Code Modulation Systems	

Table 4. Summary of Questions Allocated to Study Group XII for the Period 1977-1980[3]

Question Number	Short Title	Remarks
1/XII	Reference Equivalents of National Systems in the International Transmission Plan	
2/XII	Assessment of Service Transmission Quality	Coordination with Questions 4/II and 15/II
3/XII	Loudness Ratings of Operators' Telephone Systems and Headsets	
4/XII	Effect of Circuit Noise on Transmission Performance	
5/XII	Noise Causes for Telephone	Coordination with Questions 4/CMBD and 8/XVI
6/XII	Subscribers' Tolerance of Echo and Propagation Time	Coordination with e) of Question 10/XV
7/XII	Models for Predicting Transmission Qualities from Objective Measurements	
8/XII	Measuring the Efficiency of a Microphone or a Receiver	
9/XII	Sidetone	
10/XII	Increase in the Sensitivity of Local Systems	Documentary Question; coordination with a) of Question 5/XVI
11/XII	Limits of Intelligible Crosstalk	Coordination of 5), 6), 7) with Question 3/XVI; see also Question 6/XV
12/XII	Artificial Voices, Mouths and Ears	
13/XII	Nonlinear Distortion of Telephone Apparatus	

[3] There is a strong interest of Study Group XVI in Study Group XII's work; Study Group XII is therefore requested to keep Study Group XVI continuously informed of the progress made.

Table 4. (continued)

Question Number	Short Title	Remarks
14/XII	Effect of Attenuation Distortion on Mission Performance	Coordination with ii) of Question 1/XVI
15/XII	Measurement of Loudness Ratings	
16/XII	Return Loss Variations in Subscriber Lines and Telephone Sets	Documentary Question
17/XII	Loudspeaker Telephones	
18/XII	Transmission Performance of Digital Systems	Coordination with Question 10/XVI; see also Questions 9/XVIII and 10/XVII
19/XII	Recommended Values of Loudness Ratings	New Question; coordination with Question 11/XVI
20/XII	Devices for Protection Against Acoustic Shocks	New Question; coordination with Question 5/V
21/XII	Efficiency of Telephone Kiosks and Booths	New Question

Table 5. Summary of Questions Allocated to Study Group XII for the Period 1981-1984

Question Number	Short Title	Remarks
1/XII	Future Programme of Work	New Question
2/XII	Assessment of Service Transmission Quality	
3/XII	Loudness Ratings of Operators' Telephone Systems and Handsets	
4/XII	Effect of Circuit Noise on Transmission Performance	
5/XII	Talker and Listener Echo Effects	New Question
6/XII	Subscribers' Tolerance of Echo and Propagation Time	
7/XII	Models for Predicting Transmission Quality from Objective Measurements	
8/XII	Measuring the Efficiency of a Microphone or a Receiver	
9/XII	Sidetone	
10/XII	Desirable Transmission Characteristics of Handset Telephones	
11/XII	Transmission Degradation Introduced by Echo Control and Other Voice Operated Devices	New Question
12/XII	Artificial Voices, Mouths and Ears	
13/XII	Nonlinear Distortion of Telephone Apparatus	
14/XII	Characteristics and Effects of Attenuation Distortion	New Wording
15/XII	Measurement of Loudness Ratings	
16/XII	Return Loss Variations in Subscriber Lines and Telephone Sets	
17/XII	Loudspeaker Telephones	
18/XII	Transmission Performance of Digital Systems	

Table 5. (continued)

Question Number	Short Title	Remarks
19/XII	Recommended Values of <u>Loudness Ratings</u>	
20/XII	Devices for Protection Against Acoustic Shocks	
21/XII	Efficiency of Telephone Kiosks and Booths	
22/XII	Syllabic Compandors	
23/XII	Coupling of Hearing Aids to Telephone Receivers	New Question
24/XII	Links with Mobile Stations	New Question
25/XII	Drafting of a Handbook on Voice-Ear Measurements	New Question

Table 6. Summary of Questions Allocated to Study Group XII for the Study Period 1985-1988

Question Number	Short Title
1/XII	Future Programme of Work
2/XII	Handsfree Telephony
3/XII	Measurements on Headsets
4/XII	Effect of Circuit Noise and Interference on Transmission Performance
5/XII	Speech Synthesis/Recognition Systems
6/XII	Subscribers' Tolerance of Echo and Propagation Time
7/XII	Models for Predicting Transmission Quality from Objective Measurements
8/XII	Measuring the Efficiency of a Microphone or a Receiver
9/XII	Sidetone
10/XII	Desirable Transmission Characteristics of Handset Telephones
11/XII	Transmission Degradation Introduced by Echo Control, Compandors, and Other Voice Operated Devices
12/XII	Artificial Voices, Mouths, and Ears
13/XII	Nonlinear Distortion of Telephone Apparatus
14/XII	Characteristics and Effects of Attenuation Distortion
15/XII	Algorithms for Calculating Loudness Loss
16/XII	Return Loss Variations in Subscriber Lines and Telephone Sets
17/XII	Loudspeaking Telephones
18/XII	Transmission Performance of Digital Systems
19/XII	Recommended Values of LRs
20/XII	Prevention of Hazards and Limitations of Annoyances Caused by Abnormally High-Level Signals
21/XII	Efficiency of Telephone Kiosks and Booths

Table 6. (continued)

Question Number	Short Title
22/XII	Objective Measurement of Speech Level
23/XII	Coupling of Hearing Aids to Telephone Receivers
24/XII	Links with Mobile Stations

Table 7. Summary of Questions Allocated to Study Group II for the Period 1981-1984

Question Number	Short Title	Of Interest to Other Study Groups
1/II	Application of the Instructions for the International Telephone Service and any Amendments Required	
2/II	Use of Computers to Supply Information Requested on Call Number of Telephone Subscribers in Foreign Countries	SG III
3/II	Choice and Standardization of Supplementary Services Offered to Telephone Users	
4/II	Customer Performance in Fully Automatic Working in the World-Wide Telephone Network	SG XII
5/II	Standardization of Symbols and Other Aspects of Subscriber Equipment to Meet Human Factor Needs	
6/II	Instructions for Users of the World-Wide Telephone Network	
7/II	Elements of Supplementary Service Control Procedures and Human Factor Aspects of User Indications	SG XI
8/II	Human Factors Aspects of User Inter-actions in Computer-Based Systems in International Telecommunications	SG I, SG VII, SG VIII
9/II (16/I)	Operational Aspects of Future Developments in the Maritime Mobile Service (to be studied by Joint Working Party SMM Question 1/SMM)	
10/II (17/I)	Revision of Recommendation Concerning Public Correspondence in the Maritime Mobile Service (to be studied by Joint Working Party SMM Question 2/SMM)	
11/II	International Interconnection of the Different Mobile Telephone Services	CCIR SG 8, SG XI
12/II	Development of the World Telephone Numbering Plan	SG VII, SG XI, SG XVIII

Table 7. (continued)

Question Number	Short Title	Of Interest to Other Study Groups
13/II	Review of World Routing Plan	SG XI
14/II	Use of Switched Telephone Network for Nontelephone Applications	SG I, SG III, SG VII, SG XI
15/II	Observations on Quality of International Service	SG IV, SG XII
16/II	Models for International Network Planning	
17/II	Alleviation of Transmission-Facility Failure Conditions	SG IV, SG XII
18/II	Network Management	SG III, SG IV, SG XI
19/II	Methods and Procedures for Traffic Measurements	SG XI
20/II	Methods for the Measurement and Computation of Grade of Service and Formulation of GOS Standards of International Circuit Groups	
21/II	Methods for Forecasting International Traffic	
22/II	Dimensioning of Alternate Routing Networks Taking into Account 24-hour Traffic Profiles	
23/II	Grade of Service in International Telephone Exchanges	SG XI
24/II	Grade of Service and New Performance Criteria Under Failure Conditions in International Telephone Exchanges	SG XI
25/II	Traffic and Operational Requirements for SPC (especially digital) Telecommunication Exchanges	SG XI
26/II	Preparation of a Handbook on "Service Quality, Network Maintenance and Management" (New Question, to be studied jointly by Study Groups II and IV)	SG IV

Table 8. Summary of Questions Allocated to Study Group II for the Study Period 1985-1988

Question Number	Short Title
1/II	Application of the "Instructions for the International Telephone Service" (Rec. E.141) and Any Amendments Required to These Instructions and to Other Recommendations Relating to the Operation of the International Telephone Service
2/II	Use of Computers to Supply Information Requested on Call Number of Telephone Subscribers in Foreign Countries
3/II	New International Telephone Services
4/II	Automatic International Telephone Credit Card System
5/II	Revision of Regulatory Provisions (preparation for WATTC 1988)
6/II	Spare Number
7/II	Obtaining Satisfactory Customer Performance in Using the Automatic World-Wide Telephone Network
8/II	Symbols and Pictograms to Improve Customer Performance
9/II	Elements of Telephone Control Procedures and Human Factor Aspects of Indications to the User
10/II	Human Factors Aspects of User Interactions in Computer-Based Systems in International Telecommunications
11/II	Human Factors Issues Related to the ISDN
12/II	Human Factors Considerations of Access to Telephone and Telematic Services, Through Public Terminals
13/II	Digit Button Arrangements for Use on Telephones and on Advanced Telecommunications Terminals
14/II	Spare Number
15/II	Spare Number
16/II	International Interconnection of the Different Mobile Telephone Services and the Public Switched Telephone Network
17/II	Development of the World Numbering Plan for Telephone and ISDN Application
18/II	Revisions of the International Telephone Routing Plan

Table 8. (continued)

Question Number	Short Title
19/II	Evolution of the International Telephone Routing Plan in the ISDN Era
20/II	Use of the Telephone Network for Non-Voice Application
21/II	Spare Number
22/II	Observations on the Quality of International Service
23/II	Network Management
24/II	Spare Number
25/II	(Vocabulary) Terms and Definitions of Teletraffic Engineering and International Telephone Operation
26/II	Traffic Engineering of Common Channel Signalling Networks
27/II	Network Design Alternatives for International Traffic
28/II	Methods for Forecasting International Traffic
29/II	Traffic Models and Measurements Required to Estimate Traffic Offered
30/II	Traffic Models and Measurements Required for Nonstationary Traffic
31/II	Reference Models for ISDN Traffic Engineering
32/II	Grade of Service and Performance Criteria for International Telephone Exchanges Under Failure Conditions
33/II	Traffic and Operational Requirements for SPC Telecommunications Exchanges
34/II	Spare Number
35/II	Models for Telecommunication Services
36/II	Field Gathering and Evaluation
37/II	Service Accessibility of Telecommunication Services
38/II	Retainability of Telecommunication Services
39/II	Interruption Objectives
40/II	Allocation of Accessibility and Retainability Objectives
41/II	Dependability of Telecommunication Networks

Table 9. Questions Related to Service Quality and Performance from Other Study Groups for the Period 1981-1984

Study Group Number	Question Number	Title (abbreviated slightly)
IV	3	Preparation of a Handbook on "Service Quality, Network Maintenance, and Management"
IV	8	Assessment of Network Performance
XV	2	Equipment for Digital Transmission of Sound-Programme Signals
XV	10	Echo Suppressor Improvements, Echo Canceller Specifications, and Testing Methods
XVI	1	Transmission Impairments in Evolving Networks
XVI	2	Characteristics of Leased Circuits
XVI	6	Transmission Aspects of Telephone Conference Calls
XVI	8	Consequences of the Modification of the Four Noise Clauses in Recommendation G.222
XVI	9	Echo, Propagation Time, and Stability in Telephone Connections
XVIII	7	Encoding of Speech and Voiceband Signals Using Methods Other than PCM, in Accordance With Rec. G.711
XVIII	8	Digital Speech Interpolation System
XVIII	16	Performance Characteristics of PCM Channels at Audio Frequencies

appear to be related to QOS, are in fact more concerned with traffic engineering, grade of service, and availability of service. For example, Question 15/II is titled quality of international service. However, Recommendations such as E.420, E.421, and E.421 (CCITT, 1981b) relate primarily to the call setup phase. These Recommendations deal with such things as unsuccessful calls due to wrong number, incomplete number, congestion, equipment failure, etc. While call blocking and equipment failure may be factors that influence the user's perception of the acceptability of a service, in this report we restrict ourselves to the user's perception of the intelligibility of the speech signal. Readers interested in GOS and traffic engineering, etc., are referred to the questions from other study groups listed in Tables 8 and 9, to related CCITT Recommendations (such as Series E and O; CCITT, 1981b and 1981c), and the corresponding CCITT red books to be issued in 1985.

Question 26 of SG II (see Table 7) and Question 3 of SG IV (see Table 9) deal with the preparation of a handbook on "Service Quality, Network Maintenance, and Management." Study Group IV is responsible for the chapter on maintenance while Study Group II is responsible for the chapter on quality of service. The QOS chapter will deal with user's perceptions of the service. It will define them with objectives that have been set to achieve the most satisfactory perception.

3.2.2 CCITT Recommendations for Telephone Transmission Quality

Table 10 lists the Recommendations of the CCITT (yellow books). Of primary interest in this report are the P Series (CCITT, 1981d and CCITT, 1977), the G Series (CCITT, 1981e), and the red books covering the period 1981-1984 to be issued in 1985.[4] Of lesser interest, here, is the O Series, which specifies measuring equipment. Recommendations E.100 to E.323 deal with telephone system operation including numbering plans (CCITT, 1981a). The P Series Recommendations deal with telephone QOS for international connections. The purpose of these Recommendations is to provide guidance on the control of transmission performance. The Recommendations contain performance, design, and maintenance objectives as defined in Recommendation G.102 for various transmission impairments that affect the customer opinion of transmission quality (CCITT, 1981d). Recommendation P.11 is concerned with the effect of transmission parameters on customer opinion of transmission quality. Transmission parameters of interest are:

- loudness loss
- circuit noise

---

[4]CCITT (1985), Red Book Vols. II, III, IV, V to be published, Geneva.

Table 10. Recommendations of the CCITT

Recommendation Series	Title
A	Organization of the Work of the CCITT
B	Means of Expression (Definitions, Vocabulary, Symbols, and Classification)
C	General Telecommunication Statistics
D	General Tariff Principles
E	Telephone Operation, Network Management, and Traffic Engineering
F	Telegraph Operation and Tariffs
G	Transmission on Lines, Radio Relay Systems, Radiotelephone Circuits
H	Utilization of Lines for Telegraphy and Radiotelegraphy
I	Integrated Services Digital Network
J	Radio and Television Programme Transmission
K	Protection Against Interference
L	Protection Against Corrosion
M	Maintenance of Telephony Circuits and Carrier Systems
N	Maintenance for Sound-Programme and Television Transmission
O	Specification of Measuring Equipment
P	Telephone Transmission Quality. Telephone Installations and Local Line Networks
Q	Telephone Switching and Signalling
R	Telegraph Channels
S	Alphabetical Telegraph Apparatus
T	Facsimile Telegraph Apparatus
U	Telegraph Switching
V	Data Transmission
X	New Data Networks
Z	Programming Languages for SPC Exchanges

- sidetone loudness loss
- room noise
- attenuation distortion
- group-delay distortion
- absolute delay
- talker echo
- listener echo
- nonlinear distortion
- quantization distortion
- phase jitter
- crosstalk.

Recommendation P.11 is based on information contributed in response to specific questions. Much of this information is based on the results of subjective tests in which participants have talked, listened, or conversed over telephone connections with controlled or known levels of impairments and have rated the quality on a subjective scale. Recommendation P.74 (see Appendix A) provides a recommended procedure for conducting these tests.

The following paragraphs give a brief summary of the effect of the individual impairments as described in Recommendation P.11 (CCITT, 1981d) for the study period ending in 1980.

Loudness loss

Recommendation P.11 provides recommended values of loudness loss and relates these values to representative opinion results. The speech quality is rated on the qualitative scale of excellent, good, fair, poor, and bad. Percent "good plus excellent" (good or better) and percent "poor plus bad" (poor or worse) values are provided for various values of the overall reference equivalent, which is the loudness in decibels.

Circuit noise

Circuit noise has a major effect on customer satisfaction with the circuit quality. Circuit noise, as viewed by the CCITT, includes white noise and intermodulation noise from transmission systems and interference such as impulse noise. Customer satisfaction depends on the power, frequency distribution, and amplitude distribution of the noise. Noise measurements are generally frequency-weighted, such as C-weighted or psophometric-weighted. Recommendation P.11 provides percent "good plus excellent" and percent "poor plus bad" values for various values of circuit noise. Circuit noise is a question for further study for 1981-1984 (Question 4/XII).

Sidetone loudness loss

Sidetone loudness loss is the loss in the acoustic-to-acoustic transmission path from the telephone transmitter to the telephone receiver in the same telephone set. It provides feedback to the user that the telephone set is working. Excessive sidetone loss can make a telephone set sound dead, while insufficient sidetone loss causes the sidetone signal to be too loud. In the latter case, the user is likely to reduce his speaking volume, which may make him hard to hear by the listener at the other end of the circuit. Although a sidetone loudness loss of at least 17 dB is desirable, this value is not easily achieved and values between 7 and 10.5 dB are to be expected in most cases. Sidetone is under further study (1981-1984) under Question 9/XII.

Room noise

Room noise is the background noise in the environment of a telephone set. Although it is not under the control of the transmission planner, it is a factor that must be considered. One way in which room noise can manifest itself is through leakage around the earcap of the receiver. Another is through the "nontelephone" ear. No Recommendations on room noise are provided in the CCITT Yellow Book (CCITT, 1981d). The effect of room noise is under further study (1981-1984) under Question 4/XII.

Attenuation distortion

Attenuation distortion refers to the transmission loss or gain throughout the passband relative to the transmission loss at 800 or 1000 Hz. The effect of attenuation distortion on loudness is greater at the lower end of the frequency band than at the higher end. The effect of attenuation distortion on sound articulation, is, on the contrary, more marked at the higher frequencies. The current network performance objectives for attenuation distortion are given in Recommendation G.132 (CCITT, 1981e). Annex A to Recommendation P.11 (CCITT, 1981d) provides further information on the effects of attenuation distortion including the results of subjective testing. The effect of attenuation distortion is under further study (1981-1984) under Question 14/XII.

Group-delay distortion

Group-delay distortion refers to the delay at frequencies throughout the passband relative to delay at the frequency where the delay has its minimum value. The effect of group-delay distortion is more significant for voiceband data than it is for speech. However, large amounts of group-delay distortion can cause noticeable distortion for speech signals. The effect of group-delay distortion at the upper and lower edges of the band can be described as "ringing" and "blurred" speech, respectively. However, the effect in a typical four-wire circuit is usually not serious since group-delay distortion is generally accompanied by closely related attenuation distortion, which tends to reduce the effect. Current performance objectives for group-delay distortion are provided in Recommendation G.133 (CCITT, 1981e). Annex B of Recommendation P.11 provides further information on the effects of group-delay distortion.

Absolute delay

Delays typical for terrestrial transmission systems have little effect on speech transmission quality if there is no talker or listener echo or if such echoes are adequately controlled. Satellite connections introduce larger amounts of delay but "opinion data indicates that there is little effect on the transmission quality of connections with a single satellite circuit, provided talker and listener echo are adequately controlled" (CCITT, 1981d). Caution is recommended with regard to the introduction of one-way absolute delay that is significantly greater than 300 ms. The subjects of echo, echo control, and propagation time are under further study (1981-1984) under Question 6/XII.

Talker echo

Talker echo occurs when some portion of the talker's speech is returned with enough delay (typically more than 30 ms) to make the signal distinguishable from normal sidetone. Talker echo may be caused by reflections at impedance mismatches. Talker echo is a function of the loss in the echo path. For this reason, too little loss in the circuit produces poor opinion ratings due to the echo annoyance factor. Of course, too much loss is also detrimental because of received signal strength requirements. Therefore a balance must be struck. Talker-echo performance is specified in Recommendations G.131 and G.133 (CCITT, 1981e). Echo tolerance curves are provided in the latter. The effect of echo and propagation time is under further study (1981-1984) under Question 6/XII.

Listener echo

Listener echo refers to a transmission condition in which the main speech signal arrives at the listener's end of the connection accompanied by one or more delayed versions of the signal. A common source of listener echo is a four-wire to two-wire interconnection at a hybrid junction. In such a connection reflections can occur as a result of impedance mismatches at the hybrids at each end of the four-wire path. Listener echo may be characterized by the additional loss and additional delay in the echo path relative to that in the main signal path. Recommendation G.122 (CCITT, 1981e) provides further guidance on listener echo. The effect of listener echo is under further study (1981-1984) under Question 5/XII.

Nonlinear distortion

Nonlinear distortion occurs in systems in which the output is not linearly related to the input. One major source of nonlinear distortion in telephone connections is telephone sets with carbon microphones. Syllabic compandors also may be significant contributors. Nonlinear distortion is normally more significant for data transmission than it is for speech. Annex C to Recommendation P.11 provides more information on nonlinear distortion. The effects of nonlinear distortion are under further study (1981-1984) under Question 13/XII.

Quantization distortion

This type of distortion was discussed previously relative to the IEEE draft standard. Quantization distortion occurs during the processes of converting from an analog-to-digital signal and vice versa. The CCITT recommends the use of the modulated noise reference unit discussed earlier. Subjective test results have been reported that have evaluated the effects of both circuit noise and quantization noise on customer opinion (CCITT, 1981d). Annex D to Recommendation P.11 provides additional information. The transmission performance of digital systems is under study (1981-1984) under Question 18/XII.

Phase jitter

Phase jitter occurs when the desired signal is phase or frequency modulated during transmission. The jitter appears as low-index modulation (25-200 Hz is reported by one administration; CCITT, 1981d). If such distortion is present in sufficient quantity, the transmission quality is degraded.

Crosstalk

Crosstalk occurs when the speech signal from one connection is inadvertently coupled to another connection such that the coupled signal is of sufficient strength as to be heard by one or both of the participants on the second telephone connection. Information is provided in Recommendation P.16 (CCITT, 1981d) on the intelligibility threshold for crosstalk. Numerous curves on performance as affected by crosstalk are presented in that Recommendation. A number of factors influence the effect that crosstalk has on the intelligibility of a speech signal. They include sidetone, circuit noise, room noise, coupling loss, the interfering talker's speech level, and the hearing acuity of the listener.

Multiple impairments and the use of opinion models

In the above paragraphs we have discussed 13 types of impairments that influence the user's perception of the quality of service. Each of these impairments has been discussed independently. Unfortunately, in the real world these impairments are likely to coexist. When considered jointly, the overall degradation may cause the service to be judged unacceptable even though the individual impairments do not cause unacceptable service. For this reason analytical models, such as that included in the IEEE Draft Standard 823 (Methodologies for Specifying Voice Grade Channel Transmission Parameters and Evaluating Connection Transmission Performance for Telephony), have been developed. The CCITT has also been working on the development of extensive analytical models. Supplements 3 and 4 of Volume V of the Yellow Book (CCITT, 1981d) provide information on opinion models developed by AT&T and British Telecom. Both models will be reviewed in Section 4 of this report. Contributions to the current (1981-1984) Question 7/XII on models for predicting transmission quality will be reviewed in Section 3.2.4.

3.2.3 Methods for Assessing Telephony Transmission Performance

Supplement 2 of Vol. V of the CCITT Yellow Book (CCITT, 1981d) provides descriptions of methods for assessing telephony transmission performance that are recommended by the CCITT or have been employed during Study Periods from 1968 to 1980 in studying questions assigned to Study Group XII. Five methods are described. They are:

1) loudness comparison for speech (loudness ratings)
2) articulation (AEN) ratings
3) listening opinion tests

4) conversation opinion tests
5) quantal-response detectability tests.

These methodologies are briefly described in the next paragraphs.

Loudness comparisons for speech are loudness ratings that quantify the relative level at which speech signals reach the ears of the listener. In order to standardize the measurement procedure, the talking and listening conditions are controlled in a specified manner. Circuit noise and room noise are excluded from the determination so that results are governed by the loss in the speech path. The present recommended method is given in Recommendation P.72.

Articulation (AEN) measurements are based on the measurement of the fraction of speech sounds recognized correctly. Circuit noise and room noise at specified levels should be present and the results are affected by their levels. The articulation measurement method recommended by the CCITT is described in Recommendation P.45.

Listening opinion tests are widely used methods for the subjective evaluation of speech quality. Typically the tests are conducted using speech material in the form of sentences. Appendix A reviews different methods for the subjective evaluation of speech quality.

Conversation opinion tests may be conducted either as interviews after real subscribers have made actual calls or as laboratory tests. Extreme care must be taken in the preparation and execution of laboratory conversation opinion tests. The procedure that is used by British Telecom in conducting these type of tests is described in Supplement 2 of Volume V of the Yellow Book (CCITT, 1981d).

Quantal-response detectability tests are tests used to evaluate the detectability of some property of the signal such as echo. The subjective evaluation makes use of a scale such as the one that follows:

<u>CCITT Opinion Scale 6A</u>

A Objectionable
B Detectable
C Not detectable

Scales of this type may be used in a variety of quantal-response tests such as echo, reverberation, sidetone, voice-switching, mutilation, or interfering tones.

3.2.4 Contributions to Questions 7/XII and 18/XII

Table 11 lists contributions to Study Group XII, Question 7, for the Study Period 1981-1984 (CCITT, 1984). The subject of Question 7 is the development of

Table 11. Contributions to Study Group XII, Question 7, for Study Period 1981-1984

Contribution Number	Source	Title	Date
3	People's Republic of China	Examination of the "New Algorithm"	Feb. 1981
10	People's Republic of China	High-Pass and Low-Pass Filters Tests and Parameters m, G, and S' Estimated Therefrom	Apr. 1981
123	France	A Mathematical Model for the Calculation of Opinion Scores; Case of Analogue Transmission	Feb. 1983
124*	France	Application of Information Index to "Quantizing Noise" in PCM	Feb. 1983
124E*	France	Addendum to Contribution No. 124	Mar. 1983
125*	France	Definition of the Signal-to-Noise Ratio in Adaptive DPCM; Comparison of the Theory and Subjective Tests	Feb. 1983
158	AT&T	Subjective Effects of Sidetone	June 1983
174-E	Nippon Telegraph and Telephone	Transmission Performance Objective Evaluation Model for Fundamental Factors	Nov. 1983
187	Nippon Telegraph and Telephone	Analytical Results for the Measurement Results of CCITT Laboratory Technical Report No. 711	Jan. 1984
197	Special Rapporteur - STC	Contribution to Report on the Question	Jan. 1984
220*	Administration des Telecomm. De L'URSS	Evaluation de la Qualite de Transmission de la Parole D'un Systeme Numerique comportant des Codecs Micda	Jan. 1984

*Contribution applies to Question 18/XII also.

models for predicting transmission quality from objective measurements. The contributions listed in Table 11 that appear to be of particular interest are summarized in the following paragraphs.

Contribution No. 123 describes a mathematical model for the calculation of opinion scores for analog transmission. The model consists of an information index that can be calculated from objective measurements for transmission impairments such as attenuation, noise, and distortion. The information index provides a good correlation with the mean opinion score determined in both subjective conversation and listening tests. Contribution No. 124 describes the application of this model to quantization distortion.

Contribution 125 presents a table and curves that show the correlation of mean opinion scores with various definitions of the signal-to-noise ratio for adaptive delta pulse code modulation (ADPCM) coders.

Contribution 158-E describes a subjective test conducted by AT&T to evaluate the effects of sidetone path loss and sidetone frequency response on opinion ratings. In this test 24 sidetone conditions were presented in combination with 16 different conditions of loss, noise, and talker echo. There was a total of 384 test conditions each of which was subjectively evaluated on the excellent, good, fair, poor, and unsatisfactory scale. The results of these tests have been used to formulate an extension to the AT&T opinion model for transmission quality. This opinion model is described in Supplement 3 of Volume V of the CCITT Yellow Books (CCITT, 1981d), and is summarized in Section 4.1 of this report.

Contribution 174-E describes a model developed by Nippon Telegraph and Telephone called OPINE (Overall Performance Index Model for Network Evaluation). Curves are presented that compare predicted and experimental mean opinion scores (MOS) for transmission both with and without room noise. The standard deviation between predicted and experimental MOS was 0.19.

Contribution 187 describes the subjective and objective evaluation of the loudness rating of 13 telephone systems. The objective measurements were made based upon Recommendation P.79 (CCITT, 1981d) and were reported in CCITT Technical Report 711. Curves of subjective vs objective ratings are provided in the contribution. As noted in the contribution, if the deviation between subjective and objective ratings falls within the standard deviations in the subjective rating, the objective rating should be regarded as coinciding with the subjective rating. In the results described in the contribution, only 7 out of 56 ratings exceeded this range.

Contribution 197 from the special rapporteur discusses problems that are involved in the comparison of the various models that are now available. It is stated in the contribution that it is very desirable that as many persons as possible make comparisons between models so that the Study Group can obtain a wide view of the results of using the various models.

Table 12 lists contributions to Study Group XII, Question 18, for the Study Period 1981-1984. The subject of Question 18 is the transmission performance of digital systems. As can be seen from the table, the effect of transmission impairments on the end user's subjective evaluation of quality is of interest in digital systems just as it is for analog systems. The nature of the impairments of concern differs between analog and digital systems, however. For example, in digital systems, one is concerned with bit error rate and its effect on voice quality, while in analog systems one is concerned with loudness loss. Some impairments, such as delay, are equally of concern in both analog and digital networks. Further study is required on the use of models for digital telephony systems that perform the mapping of objective measurements of transmission impairments with the user's subjective evaluation. Prior work on data standards (ANSI, 1983; Seitz, 1980) has some direct applicability for digital voice transmission.

3.3 Exchange Carriers Standards Association Activities

The Exchange Carriers Standards Association (ECSA) is a private, voluntary association formed at the time the Bell Operating Companies were divested from AT&T. The ECSA furnishes the secretariat for a standards development body known as the T1 Committee. This committee follows the American National Standards Institute (ANSI) procedures in developing standards and is accredited by that organization. Membership on a T1 subcommittee is open to all interested and affected parties. These members generate standards using voluntary consensus procedures. There are currently six major subcommittees covering various technical areas. The T1Q1 subcommittee deals with Performance Objectives. Working groups reporting to T1Q1 include:

T1Q1.1	4 kHz Voice
T1Q1.2	4 kHz Voiceband Data
T1Q1.3	Digital Circuits
T1Q1.4	Digital Packet
T1Q1.5	Wideband Program
T1Q1.6	Wideband Analog.

Table 12. Contributions to Study Group XII, Question 18, for the Study Period 1981-1984

Contribution Number	Source	Title	Date
W/7	ITT Corp.	Effects on Telephony Transmission Performance of Transcoding and Transmission Errors in Digital Connections	Mar. 1981
W/15	Canada (BNR)	Transmission Performance of Digital Attenuators	Apr. 1981
W/24	Japan (NTT)	Study on Determination of Subjectively Equivalent Noise	July 1981
W/25	Japan (NTT)	Determination of Impairment Unit Assignment for Various Digital Systems	June 1981
W/35	Study Group XVIII	Extract from the Report of the Working Party XVIII/2: "Speech Processing" (COM XVIII-No. R4)	Nov. 1981
W/36	Study Group XVI	Extract of the Reply of Study Group XVI (COM XVI-No. R 1) to Question 1/XVI (Transmission Impairments in the Evolving Network)	Nov. 1981
W/52	France	Codage MIC Differentiel a 32 kb/s--Qualité de Transmission Vocale et Unites de Degradation	Mar. 1982
W/61	Canada (BNR)	Subjectively Equivalent Noise for Linear and Carbon Microphone Originated Speech Signals	Apr. 1982
W/63	Japan (NTT)	Some Considerations on Specifications for Modulated Noise Reference Unit (MNRU)	Apr. 1982
W/64	Japan (NTT)	Subjective Experimental Results for Impairment Unit Assignment to 32 kb/s ADPCM System	Apr. 1982
W/85	ITT	The Evaluation of a Test Protocol for 32 kb/s Codecs	June 1982
W/94	USA (AT&T)	32 kb/s ADPCM-DLQ Coding	June 1982

Table 12. (continued)

Contribution Number	Source	Title	Date
W/103	Study Group XVIII	Extract from Report COM XVIII-No. R 13 - Question 16/XVIII--Performance Characteristics of PCM Channels at Audio Frequencies	Sep. 1982
W/104	Study Group XVIII	Extract from the Report of the Meeting of Working Party XVIII/2 (Speech Processing) - COM XVIII-No. R10)	Sep. 1982
W/105	Study Group XVI	Extract of the Report (COM XVI-No. R 2) of Study Group XVI, Meeting of 14-16 June 1982	Oct. 1982
W/107	U.K. (British Telecom)	Methods of Subjective Assessment of Digit Processing Using the Modulated Noise Reference Unit	Nov. 1982
W/115	Norway	Performance Requirements for the MNRU	Feb. 1983
W/119	France	Description and Method of Use of the Modulated-Noise Reference Unit (MNRU/MALT)	Feb. 1983
124-E	France	Application of Information Index to "Quantizing Noise" in PCM	Feb. 1983
124-E	France	Addendum to COM XII - No. 124-E	Mar. 1983
125-E	France	Definition of the Signal-to-Noise Ratio in Adaptive DPCM; Comparison of the Theory and Subjective Tests	Feb. 1983
W/130	Canada (Bell Northern Research)	Subjective Equivalence Functions for Consideration in a Quantization Distortion Opinion Model	Mar. 1983
W/162	COMSAT	Transmission Performance of Digital Systems	July 1983
W/164	Japan (NTT)	Subjective Evaluation Results on 32 kb/s ADPCM Standards Candidates	July 1983

Table 12. (continued)

Contribution Number	Source	Title	Date
W/165	Japan (NTT)	Additivity of Subjective Evaluation Results for Voiceband Codecs	July 1983
W/166	USA (AT&T)	Subjective Evaluation of Speech Quality for Three 32 kb/s ADPCM Codecs	July 1983
W/167	Canada (BNR)	The Speech Performance of Three 32 kb/s ADPCM Codecs	July 1983
W/178	Working Party XVIII/2	Communications to SG VII, XVI, and XVII	Dec. 1983
W/179	USSR	Design Objectives of Mixed Analogue-Digital Connections	Dec. 1983
W/181	France	Quantizing Distortion in the Tandeming of 32 kb/s Differential PCM Codecs	Jan. 1984
W/215	France	Subjective Performance Assessment of the Digital Processing Using the Degradation Category Rating Procedure	Jan. 1984
W/214	France	Subjective Evaluation of Three 32-kb/s ADPCM Coders by Means of an Absolute Category Rating Test	Jan. 1984
W/219	USSR	Determination of Units of Quality Degradation in Voice Transmission Due to Quantification Distortion (dq units) [In French]	Jan. 1984
W/220	Administration des Telecommunications de L'URSS	Evaluation de la Qualité de Transmission de la Parole d'un Systeme Numerique Comportant des Codecs Micda	Jan. 1984

These groups are addressing various performance standards questions including voice transmission performance standards and performance standards for local exchange and interexchange reference connections.

Proposed standards development projects are submitted to the T1 secretariat where a T1 advisory group reviews them and recommends appropriate action. If the consensus is to develop a standard, the project is assigned to the appropriate committee.

Since T1Q1 was established in February 1984, this subcommittee has met on several occasions and each subcommittee's work is well under way. Currently there are 10 projects being pursued in T1Q1 and its working groups.

One area of particular interest here is a joint T1Q1.1 and T1Q1.2 working group project on the Switched Exchange Access Network Standard. As stated by the chairman, the development of a standard for network access does not ensure uniform quality performance but it should contain information that could be used toward ensuring uniform quality exchange access to all interexchange carriers provided other steps are also taken.

Transmission parameters for the Exchange Access Network will be specified in the standard. Specific values, limits and statistical distribution characteristics, and specific methods of assessing performance relative to the standard will also be included. The specific parameters suggested for inclusion are loss, noise, talker echo, listener echo, frequency response, crosstalk, intermodulation, distortion, envelope delay distortion, frequency shift, absolute delay, speech clipping, phase and amplitude jitter, transients, peak-to-average ratio, and digital impairments.

The Exchange Access Network Performance Standards are expected to become a part of a larger body of Network Performance Standards covering other aspects as well as exchange access. Included in the outline of the Network Performance Standards proposal are the following standards:

- Exchange Access Network Performance Standards
- Network-Interface to Network-Interface
- Point-of-Termination to Point-of-Termination Reference Document
- Point-of-Termination to Network-Interface Standards
- National Extension Standards.

4. QUALITY-OF-SERVICE MODELS

In the previous section, the telephony QOS standards being developed by IEEE, ECSA, and CCITT standards groups were discussed. In this section, we shall discuss some model development efforts that have taken place outside of these standards bodies. These models have been developed by AT&T and British Telecom. A discussion of these delay models and customer behavior models developed at Bell Laboratories is also included.

4.1 AT&T Transmission Performance Model

Prior to divestiture, AT&T had a long history of measuring the transmission performance of the Bell System. System-wide transmission surveys were conducted in 1959, 1962, 1966, and 1969-1970. Kessler (1971) and Duffy and Thatcher (1971) reported on the 1966 and 1969-1979 surveys, respectively. These surveys were designed to characterize the analog transmission performance of the switched network. The performance measurements were made on a sample of the toll connecting trunks. For example, in the 1966 survey, performance measurements were made of 392 trunks out of an estimated 800,000 toll connecting trunks in service. These 392 trunks included 242 trunks from 25 end offices having more than 400,000 annual outgoing toll messages (AOTM) and 150 trunks from 15 end offices having less than 400,000 AOTM's. Thus trunks from both large and small end offices were included. The following transmission parameters were measured in the 1966 survey:

- 1000 Hz loss
- frequency response
- relative envelope delay
- message circuit noise
- impulse noise
- peak-to-average ratio (P/AR)
- harmonic distortion
- level tracking on compandored facilities.

The parameters measured during the 1969-1970 survey differed from previous surveys (Duffy and Thatcher, 1971). For those parameters that had been measured in previous surveys as well as the 1969 survey, comparisons were made. These comparisons indicated a general trend toward improved transmission performance. Substantial improvement in phase jitter was observed. Significant improvements were also found in envelope delay distortion, attenuation distortion slope, and 1000-Hz loss results.

In addition to the surveys of toll connecting trunks reported by Kessler (1971) and Duffy and Thatcher (1971), measurements have also been made of the subscriber loops. Information on the characteristics of loops may be found in a paper by Gresh (1969).

The information from the performance measurements of the toll connecting trunks and subscriber loops has been used by AT&T in both the development of national network transmission plans and in the development of models of the network. The AT&T planning process is described by Andrews and Hatch (1971). The transmission plan associated with the toll switching plan used by AT&T is an evolutionary plan that changes to take advantage of new technology in providing better transmission quality and new service capabilities. It has been the policy of AT&T that whenever technological advances made it possible to transmit more channels for longer distances at lower costs, part of the benefit was converted to transmission improvements such as less attenuation distortion and lower loss and noise (Andrews and Hatch, 1971). This rule has led to a network that is a compromise between the best possible performance and the least cost of the service. The increased availability of alternative equipment and services has an impact on future network planning. This was noted even in 1971 by Andrews and Hatch who had the foresight to recognize the following factors as having a major impact on transmission planning:

- o The increased use of digital transmission facilities and digital switching may cause some increased problems initially but offer promise of improved transmission in the long run.
- o The increased use of the network for data transmission will require continued effort to control delay distortion, impulse noise, phase jitter, nonlinearities, and sudden changes in amplitude and phase.
- o The increased use of loudspeaking telephone sets will require additional consideration in network planning.
- o The increased use of customer-owned station equipment and systems that are interconnected to the network will put increased emphasis on the need for well-defined transmission performance of the network. In addition, the characteristics of signals produced by customer-provided equipment must be specified and controlled so that intermodulation and crosstalk caused by these signals do not interfere with other uses of the network.
- o The increased use of satellites with their inherently long transmission delays will continue to emphasize the importance of improved echo control.

The fourth item above is particularly significant to the subject of this report. Because of this early recognition of potential problems associated with the interconnection of a variety of services and equipment it became necessary for transmission standards to be developed. The work of the IEEE and CCITT QOS standards group is based, in part, on earlier measurement surveys, transmission network planning, and models developed by AT&T. The models will now be described.

Early work in the development of the AT&T transmission performance models has been described by Sullivan (1971), Cavanaugh et al. (1976), Hatch and Sullivan (1976), Cavanaugh et al. (1980), Cavanaugh (1980), and Cavanaugh et al. (1983). The latter paper as well as Supplement 3 to the CCITT Yellow Book, Volume V (CCITT, 1981d) provide summaries of the model developed by AT&T which has evolved over the last decade.

The tests reported by Cavanaugh et al. (1983) used a five-category rating scale (excellent, good, fair, poor, and unsatisfactory). Because subjective ratings are affected by differences in the types of test (conversation or listening), the subject group, and various other factors, the AT&T concept separates the relationship between the transmission parameters and opinion ratings into two parts:

1) the transmission rating, R, as a function of the transmission parameters, and

2) the relationship between the transmission rating, R, and opinion ratings, which then can be given for each individual test.

The model is expressed in terms of a series of equations from which the transmission rating, R, can be calculated based on knowledge of the transmission values for connections of interest (Cavanaugh et al., 1983). For example, the equation for talker echo is given by:

$$R_E = 106.4 - 53.45 \log_{10} \left[(1 + D)/\sqrt{1 + (D/480)^2}\,\right] + 2.277E \tag{11}$$

where E = loudness loss (in dB) of the talker echo path, and

D = round-trip talker echo path delay (in ms).

The R ratings may then be translated into percent good or better (GOB) or poor or worse (POW) through the use of the following equations:

$$\%\ \text{GOB} = \frac{100}{\sqrt{2\pi}} \int_{-\infty}^{A} e^{-t^2/2}\, dt \tag{12}$$

where t is the integration variable for the transmission rating factor.

$$\% \text{ POW} = \frac{100}{\sqrt{2\pi}} \int_{B}^{\infty} e^{-t^2/2}\, dt \tag{13}$$

where A and B are given by the following for three different data bases:

Data Base	A	B
1965 Murray Hill Test	(R-64.07)/17.57	(R-51.87)/17.57
AT&T Long Toll Tests	(R-51.5)/15.71	(R-40.98)/15.71
CCITT Conversation Tests	(R-62)/15	(R-43)/15

Figure 8 is a plot of the percent GOB and percent POW for each of the three data bases (CCITT, 1981d).

The AT&T model provides equations for the transmission rating for the following parameters:

- o listener echo (R_{LE})
- o loss/noise/listener echo (R_{LNLE})
- o circuit noise equivalent of quantization noise (N_{Qe})
- o circuit noise equivalent of room noise (N_{Re})
- o bandwidth/attenuation distortion (R_{LNBW})
- o sidetone (R_{LNST})

The equations for these various transmission ratings and plots of the equations may be found in Cavanaugh et al. (1983) and CCITT (1981d). The model can be implemented on small desktop computers for the analysis of the performance of simple connections, but must be put on larger computers for network studies.

Figures 9 through 13 present the percent GOB for circuit noise, number of tandem A/D conversions, bandwidth, echo path loss, and reference equivalent of the echo path. The overall reference equivalent, which is the dependent variable in Figure 9 and a parameter in Figures 10 through 13, is a function of circuit noise at the input to the telephone (N_c), the circuit noise equivalent of room noise (N_{Re}), the circuit noise equivalent of quantization noise (N_{Qe}), and the reference equivalent of the overall telephone connection (L_e). The equation for the overall reference equivalent is given by (CCITT, 1981d):

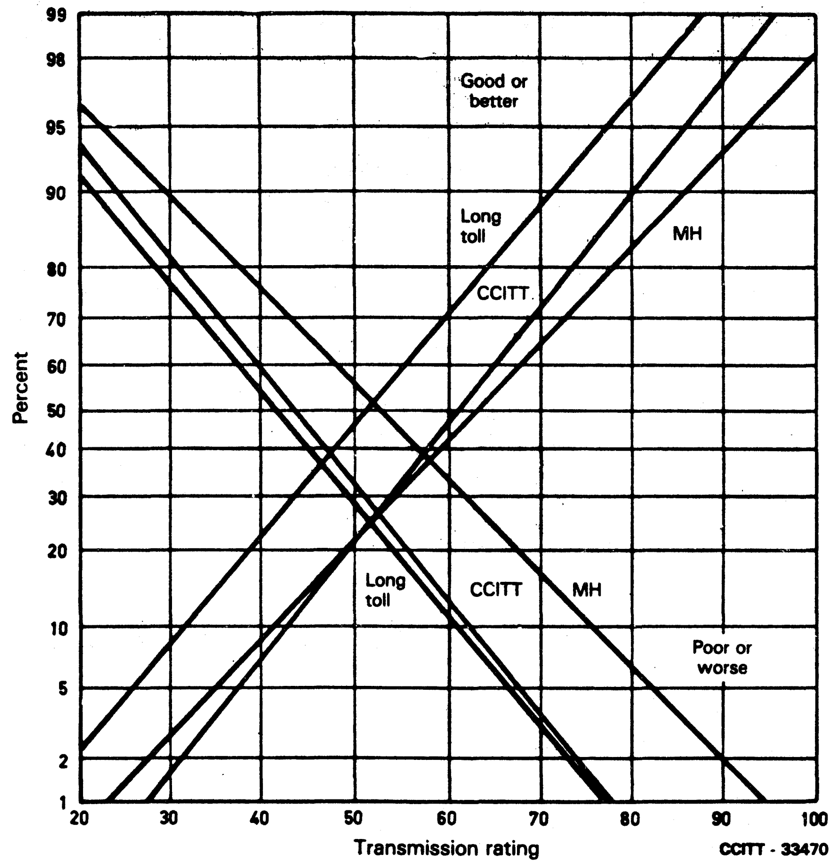

Figure 8. Comparison of opinion ratings as a function of transmission rating.

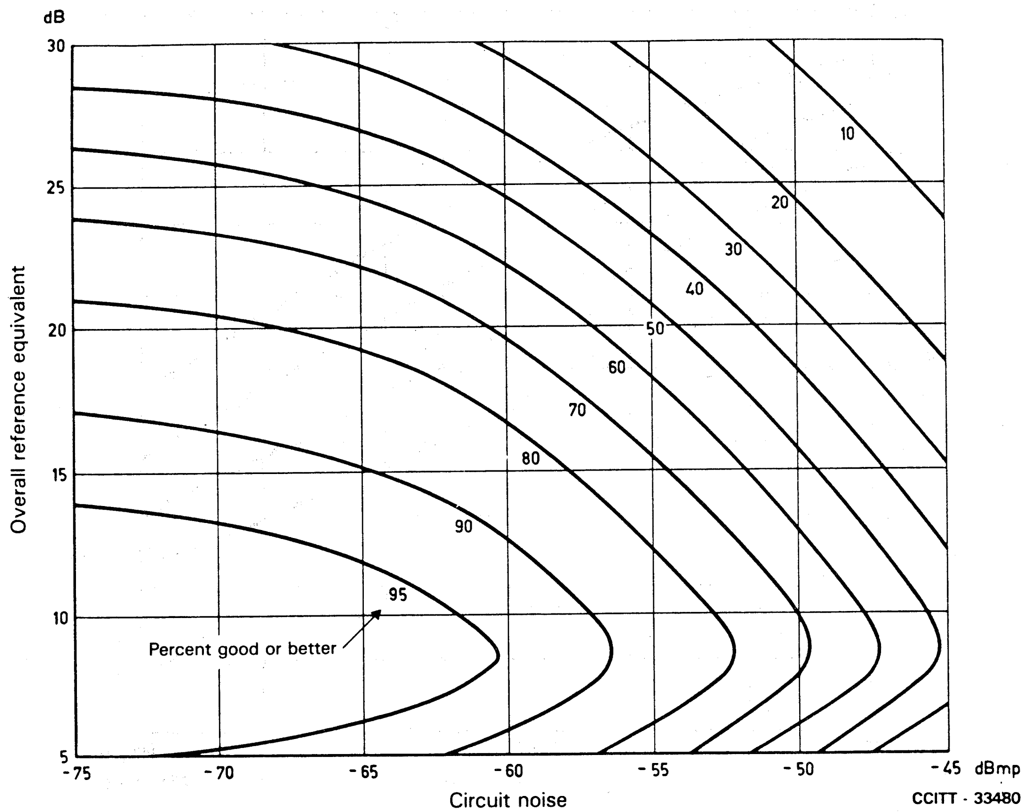

Figure 9. Opinion rating for overall reference equivalent and circuit noise.

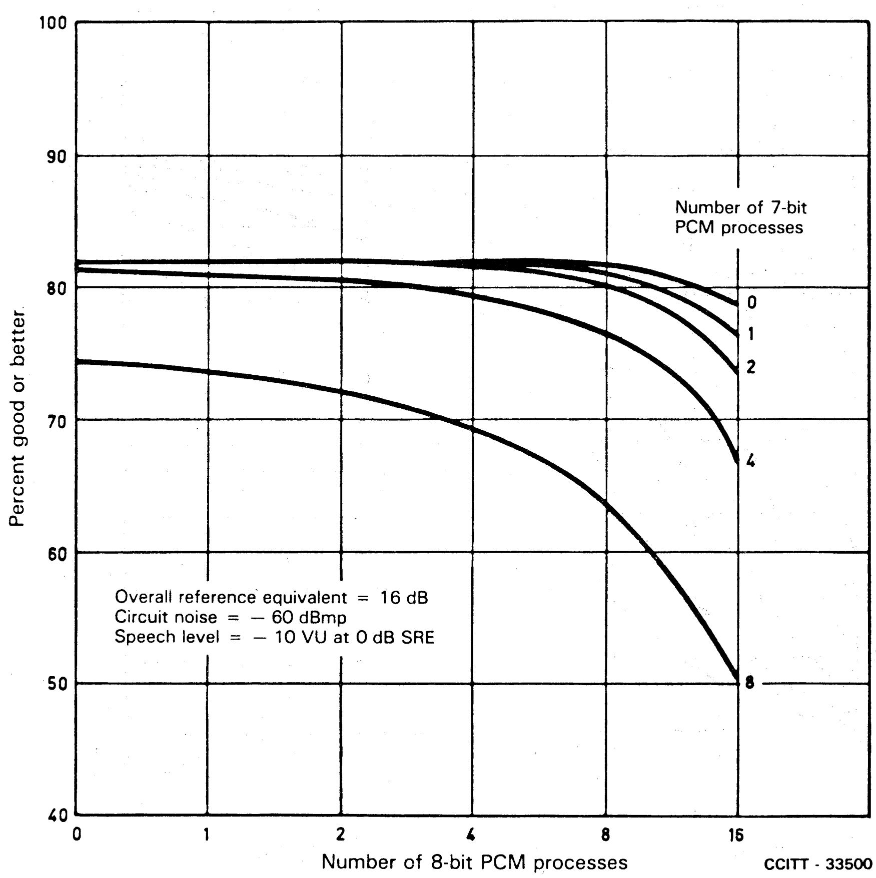

Figure 10. Opinion rating for tandem PCM processes.

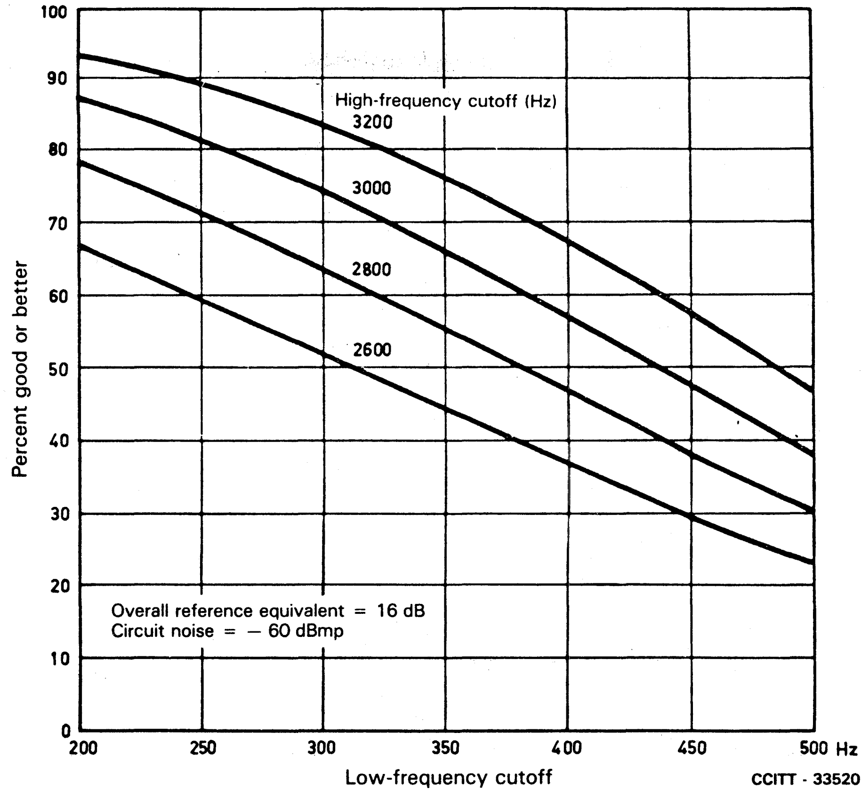

Figure 11. Opinion rating for bandwidth.

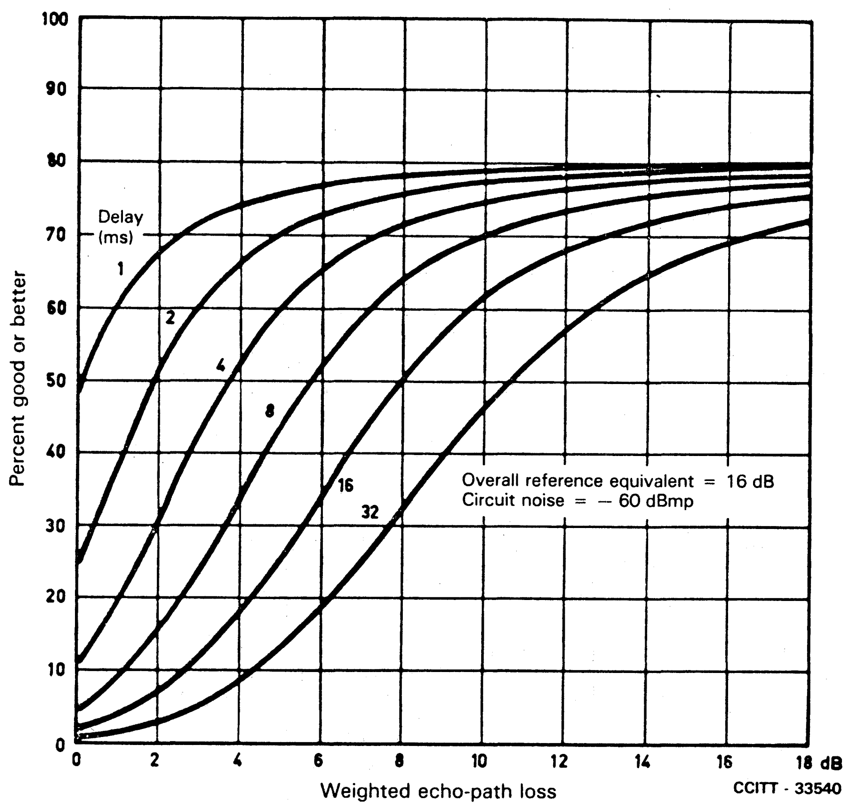

Figure 12. Opinion rating for listener echo.

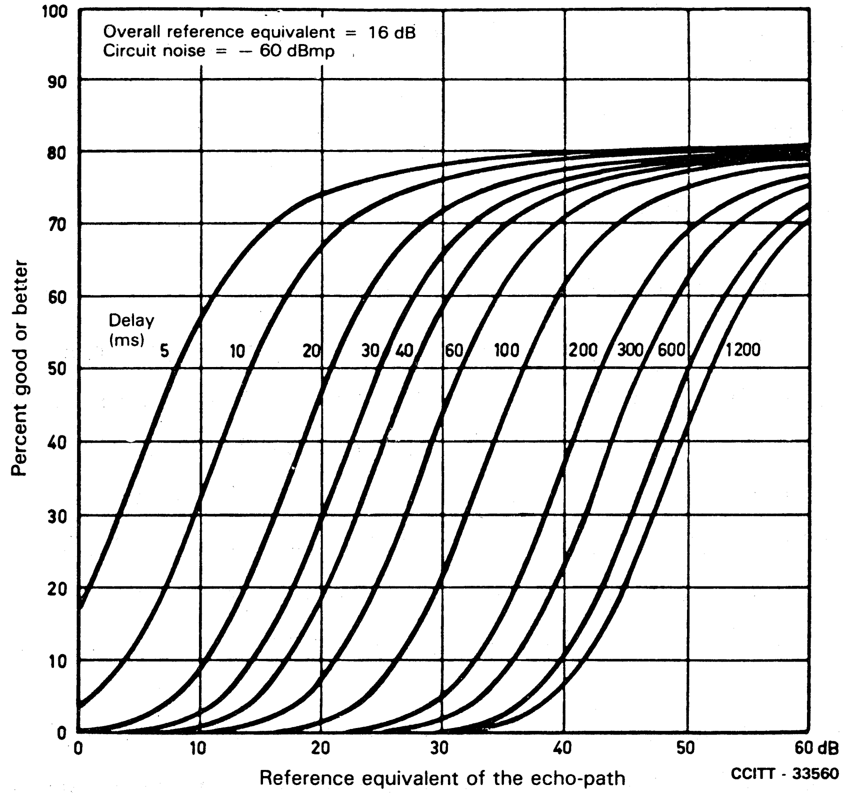

Figure 13. Opinion rating for talker echo.

$$R_{LN} = -34.88 - 2.257[(L_e - 8.2)^2 + 1]^{1/2} - 2.0294\ N_F + $$
$$+ 1.883\ L_e + 0.02037\ L_e N_F \qquad (14)$$

where $N_F = N_c \circledast N_{Re} \circledast N_{Qe}$, and

$\circledast$ indicates power addition.

While it is outside the scope of this report to discuss impairment measurement instrumentation and methods, it should be noted that these parameters are easily measured. Consider Figure 11 for example, which is a plot of percent GOB versus bandwidth. Modern microprocessor-based instrumentation is available which automatically measures and plots the frequency response of telephone circuits from 0 to 4.0 kHz. It is an easy step to translate these frequency response curves to the expected MOS using the curves provided in Figure 11.

4.2 British Telecom Transmission Performance Model

The British Telecom model is intended to predict loudness judgments, listening-effort scores, conversation-opinion scores, and vocal levels from subjective information supplied (CCITT 1981d, Supplement 4). The model deals with the subjective effects of circuit loss, attenuation-frequency distortion, circuit noise, quantizing noise, room noise, and sidetone paths for a wide range of values of these characteristics in any combination.

The model uses the mean Conversation Opinion Score (Y_c) and the Listening Effort Score (Y_{LE}). The Y_c score can take any value between 4 and 0, the scale being:

4 = excellent
3 = good
2 = fair
1 = poor
0 = bad.

The listening effort score is determined by transmitting lists of sentences at a standard input speech level over the connection. The listener votes, at a number of different listening levels, on the effort required to understand the meanings of the sentences according to the following scale:

A - complete relaxation possible; no effort required
B - attention necessary; no appreciable effort required
C - moderate effort required
D - considerable effort required
E - no meaning understood with any feasible effort

The votes are scored on a scale from 4 to 0, with the highest score given to the "A" evaluation.

The model has been implemented in computer programs and requires the following inputs:

- overall sensitivity-frequency characteristic of each transmission path and sidetone path. These sensitivities may be measured by the method described in Recommendation P.64 (CCITT, 1981d).
- noise level spectrum at each listener's ear
- average speech spectrum and average threshold of hearing.

From these data the loudness ratings are calculated. With the speech level fixed, Y_{LE} and a provisional value of Y_c are evaluated for each participant.

It is possible to estimate Y_{LE} by a process similar to processes already used in calculating loudness and articulation scores. An intermediate quantity, Listening Opinion Index (LOI), is first calculated as follows. Each elementary band in the frequency range contributes to LOI an amount proportional to the product $B_f P(Z_f)$, where B_f is a frequency-weighting factor expressing the relative importance of that elementary band for effortless comprehension, and P is a growth function applied to the sensation level Z which is evaluated for the loudness calculation. The growth function is limited to the range 0 to 1 as in articulation, but the form used is:

$$P(Z) = 10^{\left(\frac{Z + 3.8}{10}\right)} \qquad \text{for } Z < -11$$

$$= 1 - 10^{\left[\frac{0.3\,(Z + 14)}{10}\right]} \qquad \text{for } Z \geq -11 \;. \tag{15}$$

Figure 14 is a plot of (15). Figure 15 is a plot of the logarithm of B for discrete frequency bands from 100 to 8000 Hz. The Listening Opinion Index is evaluated from:

$$LOI = AD \sum_i B_i \, P(Z_i) \tag{16}$$

where A is a multiplier depending on the received speech level, and D is a multiplier depending on the received noise level.

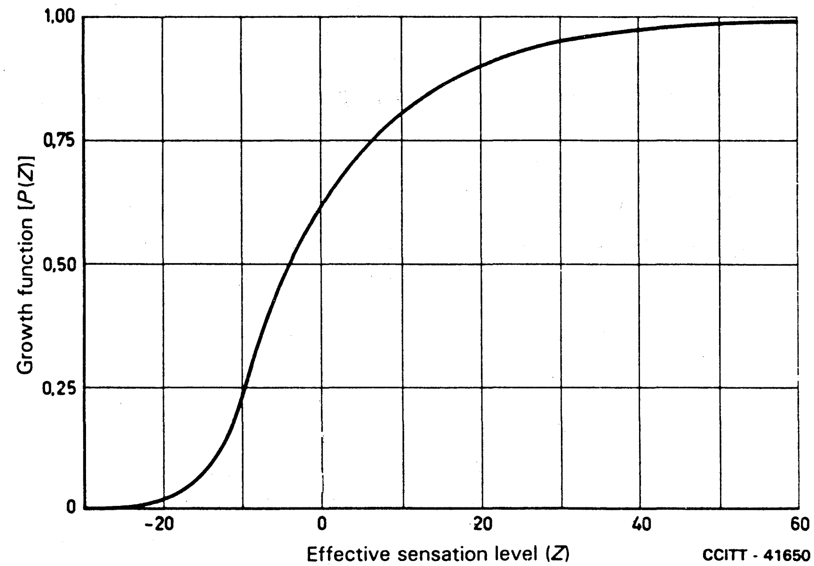

Figure 14. Growth function P(Z).

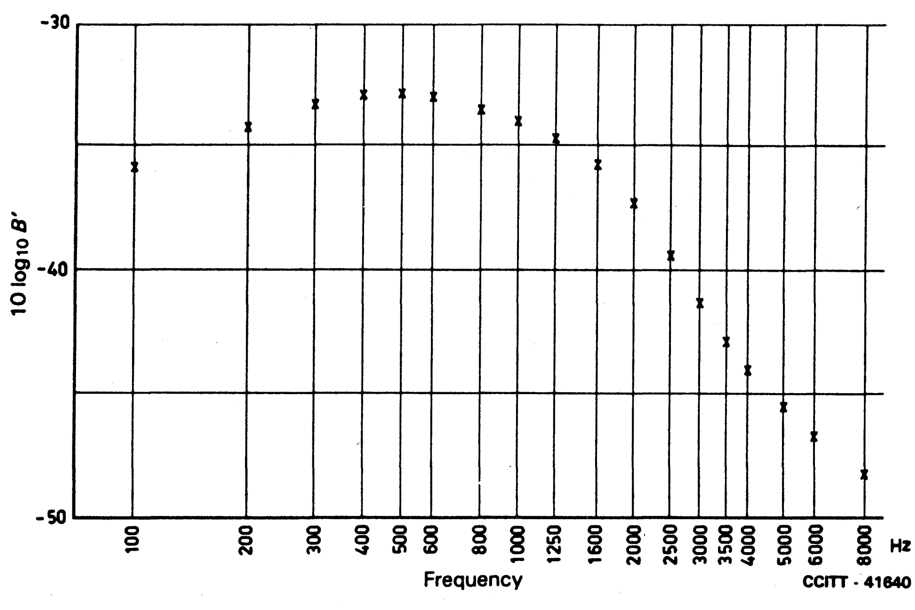

Figure 15. Frequency-weighting factor B' for Listening Opinion Index.

Figures 16 and 17 are plots of the multiplier functions A and D. The Listening Effort Score (Y_{LE}) is then calculated from the following equation:

$$\ln \left(\frac{Y_{LE}}{4 - Y_{LE}}\right) = 1.465 \left[\ln \left(\frac{LOI}{LOI_{LIM} - LOI}\right) - 0.75\right]. \tag{17}$$

Figure 18 is a plot of Y_{LE} vs LOI. Thus Y_{LE}, predicted Listening Effort Score, can be evaluated for each participant as a function of listening level.

The conversation opinion score (Y_C) can be calculated from Y_{LE} using an equation developed by British Telecom (CCITT, 1981d, Supplement 4).

British Telecom reports fair correspondence between values for Y_C obtained using the model and observed values obtained through laboratory conversation tests.

4.3 Delay, Timing Jitter, and Speech Loss Models

The models discussed previously in this report have not explicitly included the effect of delay on speech quality as perceived by the end user. Although not incorporated in any model under consideration by standards groups, there are nevertheless existing models that show the correlation between delay and subjective mean opinion scores (MOS) test scores (see Appendix A for a description of subjective MOS tests). The works of Falconer (1983) and Gruber and Strawczynski (1982) are but two examples of research performed on the effect of delay and timing jitter on the acceptability of transmission performed. This research is briefly discussed in the following paragraphs. Cox (1980) discusses time delay effects on speech intelligibility.

Falconer (1983) considers the problem of the need for carefully designed echo cancellers for use on full-duplex digital transmission at 64 kb/s on the single twisted-wire pair constituting a normal subscriber loop. Such transmission is a necessary major step toward the digitization of the loop plant. Coupling of the digital transmitters and receivers by imperfectly matched hybrids and other line discontinuities can cause echoes of the local transmitter's signal to interfere with the desired far-end signal at the receiver input. Careful design of the echo canceller is mandated in order to achieve acceptable performance. One area that must be given particular attention is that of timing jitter. Falconer analyzes the effect of this jitter on the mean-squared error of the signal entering the receiver, but does not relate it to the end-user's perception of the quality of the service whether the service is voice or voiceband data.

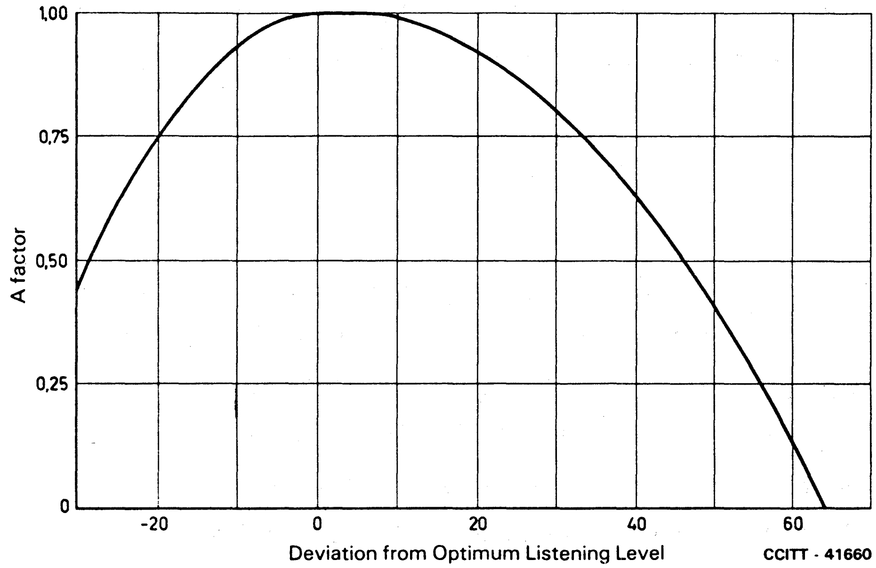

Figure 16. Effect of listening level on Listening Opinion Index.

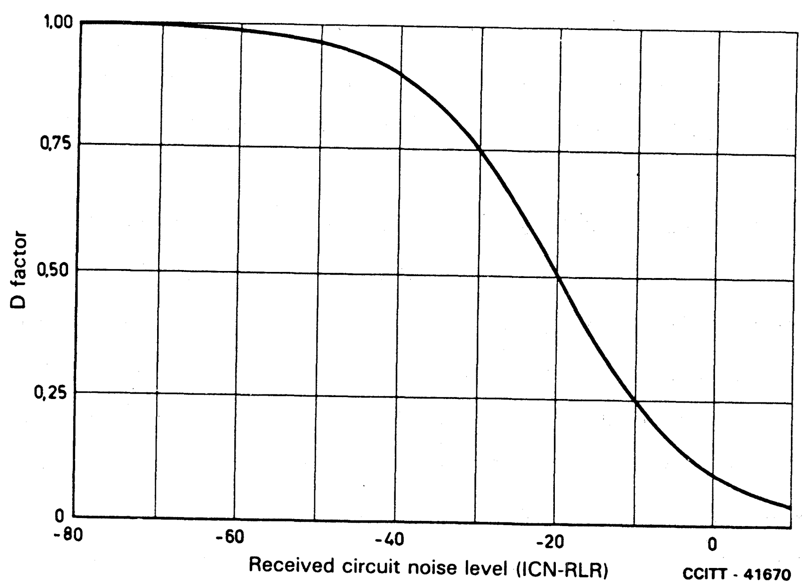

Figure 17. Effect of received noise level on Listening Opinion Index.

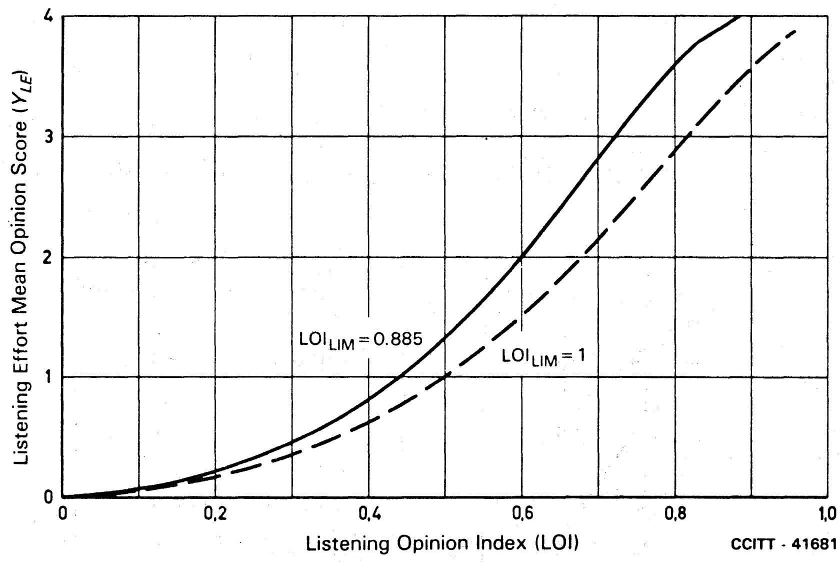

Figure 18. Listening opinion score as a function of Listening Opinion Index.

Gruber and Strawczynski (1982) provide numerous curves that show the effect of percent of speech loss, speech loss duration, talkspurt delay, and talkspurt hangover on subject MOS ratings from listening tests. A speech decision or talkspurt is indicated if the signal power exceeds a preset threshold. The speech signal may be delayed (10 ms typically) relative to the speech detector to minimize front-end clipping due to the speech detector. The speech decision is maintained if the signal power drops below the threshold for less than the hangover interval. Once the hangover interval is exceeded, the signal is classified as silence.

The focus of the paper of Gruber and Strawczynski is on speech impairments in dynamically managed voice (DMV) systems. A DMV system is one that includes speech activity detection and variable-bit-rate coding to exploit speech idle time. Examples of such systems include nonbuffered and buffered speech interpolators, store and forward (e.g., packet voice networks), integrated voice/data switches and networks and voice messaging systems.

The curves provided by Gruber and Strawczynski (1982) show the following results:

- o The effect of talkspurt delay on MOS scores is highly dependent upon the talkspurt hangover. For a talkspurt hangover of 50 ms, the MOS = 3 threshold (fair) is crossed at delays of only 50 ms. For a talkspurt hangover of 300 ms the MOS = 3 threshold is not reached until delays of almost 450 ms are encountered.

- o The MOS = 3 threshold is reached for percent of speech losses of only 1% to 5% depending upon speech loss duration.

  The percent of speech activity is also a factor that would affect the MOS ratings. For example, a speech signal having an activity factor of 90% would be degraded significantly more by a talkspurt delay of, for example, 50 ms than a speech signal having an activity factor of 50%. Gruber and Strawczynski do not show these types of curves, however.

- o Front-end clipping (FEC) is superior to mid-talkspurt clipping (MTC) for less than 20 clips per minute and talkspurt hangover of greater than 200 ms.

The third item above appears to need further investigation because other researchers have indicated that MTC is superior to FEC. Which is better apparently depends upon other parameters such as hangover time.

Although useful results have been reported in the literature on the effect of delay on mean opinion scores, these results have not been incorporated into any models being developed by the IEEE and CCITT standards groups. Delay modeling remains an open issue.

4.4 Customer Behavior Models

Liu (1980) describes long-distance dialing call completion and customer retrial behavior. Kort (1983) discusses models developed by Bell Laboratories for the customer acceptance of telephone connections in the Bell System Public Switched Telephone Network. The system of models described by Kort is a true quality-of-service model in that it models the acceptability of all three phases of a telephone call including call setup, transmission, and call completion. The system of models described by Kort includes the following:

- customer opinion models
  - call setup rating
  - transmission rating
  - call completion rating
- customer behavior models
  - abandonment/retrial behavior
  - complaint rate behavior
- network performance characterization models.

Most of the prior emphasis has been focused on transmission performance. The system of models described by Kort includes the behavior model depicted earlier in Figure 2. The various parts of this system of models will now be briefly described. Actual equations will not be given, but may be found in Kort (1983).

<u>Call setup rating</u>

The call setup rating, R_{cs}, is a function of the dial-tone delay, the post-dialing delay, the total number of digits that were dialed in all ineffective attempts that preceded the successful attempt, and the total waiting time (dial-tone delay plus post-dialing delay) of all ineffective attempts that preceded the successful attempt.

<u>Transmission performance opinion models</u>

The transmission rating for loss-noise-echo, R_{LNE}, is given by the following:

$$R_{LNE} = 0.5[R_{LN} + R_E - ((R_{LN} - R_E)^2 + 400)^{1/2}] \tag{18}$$

where R_{LN} is given by equation (14), and

R_E is given by equation (11).

Call completion rating

The call setup rating, R_{cs}, and the transmission rating, R_{LNE}, combine to form a composite call completion rating, R_{cc}. Overall opinion of long-holding-time calls is strongly dominated by transmission quality. Conversely, for short-holding-time calls, call setup impairments dominate opinion. The call completion rating has the following form:

$$R_{cc} = R_{cs}^{\alpha} \; R_{LNE}^{1-\alpha}$$

where α is a function of the conversation time.

Customer abandonment behavior model

Customers may abandon a call attempt in any of three stages of call setup: while waiting for a dial tone, during dialing, or while waiting for network response after dialing is complete. During conversation, a customer may abandon and replace a call, or simply terminate early due to poor transmission. Kort (1983) provides equations giving the probability for each of these types of abandonment.

Complaint behavior model

This model describes the complaint rate to the operator, which was determined from field-trial call-back interviews. The percent of complaints is surprisingly low even for connections that are judged to be below the good MOS rating. For example, for 10,000 calls of which there was a 50% good or better MOS score, fewer than 400 are reported to the operator (Kort, 1983).

Network performance models

Statistical models of network performance have been developed from surveys and measurements. The distribution of the noise-loss transmission rating, R_{LN}, as a function of mileage has been determined to be a normal distribution. Probabilities have been found for call setup dispositions (including ring and answer, ring/no answer, station busy, etc.). Holding times have been found to be exponentially distributed with a mean of 10 minutes for toll calls, 2.2 minutes for local business calls, and 4.5 minutes for local residence calls (Kort, 1983).

The network performance characterization models may be applied to customer opinion and behavior models to evaluate customer acceptance of the telephone network. Since network impairments are characterized by statistical distributions and the impairments accumulate using fairly complex equations, the calculation of overall performance is most easily accomplished through Monte-Carlo simulation. Kort (1983) describes this simulation process.

5. ISSUES FOR FURTHER RESEARCH

Table 13 lists several issues that need further research. While this list is incomplete, it contains some of the most significant issues related to quality of service. Some of the issues are quite broad in nature as they include not only present analog telephony networks, but digital and hybrid networks that will continue to be present during the transition to the future Integrated Services Digital Network (ISDN). Other issues are quite specific. Each of the issues listed in Table 13 will be summarized in the following paragraphs.

Table 13. Issues for Further Research

1.	Enhancement of IEEE and CCITT telephony QOS standards
2.	Objective measures of voice quality for the end user
3.	User's ability to isolate problems in a multivendor connection
4.	QOS standards for voiceband data and other services
5.	Standards for interconnected analog and digital networks and the transition to the ISDN
6.	Application of the Open Systems Interconnection Model

Enhancement of IEEE and CCITT telephony QOS standards

There are numerous subissues to the broad issue of standards enhancement. They include, but are not limited to, the following:

- Identification of additional impairments (such as delay and timing jitter) to be included in telephony QOS standards. Should IEEE 823 be revised to include more of the individual impairments listed by the CCITT?

- Determination of the impact on performance of multiple impairments. How many impairments individually and in combination can be realistically evaluated? What is the range of values of the impairments that should be evaluated (e.g., range of delay values)?

- Determination of distributions to be used in applying IEEE Standard 823 (see Jensen, 1983).

- Specification of how to measure new parameters (such as delay).

- Definition of the desired percent GOB service. What do most customers expect?

- Extension of telephony QOS standards to the call setup and call definition phases. Current emphasis is on the information transfer phase, but customer satisfaction is also dependent upon call setup time, numbering plans, accuracy in billing, etc. Can the IEEE/CCITT standards activities be extended to include customer behavior models such as those described in Section 4.4?

- Extension of the IEEE loop performance standard (P-280). Palladino and Wilkens (1983) note the limitations of the standard to two-wire analog transmission. Can the standard be extended to full-duplex digital transmission on two-wire subscriber loops or is a new standard required?

- Investigation of the method of measurement of transmission parameters. Silverthorne (1983) notes that measurement methods are being developed under a separate IEEE standards project. Are these methods complete? Do they apply only to voice?

- Identification of the interface at which the service will be defined and measured.

Silverthorne (1983) notes that future issues of IEEE Standard 823 are expected to deal with additional speech impairments and parameters (such as frequency response, listener echo, speech clipping, etc.). Since the list of individual impairments under investigation by the CCITT is more extensive than that of the IEEE, can the CCITT standards work be applied, at least in part, to future revisions of the IEEE standard? A broader question may be stated as follows: How applicable is the work expended in maintaining service quality in interconnecting international networks applicable to the domestic problem of interconnecting different interexchange carriers and local access networks?

Objective measures of voice quality

More research is needed into the development of objective measures for voice quality of service. Although much research has been performed in this area, as discussed in Appendix B, no universally accepted objective measure has been developed. The research reports referenced in Appendix B need to be studied in more detail with the idea of then choosing two or three promising objective measurement technologies for detailed investigation in the laboratory. There are numerous problems with subjective testing: differences in results depending upon male/female/and ethnic category of both speakers and listeners, language type of test (listening versus conversation), range of conditions, and completeness. Because of the large number of impairments that can be varied both individually and in combination, efficient evaluation is mandated. Objective evaluation is a desirable goal that will help attain this efficiency. The objective measure, in its development stage, must be verified through the use of parallel subjective testing, i.e., the objective measure must be correlated to the subjective opinion score in order to verify the accuracy of the objective measure. After such accuracy has been demonstrated there should no longer be a need for subjective testing.

User's ability to isolate problems

One of the thorniest problems from a user's point of view is in the isolation of a problem when many different vendors of both equipment and services are involved in the end-to-end connection. Rather than being able to go to a single vendor as in the past, the customer is faced with the prospect of having to go to multiple vendors in an attempt to isolate QOS problems. In the hypothetical (but quite likely) case, service providers may believe that their service meets all QOS standards such as the IEEE Standard 823. Yet the net result is that the service on the end-to-end connection is poor. How does the customer verify that in fact any particular vendor's service does meet standards? The problem is made more difficult because of the statistical nature of the end-to-end connection. One does not expect each long-distance call to use the same trunking each time the call is placed between any two given stations. Thus testing and verification can be a time-consuming, costly process. With multiple vendors involved, each of which may believe that they provide high quality service, the isolation of problems is difficult. While the development of standards is the necessary first step, the fault isolation problem remains difficult for the customer when many vendors are involved.

QOS for voiceband data and other services

As noted by Silverthorne (1983), the original objective of the IEEE Working Group was to develop a standard to define and specify methodologies, parameters, and performance criteria relevant to the transmission of speech and data through voiceband telecommunications networks. The current draft standard does not include voiceband data or any other type of services except voice that can be transmitted over voiceband networks. Therefore, there is a need for the extension of the standards to include such services, or the development of a new standard.

Transition to the ISDN

The problem of QOS for the Integrated Services Digital Network and the hybrid analog/digital networks that are its predecessor has not been addressed by the various standards bodies or very extensively in the literature. Gruber and Le (1983a) define the problem, but do not provide approaches to its solution except in very broad terms. The problem of ISDN QOS is complicated by the fact that not only are the parameters of interest somewhat different for voice and for data, but also that even when there is a common parameter the range of acceptable values is likely to be quite different. Consider delay in an integrated services network. Data service end users are likely to be more tolerant of delays than are voice end users. On the other hand, voice users are likely to be more tolerant of errors in transmission than are data users. Although Gruber and Le (1983a) suggest some possible approaches to the issue, it is obviously very complex and will require further research. Decina (1982) discusses the status of CCITT activity on signal processing for ISDN.

Relevance of the Open Systems Interconnection Model

The OSI model has provided an effective framework for data communications standards development. Research is needed to determine if this model can be adapted to telephony in order to assure that the complete user-user functional and logical connection is covered in standards development efforts and that complete user-oriented performance for the entire end-to-end connection can be represented in various standards (McManamon, 1984).

6. RECOMMENDATIONS

It is recommended that technical research programs be strengthened in each of the areas described below:

Participation in IEEE QOS Standards Committees

Although the IEEE Draft Standard 823 is expected to be approved in the near future, revisions to the standard are likely in order to incorporate additional transmission parameters. Some of the specific enhancements that are needed were identified in Section 5 (Issues for Further Research). The standard needs to be extended to include voiceband data as well as speech. Equal exchange access, as discussed by McManamon (1984), has a major impact on the QOS as perceived by the end user. The IEEE standards work should emphasize the equal exchange access aspect of the QOS issue.

Participation in CCITT Study Groups

Significant work has been performed by CCITT SG XII on QOS issues over the last several study periods. The focus of this work has been on analog voice networks. Little work has been done to date on QOS issues for the Integrated Services Digital Network. This is a complicated issue as noted by Gruber and Le (1983a). There are many differences in the choice of QOS parameters for various services such as voice and data. Even in the case where a parameter is equally applicable to various services (e.g., delay), the range of acceptable values is likely to be dramatically different for different services.

Development of objective voice quality measures

Much research has been conducted over the last decade in an attempt to develop an objective voice quality measure (see Appendix B). No universally accepted objective measure has been developed, however. A two-phase project is suggested for the voice quality objective measure program. Phase I of the project would be to complete the task begun in Appendix B to evaluate the results of previous research in this area. Promising objective measures and new concepts should be evaluated. Phase II of the project would be to demonstrate selected measures and any new concepts using a laboratory test facility.

The development of objective voice quality measures can take several approaches. One approach is illustrated in Figure 19. The goal is to develop a user-oriented,

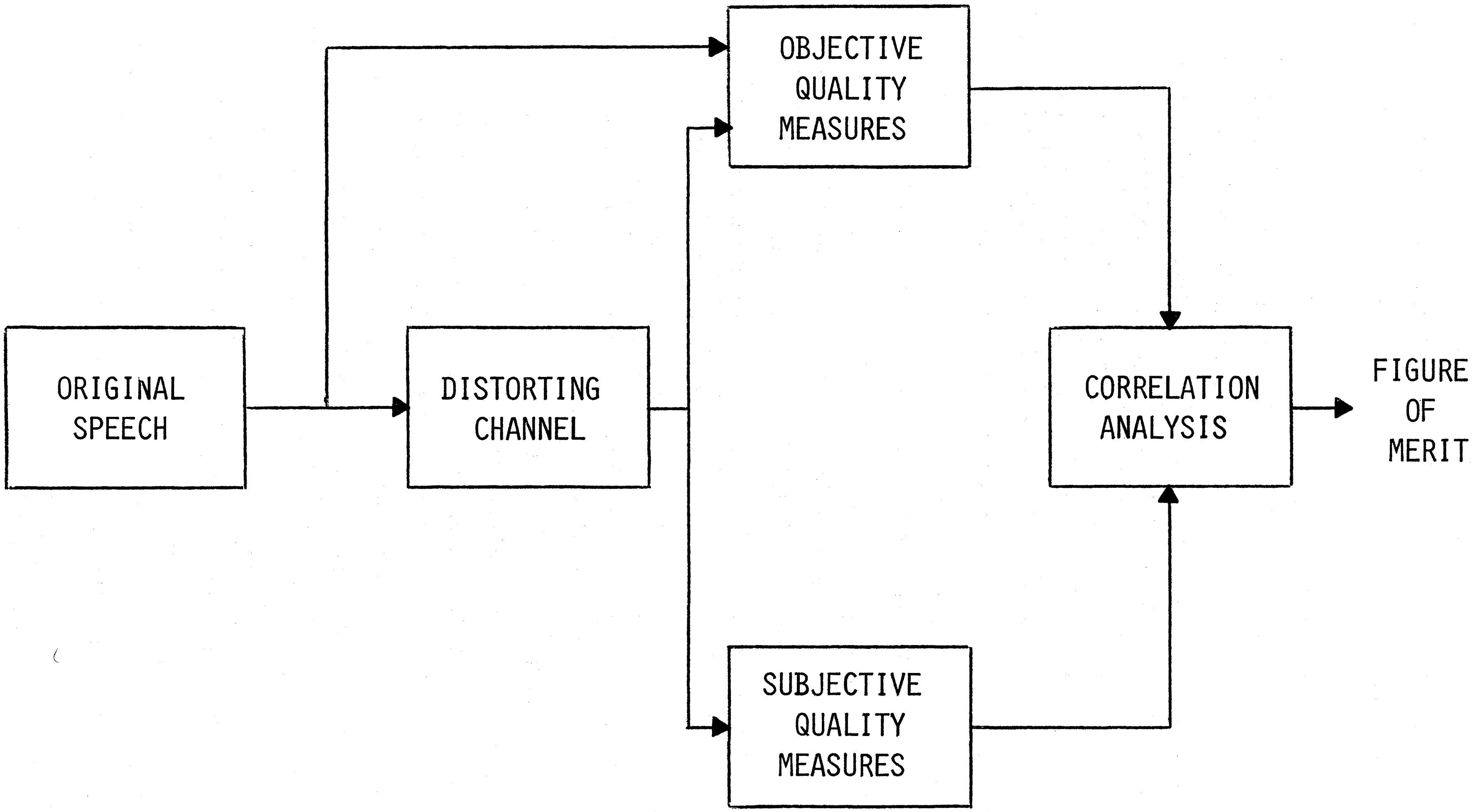

Figure 19. Approach to development of objective measures for quality of speech.

system independent, objective performance measure or figure of merit for voice. The approach suggested here is to assess the merits of various objective measures currently in use (such as spectral and amplitude distances measures or signal to noise and distortion measures) by conducting a correlative analysis with subjective measures (such as mean opinion scores). The process is depicted in Figure 19. It is entirely feasible that some entirely new objective measures may be required or at least a set of existing ones that will cover the entire range of systems being implemented.

7. REFERENCES

Andrews, Jr., F.T., and R.W. Hatch (1971), National telephone network transmission planning in the American Telephone and Telegraph Company, IEEE Trans. Commun. Tech. COM-19, No. 3, pp. 302-314, June.

ANSI (1983), American National Standard for information systems, for data communication systems, and services, user-oriented performance parameters, ANSI X3.102-1983 (ANSI, 1430 Broadway, New York, NY 10018).

Bellamy, J.C. (1982), Digital Telephony, (John Wiley and Sons, New York).

Cavanaugh, J.R. (1980), A model for the subjective effects of quantizing distortion in digital transmission of speech, ICC '80 Conference Record, Vol. 2, pp. 20.5.1-5.

Cavanaugh, J.R., R.W. Hatch, and J.L. Neigh (1980), A model for the subjective effects of listener echo on telephone connections, BSTJ 59, No. 6, pp. 1009-1060, July.

Cavanaugh, J.R., R.W. Hatch, and J.L. Sullivan (1976), Models for the subject effects of loss, noise, and talker echo on telephone connections, BSTJ 55, No. 9, pp. 1319-71, November.

Cavanaugh, J.R., R.W. Hatch, and J.L. Sullivan (1983), Transmission rating model for use in planning of telephone networks, GLOBECOM '83 Conference Record, Vol. 2, pp. 20.2.1-6.

CCITT (1969), Telephone transmission quality, local networks, and telephone sets, Recommendations of the P series, White Book, Vol. V, Geneva.

CCITT (1973), Telephone transmission quality, local networks and telephone sets, Recommendations of the P series, Green Book, Vol. V, Geneva.

CCITT (1977), Telephone transmission quality, Recommendations of the P series, Orange Book, Vol. V, Geneva.

CCITT (1981a), International Telephone service operation, Recommendations E.100-E.323, Yellow Book, Vol. II, Fascicle II.2, Geneva.

CCITT (1981b), International Telephone service network management traffic engineering, Recommendations E.401-E.543, Yellow Book, Vol. II, Fascicle II.3, Geneva.

CCITT (1981c), Specifications of measuring equipment, Recommendations of the O series, Yellow Book, Vol. IV, Fascicle IV.4, Geneva.

CCITT (1981d), Telephone transmission quality, Recommendations of the P series, Yellow Book, Vol. V, Geneva.

CCITT (1981e), General characteristics of international telephone connections and circuits, Recommendations G.101-G.171, Yellow Book, Vol. III, Fascicle III.1, Geneva.

CCITT (1984) Contribution to Report on the Question (Question 7/XII), SG XII, COM XII-197, Special Rapporteur - STC, 6 pp.

Cox, J.C. (1980), Time delay effects on speech intelligibility, Proc. ICASSP '80, Vol. 3, pp. 1034-6.

DiBiaso, L.S., (1983), Telecommunications systems performance needs for standards GLOBECOM '83 Conference Record, Vol. 2, pp. 20.1.1-7.

Duffy, F.P., and T.W. Thatcher, Jr. (1971), 1969-70 connection survey: Analog transmission performance on the switched telecommunications network, BSTJ 50, pp. 1311-47, April.

Ebert, I.G. (1984), The evolution of integrated access toward the ISDN, IEEE Comm. Mag. 22, pp. 6-17, April.

Falconer, D. (1983), Timing jitter considerations for digital subscriber loop echo cancelers, GLOBECOM Conference Record, pp. 40.8.1 to 40.8.7.

Goodman, D.J., and R.D. Nash (1982), Subjective quality of the same speech transmission conditions in seven different countries, Proc. ICASSP 1982, Vol. 2, pp. 984-7.

Gresh, P.A. (1969), Physical and transmission characteristics of customer loop plant, BSTJ 48, pp. 3337-3386, December.

Gruber, J.G., and N.H. Le (1983a), Performance requirements for integrated voice/data networks, Jour. on Selected Areas in Comm., SAC-1, December, pp. 981-1005.

Gruber, J.G., and N.H. Le (1983b), Performance requirements for integrated voice/data networks, GLOBECOM '83, Conference Record, pp. 2.2.1-9.

Gruber, J., and L. Strawczynski (1982), Subjective effects of variable delay and speech loss in dynamically managed voice systems, GLOBECOM '82 Conference Record, Vol. 3, Miami, pp. F7.3.1-F7.3.5.

Hatch, R.W., and J.L. Sullivan (1976), Transmission rating models for use in planning of telephone networks, National Telecommunication Conference Record, Vol. II, pp. 23-2-5, November.

ITS Staff (1983), A primer on integrated service digital networks (ISDN): implications for future global communications, NTIA Report 83-138, September (NTIS Order No. PB84-152586).

Jensen, R.E. (1983), Applications of the draft IEEE standard for specifying and evaluating voiceband channel transmission performance, GLOBECOM '83 Conference Record, Vol. 2, pp. 20.4.1-5.

Johnson, C.P. (1983), Post-divestiture telecommunications standards requirements--position statement, GLOBECOM '83 Conference Record, Vol. 3, pp. 48.2.1-2.

Kessler, J.E. (1971), The transmission performance of Bell System Toll Connecting Trunks, BSTJ 50, pp. 2741-2777, October.

Kort, B.W. (1983), Models and methods for evaluating customer acceptance of telephone connections, GLOBECOM '83 Conference Record, Vol. 2, pp. 20.6.1-9.

Listanti, M., and F. Villani (1983), Voice communication handling in X.25 packet switching networks, GLOBECOM '83 Conference Record, Vol. 1, pp. 2.4.1-5.

Liu, K.S. (1980), Direct Distance Dialing: call completion and customer retrial behavior, BSTJ 59, No. 3, pp. 295, March.

Maitre, X., and T. Aoyama (1982), Speech coding activities within CCITT: Status and trends, Proc. ICASSP 1982, Vol. 2, pp. 954-9.

McManamon, P.M. (1984), Telecommunications quality of service in the post-divestiture arena, report published by the Cyberlink Corporation, Boulder, CO 80303, under contract to NTIA.

Nesenbergs, M.N., W.J. Hartman, and R.F. Linfield (1981), Performance parameters for digital and analog service modes, NTIA Report 81-57, 294 pp (NTIS Order No. PB81-176455).

Palladino, J.R., and W.B. Wilkens (1983), Standard on voice frequency telephone loop performance characteristics, GLOBECOM '83 Conference Record, Vol. 2, pp. 20.51-4.

Schweizer, L. (1983), Performance of terrestrial and satellite 64 kbit/s paths: Requirements of voice and data, and standards of the future integrated services digital network (ISDN), ICC '83 Conference Record, Vol. 1, pp. A1.4.1-5.

Seitz, N.B. (1980), Interim Federal Standard 1033 Reference Manual, NTIA Report 80-55, December (NTIS Order No. PB81-174898).

Silverthorn, R.D. (1983), IEEE draft standard: Methodology for specifying and evaluating voiceband channel performance criteria, GLOBECOM '83 Conference Record, Vol. 2, pp. 20.3.1-8.

Sullivan, J.L. (1971), A laboratory system for measuring loudness loss of telephone connections, BSTJ 50, No. 8, pp. 2663-2739, October.

8. BIBLIOGRAPHY

Agrawal, A., and Wen C. Lin (1974), An on-line speech intelligibility measurement system, IEEE Trans. Acoustics, Speech and Signal Processing, ASSP-22, pp. 203-206.

Aidarous, S.E., S.A. Mahmound, and J.S. Riordon (1983), Performance analysis of a packetized voice radio network, ICC '83 Conference Record, Vol. 1, pp. B8.7.1-7.

ANSI (1969), American National Standard methods for the calculation of the articulation index, ANSI S3.5-1969 (American National Standards Institute, Inc., 1430 Broadway, NY, NY 10018), 21 pp.

ANSI (1971), USA Standard method for measurement of monosyllabic word intelligibility, ANSI S3.2-1960 (R1971) (Acoustical Society of America), (ANSI, 1430 Broadway, New York, NY 10018), 19 pp.

ANSI (1977), American National Standard for Rating Noise with Respect to Speech Interference (Acoustical Society of America, 335 East 45th St., New York, NY 10018), ANSI S3.14-1977, 6 pp.

Andrews, Jr., F.T., and R.W. Hatch (1971), National telephone network transmission planning in the American Telephone and Telegraph Company, IEEE Trans. Commun. Tech. COM-19, No. 3, pp. 302-314, June.

Avellaneda, O.A., J.F. Hayes, and M.M. Nassehi (1982), A capacity allocation problem in voice-data networks, IEEE Trans. Commun. COM-30, No. 7, July, pp. 1767-72.

Barberis, G., M. Calabrese, L. Lambarelli, and D. Roffinella (1983), Coded speech in packet-switched networks: Models and experiments, GLOBECOM '83 Conference Record, Vol. 1, pp. 2.6.1-5.

Barberis, G., and G. Tamburelli (1978), Topics in analysis and design of generalized packet communications networks, Proceedings, ICC Toronto, C3.4.1-C3.4.5 (pp. 553-557).

Barnwell, T.P., III (1980), A comparison of parametrically different objective speech quality measures using correlation analysis with subjective quality results, Proc. ICASSP '80, Vol. 3, pp. 710-13.

Barnwell, T.P., III (1980), Correlation analysis of subjective and objective measures for speech quality, Proc. ICASSP '80, Vol. 3, pp. 706-9.

Barnwell, T.P., III, A.M. Bush, and R.M. Mersereau (1977), Speech quality measurement, Georgia Institute of Technology, Atlanta School of Electrical Engineering, AD-A058 833.

Barnwell, T.P., III, and S.R. Quackenbush (1982), An analysis of objectively computable measures for speech quality testing, Proc. ICASSP 1982, Vol. 2, pp. 996-9.

Barnwell, T.P., III, R.W. Schafer, and A.M. Bush (1977), Tandem interconnections of LPC and CVSD digital speech coders, Defense Documentation Center Report AD-A060 838 from Georgia Institute of Technology, Atlanta, GA (Defense Comm. Engineering Center, 1860 Wiehle Ave., Reston, VA 22090).

Barnwell, T.P., III, and W.D. Voiers (1979), An analysis of objective measures for user acceptance of voice communications systems, Final Report DA100-78-C-0003, DCEC, Reston, VA 22090, September.

Beeson, W.R. (1977), An algorithm for determining speech intelligibility, available from Defense Documentation Center, report AD-A055 236.

Belfield, W. (1977), Performance measurements of the CVSD speech coding algorithm, available from Defense Documentation Center, report AD-A0050 163.

Billi, R., and C. Scagliola (1980), An identification method for objective quality measurements on speech waveform coders, Proc. ICASSP '80, Vol. 3, pp. 714-7.

CCITT (1982a), Study Group XII, W/85, The evaluation of a test protocol for 32 kb/s codecs, (ITT), June, 9 pp.

CCITT (1982b), SG XII, W/94, 32 kb/s ADPCM-DLQ Coding, USA--AT&T, June, 10 pp.

CCITT (1982c), Study Group XVIII, W/103, Extract from Report COM XVIII-No. R13--Question 16/XVIII--Performance characteristics of PCM channels at audio frequencies, (Study Group XVIII), September, 6 pp.

CCITT (1982d), Study Group XII, W/104, Extract from the report of the meeting of Working Party XVII/2 (Speech Processing), COM XVIII-No. R 10, Study Group XVIII, September, 14 pp.

CCITT (1982e), Study Group XII, W/105, Extract of the Report (COM XVI-No. R 2) of Study Group XVI, Meeting of 14-16 June 1982 (Study Group XVI), October, 3 pp.

CCITT (1982f), Study Group XII, W/107, Methods of subjective assessment of digital processes using the modulated noise reference unit, U.K., British Telecom, November, 6 pp.

CCITT (1983a), Study Group XII, W/115, Performance requirements for the MNRU, Norway, February, 9 pp.

CCITT (1983b), Study Group XII, W/119, Description and method of use of the Modulated-Noise Reference Unit (MNUR/MALT), France, February, 3 pp.

CCITT (1983c), Study Group XII, W/179, Design objectives of mixed analogue-digital connections, USSR, December, 8 pp.

CCITT (1983d), Study Group XII, W/178, Communication to Study Groups VII, XVI, and XVII, (WP XVIII/2), December, 1 p.

CCITT (1983e), Study Group XII, W/167, The speech performance assessment of three 32-kb/s ADPCM codecs, Canada--BNR, July, 6 pp.

CCITT (1983f), Study Group XII, W/166, Subjective evaluation of speech quality for three 32-kb/s ADPCM coders [sic], U.S.A.--AT&T, July, 7 pp.

CCITT (1983g), Study Group XII, W/165, Additivity of subjective evaluation results for voiceband codecs, Japan--NTT, July, 8 pp.

CCITT (1983h), Study Group XII, W/164, Subjective evaluation results on 32-kb/s ADPCM standard candidates, Japan--NTT, July, 9 pp.

CCITT (1983i), Study Group XII, W/162, Transmission performance of digital systems, COMSAT), July, 3 pp.

CCITT (1983j), Study Group XII, W/130, Subjective equivalence functions for consideration in a quantization distortion opinion model, Canada--BNR, March, 6 pp.

CCITT (1983k), Application of information index to "quantizing noise" in PCM, Questions 7/XII and 18/XII, SG XII, COM XII-No. 124, France, 7 pp.

CCITT (1983l), Study Group XII Digital echo suppressor with adaptive function - practical test results from studies of customer conversations, SG XII, COM XII-No. 170, Fed. Republic of Germany, 1 p.

CCITT (1983m), Application of the Information Index to quantizing noise in PCM, SG XII, COM XII-No. 124, France, 3 pp.

CCITT (1983n), Definition of the signal-to-noise ratio in adaptive DPCM; Comparison of the theory and subjective tests, SG XII, COM XII-No. 125, Feb. France, 9 pp.

CCITT (1983o), Subjective effects of sidetone, SG XII, COM XII-No. 158, AT&T, Questions: 4/XII and 7/XII, 9 pp.

CCITT (1983p), Proposal for a supplement No. XYZ to Volume V--Noise spectra. Geneva 1984 - (Quoted in Recomms. P.44 and P.45 and Question 24/XII), SG XII, COM XII-No. 140, April 1983, British Telecom, 6 pp.

CCITT (1983q) Report of the 3rd meeting of SG XII, Geneva, 30 May-3 June 1983, SG XII, COM XII, No. R 7, June, SG XII/CCITT Secretariat, 172 pp.

CCITT (1984a), Transmission performance objective evaluation model for fundamental factors, SG XII, COM XII-174, Nippon Telegraph and Telephone Public Corporation, 10 pp.

CCITT (1984b), Analytical results for the measurement results of CCITT Laboratory Technical Report No. 711 (Loudness rating of 13 telephone systems obtained subjectively and objectively using aural measured system), SG XII, COM XII-187, Nippon Telegraph and Telephone Public Corp., 3 pp.

CCITT (1984c), Loudness parameter values M and G for calculating objective RE, SG XII, COM XII-188, Nippon Telegraph and Telephone Public Corp., 8 pp.

CCITT (1984d), Extract of the Report of the Meeting, COM XVI-R 3, Geneva, 6-9 June 1983, SG XII, COM XII-203 (Jan. 1984) SG XVI, 2 pp.

CCITT (1984e), Extract of the Report of the Meeting, COM XVI-R 3, Geneva, 6-9 June 1983, SG XII, COM XII-205 (Jan. 1984), Study Group XVI, 4 pp.

CCITT (1984f), Draft Revised Recommendation P.72, COM XII-211 January, CCITT Laboratory working party, Geneva, Jan. 17-19, 18 pp.

CCITT (1984g), Proposed supplement to Volume V: Telephone Transmission Quality 1984 Part II, SG XII, COM XII-213 (Jan. 1984), Special Rapporteur for Question 3/XII, 11 pp.

CCITT (1984h), Evaluation de la qualite de transmission de la parole d'un systeme numerique comportant des codecs MICDA, SG XII, COM XII-220 (Jan. 1984), Admin. of Telecommunications, USSR, p. 5.

CCITT (1984i), Study Group XII, W/219, Determination des unites de degradation de la qualite de transmission vocale due a la distortion de quantification (unites dq), USSR, January, 5 pp.

CCITT (1984j), Study Group XII, W/215, Subjective performance assessment of digital process using the degradation category rating procedure (D.C.R.), France, January, 8 pp.

CCITT (1984k), Study Group XII, W/181, Quantizing distortion in the tandeming of 32 kb/s differential PCM codecs, France, January, 12 pp.

Chytil, M., and A. Niedzwiecki (1982), Simulated speech signal and its use in voice transmission measurements, ICC '82 Conference Record, Vol. 3, pp. 61.5.1-61.5.6.

Clarke, P.G. (1983), Status of CCITT studies on the overall performance of public data networks, GLOBECOM '83 Conference Record, Vol. 1, pp. 13.6.1-3.

Combescure, P., A. Le Guyader, and A. Gilloire (1982), Quality evaluation of 32 Kb/s coded speech by means of degradation category ratings, Proc ICASSP 1982, Vol. 2, pp. 988-91.

Coulter, D.C. (1974), Quality and acceptability testing of voice processors for military applications, Report #NRL-7773.

Coverdale, P.V. (1974), The speech voltmeter type 5B MK II, DOD Report, April.

Crochiere, R.E., and J.L. Falanagan (1982), Current perspectives in digital speech, ICC '82 Conference Record, Vol. 2, pp. 3G.3.1-3G3.5.

Crochiere, R.E., L.R. Rabiner, S.S. Joyant, and J.M. Tribolet (1978), A study of objective measures for speech waveform coders, International Zurich Seminar on Digital Communications, Zurich, Switzerland, March.

Crochiere, R.E., J.M. Tribolet, and L.R. Rabiner (1980), On the measurement of waveform coder distortion using the log likelihood ratio, Proc. ICASSP '80, Vol. 2, pp. 340-3.

Dal Degan, N. (1978), Vocoder quality: An automatic procedure to measure the performance of pitch extractors, GLOBECOM '82, Conf. Record, Miami, Vol. 3, F7.2.1.-F7.2.4 (pp. 1325-1328).

Daumer, W.R. (1982), Subjective evaluation of several efficient speech coders, IEEE Trans. Commun. COM-30, pp. 655-662, April.

Daumer, W.R., and J.R. Cavanaugh (1978), A subjective comparison of selected digital codecs for speech, BSTJ 57, pp. 3119-3165.

Daumer, W.R., and J.L. Sullivan (1982), Subjective quality of several 9.6-32 kb/s speech coders, Paper S14.12, Vol. 3, Proceedings ICASSP 82, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1709-16, May.

Defense Documentation Center (1979), Report on Speech/voice coding and standards, DDC Report No. T30263, approximately 100 pp.

Deutsche, Bundespost (1970), 4th International Symposium on Human Factors in Telephony, September 23-27, 1968, Bad Wiessee, Germany.

Dibiaso, L.S. (1977), Transmission considerations for local switched digital network, Telephony 133, pp. 40-52, October.

DiBiaso, L.S. (1979), Satellite user reaction tests - A subjective evaluation of echo control methods, NTC '79 Conference Record, Vol. 3, pp. 48.6.1-6, November.

Duffy, F.P., and R.A. Mercer (1978), A study of network performance and customer behavior during direct-distance-dialing call attempts in the U.S.A., BSTJ 57, pp. 1-33.

Farber, L., and J. Cormack (1977), Performance of voice communications systems in the presence of spread spectrum interference, Defense Documentation Center report AD-A050 844, Report to DOD Electromagnetic Compatibility Analysis Center, North Severn, Annapolis, MD 21402.

Fischer, M.J., (1979), Delay analysis of TASI with random fluctuations in the number of voice calls, DCEC-TN-15-79.

Fairbanks, G. (1958), Test of phenomic differentiation: the rhyme test, J. Acoust. Soc. Am. 30, pp. 596-600.

Gamauf, K.J., and W.J. Hartman (1977), Objective Measurement of Voice Channel Intelligibility, U.S. DoT (FAA) Report No. FAA-RD-77-153, 76 pp.

Gibson, J.D. (1980), Predictor optimization and performance improvement in DPCM, NTC '80 Conference Record, Vol. 2, pp. 36.6.1-3.6.5.

Goodman, D.J., and R.D. Nash (1982), Subjective quality of the same speech transmission conditions in seven different countries, Proc. ICASSP 1982, Vol. 2, pp. 984-7.

Griffiths, J.D. (1967), Rhyming minimal contrasts: a simplified diagnostic articulation test, J. Acoust. Soc. Am. 42, pp. 236-241.

Gruber, J., R. Vickers, and D. Cuddy (1982), Impact of errors and slips on voice service, ICC '80 Conference Record, Vol. 1, pp. 2D.5.1-2D.5.7.

Hanson, E.D. (1971), Intelligibility testing of the continuous variable slope delta (CVSD) Coder-vocoder (CODEC), Tech. Report ECOM-3393, U.S. Army Report, Ft. Monmouth, NJ.

Hartman, W.J., and S.F. Boll (1976), Voice channel objective evaluation using linear predictive coding, FAA Report No. FAA-RD-75-189 (U.S. Dept. of Transportation, FAA Systems Research and Development Service, Washington, DC 20590), August.

Hartman, W.J., and L.E. Pratt (1980), Voice performance measurements, FAA Report No. FAA-RD-80-71 (U.S. Dept. of Transportation, FAA Systems Research and Development Service, Washington, DC 20590), June.

Heggestad, H.M., R.J. McAulay, and J. Tierney (1982), Practical considerations for speech digitizing systems at rates from 64.0 to 0.6 kbps, GLOBECOM '82 Conference Reocrd, Vol. 1, B3.8.1-B3.8.6, pp. 349-354.

Heggestad, H.M., and C.J. Weinstein (1983), Voice and data communication experiments on a wideband satellite/terrestrial internetwork system, ICC '83 Conference Record, Vol. 1, pp. A1.1.1-8.

House, A.S., C.E. Williams, M.H.L. Hecker, and K.D. Kryter (1965), Articulation-testing methods: consonatal differentiation with a closed-response set, J. Acoust. Soc. Am. 37, pp. 158-166.

Hubbard, R.W., and W.J. Hartman (1974), Objective measurement techniques for evaluating voice communication channels, FAA Report No. FAA-RD-74-77, Washington, D.C., July.

IEEE (1969), IEEE Recommended practice for speech quality measurements, No, 297, (IEEE, 345 E. 47th Street, NY, NY 10017) 21 pp.

IEEE (1976), IEEE Standard test procedure for measuring longitudinal balance of telephone equipment operating in the voice band, ANSI/IEEE Std. 455-1976, IEEE Std. 455-1976 (IEEE, 345 East 47th St., NY, NY 10017), 16 pp.

Kamal, S.S., R. Matyas, R.G. Lyons, and P.P. Nuspl (1983), Application of demand assignment to mixed-services TDMA, GLOBECOM '83, Conference Record, Vol. 3, pp. 39.1.1-6.

Kawahara, H. (1981), Assessment system for service quality, Rev. Elec. Commun. Lab. (Japan) 29, No. 9-10, pp. 907-24, Sept.-Oct.

Kawahara, H., and K. Kakehi (1981), Objective calculation of ratings based on loudness, Rev. Elec. Commun. Lab. 29, Nos. 9-10, pp. 873-880.

Kitawaki, N., K. Itoh, M. Honda, and K. Kakehi (1982), Comparison of objective speech quality measures for voiceband codecs, Proc. ICASSP 1982, Vol. 2, pp. 1000-3.

Kort, B.W. (1983), Models and methods for evaluating customer acceptance of telephone connections, GLOBECOM '83 Conference Record, Vol. 2, pp. 20.6.1-9.

Kryter, K.D. (1972), Methods for the calculation and use of the articulation index, J. Acoust. Soc. of Am. 34, pp. 1689-1697.

Lacroix, A., and B. Makai (1982), A vocoder scheme for very low bit rates (quality evaluation), Proc. ICASSP 1982, Vol. 1, pp. 618-21.

Landeide, O.E. (1978), Speech quality of telephone connections (in Norwegian), Telektronikk (Norway) 74, No. 4, pp. 287-93.

Leon-Garcia, A., R.H. Kwong, and G.F. Williams (1982), Performance evaluation methods for an integrated voice/data link, IEEE Trans. Commun. COM-30, No. 8, pp. 1848-57, August.

Low, D.A. (1978), A packet speech measurement facility, a report of work by the Computer Corp. of America, available from Defense Documentation Center, report AD-A052 338.

Low, D., and J. Rothnie (1978), Packet speech measurement facility, available from Defense Documentation Center, report AD A061 346.

Manhire, L.M. (1978), Physical and transmission characteristics of customer loop plant, BSTJ 57, p. 35, January.

McDermott, B.J., C. Scagliola, and D.J. Goodman (1968), Perceptual and objective evaluation of speech processed by adaptive differential PCM, Proc. of ICASSP, Tulsa, OK, April.

Mehmet-Ali, M.K., and C.M. Woodside (1983), Optimal choice of packet size and reconstruction delay for a packet voice system, GLOBECOM '83 Conference Record, Vol. 2, pp. 28.5.1-5.

Meister, S. (1978), The diagnostic rhyme test (DRT): An Air Force implementation, RADC-TR-78-129.

Mermelstein, P., and G. Williams (1982), Network performance isssues in 32 kb/s coding of speech and voiceband data, GLOBECOM '82 Conference Record, Vol. 1, A8.2.1-A8.2.5, pp. 234-238.

Miller, G.A., G.A. Heise, and W. Lichten (1951), The intelligibility of speech as a function of the context of the test materials, Experimental Psychology, 41, pp. 329-335.

MIT (1976), Speech evaluation, Report for DCA, available from Defense Documentation Center, report AD-A041 246.

Navy, Dept. of (1976), Speech Intelligibility, MIL-STD-1472, (Navy Publications and Forms Center, Code 3015, 5801 Tabor Ave., Philadelphia, PA 19120), 2 pp.

Neal, J. (1983), Voice band data power levels on the telecommunications network, ICC '83 Conference Record, Vol. 1, pp. B3.1.1-4.

Ottinger, D.M. (1978), Objective measure of speech intelligibility using linear predictive coding (LPC) Air Force Institute of Technology, Wright-Patterson AFB, AFIT/GE/EE/78-35.

Payne, J., and P.M. McManamon (1974), Speech Quality Measurement. Presented at SETE Workshop on Automatic Test and Checkout of Communication Systems, June 3-7, 1974.

Payne, J.A., and P.M. McManamon (1973), An objective speech quality measurement of a communication channel, Office of Telecommunications Report 73-14 (NTIS Order No. COM75-10940/AS).

Rabiner, L.R., and R.W. Schafer (1978), Digital Processing of Speech Signals (Prentice Hall, Inc., Englewood Cliffs, NJ 07632).

Richards, D.L. (1967), Some comments on application of the principles described in research report No. 20117 to the rating of subscribers' sets, DoD Report, available from Defense Documentation Center, report AD 838090L, August.

Richards, D.L. (1969), The transmission characteristics of handset telephones, DoD Report, available from Defense Documentation Center, report AD 868430L, May.

Richards, D.L. (1971), Standards of telephony performance required for the future, DoD Report, available from Defense Documentation Center, report AD 890357L, May.

Richards, D.L. (1971), Transmission performance characteristics of subscriber's telephone set 16 at 31 (with regulator): Information for network planning, DoD Report, August.

Richards, D.L., and G.J.P. Barnes (1982), Estimating communication efficiency of telephone connections, ICC '82 Conference Record, Vol. 3, pp. 6I.6.1-6I.6.6.

Richards, D.L., and G.J.P. Barnes (1982), Determination of loudness losses of telephone connections, ICC '82 Conference Record, Vol. 3, pp. 6I.2.1-6I.2.4.

Richards, D.L., and G.J.P. Barnes (1982), Estimating communication efficiency of telephone connections, ICC '82 Conference Record, Vol. 3, pp. 6I.6.1-6.

Rothauser, E.H., and G.E. Urbanek (1967), Speech quality measurements, Institut fur niederfrequenztechnik, Vienna, Austria, AFCRL-69-2004, 140 pp.

Sawakuri, T., and Y. Yamamoto (1983), Transmission loss planning for digital network, ICC '83 Conference Record, Vol. 1, pp. A3.6.1-5.

Sawakuri, T., and Y. Yamamoto (1982), A plan for digital network performance objectives, ICC '82 Conference Record, Vol. 1, pp. 2D.6.1-2D.6.5.

Schafer, R.W., and J.D. Marken (1979), Speech Analysis (IEEE Press, IEEE, New York, NY), 469 pp.

Schmidt-Nielsen, A., and S. Everett (1982), A conversational test for comparing voice systems using working two-way communication links, IEEE Trans. Acoustics, Speech, Signal Process ASSP-30, No. 6, pp. 853-863, December.

Sen, T.K., and J.D. Carroll (1973), Signal-to-noise ratio as a predictor of speech transmission quality, IEEE Trans. Audio and Electroacoustics, pp. 384-387.

Sergeant, L., J.E. Atkinson, and P.G. Lacroix (1979), The NSMRL tri-word test of intelligibility (TTI), J. Acoust. Soc. Am. 65, pp. 218-222.

Spang, T.C. (1976), Low-noise echo study of the Direct Distance Dialing network, BSTJ 55, pp. 1-36, January.

Stern, T.E. (1983), A queueing analysis of packet voice, GLOBECOM '83 Conference Record, Vol., pp. 2.5.1-6.

Suda, T., Y. Yemini, H. Miyahara, and T. Hasegawa (1983), Performance evaluation of a packetized voice system--simultion study, ICC '83 Conference Record, Vol 2, pp. C5.2.1-5.

Terada, Y., and Y. Fukushima (1980), Transmission aspects on digital local exchange in analog environments, ITC '80 Conference Record, Vol. 2, pp. 35.3.1-35.3.5.

Un, C.K., and M.H. Cynn (1980), A performance comparison of ADPCM and ADM coders, NTC '80 Conference Record, Vol. 3, pp. 50.1.1-50.1.5.

Un, C.K., and J.R. Lee (1982), On spectral flattening techniques in residual-excited linear prediction vocoding, Proc. ICASSP 1982, Vol. 1, pp. 216-9.

Unrien, M., and M. Rault (1982), Errors and jitters performances of a digital network, ICC '82 Conference Record, Vol. 1, pp. 2D.4.1-5.

Voiers, W.D. (1967), Performance evaluation of speech processing devices, III. Diagnostic evaluation of speech intelligibility, Report No. SRRC-CR-67-6, Final Report (Sperry Rand Research Center, Sudbury, MA 01776), 73 pp.

Voiers, W.D. (1971), Present status of the diagnostic rhyme test, Presented at the meetings of the Groupment des Acousticiens de Langue Francaise, Groupe, "Communication Parlee," Aix-en-Provence, France.

Voiers, W.D. (1976), Methods of predicting user acceptance of voice communication systems, Final Report (Dynastat, Inc., 2704 Rio Grande, Austin, TX 78705).

Voiers, W.D. (1977), Diagnostic evaluation of speech intelligibility, (Dynastat Inc., Austin, Texas 78705).

Voiers, W.D. (1978), Speech acceptability evaluation of Dynastat: The diagnostic acceptability measure (DAM), (Dynastat Inc., 2704 Rio Grande, Austin, TX 78705), 11 pp.

Voiers, W.D. (1980), Interdependencies among measures of speech intelligibility and speech "quality", Proc. ICASSP '80, Vol. 3, pp. 703-5.

Voiers, W.D. (1982), Measurement of intrinsic deficiency in transmitted speech: The diagnostic discrimination test (DDT), Proc. ICASSP 1982, Vol. 2, pp. 1004-7.

Voiers, W.D., M.F. Cohen, and J. Mickunas (1965), Evaluation of Speech processing devices, I. Intelligibility, Quality, Speaker Recognizability, ARCRL-65-926 (Report for the Air Force Cambridge Research Labs, Bedford, MA), 200 pp.

von Alven, W.H. (1983), Designing telephone and data equipment for the new competitive environment, ICC '83 Conference Record, Vol. 1, pp. B3.3.1-5.

Xydeas, C.S., and C.C. Evci (1982), A comparative study of DPCM-AQF speech coders for bit rates of 16-32 kb/s, Proc, ICASSP '82, Vol. 3, pp. 1680-3.

Yoshikawa, M., M. Taka, I. Tokizawa, and T. Aoyama (1982), Network applications of new speech encoding technologies, ICC '82 Miami, Vol. 1, pp. A8.1.1-A8.1.5.

Zdunek, K., and L. Longley (1982), Comparison of speech digitization and transmission techniques using a subjective quality measure, ICC '82 Conference Record, Vol. 2, pp. 4G.3.1-6.

APPENDIX A: SUBJECTIVE MEASURES OF VOICE QUALITY

As noted by Goodman and Nash (1982) the quality of a speech communication link is not a well-defined parameter. There are almost as many definitions of subjective quality as there are people who measure it. Table A-1 presents factors that have been associated with speech quality by several different researchers. The table is thought to be representative of the different approaches that have been taken in an attempt to describe subjectively that illusive phenomenon called speech quality. As can be seen from the table, of the 11 factors listed, only intelligibility is common to all researchers listed. If additional researchers were included, undoubtedly the table would need to be expanded to include additional terms that describe other aspects of speech quality. While the table is limited in scope, the major point to be derived from it should be clear--there is not a commonly accepted definition of the term "quality of speech."

Some of the terminology associated with the evaluation of speech is inconsistent and even self-defining. For example, Voiers (1965) uses the word quality in reference to the "subjectively significant characteristics of speech." He then divides the subjective evaluation of speech processes into three categories: intelligibility, quality, and speaker recognizability. Thus the word quality is, in essence, used to define itself. Voiers also uses the words "quality" and "acceptability" interchangeably.

With all the differences in the definition of speech quality and the diversity of testing methods, it is not surprising that Goodman and Nash (1982) conclude that it is not possible, in general, to compare speech quality measurements from different laboratories. Yet the need for both subjective and objective speech quality measures continues to exist. The need, perhaps, is becoming greater with the trend toward all-digital networks and the use of lower bit rate codecs such as 32 kb/s adaptive differential pulse code modulation (ADPCM). It is very important to evaluate these new networks and voice digitization processes under a full range of operating conditions. This will require standardized methods of testing and the development of efficient test procedures. Although the development of standardized objective measures may be the ultimate objective, it is necessary first to start with a widely accepted definition of the speech quality terminology and how to test and evaluate the various aspects of speech quality.

The CCITT Study Group XII is commissioned to define such subjective methods. The methodology should be simple and reliable enough to allow comparison of

Table A-1. Subjective Aspects of Speech Quality Used by Representative Researchers

	Acceptability	Annoyance factors	Clarity	Fidelity	Intelligibility	Loudness	Naturalness	Pleasantness	Preference	Quality	Speaker Recognizability
Barnwell and Voiers	X				X			X			
IEEE					X	X			X		X
Nesenbergs et al.	X	X		X	X		X			X	X
Payne and McManamon			X	X	X	X	X				X
Rothauser					X	X			X		X
Voiers					X					X	X

various codec algorithms for various languages and various speakers. As noted by Maitre and Aozama (1982), a standardization of such testing methods is urgently needed. Nesenbergs et al. (1981) state that the continuous change in subjective testing demonstrates the lack of stability of the testing methods. They further state that no widely accepted, inexpensive, and easily administered and interpreted subjective tests are now available. They also express pessimism that such tests will appear in the near future.

A.1 Subjective Testing Concepts

Subjective testing procedures are based on drawing from a population of potential users and measuring the reactions of these subjects to the speech signal. These reactions must be quantified and then averaged, or processed, to arrive at a measure of user acceptance or preference. Much has been written about these basic testing philosophies (see Barnwell and Bush, 1977; Barnwell and Voiers, 1979; Rothauser and Urbanek, 1967; and IEEE, 1969, for example). The basic concepts are described in the following paragraphs.

Iso-Preference Testing

Iso-preference testing involves the use of a known, agreed-upon reference signal condition for use as a comparison in judging an unknown signal (Barnwell and Bush, 1977). The agreed-upon conditioning must be defined so that the test signal can be found to be equally acceptable as the reference signal. The iso-preference level of the speech test signal is defined as the signal-to-noise ratio of the speech reference signal at which preference votes of a listener group are equally divided.

Relative Reference Testing

Relative reference testing involves comparisons, done independently, with each of several reference conditions (Barnwell and Bush, 1977). The reference conditions are used to establish a scale of performance, and an unknown signal can be ranked on this scale. The subjective scale of the reference must be agreed upon a priori. The commonly used Mean Opinion Score (MOS), to be discussed later, is an example of relative reference testing.

Absolute Reference Testing

Absolute reference tests are those in which the subjects performing the test are required to give an absolute numerical evaluation to the properties described in the test format (Barnwell and Bush, 1977). The Diagnostic Acceptability Measure (DAM), to be described later, is an example of absolute reference testing.

Isometric Testing

In isometric testing the listener is required to provide a simple, direct, subjective evaluation of the acceptability of a sample speech transmission (Barnwell and Voiers, 1979). For example, he may be asked to assign a score from 1 to 100 for the overall acceptability of a speech signal. Isometric testing differs from absolute reference testing in that in the latter, the subject evaluates a number of properties of the signal in addition to the overall evaluation of the signal.

Parametric Testing

In parametric testing the subject is asked to make judgments with respect to specific features of the speech signal under consideration (Barnwell and Bush, 1977). This is done ideally without regard for the subject's personal affective reactions to these qualities. Hence, the parametric approach serves to reduce the sampling error associated with individual differences in taste (Barnwell and Voiers, 1979).

Category Judgment Method

In the category judgment method, the quality of a test signal is described by the listeners in terms of five response categories: excellent (5), good (4), fair (3), poor (2), and unsatisfactory (1). The test starts with a familiarization period during which the test signals are introduced to listeners, and one or two reference signals are presented of which the "correct" category evaluation is identified to the listeners. In the evaluation phase, the test signals are presented in a random order and the listeners are asked to choose the one category that corresponds to their quality impression of the presented speech signal. From the listeners' responses, a mean opinion score or cumulative preferences can now be calculated (Rothauser and Urbanek, 1967).

Absolute Preference Judgment

The absolute preference judgment method differs from the category judgment method in that listeners evaluate test signals in terms of numbers from 0 to 10 rather than in the five response categories (Rothauser and Urbanek, 1967).

There are numerous problems encountered with subjective testing. The results of the testing are very much dependent upon the ethnic and cultural origin of both the speaker and listener, the gender of the speaker and listener, and the test material used. The IEEE (1969) presents a list of phonemically balanced sentences for use in quality of speech evaluation.

Miller et al. (1951) list three classes of variables in articulation tests, namely: personnel, test materials, and equipment. Personnel include both talkers and listeners; test materials include syllables, words, sentences, and continuous discourse; equipment includes the rooms, amplifiers, codecs, etc. Miller et al. discuss three aspects of the context in which a word might be heard:

1) context supplied by the knowledge that the test item is one of a small number of items

2) context supplied by the items that precede or follow a given item of a word or sentence

3) context supplied by the knowledge that the item is a repetition of the immediately preceding item.

All three kinds of context enable the listener to limit the range of alternatives from which he selects his response. For example, a word in a sentence must be one of a relatively few number of words that make the continuation both logically and grammatically correct.

Miller et al. provide curves that give the percent words correct vs. signal-to-noise (S/N) ratio for both words in isolation and words as part of a sentence. For example, at the 80% word recognition level, there is a 15 dB difference in the S/N level required for words spoken in isolation as compared to words spoken in sentences. The effect of words spoken in repetition is less dramatic. At the 80% word recognition level, there is about a 5 dB difference in required S/N level for a single occurrence as opposed to a triple repetition of the word.

A.2 Subjective Conversation Tests

Much of the subjective evaluation of speech quality reported in the literature is based solely on listening tests. Another order of magnitude of complexity is added when conducting conversation tests. The complexity of this type of testing is described in CCITT (1981, Supplement 2). The CCITT Recommendation P.74 recommended method for the evaluation of speech quality is the use of conversation tests while listening tests are described as supplemental methods that are permissible under certain conditions (CCITT, 1981). Listening-only tests may be misleading when assessing the effects of a factor, like circuit noise, when the magnitude of the degradation caused is substantial. In any case, the CCITT recommends sufficient comparison with the results from full conversation tests before the results from listening-only tests are accepted as being reliable. No specific methodology is prescribed by Recommendation P.74 for conversation tests. Rather, the method utilized by British Telecom as given in Supplement 2 of Volume V of the Yellow Book is referred to.

A method of conducting conversation tests is through the use of service observation procedures. Users are questioned as to their evaluation of the service at the completion of a normal telephone call. One disadvantage of this is that there is little control over the tests when using the service observation method. The AT&T Company has overcome this disadvantage through the use of a system called Sybil. According to this method, members of the staff of Bell Laboratories volunteer to allow a small proportion of their ordinary internal calls to be passed through special arrangements that modify the normal quality of transmission according to a test program. If a particular call has been treated, the volunteer is asked to vote by dialing one of a set of digits to indicate his opinion. In this way all results are recorded by the controlling computer and complete privacy is retained.

The CCITT Recommendations P.77 and E.125 provide methods for making telephone users' surveys. Questionnaires are provided in Recommendation E.125. Recommendation P.77 describes how the results of these surveys of customers making ordinary telephone calls are to be analyzed.

Other types of conversation tests are conducted in the laboratory, which require that both parties understand what information is being conveyed by the other party. For example, a stock brokerage scenario has been used in which one subject plays the role of the broker and the other plays the role of the client. British Telecom uses a conversation test in which subjects are required to describe the picture on a postcard. In any of these conversation tests, the evaluation criteria are related to how well and how easily the required information is conveyed from one party to the other and vice versa.

Richards and Barnes (1982) compare results between conversation tests conducted in the laboratory and a speech quality prediction model called TCAM (Telephone Connection Assessment Model). Laboratory experimental results agreed well with results predicted by TCAM.

Schmidt-Nielsen and Everett (1982) describe a conversational test that was developed by the Naval Research Laboratory. The test is designed to assess the usability of a system using a two-way communications task. The general format of the test consists of a communication task requiring an exchange of information between the participants, followed by an evaluation of the ease or difficulty of using the voice system.

A.3 Commonly Used Subjective Listening Tests

As noted by Nesenbergs et al. (1981), subjective testing methodologies have been continuously evolving over a number of years. In the following paragraphs we shall describe briefly some of the more commonly used subjective listening tests.

Mean Opinion Score (MOS)

The mean opinion score is one of the most commonly used subjective evaluation techniques. The MOS tests are listening-only tests in which the listener categorizes the speech quality into one of five categories: excellent (5), good (4), fair (3), poor (2), and bad or unsatisfactory (1). The numbers in parentheses are the numerical values the analyst associates with each of the subjective categories used by the listener. The MOS is simply the arithmetic mean of all such scores assigned by the several listeners of a given speech signal. Listening tests involving mean opinion scores are widely used in the evaluation of codecs (see Daumer and Cavanaugh, 1978; Daumer and Sullivan, 1982; and Svean, 1982, for example).

Diagnostic Rhyme Test (DRT)

The diagnostic rhyme test for speech intelligibility was initially developed for the U.S. Air Force Electronic Systems Division by Dynastat, Inc. In recent years the DRT has emerged as a Department of Defense standard for testing the intelligibility of voice communications terminals (Meister, 1978). The DRT is a test for intelligibility (rather than user acceptance, or quality) of voice systems. Although related, it is important to note the distinction between intelligibility and quality. It is possible for a voice system to be perfectly intelligible and yet have an unnatural sound or not permit speaker recognition. Intelligibility is a requirement for all voice systems. Naturalness and speaker recognition may or may not be a requirement depending upon the application. For example, speaker recognition is not a requirement in air traffic control communication systems, but it is a requirement in public telephone systems. The DRT is a measure of the intelligibility of a voice system while the diagnostic acceptability measure (DAM), to be discussed later, is a measure of the quality (including naturalness and recognizability) of a voice system.

The DRT is structured to allow a detailed analysis of the ability of a system to reproduce certain classes of phonemes. These classes are characterized by the following speech attributes: voicing, nasality, sustention, sibilation, graveness, and compactness. The test provides a means of measuring the performance of the voice system for each state, present or absent, of each of these six attributes as well as total intelligibility. Using these results, specific weaknesses in voice systems can be pinpointed.

The DRT test material consists of a set of pairs of words that rhyme. The word pairs have been carefully selected to measure one or more of the six voice attributes. In diagnostic rhyme tests, the listeners are required to select which of the words were spoken. The test consists of 232 words uttered by each of two speakers. The listener is given a test booklet and is asked to strike out which word or word pair was actually spoken.

Details of the DRT may be found in Meister (1978), Voiers et al. (1965), and Voiers (1967). Examples of applications of the DRT may be found in Belfield (1977) and Grahn et al. (1978). Voiers (1980) discusses the interdependence of speech intelligibility using the DRT and speech "quality" using the DAM. There are other rhyme tests that have been developed. Examples are the Fairbanks Rhyme Test (FRT) (Fairbanks, 1958) and the modified rhyme test (MRT) (House et al., 1965).

Diagnostic Acceptability Measure (DAM)

The diagnostic acceptability measure is a measure of the "quality" of a speech signal. Here quality refers to the user's perception of the speech signal itself, the background, and the total effect. Table A-2 lists the perceived acoustic traits associated with each of the three parts of the DAM. The 20 items

Table A-2. Perceived Acoustic Traits Used in DAM

<u>The Speech Signal</u>

Fluttering
Twittering - Pulsating

Muffled
Smothered - Low

Distant
Small - Compact

Rasping
Scraping - Grating

Thin
Tinny - High

Unnatural
Mechanical - Lifeless

Babbling
Chortling - Slobbering

Irregular
Spasmodic - Fitful

Nasal
Whining - Droning

Interrupted
Intermittent - Chopped

<u>The Background</u>

Hissing
Simmering - Fizzing

Chirping
Cheeping - Clicking

Roaring
Rushing - Gushing

Crackling
Scratching - Staticky

Buzzing
Humming - Whirring

Rumbling
Thumping - Thudding

Bubbling
Gurgling - Percolating

<u>The Total Effect</u>

Intelligible
Understandable - Meaningful

Pleasant
Rich - Mellow

Acceptable

listed in the table are each evaluated using a 100-point rating scale. Ten of the items are concerned with the acceptability-related perceptual qualities of the speech signal. Seven items are concerned with the perceptual qualities of the background. Three items are concerned with the perceived intelligibility, pleasantness, and overall acceptability of the total effect.

The DAM evolved from earlier quality evaluation tools developed by Dynastat, Inc., under contract with the Defense Communications Agency. The earlier tools were the Paired Acceptability Rating Method (PARM) (Voiers, 1976; Grahn et al., 1978) and the Quality Acceptance Rating Test (QUART) (Voiers, 1976; Belfield, 1977).

Details of the DAM may be found in Voiers (1977). An example of the use of the DAM may be found in Zdunek and Longley (1982).

Consonant Recognition Test (CRT)

In the CRT a jury of 10 individuals is asked to write on an answer sheet the initial consonants of a series of consonant-vowel-consonant monosyllablic words that have been played through the system under test (Hanson, 1971). There are 10 lists of such words and there are 36 words to each list. The voices of nine different speakers are interleaved in the reading of each list to overcome any bias a particular system may have for certain speakers.

Degradation Category Rating (DCR)

The DCR measurements are described by Combescure et al. (1982). The main features of the DCR method follows. The test is presented to listeners as pairs (A-B) of sentences or repeated pairs of sentences (A-B-A-B) where A is a reference sentence and B is the same sentence produced by the codec. The purpose of the reference sentence is to anchor each judgment of listeners. In the DCR procedure, the reference is a high-quality speech signal of the same bandwidth as that of the codec being tested.

A five-point scale is defined as follows:

5 - degradation is inaudible
4 - degradation is audible but not annoying
3 - degradation is slightly annoying
2 - degradation is annoying
1 - degradation is very annoying

Each codec is evaluated by means of judgments upon four talkers reading two sentences. At least two male and two female talkers are required because the quality of codecs is speaker-dependent. The sentences used are selected from a list of 10 phonetically balanced sentences.

Combescure et al. (1982) conclude that the DCR results in higher sensitivity than the IEEE Absolute Category Rating (ACR) test.

Diagnostic Discrimination Test (DDT)

Conventional intelligibility tests do not normally permit a distinction to be made between intrinsic deficiencies of speech processing systems in which irrevocable loss of useful information occurs and cosmetic deficiencies in which there is simply a failure to render the transmitted speech into a perceptually usable form (Voiers, 1982). The DDT offers one remedy for this situation. It draws from the same speech materials as the Diagnostic Rhyme Test, but requires the listener simply to judge whether the members of pairs of test words are the same or different. A correct response thus indicates that information regarding the state of one of six distinctive features (voicing, nasality, sustention, sibilation, graveness, and compactness), even when the test words are not recognizable as such. The DDT does not require a recognition response from the listener. It requires only that he judge whether two utterances, which may or may not differ with respect to the state of a single distinctive feature, are the same or different words. The listener is not required to identify the words--only state whether they are the same word or different.

Diagnostic Communicability Test (DCT)

The Diagnostic Communicability Test (Schmidt-Nielsen and Everett, 1982) was developed by Dynastat, Inc., and is based on a stock-trading game. The set of stocks assigned to each person varies from game to game, so that the same task can be reused indefinitely. This makes it possible to train and maintain a test crew with relatively stable performance, which increases the comparability of tests conducted at different times. The rules for trading are highly structured, and once the game is learned, task difficulty does not vary from game to game but is relatively unaffected by differences in skill or ability.

The test uses a crew of five trained participants who play the game for about 5 minutes after which each player rates the system on a questionnaire having 15 rating scales. The choice of the number of participants (5) and the rules for trading have been optimized to make maximum use of the communication channel and the participants' time in evaluating the system. The need for five participants does require conferencing capability in setting up the tests and limits the situations in which the test can be used. It is possible to conduct the test with fewer than five participants, but the stock game becomes uninteresting with only three people and insufferably boring with two. The use of multiple rating scales provides more information about the performance of the system than a single scale would. Communicability scales include such attributes as difficulty in hearing, understanding, and recognizing other talkers as well as background interference. Compensatory behaviors such as talking more carefully, louder, or slower are assessed; and personal reactions include effort, irritation, fatigue, and acceptability.

Loudness Preference and Listening Effort Tests

Two subjective criteria that are commonly used are loudness preference and listening effort (CCITT, 1981) for which the following scales are used:

- Loudness preference scale:

  A - much louder than preferred
  B - louder than preferred
  C - preferred
  D - quieter than preferred
  E - much quieter than preferred

- Listening effort scale:

  A - complete relaxation possible; no effort required
  B - attention necessary; no appreciable effort required
  C - moderate effort required
  D - considerable effort required
  E - no meaning understood with any feasible effort.

A.4 References

Barnwell, T.P., and A.M. Bush (1977) Speech quality measurement, produced for the Defense Communication Agency under contract F-30602-75-C-0118. Available from NTIS by Order No. ADA058833.

Barnwell, T.P., and W.D. Voiers (1979) An analysis of objective measures for user acceptance of voice communications systems, Report prepared for Defense Communications Agency by Dynastat, Inc., No. DA100-78-C-0003.

Belfield, W. (1977), Performance measurements of the CVSD speech coding algorithm, Defense Communication Engineering Center TN6-77. Available from NTIS by Order No. ADA050163.

CCITT (1981), Telephone transmission quality, Recommendations of the P series, Yellow Book, Vol. V, Geneva.

Combescure, P., A. Le Guyader, and A. Gilloire (1982), Quality evaluation of 32 kb/s coded speech by means of degradation category ratings, Intl. Conf. on Acous., Speech, and Sig. Processing, pp. 988-991, Paris, April.

Daumer, W.R., and J.R. Cavanaugh (1978), A subjective comparison of selected digital codecs for speech, BSTJ 57, pp. 3119-3165, November.

Daumer, W.R., and J.L. Sullivan (1982), Subjective quality of several 9.6-32 kb/s speech coders, Intl. Conf. on Acous., Speech, and Sig. Processing, pp. 1709-1703, Paris, April.

Fairbanks, G. (1958), Test of phonemic differentiation: the rhyme test, J. Acous. Soc. Am. 30, pp. 596-600.

Goodman, D.J., and R.D. Nash (1982), Subjective quality of the same speech transmission conditions in seven different countries, Intl. Conf. on Acous., Speech and Sig. Processing, pp. 984-987, Paris, April.

Grahn, N.D., J.W. Vest, and W.L. Belfield (1978), Wideband test and evaluation final report, Defense Communications Engineering Center TN 8-78.

Hanson, E.D. (1971), Intelligibility testing of the continuous variable slope delta (CVSD) coder-decoder (CODEC), U.S. Army Electronics Command Tech. Report ECOM-3393, Ft. Monmouth, NJ.

House, A.S., C.E. Williams, H.L. Hecker, and K.D. Kryter (1965), Articulation test methods: consonantal differentiation with closed response set, J. Acous. Soc. Am. 37, pp. 158-166.

IEEE (1969), IEEE Recommended practice for speech quality measurements, IEEE No. 297, June.

Maitre, X., and T. Aozama (1982), Speech coding activities within CCITT: status and trends, Intl. Conf. on Acous., Speech, and Sig. Processing, pp. 954-959, Paris, April.

Meister, S. (1978), The diagnostic rhyme test (DRT): an Air Force implementation, Rome Air Development Center Report TR-78-129. Available from NTIS by Order No. AD-A060.

Miller, G.A., G.A. Heise, and W. Lichten (1951), The intelligibility of speech as a function of the context of test materials, Experimental Psychology 41, pp. 329-335.

Nesenbergs, M., W.J. Hartman, and R.F. Linfield (1981), Performance parameters for digital and analog service modes, NTIA Report 81-57. Available from NTIS by Order No. PB81-176455.

Payne, J., and P. McManamon (1974), Speech quality measurement, presented at SETE Workshop on Automatic Test and Checkout of Comm. Sys., June.

Richards, D.L., and G.P. Barnes (1982), Estimating communication efficiency of telephone connections, Intl. Conf. on Comm., Paper No. 61, Philadelphia, June.

Rothauser, E.H., and G.E. Urbanek, (1967), Speech quality measurements, Institut Für Niederfrequenztechnik, Vienna, Austria, Available from NTIS by Order No. AD682334.

Schmidt-Nielsen, A., and S. Everett (1982), A conversational test for comparing voice systems using working two-way communication links, IEEE Trans. Acous., Speech, and Sig. Processing ASSP-30, No. 6, pp. 853-863, December.

Svean, J. (1982), Quality measurements on speech coders for mobile radio, Proceedings of Intl. Conf. on Acous., Speech, and Sig. Processing, pp. 1700-1703, Paris, April.

Voiers, W.D. (1965), Performance evaluation of speech processing devices, Vol. II The role of individual differences, prepared for AF Cambridge Research Laboratories under contract AF19(628)-4987, Report No. AFCRL-66-24.

Voiers, W.D. (1967), Performance evaluation of speech processing devices, Vol. III: diagnostic evaluation of speech intelligibility, prepared for AF Cambridge Research Laboratories under contract AF19(628)-4987. Available from NTIS by Order No. AD 650158.

Voiers, W.D. (1976), Methods of predicting user acceptance of voice communication systems, final report (Dynastat, Inc., 2704 Rio Grande, Austin, TX 78705).

Voiers, W.D. (1977), Diagnostic evaluation of speech intelligibility, Speech Intelligibility and Recognition, M.E. Hawley (ed.) (Dowden, Hutchinson and Ross; Stroudsberg, PA).

Voiers, W.D. (1980), Interdependencies among measures of speech intelligibility and speech quality, Proceedings of Intl. Conf. on Acous., Speech, and Sig. Processing, pp. 703-705, Denver, April.

Voiers, W.D. (1982), Measurement of intrinsic deficiency in transmitted speech: the diagnostic discrimination test (DDT), Proceedings of Intl. Conf. on Acous., Speech, and Sig. Processing, pp. 1004-1007, Paris, April.

Voiers, W.D., M.F. Cohen, and J. Mickunos (1965), Evaluation of speech processing devices, Vol. I: Intelligibility, quality, speaker recognizability, prepared for AF Cambridge Research Laboratories under contract AF19(628)-4195, available from NTIS by Order No. AD 627320.

Zdunek, N., and L. Longley (1982), Comparison of speech digitization and transmission techniques using a subjective quality measure, Intl. Conf. on Comm., paper 4.6.3, Philadelphia, June.

APPENDIX B: OBJECTIVE MEASURES OF SPEECH QUALITY

A vast amount of research has been conducted over the last decade or more in an attempt to develop an objective measure of speech quality that has a high degree of correlation with subjective measures such as those discussed in Appendix A. It is not our intent in this brief appendix to summarize all of the research that has been conducted in this field. Such an effort would result in a major report in itself. Rather, it is our intent in Appendix B to validate and support the statement made in the body of this report that there remains an unsatisfied requirement for an objective measure of speech signals.

Maitre and Aoyama (1982) state the following:

> "To our knowledge, no well agreed objective method presently enables (one) to evaluate and compare precisely the quality of speech of a large number of various coding algorithms, even if objective methods are under study within CCITT Study Group XII."

They further state in regard to the CCITT efforts to define new standards for low bit-rate digital voice that:

> "It is only when reliable quality evaluation methods for speech and other voice-band signals are defined, that it will be possible to compare various (codec) algorithms and to choose the best one for each application. However a quick CCITT (codec) standardization is necessary, to avoid the proliferation of noncompatible systems, which would finally limit the potential use of these techniques in telecommunication networks."

Extensive research in the area of objective speech measurement techniques, speech signal processing, and speech codecs has been conducted by T.P. Barnwell and his colleagues at the Georgia Institute of Technology (see Barnwell and Quackenbush, 1982; Barnwell, 1980a; Barnwell, 1980b; Marr and Barnwell, 1980; Hodges et al., 1980; Papamichalis and Barnwell, 1980; and Barnwell and Voiers, 1979). Barnwell and Quackenbush (1982) made a comparative evaluation of six of the most commonly used objective measures. They concluded that:

> "... none of the objective measures performed very well."

The crux of the issue is that there is an immediate need for an objective voice measure to 1) support CCITT low-bit-rate standardization activities as noted by Maitre and Aoyama, and 2) support IEEE standardization activities related to voice-band quality of service. Yet, as concluded by Barnwell and Quackenbush (1982) no objective voice measure currently exists.

B.1 Comparison of Objective Speech Quality Measures

Many objective measures of speech quality have been developed over the last decade or more. One of the objectives of the research reported by Barnwell and Quackenbush (1982) is "to use the techniques for evaluating objective measures to design new and better objective measures." One can, therefore, expect the list of postulated objective measures to grow.

Before one can compare objective measures, it is important to define what is expected of a satisfactory objective speech quality measure. Nesenbergs et al., (1981) discuss the requirements as follows. Any system performance measure should satisfy a certain minimal set of criteria. These may be specified as follows. (1) Reliability: This specifies that anyone making the same measurement at different times should get the same answer. (2) Repeatability: This is the property that a measurement be specified sufficiently so that measurements made by one group can be repeated or verified by another group. (3) Usability: This property involves several factors such as time, cost, and complexity. The criteria that we strive for are short test time requirements (e.g., 1 hour for a single measurement, including set up, calibration, etc., with shorter time requirements for additional measurements), low recurring costs, and simplicity (e.g., only one instrument needed for all system inputs and one needed for all output measurements). The instrument may be internally complex, but the use of it should be simple. (4) System independence: the parameters chosen should be as widely applicable as possible, and therefore not dependent on the type of system being tested. However, the interpretation of the parameter values may (and probably will) depend on a priori knowledge of the system. (5) User oriented: We list this last, because it depends greatly on the users' experience with their system, the expected service, and with the measurements chosen, but the performance measure must provide significant correlation with subjective measures including various forms of intelligibility, speaker recognition, and other voice transmission quality features.

System independence has been a difficult objective to accomplish. Some objective measure that may be satisfactory for the evaluation of one codec may be entirely inappropriate for the evaluation of another. Maitre and Aoyama (1982) state the "objective measures, such as signal-to-noise ratio, classically used to evaluate PCM quality are no longer valid to evaluate the quality for speech of more sophisticated schemes such as delta modulation or differential coding, due to the sensitivity of performance to each speaker and due to some masking effect of the quantizing noise."

Barnwell and Bush (1977) made a comparison of several objective measures. They reported a two-part experimental study of the relationship between a number of objective measures and the subjective acceptability measures obtained from the Paired Acceptability Rating Method (PARM). In the first part of the study, controlled distortions were applied to speech samples in order to measure the resolving power of the candidate objective measures on these types of distortions. In the second part, the candidate objective measures were applied to speech samples from the same systems on which PARM tests were run, and the statistical correlation between the subjective and objective measures were studied. Objective measures examined included spectral distance measures (several linear predictive coding (LPC) based spectral distances, LPC error power ratio, and cepstral distance measures), pitch comparison measures, and noise power measures. Controlled distortions were formant bandwidth, frequency, pitch, low-pass bandwidth, and additive noise. Correlations with subjective test data ranged from approximately 0.2 to 0.8. Barnwell and Bush (1977) conclude that a number of objective measures, particularly spectral distance metrics, offer considerable promise in predicting subjective quality results, but that none provides adequate performance at this time.

Another comparison of objective measures is provided by Barnwell and Voiers (1979). In this study, correlation analyses were made between several objective measures and the DAM subjective measure. The best spectral-distance measures, noise measures, parametric-distance measures, frequency-dependent measures, and composite-distance measures were selected from those measures tested.

Barnwell's work is further documented in several papers presented at the International Conference on Acoustics, Speech, and Signal Processing (Barnwell, 1980a; Barnwell, 1980b; Barnwell and Quackenbush, 1982). As described by Barnwell and Quackenbush (1982), the four objectives of their test program are:

1) design of a general test procedure for estimating the performance of objective speech quality measures

2) comparative study of 1500 parametric variations of commonly used objective speech quality measures

3) design of a new complex objective speech quality measures

4) development of techniques for the design of improved speech codecs.

The general conclusion reached by Barnwell and Quackenbush (1982) is that "none of the objective measures tested performed very well, and, by inference, most of the measures in current usage exhibit this same general performance flaw."

Nesenbergs et al. (1981) investigated various speech quality measures. The principal user-oriented performance characteristics they considered were (1) intelligibility, (2) speaker recognition, and (3) user acceptability. For intelligibility testing they recommended the following four objective measures:

(1) The normalized energy measure developed by Gamauf and Hartman (1977), because it shows good correlation with subjective scores over a wide range of system conditions. This is a measure of a modification of the energy ratio.

(2) A short-term signal-to-noise ratio (Barnwell and Voiers, 1979), because it shows excellent correlation with subjective scores over certain limited ranges of system conditions.

(3) A band-weighted signal-to-noise ratio (e.g., Steeneken and Houtgast, 1979), because it shows good correlation with subjective scores over the range of system conditions for which it has been tested.

(4) The log-area ratio (Barnwell and Voiers, 1979) because it is easily computed from parameters derived when calculating the normalized density and has shown fair correlation with subjective scores.

For voice recognition, Nesenbergs et al. (1981) found few subjective measurements for comparison with objective scores, and noted further that no objective measures have been developed. The research area of computer recognition and verification of human voice is currently a fertile area of research and is applicable to the problem of the development of objective measures of speech quality. A review of research in the area of computer recognition of speech is outside the scope of this report, however.

Nesenbergs et al. also found that, for user acceptance, the measures given for intelligibility have also been shown to be correlated with subjective parameters, which have been shown to be correlated with user acceptance (Barnwell and Voiers, 1979; Voiers, 1976) over a very limited set of conditions.

The <u>three parameters</u>, <u>intelligibility</u>, <u>voice recognition</u>, and <u>acceptability</u>, chosen as voice performance descriptors, <u>are not independent</u>. Thus, a system with low intelligibility usually would not have good voice recognition properties and would not be acceptable. On the other hand, a system may be unacceptable because of long delays in the access phase or the transfer phase even though the intelligibility and voice recognition are outstanding. The objective measures are chosen to quantify these parameters. Table B-1 shows the relationship of objective measures to the subjective measures. The primary application means that the objective measure alone is a good predictor of the subjective parameter, while the secondary application implies that other measures are also needed to predict accurately the subjective interpretation.

Subsequent to the report by Nesenbergs et al. (1981), which was based on a review of a limited amount of research involving comparisons of objective and subjective speech quality measures, there have been a few other papers in the literature that report such comparisons (see for example, Barnwell and Quackenbush, 1982, and Viswanathan et al., 1983). The work by Barnwell was discussed earlier. Results from Viswanathan's testing were incomplete as of the time of his paper (1983).

B.2 Summary of Objective Speech Quality Performance Measures

Descriptions of the most widely used objective speech quality performance measures may be found in Nesenbergs et al. (1981) and Barnwell and Voiers (1979). There will be no attempt to repeat these descriptions in this section. Table B-2 is provided as a guide to the interested reader who is seeking detailed information on the numerous objective measures that have been developed over the years. Specific references are given for each of 18 objective measures. It is not claimed that either the list of objective measures or the associated list of references is complete. Rather, the intent of the table is to serve as a preliminary guide to those needing additional information on the most widely used objective measures.

Table B-1. Principal Application of Objective Measures to Subjective Interpretation

Objective Measure \ Subjective Interpretation	Intelligibility	Acceptance	Speaker Recognition
Normalized Energy	1	2	*
Log Area Ratios	1	2	*
Short Term S/N	1	2	*
Band Weighted S/N	1	2	*
Speaker Recognition	---	---	1

(1) Primary Application
(2) Secondary Application
* Applicability not known

Table B-2. Objective Speech Quality Measures

Category	References
Short-Term Signal-to-Noise Ratio	Barnwell and Voiers (1979) Nesenbergs et al. (1981) Kitawaki et al. (1982) Zdunek and Longley (1982) Viswanathan et al. (1983) Goodman and Sundberg (1982)
Band-Weighted Signal-to-Noise Ratio and Articulation Index	Zdunek and Longley (1982) Ottinger (1978) ANSI (1969) CCITT (1981) O'Brien and Busch (1969) Payne and McManamon (1973) Payne and McManamon (1974) Hubbard and Payne (1974) Hartman and Boll (1976) Gamauf and Hartman (1977) Steeneken and Houtgast (1979) Nesenbergs et al. (1981) Barnwell and Voiers (1979)
Signal Plus Noise and Distortion to Noise and Distortion (SINAD)	Zdunek and Longley (1982)
Signal-to-Noise Ratio	Kitawaki et al. (1982) Zdunek and Longley (1982) Barnwell (1980a)
Normalized Energy Measure	Gamauf and Hartman (1977) Nesenbergs et al. (1981) Hartman and Pratt (1980)
Total Harmonics Distortion	Zdunek and Longley (1982)

Table B-2. (continued)

Category	References
Transient Intermodulation Distortion	Zdunek and Longley (1982)
Signal Transmission Index	Zdunek and Longley (1982) Steeneken and Houtgast (1979)
Log-Area Ratios (LAR)	Barnwell and Voiers (1979) Nesenbergs et al. (1981) Viswanathan et al. (1983)
Spectral Distortion	Kitawaki et al. (1982)
LPC Cepstrum Distance	Kitawaki et al. (1982)
COSH	Kitawaki et al. (1982)
Likelihood Ratio	Kitawaki et al. (1982) Viswanathan et al. (1983)
Weighted Likelihood Ratio	Kitawaki et al. (1982)
Reflection Coefficient Measure (RFC)	Viswanathan et al. (1983)
Linear Spectral Distance (LSD)	Viswanathan et al. (1983) Barnwell (1980a) Barnwell and Voiers (1979)
Spectral Distance (frequency variant and invariant)	Barnwell (1980a) Viswanathan et al. (1983) Lacroix and Makia (1982)
Composite Measures	Barnwell and Voiers (1979)

B.3 References

ANSI (1969), American National Standard Methods for the calculation of the articulation index, ANSI S3.5-1969.

Barnwell, T.O. (1980a), A comparison of parametrically different objective speech quality measures using correlation analysis with subjective quality results, Intl. Conf. on Acous., Speech, and Sig. Process., pp. 710-713, Denver, April.

Barnwell, T.P. (1980b), Correlation analysis of subjective and objective measures for speech quality, Intl. Conf. on Acous., Speech, and Sig. Process., pp. 706-709, Denver, April.

Barnwell, T.P., and A.M. Bush (1977), Speech quality measurement, Georgia Institute of Technology Report. Available from NTIS by Order No. ADA058833.

Barnwell, T.P., and S.R. Quackenbush (1982), An analysis of objectively computed measures for speech quality testing, Intl. Conf. on Acous., Speech, and Sig. Process., pp. 996-999, Paris, April.

Barnwell, T.P., and W.D. Voiers (1979), An analysis of objective measures for user acceptance of voice communications systems, Georgia Institute of Technology report prepared for Defense Communications Engineering Center under contract DA100-78-C-0003.

CCITT (1981), Telephone transmission quality, Recommendations of the P Series, Yellow Book, Vol. V, Geneva.

Gamauf, K.J., and W.J. Hartman (1977), Objective measurement of voice channel intelligibility, Report No. FAA-RD-77-153, prepared by U.S. Dept. of Commerce, Office of Telecommunications, for Federal Aviation Admin., Washington, DC 20590.

Goodman, D.J., and C.E. Sundberg (1982), Combined source and channel coding for matching the speech transmission rate to the quality of the channel, Global Comm. Conf. Record, paper B.3.2.

Hartman, W.J., and S.F. Boll (1976), Voice channel objective evaluation using linear predictive coding, Report FAA-RD-75-189, prepared by U.S. Dept. of Commerce, Office of Telecommunications, for Federal Aviation Admin., Washington, DC 20590.

Hartman, W.J., and L.E. Pratt (1980), Voice performance measurements, Report FAA-RD-80-71, prepared by U.S. Dept. of Commerce, National Telecommunications and Information Administration, for Federal Aviation Admin., Washington, DC 20590.

Hodges, C.J.M., T.P. Barnwell, and D. McWhorter (1980), The implementation of an all digital speech synthesizer using a multimicroprocessor architecture, Intl. Conf. on Acous., Speech, and Sig. Process., pp. 855-858.

Hubbard, R.W., and J.A. Payne (1974), Computer methods for the objective evaluation of speech communication systems, Office of Telecommunications Report 74-29 (NTIS Order No. COM 75-10815/AS).

Kitawaki, N., K. Itoh, M. Honda, and K. Kakehi (1982), Comparison of objective speech quality measures for voiceband codecs, Intl. Conf. on Acous., Speech, and Sig. Process., pp. 1000-1003, Paris, April.

Lacroix, A., and B. Makia (1982), A vocoder scheme for very low bit rates (quality evaluation), Intl. Conf. on Acous., Speech, and Sig. Process., pp. 618-621, Paris, April.

Maitre, X., and T. Aoyama (1982), Speech coding activities within CCITT: status and trends, Intl. Conf. on Acous., Speech, and Sig. Proc., pp. 954-959, Paris, April.

Marr, J.D., and T.P. Barnwell (1980), Two-dimensional prediction of area functions for coding of LPC speech parameters, Intl. Conf. on Acous., Speech, and Sig. Process., pp. 11-14, Denver, April.

Nesenbergs, M., W.J. Hartman, and R.F. Linfield (1981), Performance parameters for digital and analog service modes, NTIA Report 81-57 (NTIS Order No. PB81-176455).

O'Brien, P.J., and A.C. Busch (1969), Effects of selective system parameters of communications intelligibility, Report No. NA-69-21, Federal Aviation Administration, Atlantic City, NJ 08405.

Ottinger, D.M. (1978), Objective measure of speech intelligibility, Air Force Institute of Technology report, available from NTIS by Order No. AD-A064-727.

Papamichalis, P.E., and T.P. Barnwell (1980), A dynamic programming approach to variable rate speech transmission, Intl. Conf. on Acous., Speech, and Sig. Process., pp. 28-31, Denver, April.

Payne, J.A., and P.M. McManamon (1973), An objective speech quality measurement of a communication channel, OT Report 73-14 (NTIS Order No. COM75-10940/AS).

Payne, J., and P.M. McManamon (1974), Speech quality measurement, presented at SETE Workshop on Automatic Test and Checkout of Comm. Sys., June.

Steeneken, H.J.M., and T. Houtgast (1979), A physical method for measuring speech-transmission quality, Institute for Perception Report 1979-1, (P.O. Box 23, 3769 ZG Soesterberg, Netherlands).

Viswanathan, V.R., W.H. Russel, and A.W.F. Huggins (1983), Objective speech quality evaluation of mediumband and narrowband real-time speech codecs, Intl. Conf. on Acous., Speech, and Sig. Process., paper no. 12.6 Boston, April.

Voiers, W.D. (1976), Methods of predicting user acceptance of voice communications systems, final report, (Dynastat, Inc., 2704 Rio Grande, Austin, TX 78705).

Zdunek, K., and L. Longley (1982), Comparison of speech digitization and transmission techniques using a subjective quality measure, Intl. Conf. on Comm., paper 4G.3, Philadelphia, April.

FORM **NTIA-29**
(4-80)

U.S. DEPARTMENT OF COMMERCE
NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

	1. PUBLICATION NO.	2. Gov't Accession No.	3. Recipient's Accession No.
	NTIA Report 85-188		

4. TITLE AND SUBTITLE
Voiceband Quality-of-Service Issues in the Post Divestiture Environment

5. Publication Date
December 1985

6. Performing Organization Code
NTIA/ITS

7. AUTHOR(S)
James A. Hoffmeyer

9. Project/Task/Work Unit No.
910-2103

8. PERFORMING ORGANIZATION NAME AND ADDRESS
National Telecommunications & Information Administration
Institute for Telecommunication Sciences
325 Broadway
Boulder, CO 80303

10. Contract/Grant No.

11. Sponsoring Organization Name and Address

12. Type of Report and Period Covered

13.

14. SUPPLEMENTARY NOTES

15. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

This report discusses quality-of-service (QOS) issues for telephone networks. Deregulation and divestiture have fostered increased competition in the United States in the telephone equipment and service industries. There are many economic, policy, and technical issues that remain to be solved as the result of the plethora of equipment and services now available. This report addresses the technical problems associated with the interconnection of equipment from many vendors. In order to maintain a satisfactory quality of service to the end user, performance standards must be developed, approved, and implemented. The work of IEEE, CCITT, and ANSI-accredited standards groups responsible for telephone QOS is reviewed. The problem of interconnecting different national networks in the international community is seen to be analogous, in part, to the problem of interconnecting the numerous public and private networks within the United States. Although progress has been made by both national and international telephone QOS standards groups, unsolved issues remain. Principal among these are the development of objective measures of voice quality, the mapping of these objective measures into five levels of quality, enhancement of IEEE and CCITT telephony QOS standards (including the development of standards for the transmission of data on voiceband networks), and the development of QOS standards for Integrated Services Digital Networks. These are discussed in this report along with recommendations for new programs that would contribute to their resolution.

16. Key Words *(Alphabetical order, separated by semicolons)*
competition; divestiture; objective quality evaluation; quality of service; subjective quality evaluation; telephone systems standards

17. AVAILABILITY STATEMENT	18. Security Class. *(This report)*	20. Number of pages
☒ UNLIMITED.	Unclassified	132
☐ FOR OFFICIAL DISTRIBUTION.	19. Security Class. *(This page)* Unclassified	21. Price:

☆U.S. GOVERNMENT PRINTING OFFICE:1986—676-002/40007