User-Oriented Measures of Telecommunication Quality

Objective, perception-based measures of telecommunication service quality are being standardized in national and international committees.

Neal Seitz, Stephen Wolf, Stephen Voran, and Randy Bloomfield

Over the past decade, advances in digital information transmission, switching, processing, and storage technologies have created the potential for a global information infrastructure that could empower enterprises and enrich societies worldwide. Paradoxically, the same technology advances have created a bewildering overabundance of product alternatives and an industry environment of unprecedented complexity, making timely realization of the infrastructure and its benefits problematical. With deregulation and the emergence of new telecommunication providers in many countries, national and international standards committees have assumed increased responsibility for the cooperative planning of new technology development -- and the matching of multi-vendor service offerings with user needs. One important focus of this effort has been the standardization of user-oriented, technology-independent measures of telecommunication service quality. The standardized measures will be used by providers in designing and implementing telecommunication systems and services, and by users in defining telecommunication requirements and selecting the products that most effectively meet them.

Standards committee work towards this end has progressed in three broad phases. In the first phase, participants defined the basic concepts that underline the user-oriented approach to telecommunication quality assessment. In the second phase, participants developed a set of generic user-oriented quality measures for call processing and data transfer functions, and applied these generic measures in deriving technology-specific performance parameters and measurement methods for packet-switched networks and integrated services digital networks (ISDNs). In the third phase, still in progress, participants are developing user-oriented quality measures for video and voice communications. User-perceivable video and voice quality impairments are being quantified by objective measures chosen for their correlation with carefully collected and numerically quantified human reactions to the transmitted images and sounds. Planned future work will extend and consolidate the work of all phases and define integrated measures of user-perceived quality for multimedia services.

Basic Concepts

Telecommunication services exist to fulfill the needs of users. It is therefore important to specify and measure the quality of telecommunication services using parameters that accurately and concisely express the user's satisfaction (or dissatisfaction) with the delivered service. Such parameters are described as *user oriented*, or when the user is a human, *perception based*. User-oriented quality of service parameters differ from the provider-oriented network performance parameters traditionally used in network design and operation both in where they are applied and in how they are defined.[1]

The user-oriented parameters are applied at *end-user interfaces* which are typically more inclusive than the jurisdictional or regulatory interfaces that separate network provider and customer

[1] The terms "quality of service" and "network performance" are used contextually in this paper to distinguish the user-oriented and provider-oriented perspectives and parameters.

premises equipment. In particular, when the user is a human, the end-user interface includes as much of the terminal as possible. Examples of human end-user interfaces include telephone handsets, video cameras and displays, and keyboards. When the user is non-human, as in the case of a computer application program that processes communicated information, the end-user interface is typically the functional interface between the application program and the associated communications platform. Applying the parameters at the end-user's interface ensures that the parameters are observable and relevant to the user.

End-users of the network services are not intrinsically interested in how the services are implemented (whether by circuit or packet switching, for example), with the network-internal causes of externally-observable service degradations, or with any aspect of the network's internal architecture or design. However, users are vitally interested in limiting the observable effects of network imperfections, and comparing network service alternatives, in terms of certain universal user performance concerns. If user-oriented parameters are defined in terms of generic reference events or signals that can be observed at any end-user interface, rather than in terms of the protocol-specific interface events or signal encodings used by a particular provider, the parameters can be made *technology independent*. Technology-independent parameters enable the end-user's quality of service concerns to be quantified in a uniform way irrespective of how a particular telecommunications capability is provided. Such parameters can thus be used as a "common denominator" in service comparison.[2]

Call Processing and Data Communication Quality

Early U.S. work on user-oriented quality of service standards was conducted in the American National Standards Institute (ANSI) accredited Information Systems Subcommittee on Data Communications (X3S3). This work led to the development of two American National Standards: ANSI X3.102-1983 [2] and ANSI X3.141-1987 [3]. Related international work was conducted in the International Telegraph and Telephone Consultative Committee (CCITT) Study Groups on Data Communications (Study Group VII) and Integrated Services Digital Networks (Study Group XVIII). This work produced two new CCITT Recommendations: X.140 in 1984 [4] and I.350 in 1988 [5]. Collectively, these four standards provided a foundation that has been very useful in subsequent, more specific performance assessment studies.

Development of the user-oriented measures began with a comprehensive survey of then-existing provider-oriented performance parameters. Over 100 provider-oriented parameters were identified, classified, and evaluated to establish an understanding of existing practice from which the more generic measures could be deduced. Candidate generic measures were evaluated on the basis of their relevance to the subjective perceptions of human users. The goal of technology independence led standards developers to seek an unambiguous, yet extremely general, definition of the basic telecommunication service functions; and a simple, direct way of relating the possible outcomes of any attempt to perform a discrete function to distinct (ideally orthogonal) user quality concerns. Efforts to fulfill these needs resulted in the development of a matrix framework that has, with minor refinements, become a principal means of addressing the problem of performance description in both U.S. and international standards committees. The matrix comprises three rows and three columns, as illustrated in Table 1. The three rows identify

[2] This can be done either for alternative services connecting a particular user pair or (by aggregation) for a service connecting any specified population of interest. User-oriented parameters may be employed in conjunction with pricing and other information to more comprehensively describe overall customer satisfaction with telecommunication services, including non-technical factors such as sales, provisioning, billing, operation, repair, and technical support [1].

three primary telecommunication functions: access, user information transfer, and disengagement. The three columns identify the three fundamental concerns, or criteria, that are of principal interest to users in assessing the performance of any function: speed, accuracy, and dependability. The three functions correspond to the three consecutive phases in a connection-oriented telecommunications session, but are also applicable to connectionless services.[3] The three criteria correspond, respectively, to the three mutually-exclusive and jointly exhaustive outcomes that may be experienced in any individual attempt to perform a discrete function: successful performance, incorrect performance, and nonperformance. This correspondence gives the criteria an important property of orthogonality.[4]

Table 1 Matrix Framework and Representative User-oriented Parameters

CRITERION FUNCTION	SPEED	ACCURACY	DEPENDABILITY
ACCESS	• ACCESS TIME	• INCORRECT ACCESS PROBABILITY	• ACCESS DENIAL PROBABILITY
USER INFORMATION TRANSFER	• BLOCK TRANSFER TIME • USER INFORMATION BIT TRANSFER RATE	• BLOCK ERROR PROBABILITY • BLOCK MISDELIVERY PROBABILITY	• BLOCK LOSS PROBABILITY
DISENGAGEMENT	• DISENGAGEMENT TIME	• DISENGAGEMENT FAILURE PROBABILITY	

The fundamental nature of the three functions and the orthogonality of the three criteria made the 3x3 matrix a useful framework in which to identify and define user-oriented quality of service parameters for call processing and data communications. Standards developers considered each function/criterion pair in succession, and identified one or more user-oriented parameters to describe service impairments wherever a particular combination of the two attributes was judged to be of distinct user concern. A representative subset of the user-oriented parameters is illustrated in the context of the matrix framework in Table 1.[5]

Any definition of a data communication quality of service parameter is an algorithm that describes exactly how certain specified information transfers should be counted, timed,

[3] Use of connectionless services frequently involves access to and disengagement from a network entity even though no end-to-end connection is established.

[4] The same basic criteria have been used in deriving performance parameters for other telecommunications functions (e.g., billing) and non-telecommunications functions (e.g., file access or mathematical computation in ADP).

[5] Bit-oriented user information transfer parameters were also defined to facilitate quality comparison. Ancillary parameters were defined to describe the influence of user performance delays on the primary “speed” parameter values. Availability parameters were added later along the lines described in [6].

compared, or otherwise processed to produce a parameter value. The specification of such algorithms is particularly challenging in the case of user-oriented parameters because the algorithm inputs cannot be described in terms of particular protocol-specific interface events. The alternative is to define the user-oriented parameters on the basis of abstract, technology-independent reference events -- but define those reference events in such a way that each can be associated with a corresponding technology-specific interface event at any user interface of interest. The technology-independent reference events are a useful abstraction that allows the performance analyst to relate technology-specific interface events to user concerns.

An example will help to clarify the nature of the technology-independent reference events. In any data communication transaction, the objective is to transport user-defined units of information (generically called “blocks”) between end users. In such transport, each block must first pass from the physical possession and control of the source user to that of the telecommunication system; and then must pass from the physical possession and control of the system to that of the destination user. The specific interface protocols and events that control and effect these block transfers are system dependent, but for any given protocol, it is always possible to identify some distinct pair of complementary interface events that correspond to the generic “start of block transfer” and “end of block transfer” reference events.

Many existing physical interfaces (and interface protocols) were considered in defining the technology-independent reference events to ensure that the user-oriented parameters would not be restricted in application. The reference events ultimately selected are described in the following section, and are more fully defined in conjunction with the user-oriented parameters in [2]-[5]. Examples of application of the user-oriented parameters to various types of interfaces are provided in [7].

Each of the parameters illustrated in Table 1 provides a means of describing how a particular telecommunication system limitation or failure is perceived by (and impacts) the user. As an example, the speed, accuracy, and dependability concerns associated with the access function are expressed, respectively, by the user-oriented parameters *access time, incorrect access probability,* and *access denial probability. Access time* describes the delay a user experiences, after requesting that a particular information transfer capability be established, before the information transfer is actually enabled to begin. *Access time* is normally perceived by the user as a source of inefficiency, since time spent waiting for access provides no benefit and cannot normally be used for other activities. Excessive access delay can also reduce user effectiveness through data aging. In essence, *access time* is a price users pay for the economic benefit of sharing an expensive resource with others. *Incorrect access probability* expresses the likelihood that a telecommunication system will transmit user information on an improper path as a result of a system error during the access process. Incorrect access can significantly impact both the effectiveness and the integrity of user operations. For example, it can cause a source user to believe that transmitted information has been delivered to the intended destination when in fact, it has not; it can also be an indirect cause of user information misdelivery. *Access denial probability* describes the likelihood that the telecommunication system will refuse, or “block,” a user’s access request. The negative consequences of access denial are similar to those of excessive access delay, but a series of access denials is more detrimental to the user than a single access delay of equivalent duration because of the additional effort required to initiate repeated call attempts. There is a definite buildup of dissatisfaction with repeated access denials in the case of human users.

Similar relationships between parameter values and user perceptions can be described for each of the other generic quality of service parameters. These functional relationships can be made quantitative for any given user application (or family of related applications), as illustrated in [2] and [7]. Experimental measurement programs in which the user-oriented parameters were applied in characterizing performance between various types of user/system interfaces are summarized in [8]-[11].

Application to Packet-Derived Technologies

Packet switching and its broadband derivative technologies (ATM and frame relay) offer substantial opportunities for network efficiency improvement through adaptive bandwidth allocation and network resource sharing -- but at a cost of much greater complexity in network planning and operation. In view of this complexity, it is not surprising that network performance studies have played a major role in the development, standardization, and implementation of the packet-switched and packet-derived technologies. The user-oriented quality of service parameters defined in the generic standards described earlier provided a useful basis for development of the packet-switched service performance descriptors and are proving to be of similar use in the specification, implementation, and optimization of emerging ATM and frame relay services.[6] Definition of technology-specific network performance parameters by specialization of corresponding technology-independent quality of service measures makes it much easier to establish quantitative relationships among them.

Figure 1 illustrates the protocol-specific reference model that was used in defining the packet-switched service performance parameters. The model divides an end-to-end virtual connection into basic sections whose boundaries are associated with X.25 or X.75 interfaces. Two general types of basic sections are identified: a packet network, delimited on each side by a three-layer X.25 (access) or X.75 (internetwork) protocol stack, and a physical circuit that connects each network to customer data terminal equipment (DTE) or an adjacent network. These elements are called “network sections” and “circuit sections,” respectively.

The reference model also defines a set of protocol-specific interface events. Each event corresponds to the transfer of a particular type of packet across the physical interface between a network section and a circuit section, and is defined on the basis of a corresponding X.25 or X.75 interface state transition. Figure 1 illustrates the two basic types of interface events and shows how each event type is to be time-stamped in a measurement.

The protocol-specific interface events were associated with corresponding technology-independent reference events as illustrated in Table 2. (The table addresses the special case of a single packet network bounded by X.25 interfaces; the X.75 events are similar.) This association of events specialized the user-oriented parameters to the X.25/X.75 interfaces, and in so doing, identified and defined a set of candidate packet-switched service performance parameters. The candidate parameters and their definitions were reviewed and refined by packet switching experts in the national and international standards committees with the results summarized in Figure 2. Five candidate access and disengagement parameters were adopted as packet virtual call “set-up” and “clearing” parameters. Two candidate user information transfer speed of service parameters were adopted as packet data “throughput” and “transfer delay” measures. Candidate measures of data error, misdelivery, and loss probability were combined in a single summary parameter, “residual error ratio.” Technology-specific parameters were added to allow a more precise description of performance associated with X.25/X.75 “reset” and “premature disconnect” events.[7] Packet-switched service availability parameters were defined separately and are described in [6].

[6] The relevant standards are presented in [12]-[20].

[7] Technology-specific reset and premature disconnect “stimulus” probability measures were defined in addition to the illustrated parameters to identify protocol violations.

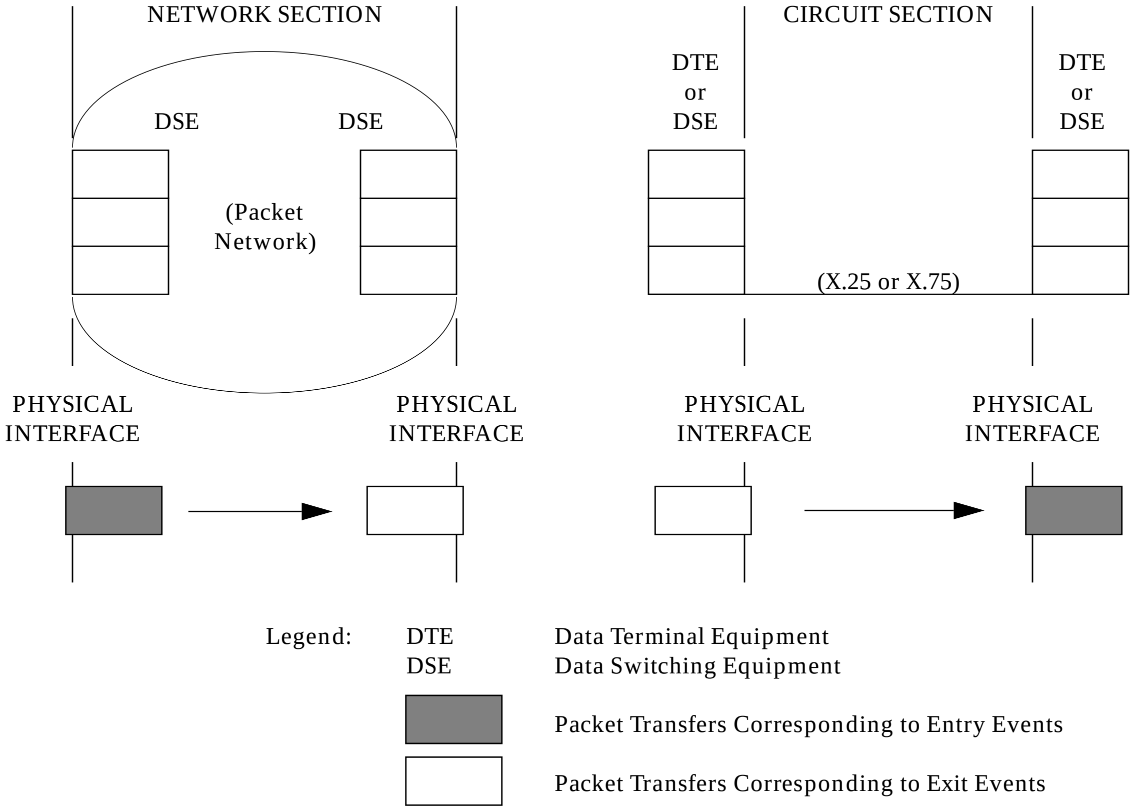
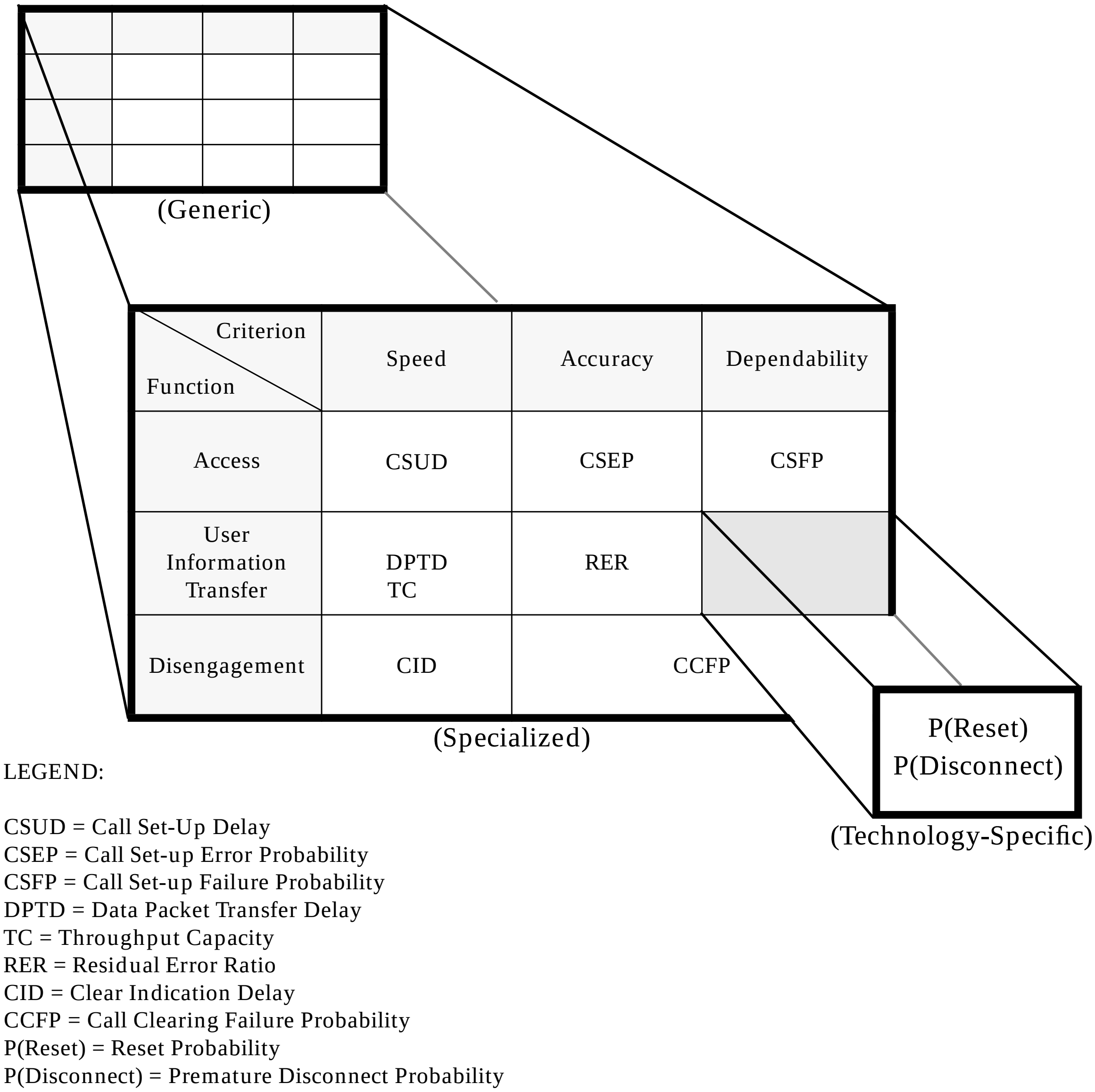


Figure 1 Interfaces and Events for Protocol-Specific Model used in Packet-Switched Service Performance Standardization

Table 2 Example Relationships between Protocol-Specific and Technology-Independent Events

Technology-Independent Event	X.25 Interface Event
Access Request	Call Request Packet Enters Calling Side DCE
Nonoriginating User Commitment	Call Accepted Packet Enters Called Side DCE
System Blocking Signal	Clear Indication Packet Exits Calling Side DCE (During Call Set-up)
User Blocking Signal	Clear Request Packet Enters Called Side DCE (During Call Set-up)
Start of Block Transfer	Data Packet Enters Source Side DCE
End of Block Transfer	Data Packet Exits Destination Side DCE
Disengagement Request	Clear Request Packet Enters Calling or Called Side DCE (During Data Transfer)
Disengagement Confirmation	Clear Confirmation Packet Enters Cleared Side DCE or Exits Clearing Side DCE



Criterion / Function	Speed	Accuracy	Dependability
Access	CSUD	CSEP	CSFP
User Information Transfer	DPTD TC	RER	
Disengagement	CID	CCFP	

LEGEND:

CSUD = Call Set-Up Delay
CSEP = Call Set-up Error Probability
CSFP = Call Set-up Failure Probability
DPTD = Data Packet Transfer Delay
TC = Throughput Capacity
RER = Residual Error Ratio
CID = Clear Indication Delay
CCFP = Call Clearing Failure Probability
P(Reset) = Reset Probability
P(Disconnect) = Premature Disconnect Probability

Figure 2 Evolution of the Packet-Switched Parameters

A similar approach has been followed in defining performance parameters for ATM cell transfer in Broadband ISDNs. A general reference model was defined to serve as a basis for ISDN performance description. The model identifies two particular types of physical interfaces (called measurement points, MPs) at which internationally standardized ISDN protocols may be observed: an MPT, associated with the CCITT-standardized “T reference point” at which terminal equipment is connected to an ISDN; and an MPI, associated with an international switching center that terminates a national ISDN. Cell transfer reference events (analogous to the X.25/X.75 interface events) were defined to identify performance-significant ATM cell transfers at the MPs. Standards participants familiar with the user-oriented performance parameters then defined four ATM user information cell transfer performance parameters by reference event association: cell transfer delay, cell error ratio, cell loss ratio, and cell misinsertion rate.[8]

[8] Three ATM-specific performance parameters were also defined: severely errored cell block ratio (SECBR) and two measures of cell delay variation (1-point and 2-point CDV).

Current Research on Video and Voice Quality Measures

A major focus of current work in telecommunication performance standardization is the development of user-oriented, technology-independent quality parameters for video and voice communication services. The call processing functions supporting switched video and voice services are similar to those supporting data services, so that the same access and disengagement parameters can be applied. However, this is not true in the case of user information transfer: there is no simple relationship between the values for data communication quality of service parameters and the perceived quality of digitally transmitted video and voice signals. Conventional measures of analog signal reproduction are also inapplicable to advanced digital video and voice services. Digital compression can introduce fundamentally different kinds of impairments than are created by traditional waveform reproduction. Examples of compression related impairments in video communications are “blocking,” image persistence, and jerky motion. Compression artifacts in voice communications include speech clipping and phonemic distortion. An additional complication arises from the fact that signal content plays a crucial role in determining the amount of compression that is possible, and the severity of compression artifacts. Thus, in general, the user-perceived quality of a digital video or voice transmission system is a function of both the system and the input signal.

The “truest” and most fundamental quality measures for digital video and voice services are the subjective responses of human viewers and listeners to the delivered images and sounds. Subjective viewing and listening tests attempt to quantify that “truth.” Unfortunately, subjective tests are impractical to use in all but a few limited situations. In addition, the repeatability and accuracy of subjective tests are difficult to quantify in a practical test program. Nonetheless, subjective tests remain the only viable reference point for validating objective measures.

The need for in-service measurements (based on the end-user’s information) that correlate with the perception of transmission quality has led telecommunications standards developers to seek new, fundamentally different video and voice quality measures that are objective but perception based. Such measures are derived from electrical and mathematical properties of the digitized input and output signals, and thus are implementable in automated test equipment; but they are chosen on the basis of their correlation with the subjective video or voice quality assessments of human users (e.g., viewer or listener panels). They achieve technology independence by recognizing important perceptual attributes and subsequently predicting human reactions to imperfections in received acoustic or visual information.

In the U.S., the video quality studies are being conducted in the ANSI accredited Telecommunications Working Group T1A1.5; the voice quality studies are being conducted in a companion group, T1A1.6. Related international studies are being conducted in ITU-T Study Groups 15 and 12, respectively.

Development Process

Figure 3 illustrates the process that is being used to develop and validate these new objective quality measures. Consistent with the need for in-service measurement, the video-voice source material is selected to be representative of the actual end-user’s applications. This source material is passed through relevant transmission systems. The resulting impaired destination material is evaluated in subjective tests. Objective measures, extracted from the digitized video and voice signals, are compared to the subjective test responses using statistical analysis techniques. Only objective measures that accurately predict the subjective responses over a suitable range of test conditions are considered for standardization.

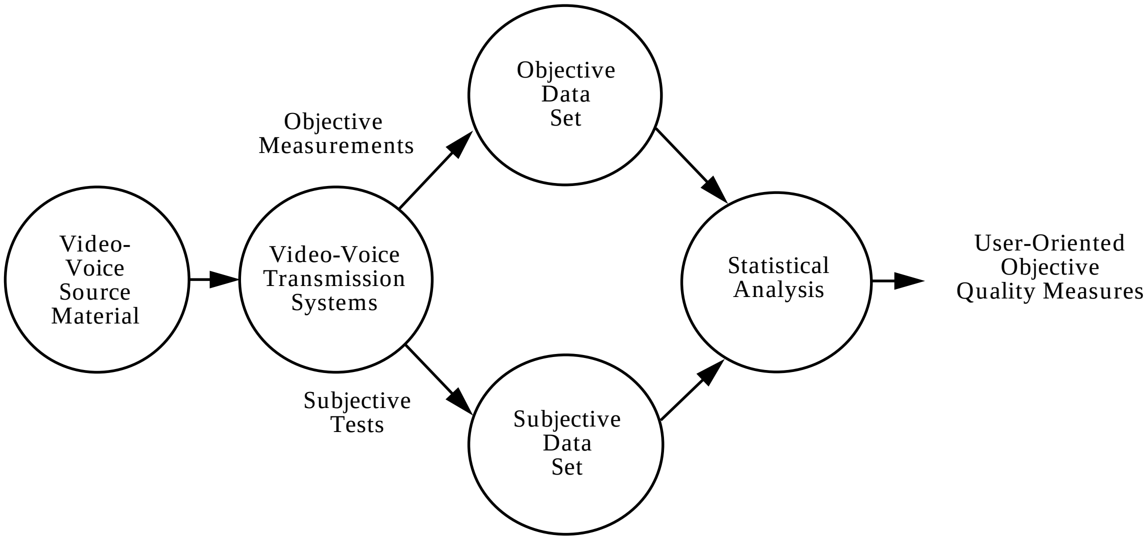

Video-Voice Source Material
Video-Voice Transmission Systems
Objective Measurements
Objective Data Set
Subjective Tests
Subjective Data Set
Statistical Analysis
User-Oriented Objective Quality Measures

Figure 3 Development Process for Perception-Based Quality Measures

Video Quality

In-Service Perception-Based Video Quality Measurement

The development process illustrated in Figure 3 has been used to develop and test a new set of technology-independent video quality measures that are applicable to digital as well as traditional analog video transmission systems. Figure 4 presents a conceptual block diagram of an in-service perception-based video quality measurement system that is being developed. The measurement system is composed of two sub-systems -- a source instrument and a destination instrument. The source instrument attaches non-intrusively to the source video and extracts a set of source features that can be used as a reference to quantify perceptual video quality changes. The destination instrument attaches non-intrusively to the destination video and extracts an identical set of destination features. An objective quality estimate of the transmission system quality can then be obtained by comparing the source features with the corresponding destination features. Experiments have shown (see [21]-[24]) that the perception-based video quality features can be communicated with a very low bit rate, much less than the bit rate of the source video. Thus, these quality features can be easily and economically transferred between the source and destination instruments, which may be separated by many thousands of miles. The referenced experiments have demonstrated excellent correlations between objective and subjective test data using quality features that can be communicated at a continuous bit rate of only about 1200 b/s. These results seem to be technology independent in that they are applicable to an extremely wide range of video systems, including very high quality 45 Mb/s studio NTSC systems. Prior video quality measures that use the entire source and destination video, such as the mean squared error between source and corresponding destination video images, require perfect copies of the source and destination video images. Thus, these prior measures are not practical unless the source video and destination video are geographically co-located (and available at the same point in time for storage and retrieval applications), a requirement that is not normally met when testing the end-to-end performance of video transmission systems.

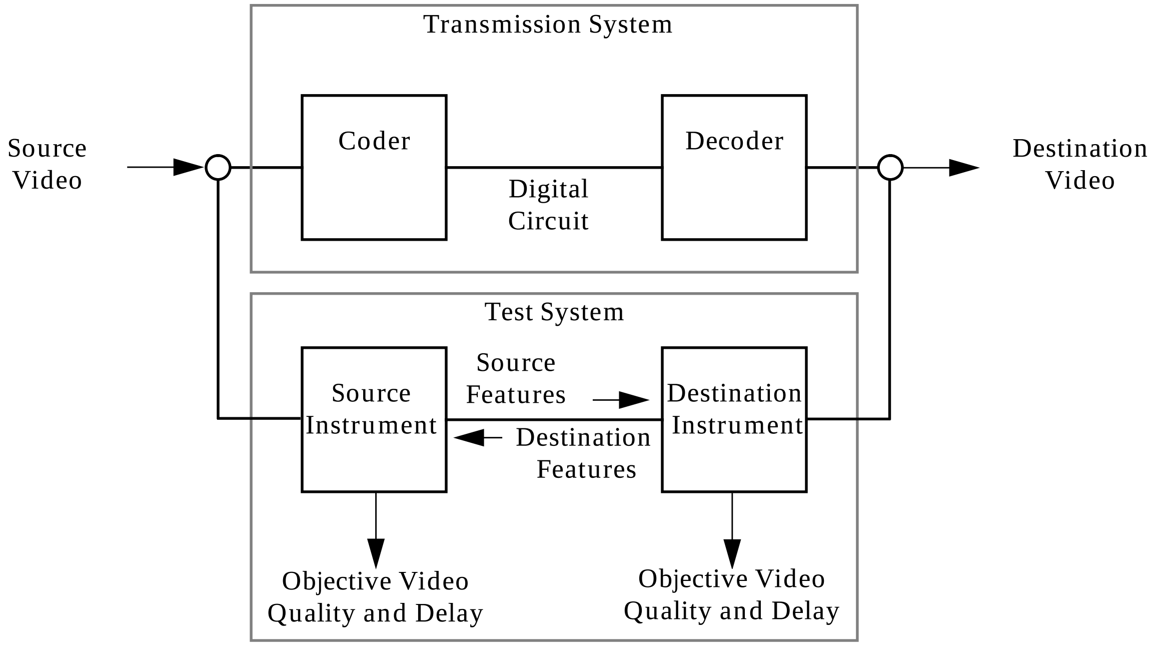


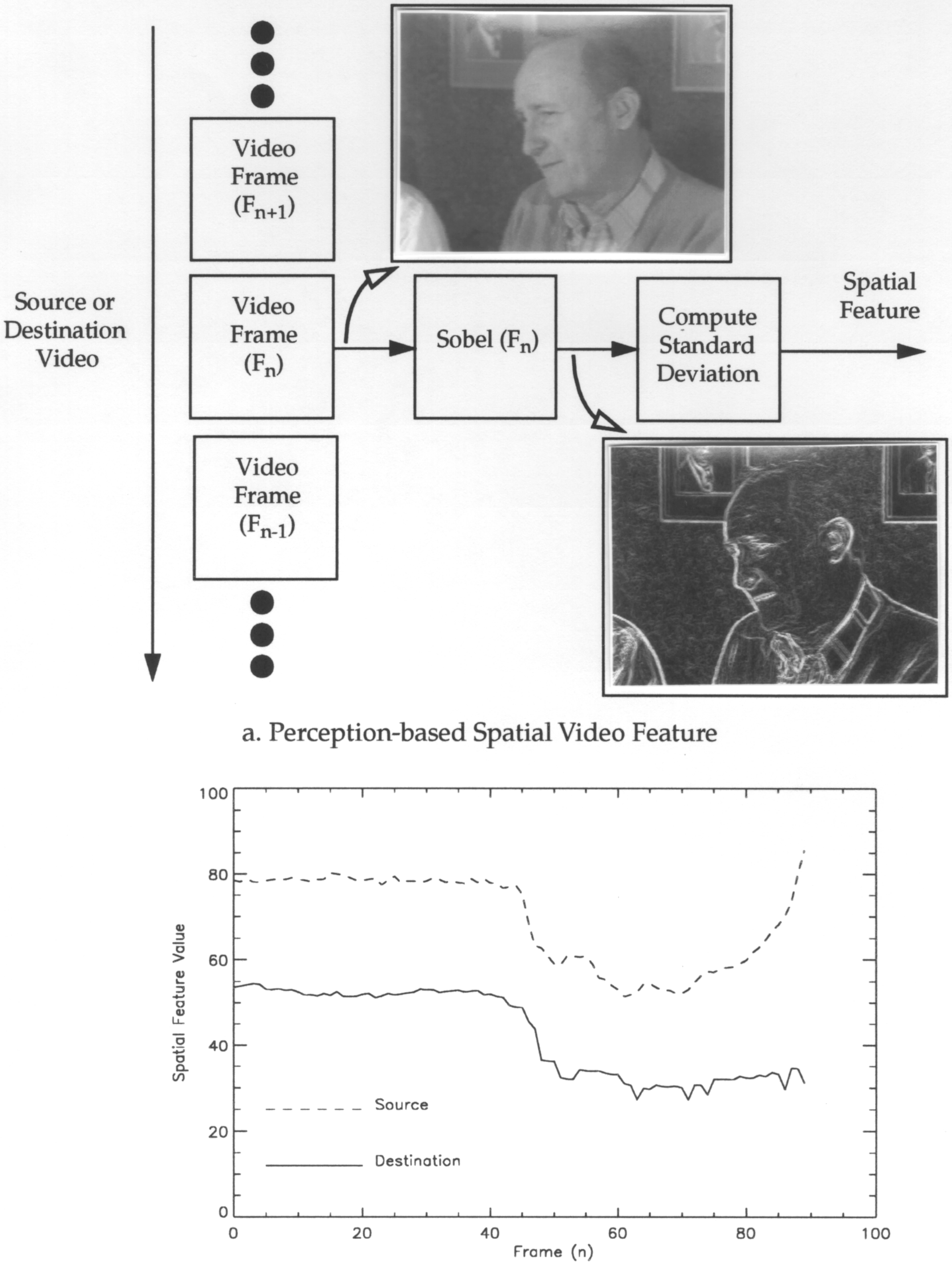
Figure 4 In-Service Perception-based Video Quality Measurement System

If the transmission system shown in Figure 4 is used for interactive (two-way) communications, then video delay is another important quality attribute. Fortunately, the low bit-rate features that quantify the quality of the video scene also quantify its motion. These features can be used to compute the in-service one-way video delay of a transmission system via a time alignment correlation process (see [24], [25]). Such a method can be used in assessing the suitability of video transmission systems for particular applications.

Examples of Low-Bandwidth Perception-based Quality Features

Two examples of low bandwidth perception-based video quality features will be given. One of these measures spatial distortions in the video and the other measures temporal distortions. Designers of modern video transmission systems often trade spatial and temporal performance aspects, and some have even made these trade-offs user selectable. For instance, in low bit-rate applications like videophone, users are sometimes given the option of seeing clear pictures at very low frame rates or blurred pictures at higher frame rates. The optimal spatial-temporal setting may depend upon cultural, personal, or application specific preferences.

Figure 5a illustrates one low bandwidth perception-based spatial video feature. Each video frame at time n, denoted F_n, is digitally sampled (e.g., see [26]) and passed through a Sobel filter [27]. The Sobel filter enhances the edge information in the video scene. High contrast edges that are well focused will produce high amplitude Sobel-filtered values. Blurring of these edges by the transmission system will cause a corresponding decrease in the amplitude of the Sobel-filtered values. This edge sharpness information can be reduced to a single number per frame by computing the standard deviation of all the samples in the Sobel-filtered frame. By extracting and comparing the same spatial feature from the source and destination video, one can determine the amount of spatial blurring introduced by the transmission system. Figure 5b gives an example time history of this spatial feature for the source video and the destination video after passing through a low bit rate digital circuit. As can be seen from the figure, the destination video is blurred in comparison to the source. The amount of blurring is quantified by the source spatial feature value minus the destination spatial feature value, normalized by the former. The greatest blurring occurs during a camera pan in the latter half of the time history.



a. Perception-based Spatial Video Feature

b. Example Time History

Figure 5 Perception-based Spatial Video Feature and Example Time History

To compute a low bandwidth perception-based temporal feature, the video frame at time n (F_n) is subtracted pixel by pixel from the video frame at time n-1. The difference frame (F_n - F_{n-1}) contains non-zero values when motion is present. Computing the standard deviation of all the pixel values in the difference frame reduces the motion information to a single number per frame. The resulting temporal feature measures the flow of motion in the video. By extracting and comparing this temporal feature from the source and destination video, one can determine the amount of motion distortion introduced by the transmission system.

The temporal feature is computed thirty times each second. Video transmission systems that transmit fewer than thirty frames per second generate discontinuous or spiked feature sequences due to the interspersed zeros. For example, at 15 frames per second, every other feature value will be zero. Frame rate is often a dynamic parameter, and the resulting variable levels of jerky motion can be quantified by appropriate processing of the source and destination temporal feature sequences. Other motion impairments introduced by modern video transmission systems are noise and/or blocks in the destination video, including those resulting from digital transmission errors. The severity of these motion distortions for source video with smooth motion (i.e., no scene changes) is quantified by the logarithmic ratio of the destination to source temporal feature values. If abrupt scene changes are present in the source video, allowances may have to be made for visual masking effects (i.e., there is a reduced sensitivity to scene detail immediately before and after a scene change).

Research is continuing to expand and refine the low-bandwidth perception-based features described above. Some simple variations of the above spatial and temporal features have proven useful. More localized measures of the spatial and temporal features may be computed by using image sub-regions of various shapes (e.g., vertical strips, horizontal strips, small rectangular areas, or even motion/still segmented areas). For instance, by selecting and computing the temporal feature using only the horizontal strip of the difference frame with the minimum amount of motion (i.e., the most stationary background), one can obtain a better estimate for the perceptual effects of line-oriented noise [24].

Correlation of Perception-based Quality Parameters with Subjective Viewer Responses

To be of value, perception-based objective quality measures must be well correlated with subjective viewer responses. Tests of correlation between proposed objective measures and subjective responses have been conducted in a number of video experiments. Standard industry procedures have been used for the collection of the subjective viewing responses [28]. From these subjective viewing responses, a subjective mean opinion score was computed, where a rating of 5 represents excellent quality (or imperceptible distortions) and a rating of 1 represents bad quality (or very annoying distortions). The perception-based features described above were measured using the same source and destination video as the subjective tests. Objective quality parameters, derived from the perception-based features, were then used to predict the subjective viewer responses. Representative results for two recent experiments will be described.

The first video experiment, described in detail in [21]-[23], utilized a very wide range of source video and transmission system impairments. The source video was selected to span widely varying amounts of spatial detail and temporal movements so that the transmission system would be subjected to various degrees of coding difficulty. Examples of the 36 source video test scenes that were used included still shots, video conferences, motion graphics, and sports events. The 28 transmission systems that were used in the experiment included 11 digital video codecs (coder-decoders) from 7 manufacturers operating at bit rates from 56 kb/s to 45 Mb/s over simulated digital circuits with controlled error rates, in addition to the following analog transmission systems not represented diagrammatically by Figure 4: the no impairment system, NTSC encode / decode cycles, VHS and S-VHS record / playback cycles, and a system with noise. In all, 132 combinations of scenes and transmission systems were evaluated. The coefficient of correlation between the subjective and objective scores was .92 and the RMS error was .53 quality units. This result is quite remarkable considering the wide range of source video test scenes and transmission systems used in the experiment.

The second video experiment, described in detail in [24], was performed to test contribution quality (i.e., studio-to-studio) 45 Mb / s transmission systems. These high-end video transmission systems are typically used by broadcasters. The source video for the test consisted of 10 test scenes with various amounts of spatial detail and temporal movements. Several of the test scenes were very rich in information content. The 38 transmission systems that were used in the experiment included four 45 Mb / s codecs configured for multiple passes (one, two, three, four, five, and six passes), four 45 Mb / s codecs operating over simulated digital circuits with controlled error rates, and several analog transmission systems including the no impairment system, and a system with noise. In all, 130 combinations of scenes and transmission systems were evaluated. The coefficient of correlation between the subjective and objective scores for this experiment was .92 and the RMS error was .33 quality units. This excellent result, together with the similarly excellent result from the first experiment, suggest that the perception-based quality features are measuring fundamental perceptual quantities and thus are applicable to a wide range of video services and transmission systems.

Forming A Composite Subjective and Objective Rating for a Video Transmission System

The dependency of perceived video quality on the input video scene content played a strong role in the development of the perception-based video quality assessment system. For some low bit-rate transmission systems of the type illustrated generically in Figure 4, this perceived quality can vary from 1 to 5 subjective mean opinion score points depending on the input video scene. This raises a question as to how one should form an overall benchmark or composite subjective and objective rating for a video transmission system. One method of doing this would be to test the transmission system using some ensemble of test scenes. Then one could compute an average subjective or objective rating for the transmission channel by averaging the individual ratings for each of the test scenes. Ideally, this ensemble of scenes would be all the scenes that the end-user is using for his or her application (with the appropriate probability weighting for each scene). This is where the concept of in-service monitoring really shines, since an in-service video quality assessment system can measure the objective quality of each and every scene and then perform some kind of scene-averaging on the results.

Figure 6 shows the results of the scene-averaging process for the first video experiment. Here, the objective and subjective scores for the individual test scenes have been averaged to produce one point for each of the 28 transmission systems. In Figure 6, the coefficient of correlation between the averaged subjective and objective scores is .98 and the RMS error is .25 quality units. A similar averaging for the second experiment yielded a coefficient of correlation between the averaged subjective and objective scores of .97 and an RMS error of .23 quality units. These two scene-averaging results indicate that the objective measurement errors are random with respect to scene content, and thus one can realize an improvement in measurement accuracy by averaging the objective scores across scenes. However, the scene-averaging process is not without drawbacks since if only the scene-averaged subjective or objective scores were reported (as is typically done when reporting benchmarks), one could no longer determine how the transmission system performed for a particular test scene. To overcome this problem, another scene-averaging concept that seeks to preserve some scene dependency information has been developed. This is the concept of scene categories (e.g., "head and shoulders"). Here, objective and subjective scores for all scenes within a scene category are averaged and results are reported for each scene category. Appropriate test scenes and scene categories would be established by standards bodies and made available to the public for testing purposes.

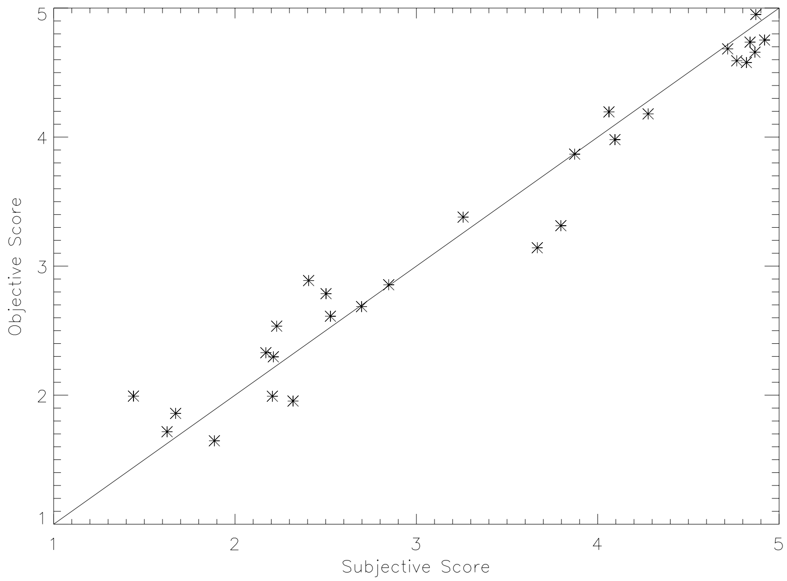


Figure 6 First Video Experiment, 28 Transmission Systems

Voice Quality

As indicated earlier, the development process described in Figure 3 can produce objective perception-based measurements of voice transmission quality. An ideal objective voice quality measurement system would give accurate and repeatable results for female and male speakers of all ages speaking all languages of interest. In addition, an ideal measure would provide meaningful outputs across a wide range of voice transmission systems -- from high quality waveform coders to low bit rate coders operating over poor quality radio links. While such a universal objective voice quality assessment tool seems to be an elusive goal, some objective measurements that have the desired attributes over a limited range of test conditions have been identified. This section describes some of those measurements, and then outlines the principles behind the ongoing work focused on incorporating more complete models of auditory perception to form objective audio quality measurements.

For waveform coders operating over digital channels with relatively few bit errors, an objective measurement called segmental signal-to-noise ratio (SNR) is useful [29]. The segmental SNR is calculated by averaging a series of local SNR calculations each taken over a 15 to 20 mS interval of a voice signal. When non-waveform coding techniques are used in a voice transmission system, SNR's do not provide a useful objective measurement of quality.

Other objective measurements appear to be effective over a wider range of test conditions. Several of these measurements and a technique for combining them are briefly described here.

Cepstral coefficients create a compact physiological model for the speech generation process. One can compute cepstral coefficients for the voice signals that are input to and output from a voice transmission system under test. The Euclidean distance between these two sets of coefficients forms the objective measurement called cepstral distance [30]. An objective measurement called the coherence function uses a normalized version of the cross-power spectrum between input and output voice signals to generate coherent and non-coherent power distributions for the voice system under test [31]. These distributions are combined via frequency selective hearing thresholds and weighting functions to generate an objective measure of the voice system under test. The information index is an objective voice quality measurement that uses the distance between input and output power spectra as its starting point. This distance is used to form a signal-to-distortion ratio. Frequency dependent weights are then applied to this ratio and information theoretic relations are used to generate the final objective quality measurement. While a detailed analysis of cepstral distance, coherence function, and information index is beyond the scope of this article, it is safe to say that each of these measurements has both strengths and weaknesses.

A technique for constructively combining strengths of these three measurements to develop a single composite measurement is proposed in [32]. This pattern recognition based technique utilizes prior knowledge of the statistical distributions of the three objective measurements and the relationship between these distributions and subjective voice quality on a five point mean opinion score scale. This technique was recently used in conjunction with a CCITT organized test of 16 kb/s voice codecs utilizing code excited linear prediction (CELP). For comparison purposes the tests also included 64 kb/s pulse code modulation (PCM) voice codecs and 32 kb/s adaptive differential pulse code modulation (ADPCM) voice codecs. A reference condition known as the modulated noise reference unit (MNRU) was also included in these tests. In total, 26 different voice systems were tested using 144 test sentences in each of five languages. The results of the pattern recognition based mapping of cepstral distance, coherence function and information index were then compared with corresponding subjective test results. Correlation coefficients between the composite objective measures and the subjective mean opinion scores greater than .9 were relatively common, and correlations greater than 0.97 were observed. On average, these correlations are similar to those reported for video measures. The work described above has been carefully studied by the former CCITT Study Group XII and appears in Annex G to Supplement 3 of their Recommendation P.11. The work has also been considered by the ANSI accredited T1A1.6 group and will appear as a technical report from that telecommunication standards working group.

To date, most objective voice quality measurements have stemmed more from an understanding of voice coding and transmission techniques than from an understanding of human perception and judgment. The incorporation of more complete models for human auditory perception and judgment may result in objective voice quality measures that come closer to the desired goals of breadth of applicability and correlation with subjective assessments. Key elements of perceptual models are the limited frequency resolution and temporal resolution of the human auditory system as well as its nonlinear transfer characteristics.

An example of a model for the limited frequency resolution of the human hearing process follows. It has been established that one’s ability to resolve neighboring frequencies is nearly constant for frequencies below 500 Hz and decreases markedly for frequencies above that point. This effect can be modeled by a nonuniform bank of overlapping band-pass filters. Figure 7 shows the response of a set of 16 such filters covering the interval from DC to 4 KHz. On the psychoacoustic frequency scale known as the Bark scale, these filters are spaced at one Bark intervals and each filter has a bandwidth of one Bark. If the energy of a voice signal that passes through each of these filters is accumulated over a short time interval, the resulting set of 16 values provides an approximation to the neurological stimulation that is generated by that voice signal at that time. Separate sets of values can be calculated for the voice signals that are input to and output from the voice transmission system under test. It remains to measure the distance between these two “perceptual stimulus vectors” in a meaningful way.

While fairly explicit models of the auditory processes exist, the human judgment or comparison process remains mostly a mystery. Thus, the most appropriate way to measure the distance between two “perceptual stimulus vectors” remains an open research topic. In one study, a filter bank much like that described in Figure 7 was used to perceptually transform the input to and output from a voice system under test [33]. The filter bank was followed with non-linearities thought to further model the human auditory processes. The Euclidean distance between the resulting “perceptual stimulus vectors” was measured to generate a single objective quality value. These values correlated quite well with the results of the subjective tests over a fairly wide range of voice coders. While these results are encouraging, is clear that much additional work can be done to enhance the performance of objective voice quality measurements rooted in perceptual models.

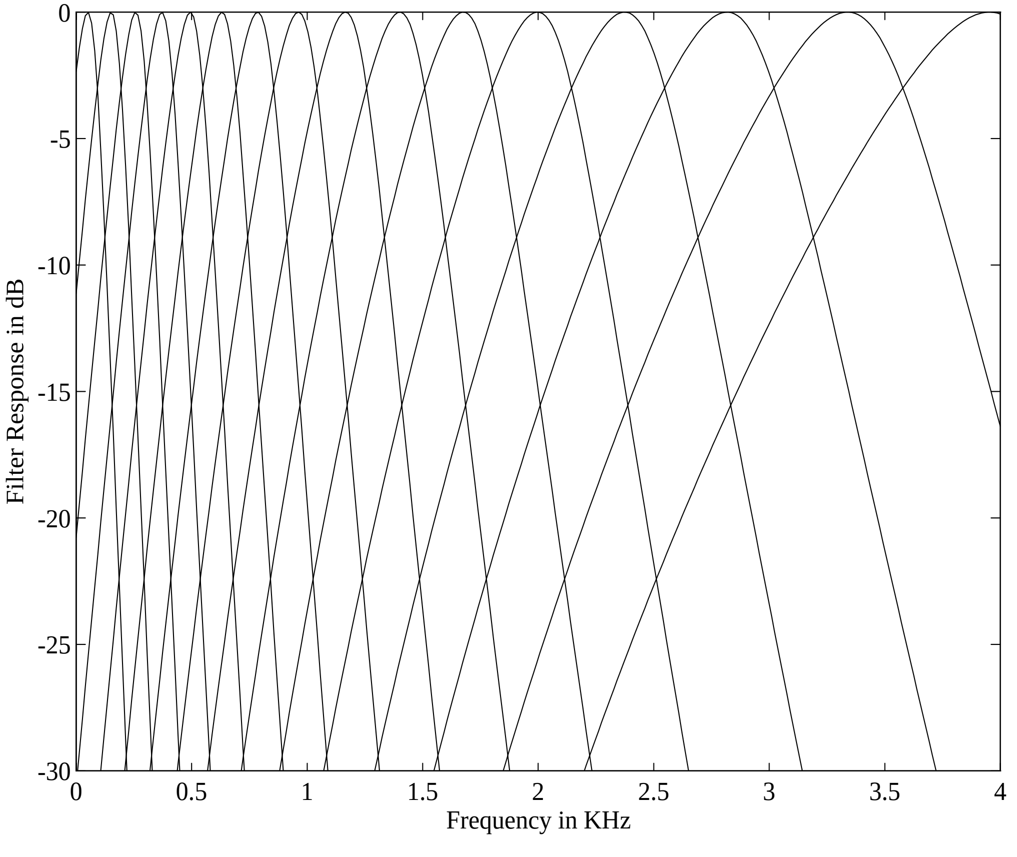


Figure 7 Filter Bank Model for Auditory Frequency Resolution

The psychoacoustic literature contains perceptual models that extend out to 20 KHz. The incorporation of these models into new objective audio quality assessment tools could allow for reliable objective assessment of wideband speech and music transmission systems. Already, these wideband perceptual models are being used to encode high quality music into a minimum of bandwidth by exploiting masking properties of the human ear. Many low bit rate digital voice coders have been designed to operate over radio channels which can introduce a significant number of bit errors. Depending on how many and which bits are errored, these voice coders can create some rather unusual sounds. It is hoped that voice quality assessment tools that utilize perceptual models will be able to more accurately characterize the perceptual impact of these bit errors and the bursts of sound that they create.

Ongoing Work and Future Challenges

The application and elaboration of user-oriented, technology-independent quality of service parameters is continuing in both national and international standards committees. The user-oriented data communication quality parameters are currently being applied in the development of protocol-specific performance parameters for frame relay services. Likely future applications are in B-ISDN call processing and intelligent network/universal personal telecommunication (IN/UPT) performance description. These future applications will require an expansion of the 3x3 matrix to encompass additional functions. A major future challenge would be to make the user-oriented data communication quality parameters fully perception-based, like their video and voice counterparts. This would require the quantification of user responses to service quality impairments for a representative set of data applications analogous to the video scenes used in developing the video quality measures.

In the video and voice areas, ongoing work is focused on optimizing the basic electrical measures and their mappings with human perceptions through additional research and objective/subjective parameter correlation experiments. Near-term future studies are expected to focus on specializing the perception-based objective measures to improve their correlation with subjective perceptions for particular types of information content (e.g., high-resolution medical imagery vs. sports action, language differences, music vs. voice) and particular evaluation conditions (e.g., untrained vs. expert viewers). The development of low-bandwidth perception-based quality features for technology independent measurement of in-service quality opens up a host of possibilities for end-users and telecommunication systems designers. Since these perception-based quality features can be easily and economically communicated throughout the telecommunication network, possible future uses include automatic quality monitoring, fault detection, optimization of network resources to assure a constant or minimum level of received quality, and dynamic allocation of bandwidth between the audio, video, and data components of multimedia transmission systems.

The most compelling (and demanding) long-term challenge for the developers of user-oriented data, video, and voice performance measures will be to integrate their work in two orthogonal “dimensions” as illustrated in Figure 8. “Horizontal” integration will be required to create objective, perception-based performance measures for multi-media (e.g., combined voice, video, and data) services. The multi-media communication performance measures will not be a simple superposition of their single-medium counterparts because human reactions to multi-media impressions depend on the interactions among media. A simple example of this is lip synchronization. Successful “vertical” integration would, at last, give network providers the long-sought means of relating network performance impairments (or enhancements) with the subjective perceptions (and practical application decisions) of their customers.

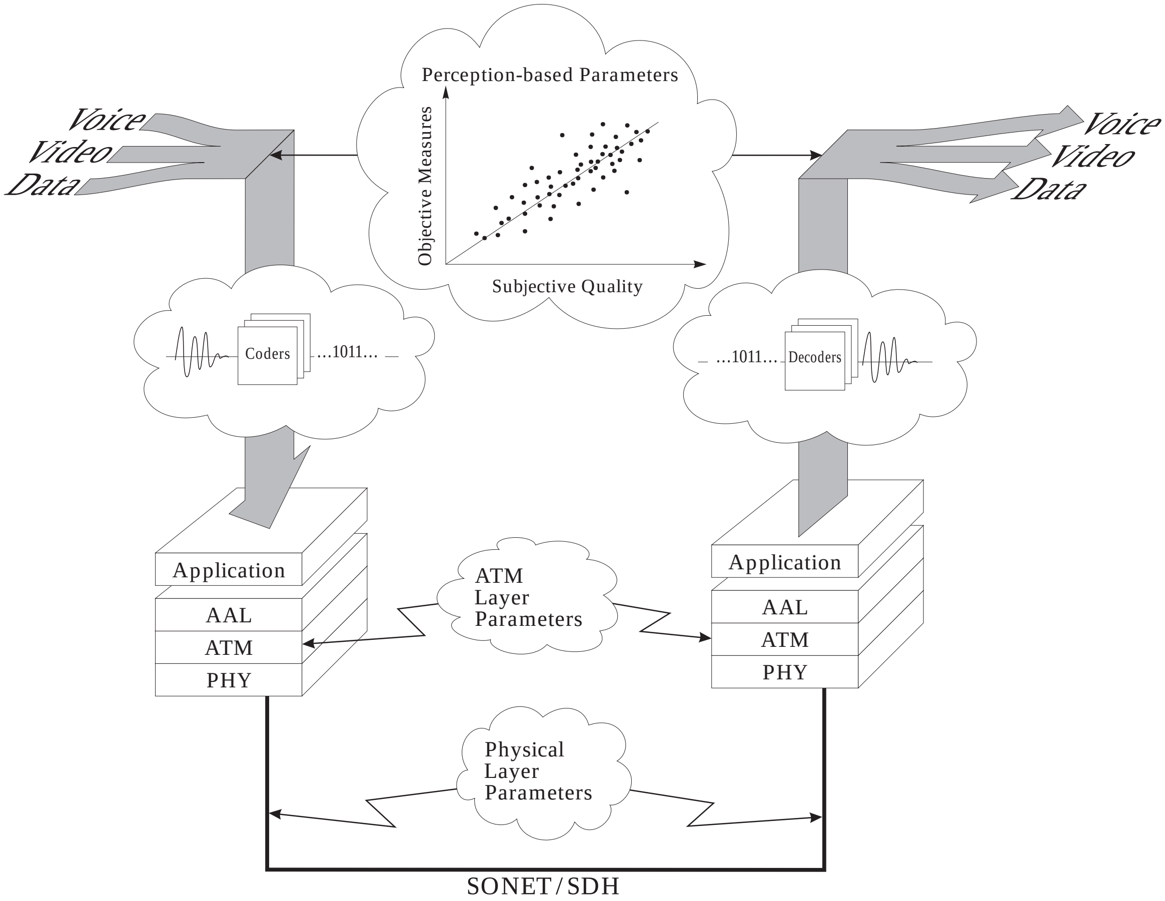


Figure 8 Possible Future Integration of Performance Measures

Acknowledgments

The authors gratefully acknowledge the contributions of many individuals to the successful development of the user-oriented performance measures and standards described in this paper. Our particular thanks go to Dave Wortendyke, Ken Spies, Marty Miles, Arthur Webster, Coleen Jones, Margaret Pinson, Robert Kubichek, David Atkinson, Connie Sholl, and Paul King, who first implemented the proposed measures within the Institute for Telecommunication Sciences; Dana Grubb, Ken Glossbrenner, and Art Reilly, who made important technical and administrative contributions to their elaboration in national and international standards committees; and Dennis Bodson, who provided key resources to support the research as NCS Deputy Manager and Chairman of the Federal Telecommunications Standards Committee.

References

[1] J.S. Richters and C.A. Dvorak, “A Framework for Defining the Quality of Communications Services,” IEEE Communications Magazine, vol. 26, no.10, October 1988, pp. 17-23.

[2] American National Standard for Information Systems - Data Communication Systems

and Services - User-Oriented Performance Parameters, ANSI X3.102-1983, American National Standards Institute, Inc., New York, NY.

[3] American National Standard for Information Systems - Data Communication Systems and Services - Measurement Methods for User-Oriented Performance Evaluation, ANSI X3.141-1987, American National Standards Institute, Inc., New York, NY.

[4] CCITT Recommendation X.140, General Quality of Service Parameters for Communication via Public Data Networks, Volume VIII, Fascicle VIII.8, CCITT Blue Books.

[5] CCITT Recommendation I.350, General Aspects of Quality of Service and Network Performance in Digital Networks, Including ISDN, Volume III, Fascicle III.8, CCITT Blue Books.

[6] K.C. Glossbrenner, The Availability and Reliability of Switched Services, IEEE Communications Magazine, 1993.

[7] N.B. Seitz and D.S. Grubb, American National Standard X3.102 User Reference Manual, NTIA Report 83-125, October 1983.

[8] N.B. Seitz, D.R. Wortendyke, and K.P. Spies, User-Oriented Performance Measurements on the ARPANET, IEEE Communications Magazine, pp. 28-44, August 1983.

[9] Spies, K.P., D.R. Wortendyke, E.L. Crow, M.J. Miles, E.A. Quincy, and N.B. Seitz, User-Oriented Performance Evaluation of Data Communication Services: Measurement Design, Conduct, and Results, NTIA Report 88-238, August 1988.

[10] Bloomfield, R.S., and K.P. Spies, User-Oriented Performance Evaluation of Data Communication Services: Measurement Design, Proceedings of the IEEE International Conference on Communications, June 11-14, 1989.

[11] Bloomfield, R.S., and K.P. Spies, User-Oriented Performance Evaluation of Data Communication Services: Measurement Results, Proceedings of the IEEE International Conference on Communications, June 11-14, 1989.

[12] CCITT Recommendation X.134, Portion Boundaries and Packet Layer Reference Events: Basis for Defining Packet-Switched Performance Parameters, Volume VIII, Fascicle VIII.3, CCITT Blue Books.

[13] CCITT Recommendation X.135, Speed of Service (Delay and Throughput) Performance Values for Public Data Network When Providing International Packet-Switched Service, Volume VIII, Fascicle VIII.3, CCITT Blue Books.

[14] CCITT Recommendation X.136, Accuracy and Dependability Performance Values for Public Data Networks When Providing International Packet-Switched Service, Volume VIII, Fascicle VIII.3, CCITT Blue Books.

[15] CCITT Recommendation X.137, Availability Performance Values for Public Data Network When Providing International Packet-Switched Service, Volume VIII, Fascicle VIII.3, CCITT Blue Books.

[16] CCITT Recommendation X.138, Measurement of Performance Values for Public Data Networks When Providing International Packet-Switched Services, COM VII-R 39-E, April 1992.

[17] CCITT Recommendation X.139, Echo, Drop, Generator and Test DTE's for Measurement of Performance Values in Public Data Networks When Providing International Packet-Switched Services, COM VII-R 39-E, April 1992.

[18] American National Standard for Telecommunications - Packet-Switched Data

Communication Service - Performance Parameters, T1.504-1989, American National Standards Institute, Inc., New York, NY.

[19] American National Standard for Telecommunications - Packet-Switched Data Communication Service - Performance Measurement Methods, T1.504a-1991, American National Standards Institute, Inc., New York, NY.

[20] American National Standard for Telecommunications - Packet-Switched Data Communication Service - Performance Objectives, T1.504b-1993, American National Standards Institute, Inc., New York, NY.

[21] S.D. Voran and S. Wolf, "The development and evaluation of an objective video quality assessment system that emulates human viewing panels," International Broadcasting Convention, Conference Publication Number 358, Amsterdam, The Netherlands, July, 1992.

[22] Arthur A. Webster, Coleen T. Jones, Margaret H. Pinson, Stephen D. Voran, Stephen Wolf, "An objective video quality assessment system based on human perception," SPIE Human Vision, Visual Processing, and Digital Display IV, Volume 1913, San Jose, California, February, 1993.

[23] Stephen Voran and Stephen Wolf, "An objective technique for assessing video impairments," IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, B.C., Canada, May, 1993.

[24] Stephen Wolf and Arthur Webster, "Objective and Subjective Video Performance Testing of DS3 Rate Transmission Channels," American National Standards Institute (ANSI) Accredited Standards Working Group T1A1 contribution number T1A1.5/93-060, April, 1993. Available from Standards Committee T1 Telecommunications Secretary, 1200 G. Street, N.W., Suite 500, Washington, D.C., 20005.

[25] Coleen Jones, Stephen Wolf, and Margaret Pinson, "Preliminary results of one-way video delay measurement algorithms," American National Standards Institute (ANSI) Accredited Standards Working Group T1A1 contribution number T1A1.5/92-139, July, 1992. Available from Standards Committee T1 Telecommunications Secretary, 1200 G. Street, N.W., Suite 500, Washington, D.C., 20005.

[26] CCIR Recommendation 601-2, "Encoding Parameters of Digital Television for Studios," Recommendations and Reports of the CCIR International Radio Consultative Committee, Volume 11, part 1, 1990.

[27] R.C. Gonzalez and P. Wintz, *Digital Image Processing*, 2nd Edition, Addison-Weslsy Publishing Co., Reading, Massachusetts, 1987.

[28] CCIR Recommendation 500-4, "Method for the Subjective Assessment of the Quality of Television Pictures," Recommendations and Reports of the CCIR International Radio Consultative Committee, Volume 11, part 1, 1990.

[29] N. Kitawaki, H. Nagabuchi, and K. Itoh, "Objective quality evaluation for low-bit-rate speech coding systems," IEEE Journal on Sel. Areas in Communications, vol. 6, no. 2, pp. 242-248, Feb., 1988.

[30] Bell Northern Research, "Re-evaluation of the objective method for measurement of non-linear distortion," Contribution to CCITT, COM XII-175-E, June 1987.

[31] J. Lalou, "The information index: an objective measure of speech transmission performance," Ann. Telecommun., vol. 45, no. 1-2, pp. 47-65, 1990.

[32] R. Kubichek, D. Atkinson, and A. Webster, "Advances in objective voice quality assessment," IEEE Global Telecommun. Conference, Phoenix, AZ, Dec. 1991.

[33] S. Wang, A. Sekey, and A. Gersho, “An objective measure for predicting subjective quality of speech coders,” IEEE Journal on Sel. Areas in Communications, vol. 10, no. 3, pp. 819-829, June., 1992.