Experimental Characterization and Modeling of the QoS for Real Time Audio and Video Transmission

José Roberto de A. Amazonas, Carlos Rebolledo Barra

Abstract— **Multimedia applications have intrinsic quality of service (QoS) requirements. However, a quantitative evaluation of the QoS offered by the network is rarely realized. Furthermore, the traditional methods for ensuring an end-to-end QoS rely on mechanisms provided by the distribution network. In this paper we propose a new quantitative QoS model that is validated by extensive experimental characterization of the QoS dependence on traffic parameters. This model is then used to design a QoS restoration technique that is independent of special resources available in the distribution network.**

Index Terms— **QoS, characterization, modeling, restoration.**

I. INTRODUCTION

THE Internet has been growing rapidly with respect to the number of users and the amount of traffic has been recognized as an important infrastructure for information in social and business use. So, although the initial and main issue of the Internet has been its connectivity and transmission capacity, attention has recently been paid to its quality too. The traffic conveyed by the Internet is generated by a wide variety of applications, which have different characteristics and different quality requirements. Thus, Quality of Service (QoS) and measurements are crucial in controlling and managing QoS and provisioning networks. In recent years, different approaches have been used to tackle such issues. Most approaches rely on mechanisms provided by the distribution network, i. e., the Internet.

Closer to the objectives of this work, are the proposals of communications resources management. In [1], the authors' middleware QoS management approach encapsulates QoS behaviors as software components. Using the Corba component model, they build these specialized QoS components and combine them to produce a comprehensive management system that maintains QoS. In [2], a rule-based self-management control system for data streaming applications is presented. [3] introduces an abstract model for resource mechanisms, called Virtual Resource Tree (VRT). The VRT emphasizes scalability as an important role in multimedia applications scenarios, imposing distributed management of heterogeneous resources and the support for adapting the different underlying management algorithms. [4] addresses the fact that audio and video QoS requirements can become

more complex when applications are executed in different platforms including different technologies, and proposes a QoS management and specification model to support adaptation in middleware for distributed multimedia systems. Adaptation is also a key feature for a wireless pervasive computing system. The mobility of devices coupled with the unpredictability of wireless networks results in fluctuations in QoS as the user moves around or other user traffic varies. To maintain the minimum level of QoS desired by each individual user, a pervasive system needs to be dynamic and adaptive to changes in QoS [5].

Last but not least, the subjective perception of QoS is a major concern for VoIP and video applications. Presently, VoIP quality is not as good as achieved in traditional telephone networks due to the data loss and delay in IP networks. In [6], a perceptual quality-driven scheduling scheme is proposed and in [7], a real-time non-intrusive VoIP evaluation procedure is demonstrated. [8] analyses the metrics used to evaluate the performance perceived by users of video streaming applications and introduces some new metrics.

The available models that try to translate the subjective perception of the quality of a service into a mathematical function of measurable network parameters suffer from low inherent credibility and high complexity to be useful as a basis of an on-line QoS restoration algorithm. Furthermore, the standardized models were conceived as networks planning tools, i.e., their objectives are to provide means for identifying physical impairments that degrade the QoS and not how to get best possible QoS from a given network under real operational conditions.

Despite the huge amount of papers dealing with real-time multimedia transmission systems QoS, to the best of our knowledge, the approach of this paper is quite unique in mapping the subjective perception into a mathematical framework and applying this new model for on-line QoS restoration without changing the distribution network.

In this paper we address the problem of ensuring end-to-end QoS by an innovative perspective made of the following steps:

- definition of session and service qualities;
- proposal of a mathematical formulation for session and service qualities based on measurable parameters;
- proposal of a QoS restoration technique based on a finite state machine modeling of a service;
- extensive experimental characterization of QoS behavior for real time transmission of audio services, both interactive and streaming, and video streaming;

- modeling of QoS behavior based on the mathematical formulation and the experimental data;
- suggestion of an algorithm implementation of QoS restoration for audio and video streaming that does not rely on special resources provided by the distribution network.

After this brief introduction, Section II discusses some of the available voice quality assessment methods. In Section III the concepts, mathematical formulation and finite state machine based restoration technique of QoS are introduced. In Section IV the QoS experimental characterization is presented. Section V validates the proposed model. In Section VI an algorithm for restoring QoS is suggested. In Section VII the conclusions and future work are summarized.

II. USER ASSESSMENT OF VOICE QUALITY

When users talk about the quality of voice carried over an interactive voice connection, they are generally trying to describe their reaction to, or satisfaction with, one of two attributes of a voice service [1]:

- *Connection quality.* Determined by what is heard over the connection.
- *Connection usability.* Determined by what is experienced in conversational exchanges over the connection.

In listening to a connection to assess connection quality, users will, for example, be conscious of whether there is perceptible echo; the distant speakers are easily heard, readily understood, and sound natural; the connection is degraded by distracting impairments; and the voice is good enough to enable recognition of distant speakers and to detect nuances in their articulation.

In contrast, when assessing the usability of a telephone voice connection, users will determine whether the overall effect of all impairments and conditions on that connection was severe enough to interfere with natural conversational rhythms and speech patterns or was otherwise distracting to the flow of information.

User perception of the quality of a voice service with respect to these two attributes is almost a primitive notion, constituting something that is inescapably subjective and dependent on the tastes, dislikes, and expectations of each individual user.

A viable measure of connection quality is one that will provide some basis for addressing the following users' concern: Will what I hear over the voice connections sound all right?

There are a variety of measures of quality that have been proposed or employed for addressing this question over the years. Nearly all such measures today, however, are based on responses elicited from test subjects by

1) Defining categories of possible subjective descriptions of the quality of a voice connection, such as excellent, good, fair, poor.
2) Conducting tests in which participants listen to voice signals and assign subjective responses to samples.

In early testing of this kind, the measures used were simply proportions of the sample responses that fell, or were expected

[1]This section is an excerpt from [9]

TABLE I
OPINION SCALES FOR SUBJECTIVE TESTS

	AT&T 1960s	ITU-T P.800 1996	Fractional scores		Interpretation
Excellent	4	5	3.5	4.5	Very good
Good	3	4	2.5	3.5	Fair to good
Fair	2	3	1.5	2.5	Fair to poor
Poor	1	2			
Bad	0	1			

to fall, into certain categories. Thus, for example, outputs from the earliest attempts to predict voice quality with the so-called loss/noise grade of service model developed by AT&T in the mid 1960s were simply expected percentages:

- %GoB. Represents the percentage of calls rated as "good" or better.
- %PoW. Represents the percentage of calls rated as "poor" or worse.

A. Mean Opinion Scores

The most widely used and accepted measure of user perception of voice today is, however, a mean opinion score (MOS). Such a MOS is a value created by:

- Assigning numerical values to each of the subjective descriptors.
- Calculating the average of the corresponding numerical values.

The particular values assigned to each category might, in fact, be anything. However, the convention adopted by the ITU-T in Recommendation P.800 [10] is the five-value scale shown in Table I. This table also shows other scales that have been used. The 0 to 4 scale in the table, for example, is the original quality scoring scheme used by AT&T in the 1960s. The fractional score scales are ones that have been allowed or explicitly adopted in various subjective tests.

Because of the formal definition and widespread use of the abbreviation MOS, there is a tendency to assume that a mean opinion score is a stable, immutable measure of voice quality which always means the same thing. However, a particular value such as 3.75, cannot be meaningfully interpreted without at least a description of the context of its calculation, and probably not even then without a known comparative reference for that context.

If we are going to be able to interpret the MOS derived from a particular test, we must, as a minimum, also discriminate the scale, type of test (listening or conversational), method of sampling, and convention for handling shaded evaluations used in that test. However, even with all that information, we will not have an adequate basis for gleaning information as to likely user satisfaction from the MOS reported, unless the MOS for a particular service being evaluated is complemented with a similarly derived value of MOS for a known service. The reason is that absolute MOS values, even when they are derived from tests conducted under a particular, known test protocol, are meaningless.

TABLE II
LISTENING EFFORT SCALE DEFINED IN ITU-T P.800

Effort required to understand meanings of sentences	Score
Complete relaxation; no effort required	5
Attention necessary; no appreciable effort required	4
Moderate effort required	3
Considerable effort required	2
No meaning understood with any feasible effort	1

TABLE III
CRITERIA FOR CONNECTION USABILITY DEFINED FOR TEST SUBJECTS IN
SERVICE ATTRIBUTE TESTS

Reported effect	Description
Unusable (U)	Connection was so poor that the only viable option would have been to hang up and try again
Difficult (D)	One party or the other had to ask for repetition of information, speak louder to be heard, slow down and articulate more clearly, or otherwise adopt an unnatural conversational pattern in order to maintain the conversation.
Irritating (I)	No actual difficulties were noticed, but the impairments or problems were great enough to represent a distraction or an irritant.
Noticeable (N)	Some impairments or problems were noticeable, but were neither great enough nor persistent enough to materially affect the conversation.
None (O)	Either no problems were noticed, or they were so slight or sporadic as not to represent any meaningful effect on the conversation.

B. Measurement of Connection Usability

In contrast to the question of how good transmitted voice sounds, which is addressed by measures of connection quality, the principal concern with connection usability is expressed in the question, Will we be able to maintain natural, clearly understood conversational exchanges, without distraction, or interruptions caused by needs to adjust to the connection?

Unlike mean opinion scores, which are widely used to discuss connection quality, there are no widely used or accepted measures that have been defined to directly address this particular question. ITU-T P.800 does define a scheme for scoring listening effort with the descriptors shown in Table II. This represents recognition, at least, of the important distinction between what is heard and what is experienced and the need for different kinds of measures to capture that distinction.

However, if the earlier description of the processes by which user assess connection usability is accurate, then what is needed in this context is something that defines bad calls in terms of overall experience and measures the frequency with which users are likely to perceive their occurrence. This kind of measure is exemplified by the proportion of calls that have been rated for overall effect of whatever was experienced regarding the ability to maintain a smooth, natural flow of conversation.

One scale for describing such effects is defined by the criteria shown in Table III. These definitions are taken from the protocols for service attribute tests (SATs). The associated measure, P[UDI], is defined as the proportion of calls that SAT subjects rated, or would have rated, as being "unusable", "difficult", or "irritating" according to that scale.

Although P[UDI] is specific to the service attribute test, it can be treated as exemplar for any analogous measures that can be used to gauge likely user perception of usability of voice connections. Such measures are distinguished by two characteristics:

1) Rather than reflecting what was heard over a voice connection that might cause disruptions to the flow of conversation, the subjective assessments reflect the effects of everything experienced on a connection on the natural flow of conversation and exchange of information.
2) The associated quantifier is a proportion, rather than a mean opinion score.

It is possible to have a high MOS, indicating satisfaction with what is being heard, and still have a high P[UDI], indicating user dissatisfaction with what is experienced.

C. User-Perception Models

A carefully designed subjective user test is the recommended vehicle for direct quantification of measures of voice quality.

In the case of user-perception models, the principal function of the model is to enable us to predict likely user perception of voice quality without having to elicit subjective evaluation from users of the service.

The principal criterion os assessing the utility of a user-perception model is, then, the model's reliability, as gauged by the probability that the answer to the questions will be correct. Since the answers to these questions are qualitative, the question of reliability is not one of accuracy of whatever numbers are produced by the model, but whether the interpretation of those numbers produces correct answers to the target question. In practical terms, this implies that the only real test of reliability of a user-perception model is the demonstration of its success in producing correct answers.

Some criteria that can be used to gauge the likely reliability of a user-perception model when there is no history of its success in the intended application. The most common of these are tests for completeness, consistency, and correlation, which comprise part of the basis for determining whether a model is *scientifically defensible.*

The utility of a user-perception model for some purposes will depend on inherent credibility, extensibility, and manipulability of the model. It is clear that user-perception models have in the past been evaluated and endorsed without any consideration of extensibility. for example, the PESQ (Perceptual Evaluation of Speech Quality) model is endorsed in ITU-T Recommendation P.862 [11] as "...an objective method for predicting the subjective quality of 3.1 kHz (narrow-band)

handset telephony and narrow-band codecs." The model is endorsed therein as a means of answering the question. "how good will speech carried via a handset over a particular codec sound?" It was later endorsed as the preferred model for gauging fidelity of speech signals exposed to dropped frames during transport via a VoIP service. At the same time, ITU-T P.862 declares, in effect, that the endorsed model is definitely not extensible, by stating that:

> It should also be noted that the PESQ algorithm does not provide a comprehensive evaluation of transmission quality. It only measures the effects of one-way speech distortion and noise on speech quality. The effects of loudness loss, delay, sidetone, echo, and other impairments related to two-way interaction (e.g., center clipper) are not reflected in the PESQ scores. Therefore, it is possible to have high PESQ scores, yet poor quality of the connection overall.

This quotation demonstrates the sense in which a model, in this case the PESQ, may not be extensible with respect to applicability. A model that is readily extensible with respect to new data is one in which any factors that are assumed or derived from empirical data represented as parameters in the model that can be readily changed.

D. Service Attribute Tests

One of the most versatile tools for measurement of likely user perception of voice quality is a carefully constructed subjective user test, whose design:

- Achieves operational realism.
- Employs efficient sampling strategies and tester orientation.
- Produces a basis of comparison of both connection quality and connection utility of the service tested with a familiar service used by the testers.

The particular protocol chosen to illustrate design criteria for subjective user tests is the service attribute test (SAT). This test was developed by Satellite Business Systems in the early 1980s, because subjective tests in use at the time were found to be inadequate for purposes of analyzing long-distance voice services employing satellite transport. Attempts to use the then-standard subjective test protocols to try to determine why a voice service that was lossless and virtually noise-free was not acceptable to users pointed up the necessity to expand subjective testing in two ways. The the-standar tests were showing user dissatisfaction, but providing little insight as to why the users were the connection quality lower when it was expected to be higher. This demonstrated the requirement for designing subjective tests that would consistently elicit information as to what the users were reacting to. In addition what information was being derived from free-form comments was suggesting that some of the dissatisfaction was due to greater transmission delay over the satellite, which had little to do with how clear the voice was. This suggested the necessity to include user assessments of what we have characterized as connection usability, as well as the assessments of connection quality that were being collected. The resultant subjective test

tool achieved the kind of inherent credibility, extensibility, and manipulability of results as features that enhance the utility of user-perception models.

When a test call is connected, the test subject is expected to hold a brief conversation with the distant party. when the called party is not a test subject, the test subject reporting the call is instructed to end each conversation by soliciting and recording answers to two classical subjective user test questions:

- Did you have any difficulty talking or hearing this call?
- How would you rate the quality of this call?

The test subjects are instructed to hang up the phone, and only then report their assessments of what they heard and experienced in the required formats.

The first element of that reporting is a rating of the presence and severity of impairments to the quality of what was heard over the connection. The types of impairments to be listened for and reported may vary. However they will be set for any particular SAT before the testing begins by defining and describing for the test subjects readily recognizable manifestations of conditions expected to degrade the quality of the connections sampled.

As suggested, the particular impairments defined and described in a SAT can be almost anything that is a readily recognizable transmission artifact that can be clearly and unambiguously described for the test subjects. The important thing in their selection is that they represent something that is expected to occur in the services being tested and are described in such way as to pose no problems for training test subjects in their recognition.

E. Speech Distortion Measurement

The implications from studies of speech and hearing were recognized far earlier than they could be exploited for measurement and evaluation of voice quality. As late as 1980, the best attempts to measure some of the more complex types of waveform distortion that might affect user perception of voice quality were based on mechanical simulators like the artificial ear for telephonometric measurements patented in 1978. More robust measurements capabilities had to await the evolution of microcomputing technology that could handle the massive computational workloads involved in capturing, storing, and processing speech waveforms to quantify complex measures of waveform distortion.

As those capabilities have become available, however, the development of models that would predict user perception of speech quality based on analysis of deformations of speech waveforms has become a distinct possibility. At the core of these efforts have been attempts to define measures of speech distortion that;

1) Can be quantified by processing of digital images of complex waveforms.
2) Reflect differences between natural and telephonic speech that affect what users hear without rendering speech unintelligible.
3) Can be shown in subjective user tests to exhibit a high correlation between values of the measures and user assessments of voice quality.

TABLE IV
TYPES OF SPEECH DISTORTION MEASUREMENT

	Active (based on comparison of what was received with what was transmitted)	Passive (based only on what is received, without knowledge of what was transmitted)
Electroacoustic	Multifrequency tone tests	Analysis of statistics of rates of change in PCM signals
Psychoacoustic	Perceptual Speech Quality Measure (PSQM) Perceptual Analysis Measurement System (PAMS) Perceptual Evaluation of Speech Quality (PESQ)	Extensions of PAMS

Various efforts in this direction have resulted in four different kinds of measurements techniques, depending on whether the measurements involved are psychoacoustic or electroacoustic and on whether the data acquisition is active or passive. These distinctions are described in Table IV, together with the names of some of the particular techniques of each approach.

The Perceptual Speech Quality Measure (PSQM) is distributed as ITU-T Recommendation P.861 [12] for measuring the quality of speech codecs. The purpose of the recommendation was to set standards for objectively gauging the likely user perception of quality of voice carried via different codecs under a variety of conditions. As a measure of likely subjective perception of quality, the PSQM measure by itself can serve as a tool for analysis of relative quality.

A psychoacoustic measurement technique very similar to the PSQM was the Perceptual Analysis Measurement System (PAMS). The basic PAMS measurement approach differs from the PSQM only in the details of the methodology. Both utilize transforms that purport to produce measures based on a comparison of the source and test signals as they would be heard, rather than the way that they would be represented in an electroacoustic waveform. Both rely on subjective user testing under controlled conditions to establish the relationship between the measures calculated and the predicted MOS.

In February of 2001 the improvements and refinements of both PSQM and PAMS were integrated to produce a new standard, described in ITU-T Recommendation P.862 [11]. The new system was called the Perceptual Evaluation of Speech Quality (PESQ) and described as "an objective method for end-to-end assessment of narrow-band telephone networks and speech codecs".

The new PESQ promulgated in ITU-T P.862 was, in fact, a combination of the perceptual models developed for the PSQM, as revised and improved through 1999, and the process for detecting and correcting losses of synchronization between the source and test signal developed for the PAMS.

Although the incorporation of the PAMS methodology for handling variability in the synchronization of source and test signals enabled the application of the PESQ to packet-switched connections, the approach taken in the psychoacoustic test procedures have four characteristics in common that limit their applicability, namely:

1) Requirements for level alignment.
2) Combination of effects.
3) Comparison of received and transmitted signals.
4) In a packet-switched network, a single value of a measure like those calculated in these test procedures cannot be meaningfully interpreted. The problem is this. In a circuit-switched network the conditions that degrade quality, such as noise or a high bit error rate, persist for the duration of the call, producing effects that are equally distributed among all parts of the waveform sampled. In packet-switched networks the principal contributor to speech distortion is dropped frames, which are at best randomly distributed discrete events and may, in fact, be manifested as groups of dropped frames or short periods in which many are dropped. Because of this, the pattern of dropped frames in any particular sample may be one in which the effects of dropped frames are not equally distributed between periods of speech and periods of silence. Since the psychoacoustic measurements cited here all ascribe a different weight to disturbances of speech and disturbances of silence, this means that the same underlying conditions can result in substantially different measures for the same sample.

What these limitations imply is that none of the P-family of psychoacoustic measures (PSQM, PAMS, PESQ) should be used as a stand-alone gauge of quality of packet-switched voice services. Even in the pure packet-switched environment where they can be reliable indicators of emergence of problems, particularly those attributable to dropped packets, these measures are not complete. The monitoring and assessment of likely user perception of quality still must be complemented with measures that reflect the likelihood that users are experiencing round-trip delays great enough to disrupt conversational patterns. Moreover, in any services in which the calls are originated in the PSTN and passed into a packet-switched network for onward delivery, the quality of voice can be degraded by exactly the effects that are not reflected in the measures. This means that these measures are simply inadequate for purposes of assessing the quality of voice in hybrid transport networks.

F. Multiple-Effects Models

The broad objective of the modeling efforts described in the preceding section was to produce measures of the deformation of speech waveforms that might serve as reliable indicators of the likely user perception of quality of voice. However, these measures reflect only some of the effects of telephone transmission that may degrade the quality of what is heard. The measures reflect the overall effects of waveform deformations caused by such things as signal filtering, codec encoding, errors in digital transmission, dropped packets, and noise. however, they are completely invariant with respect to other

TABLE V
ANCHOR POINTS FOR THE LOSS/NOISE TRANSMISSION RATING

Date	Loudness loss, dB	Circuit noise dBrnC	Transmission rating
1976	15	25	80
	30	40	40
1983	10	25	80
	25	40	40

characteristics, such as low volume and and echo, that have been demonstrated in SATs to affect user perception of quality of what is heard. Moreover, because those measures are calculated independently of the average round-trip delay, they can give no indication of possible degradation of usability of service from disruption of normal conversational patterns.

One of the earliest multiple-effects models was the classical loss/noise grade of service model developed at AT&T Bell Labs. The principal motivation for the development of the loss/noise grade of service model was the inadequacy of the measure reflecting the combined effects of loss and noise in use at the time.

The indicator produced by Bell Labs was the loss/noise transmission rating, denoted R_{LN}. It was created by "anchoring" the indicator at two points that were to be used as boundary conditions for fitting other data. The points originally chosen are shown in Table V, together with those eventually adopted as the standard by 1983.

In this scheme of things, any subjective test with mean opinion scores calculate for the 1976 anchor conditions, $\mu[15, 25]$, $\mu[30, 40]$ would, then, be fit to the transmission rating by setting

$$80 = R_{LN}[15, 25] = a + b\mu[15, 25]$$

and

$$40 = R_{LN}[30, 40] = a + b\mu[30, 40]$$

and solving for a and b.

Application of these definitions and fitting the data captured by the Bell Labs experiment produced tha classical loss/noise transmission rating scale:

$$R_{LN} = 147.76 - 2.257[(L_e - 7.2)^2 + 1]^{0.5} - 2.009N_F + 0.02037(L_e)(N_F) \tag{1}$$

where L_e is loudness loss expressed in positive decibels, and N_F is the circuit noise N in dBrnC, increased by the power addition of 27.27 dBrnC.

Eq. (1) illustrates a characteristic found in many multiple-effects models for user perception of quality: the model in general, and the formulas for calculating relevant measures, in particular, are *intuitively obscure*.

Eq. (1), for example, provides us with a painfully precise formula for the calculation of the loss/noise transmission rating that also happens to be completely arcane. The equation does suggest that R_{LN} decreases linearly as loss/noise increase, but the decrease is ameliorated by an amount that gets bigger as both the loudness loss and circuit noise increase. However,

there is absolutely nothing in the structure of that formula that conveys an intuitive notion of how the combination of loudness loss and circuit noise shapes user perception of voice quality. There is no "natural" content in an indicator whose best value is 147.76, rather than some simple value like 100, which might suggest a grading scale.

The basic approach of defining transmission ratings for various different factors expected to affect user perception of voice quality and producing indicators of their combined effects has since been extended to numerous other factors.

The most recent extension of this kind has resulted in the so-called E-model, adopted by the ITU-T in Recommendation G.113 [13] as an acceptable "computational model for use in transmission planning". In this model, the overall transmission rating R is defined by Equation 2:

$$R = R_0 - \text{Is} - \text{Id} - \text{Ie} - \text{A} \tag{2}$$

where

- R_0 = transmission rating based on the signal-to-noise ratio;
- Is = effects of a combination of impairments that "occur more or less simultaneously with the voice signal";
- Id = effects of impairments due to delay;
- Ie = degradation of quality caused by low bit rate codecs;
- A = "compensation of impairment factors when there are other advantages of access to the users".

The extensions and modifications of the original transmission rating factor result in a transmission rating factor whose values now range between 0 and 100 and depend on the values of many measures and indicators.

However, if the descriptor *intuitively obscure* means "hard to grasp at the level of apprehending what is going on", the E-model is intuitively opaque! Its validity must be accepted at face value on the basis of the 30 years of research that have gone into its development and its endorsement as an evolving standard by ITU-T [14].

While the E-model has low inherent credibility, it does have two features that make it attractive as the starting point for producing user-perception models for packet-switched voice services. The first is that the E-model reflects the combined effects of nearly of all factors to affect the incidence and severity of manifestations that shape user perception of quality. The other feature that makes the E-model attractive is that it can easily be extended to include new effects as they are recognized. such ready extensibility means, in particular, that the model can be tailored to incorporate dropped frame rates as necessary to make it applicable to packet-switched voice services.

III. A MATHEMATICAL FORMULATION OF QoS BEHAVIOR

In the preceding section an extensive presentation of voice quality assessment models has been presented. A similar description could have been made for other kinds of services, but due to limitations of space it has been omitted. It has become clear that there is not a final word as far as the best model to grasp all factors that affect the voice quality, with a computational complexity compatible with real time

requirements, is concerned. It is also important to realize that those models are aimed at networking planning. They were not developed envisaging obtaining the best QoS from a network at continuously changing operational conditions. This being so, it is licit to postulate a new conceptual model to be empirically validated.

The transmission of multimedia content concerns three kinds of information: data, audio and video (including static and dynamic images). The traffic generated by each type of information is different, being, respectively, classified as: bursty, Constant Bit Rate (CBR) and Variable Bit Rate (VBR).

The transport of any type of information is designated as a service. Thus, there are data, audio and video services and to each of them is assigned a Quality of Service, expressed as $QoS_{data-service}$, $QoS_{audio-service}$, and $QoS_{video-service}$.

Application is a term used to refer to a high-level software as, for example, a videoconference, a corporative remote training, and so on. These applications, in general, use multimedia content and, by consequence, need data, audio and video services.

When an user runs an application, he or she opens a session that consists of a set services required by the application. A session also has a quality of service assigned to it, and is designated as $QoS_{user-session}$. It is also necessary to consider that several users may be running the same application at the same time. However, this does not mean that all sessions have the same $QoS_{user-session}$, because different users may have different available resources: codecs, available bandwidth, bit error rate on the access channel, etc. This aspect is illustrated on Fig. 1 where three users participate in a teleconference.

User 1	User 2	User 3
Audio: CD	Audio: 64 kbps	Audio: G.709
Video: 384 kbps	Video: NA	Video: 64 kbps
Access: wired	Access: ADSL	Access: wireless
At 2.048 Mbps		
Session 1	Session 2	Session 3
$QoS_{user-session1}$	$QoS_{user-session2}$	$QoS_{user-session3}$

Fig. 1. Example of three users participating in a teleconference

Fig. 1 shows that each user has different media exhibition or reproduction available resources. Besides, each of them may be connected to different access networks and they may have different available networks resources. This being so, in general, for a given application with N simultaneous users, one can expect:

$$QoS_{user-session_i} \neq QoS_{user-session_j}$$

for $i \neq j$. The purpose of ensuring end-to-end QoS can then be stated as a local objective of ensuring the best $QoS_{user-session_i}$, for $1 \leq i \leq N$. In this context, the term application QoS is not used.

The $QoS_{user-session}$ is given by (3):

$$QoS_{user-session} = q\left(QoS_{data-service}, QoS_{audio-service}, QoS_{video-service}\right) \quad (3)$$

where $q(\cdot)$ is a function to be determined.

On the other hand, $QoS_{data-service}$, $QoS_{audio-service}$ and $QoS_{video-service}$ are functions of:

- Packet Delay Transfer (PDT): the time it takes a packet to get across the packet-switched network to its destination;
- Packet Delay Variation (PDV): the variability in PDT;
- Throughput (THRU): the rate of correct packets delivered to their destinations;
- Packet Loss Rate (PLR): the frequency with which packets do not get to their destination in time to be used.

The first step concerning the QoS quantification is to determine each $QoS_{service}$ dependence on the services parameters PDT, PDV, THRU and PLR.

It is reasonable to assume that the QoS for any kind of service follows the curves illustrated in Fig. 2. For example, if we consider an audio service, if the throughput is very low (THRU < V2), refer to the upper part of Fig. 2, it is most likely that the sound will be so often interrupted to become ununderstandable. For this case, the value zero is assigned to the quality of service. At the other end (THRU > V1), the audio reproduction is perfect and the subjective perceived quality does not increase by further increasing the throughput. For this case, the value one is assigned to the quality of service. For throughput values between V2 and V1, we define the QoS transition region. At this point, its exact behavior is not known and, for the sake of simplicity, it is assumed to be linear. The exact values of V1 and V2, and the transition region shape were experimentally determined (see Section III). However, even though the exact shape of the transition region is not linear, it is good enough for the objective of implementing a QoS restoration algorithm.

In the lower part of Fig. 2, it is shown that the QoS decreases as PDT, PDV and PLR increase. If the parameter value is less than V1, the subjective perceived quality does not increase anymore and the value one is assigned to QoS. If the parameter value is greater than V2, the service becomes useless and the value zero is assigned to QoS. It is important to realize that for each parameter (PDT, PDV and PLR) it is necessary to determine the correspondent pair (V1, V2).
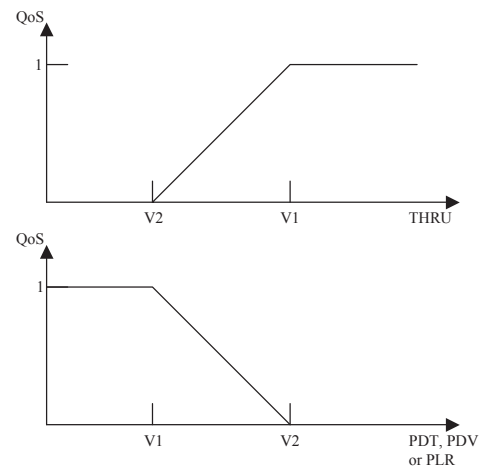

Fig. 2. QoS dependence on THRU, PDT, PDV, and PLR

Similar arguments can be used for video transmission and

the curves presented in Fig. 2 are also representative of this kind os service.

The parameters THRU, PDT, PDV and PLR affect simultaneously the QoS. We propose the conceptual model represented in Fig. 3 to describe the $\text{QoS}_{\text{service}}$ dependence on the services parameters. Such model is empirically validated in this work.
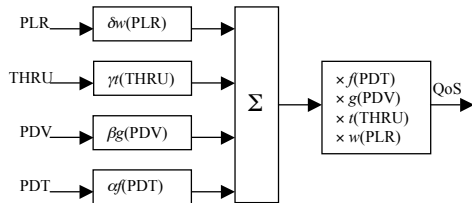

Fig. 3.   Proposed QoS model

According to Fig. 3, the expression for $\text{QoS}_{\text{service}}$ is given by (4):

$$\text{QoS}_{\text{service}} = [\alpha f(\text{PDT}) + \beta g(\text{PDV}) + \gamma t(\text{THRU}) + \delta w(\text{PLR})]$$
$$\times f(\text{PDT}) \times g(\text{PDV}) \times t(\text{THRU}) \times w(\text{PLR}). \quad (4)$$

α, β, γ and δ are positive weighting coefficients in the range $0 \leq \alpha, \beta, \gamma, \delta \leq 1$, and $\alpha + \beta + \gamma + \delta = 1$.

The functions $f(\cdot)$, $g(\cdot)$, $t(\cdot)$ and $w(\cdot)$ are of the type proposed in Fig. 2, and must be determined for each of the services.

In the $\text{QoS}_{\text{service}}$ expression, (4), the additive part represents positive contributions from each parameter to the QoS. However, this contribution is not a complete model. For example, consider a case where THRU = 0. The additive part would still provide a $\text{QoS}_{\text{service}} \neq 0$, that is not possible. So, a product term must be included.

For a multimedia session, following a similar reasoning, the $\text{QoS}_{\text{user-session}}$ could be expressed as seen in (5):

$$\text{QoS}_{\text{user-session}} = A \times \text{QoS}_{\text{data-service}} + B \times \text{QoS}_{\text{audio-service}}$$
$$+ C \times \text{QoS}_{\text{video-service}}, \quad (5)$$

where A, B and C are positive weighting factors in the range $0 \leq A, B, C \leq 1$, and $A + B + C = 1$.

With the proposed model, the evaluation of $\text{QoS}_{\text{user-session}}$ is a simple task, once the $\text{QoS}_{\text{service}}$ of all services is known. The parameters PDT, PDV, THRU and PLR can be dynamically obtained by means of network characterization [15]. Nevertheless, such procedure is not enough to perform a QoS control. In order to do so, it is necessary to identify the variables that are capable of influencing each of the parameters. For example, a THRU decrease may be caused by network congestion, reduction of the available bandwidth, increase of the bit error rate, etc. Among the mentioned variables, some of them are controllable, while others are non-controllable. For example, an increase of the bit error rate in a wireless access scenario may be caused by degradation of radio waves propagation conditions. However, decreasing either the transmission rate

or the packet size it is possible to exercise some control over the bit error rate.

The QoS control can, then, be based on the following steps:

- Session description - identifies all session available resources and the kind of access (wired or wireless);
- Variables identification - identify all variables that affect PDT, PDV, THRU and PLR. The variables are classified into controllable and non-controllable;
- Services modeling - for each required service, and for a particular session, a QoS model is proposed in terms of a finite state machine, as illustrated by Fig. 4.
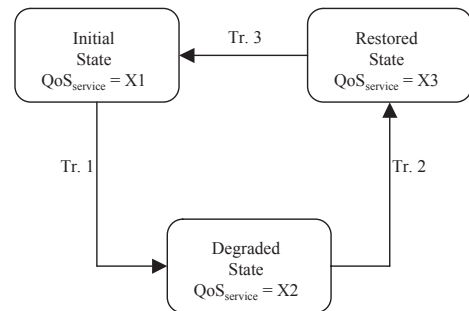

Fig. 4.   Finite state machine-based QoS service model

Other models can also be used as, for example, a stochastic process analysis based on Petri nets, or Markov model [16], [17], but, in this work we have chosen to model the QoS in terms of a finite state machine that monitors the QoS degradation according to the metrics proposed in (4).

In Fig. 4, the Initial State represents the state assumed by the service just after the Call Admission Control (CAC) algorithm has accepted the call. For the Initial State, $\text{QoS}_{\text{service}} = X1$. After a while, the service goes to the QoS Degraded State (transition 1), due to a change of a non-controllable variable. In this state, $\text{QoS}_{\text{service}} = X2 < X1$. As the objective of this work is to ensure the best possible end-to-end QoS, the service model shows that there is a possibility of going to a new state (transition 2), changing the value of a controllable variable. The restoration algorithm is executed and the service enters the QoS Restored State, where $\text{QoS}_{\text{service}} = X3 > X2$. It is most likely that $X3 < X1$. This state is considered a new Initial State. If the network conditions change again, there is a possibility that the service returns to the state where $\text{QoS}_{\text{service}} = X1$ (transition 3). It is also possible to enter a state where $Xn > X1$.

There is no doubt that each different scenario (session) may be represented by different models and, in general, the models are much more complex than that shown in Fig. 4. It is also possible that leaving a QoS degraded state can only be done by changing a non-controllable variable. This situation, however, does not invalidate this proposal because, whenever possible, the best QoS is achieved.

It is also important to mention that different implementations of the QoS controller may be proposed, namely:

- the Qos controller executes all actions without any intervention from the application or end-user;

- the QoS controller notifies the application or end-user of a proposed action and the application/end-user accepts it or not;
- the QoS controller notifies the application or end-user of possible actions and the application/end-user chooses what to do.

IV. QoS Experimental Characterization

A. Methodology

The experimental characterization performed in this work follows the principles of SATs. For all kinds of services evaluated in this work, audio streaming, interactive audio and video streaming, the same steps are taken, namely:

1) A small number of test subjects is chosen: typically four.
2) The test consists on transmitting information under controlled degraded conditions, as described in the following sections.
3) Without rating the service, the test subjects are asked to identify what are the annoying factors of a transmission: interruptions, long delays, blurred images, etc.
4) The test subjects are instructed to rate each experiment in a scale ranging from 0 to 100 (or equivalently from 0 to 1) and to inform the annoying factor according to the list prepared in the preceding step.
5) The experiment is run five times for each test subject in different days and environmental conditions.
6) The average score is computed for each experiment and taking into account the informed annoying factors it is constructed a table relating QoS scores and the corresponding user perceived factors.
7) The experimental data feeds the conceptual model and the parameters are computed by best fitting procedures.

In all experiments the QoS scale range was from 0 to 1, except when the codecs were changed because it was possible to perform a direct comparison among them. In this case, as it will be further explained, a reference condition was chosen as QoS = 1.

This methodology is very easy to be repeated and allows to devise algorithms to be run at the end sides of a connection in order to shape an application's behavior under adverse operational conditions to get the best possible QoS.

B. Testbed Topology

The testbed topology consists of three nodes (computers) numbered from 0 through 2 as shown in Fig. 5. This testbed is used for all experiments concerning QoS degradation.

The functions of each node are:

- node 0 is the traffic source. Traffic is injected by the media server in the form of a stream or by a microphone (interactive audio). In the full-duplex interactive audio case, the sound may be heard by means of a headset or a loud-speaker. At this node, the network analyser captures the packets and stores their sizes, protocols, instant of capture and quantity.
- node 1: at this node is installed the network emulator that can modify the transmission parameters: PDT, PDV, BW,
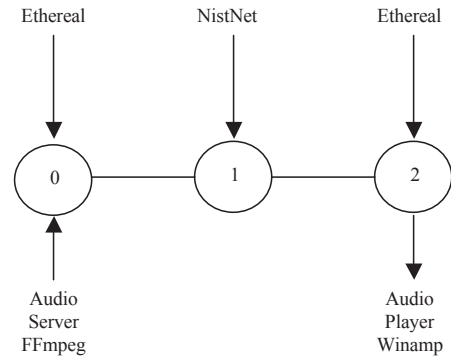

Fig. 5.   Testbed topology used in the QoS experimental characterization

THRU and PLR. It is also possible to define the route the packets will follow.
- node 2 is the traffic destination where the media are played. The audio stream can heard by means of a headset or a loud-speaker. The video stream can be displayed on the computer screen. In the full-duplex interactive audio case, the microphone can inject the audio traffic. At this node, the network analyser performs the same functions as at node 0.

The NistNet [18] and the Ethereal [19] were chosen as network emulator and analyser, respectively. Both are general purpose freeware softwares, widely accepted by the network research community.

C. Characterization Tools

The tools used for the QoS experimental characterization include audio and video servers and players, and network emulator and analyser.

1) The Audio and Video Player: As audio and video player the NullSoft's Winamp [20] has been chosen because it can play and organize tens of media files in audio, video and MP3 CDs or other files. The audio formats are: CDA, MP3, MP2, MP1, AAC, NSA, OGG, WMA, MID, MIDI, RMI, KAR, MIZ, MOD, MDZ, NST, STM, STZ, S3M, S3Z, IT, ITZ, XM, XMZ, MTM, ULT, 669, FAR, AMF, OKT, PTM, WAV, VOC, AU, SND, AIF, AIFF. The video formats are: MPG, MPEG, M2V, AVI, ASF, WMV, NSV. The playlist formats are: M3U, PLS.

2) The Audio and Video Servers: The servers choice has been made among those available in the market place, both freeware and commercial ones as well. It has been observed that some servers show a behavior that is dependent on the network emulator parameters. In some cases, a small variation of the network emulator configuration could produce an abrupt QoS degradation or the transmission would be totally interrupted. So far, it has not been found a reasonable explanation for this interaction that prevents a correct QoS experimental characterization.

As interactive audio server, the choice has fallen on the SJPhone [21]. It allows to communicate with any PC, PDA and IP phone. It provides support for SIP and H.323 protocols, and full-duplex communication. Besides, it presents an initial

delay that is compatible with those accepted in telephony, it keeps the connection up while a network QoS degradation is emulated by the application of restrictions on the transmission parameters imposed by the NistNet network emulator.

As stream audio server, the choice has fallen on the FFmpeg [22] that is a component of the FFmpeg Streaming Multimedia System along the FFserver. The FFmpeg Streaming Multimedia System is a complete audio and video diffusion system through the Internet over any Linux/Unix platform. It can code different formats in real time: MPEG1, MPEG4, H.263, AC3, ASF, AVI, REAL, MJPEG AND FLASH. It was chosen because it allows MP3 recording in different coding rates and it keeps the connection up while a network QoS degradation is emulated by the application of restrictions on the transmission parameters imposed by the NistNet network emulator.

As stream video server, the choice has fallen on the Windows Media Server 2003 (WMS) [23]. It is the server component of the Windows Media and allows to distribute digital midia content through corporate Intranets and the Internet. It also provides support for transforming audio and video files to its own WMV format. It was chosen because it allows to choose the desired transport protocol for different coding rates, for the same screen aspect ratio without interrupting the transmission when the transmission parameters are modified by the NistNet.

D. Audio Stream: Scenario Characterization and QoS Metrics Assignment

The main objective is to identify the parameters controllable by the end-user that allow to restore a degraded QoS to a better level. So, it is necessary to establish criteria to assign objective values to a degraded QoS, i. e., a metrics; to experimentally verify the QoS dependence on the transmission parameters: PDT, THRU, PDV, PLR and BW; to determine the values of V1 and V2 (refer to Fig. 2), and the intermediate values of QoS according to the chosen criteria.

As already mentioned, the FFmpeg has been chosen for the audio stream characterization. The tests conditions were:

- transport protocol: TCP;
- RTP and RTSP protocols were not implemented;
- signaling protocol: SIP;
- audio coding: MPEG.

For the QoS quantification in the audio stream scenario the number of interruptions (clippings) in the received audio was considered by the average listener as the main parameter that determines the quality. The initial delay after the emitter has initiated the transmission was not considered a major issue.

These tests are initiated by a subjective perception of an CD quality audio stream reception under the best listening conditions. Next, the condition of transmission of the link between the server (node 0) and the client (node 2) is degraded modifying the PDT, PDV, PLR and BW parameters by means of the NistNet, and changing the datagrams coding rate. A subjective evaluation of the new reception condition is made and, at the same time, an objective equivalence of the audio degradation is proposed based on measured parameters that were modified as consequence of the new network conditions.

The emulator program was initially configured with PDT, PDV and PLR equal to zero, and without any bandwidth (BW) limitation. In this case, the program is not interfering on the network parameters. The audio player Winamp has been set to work without any buffer. However, a residual buffer of 16 kB is always present. These settings provide the reference QoS = 100%.

The following tests imposed limitations on PDT and BW. These are the parameters most affecting the QoS.

The number of interruptions (INTERR) in 30 s as function of BW is easy to be counted (see Fig. 6) and was considered the main variable to establish a QoS metrics that assigns a grade between 0 and 100% to a subjective QoS perception.
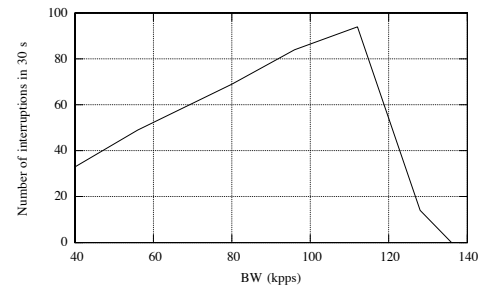

Fig. 6.   Number of interruptions as function of BW

In Fig. 6, it can be observed that narrower is the bandwidth, smaller is the number of interruptions in 30 s, and they increase up to a BW equal to 112 kbps. However, in this region, less interruptions also imply a longer time without any sound, and by consequence a worse user perceived QoS. For BWs larger than 112 kbps, the number of interruptions decreases again, improving the quality of the received audio, as the time intervals with sound also increase. For BW $\geq$ 136 kbps the audio is not interrupted anymore, maintaining QoS at 100%.

When assigning a QoS metrics to characterize the dependence on the BW, it was perceived a linear correlation between the number of interruptions in the received audio, the BW and QoS subjective perception. Such relationship is shown in Table VI.

The QoS metrics assignment to the dependence on BW is illustrated by Fig. 7. It can be seen that the dependence of QoS on BW is pratically linear. This same behavior was observed in the dependence of QoS on THRU.
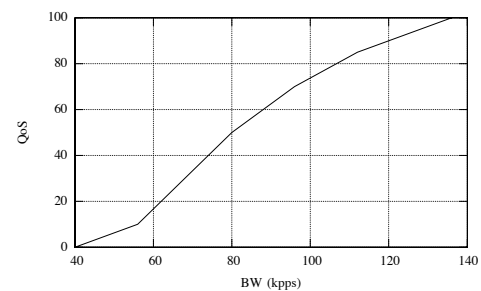

Fig. 7.   QoS dependence on BW for the audio stream scenario

The dependence of QoS on PDT is shown in Fig. 8

TABLE VI
METRICS ASSIGNMENT CHARACTERIZING THE DEPENDENCE OF AUDIO
STREAM QoS ON BW

BW (kbps)	INTERR in 30 s	QoS (%)	Subjective Perception
40	33	0	Sound not understandable; very short time intervals with sound between interruptions.
56	49	10	Sound begins to be recognized.
80	69	50	Sound is identified but there are long silence intervals between interruptions.
96	84	70	Shorter silence intervals but too many interruptions.
112	94	85	Very short silence intervals and many interruptions.
128	14	95	Long time intervals with sound and few interruptions.
136	0	100	Sound without interruptions and nice to be heard.

(solid line). This curve has been inferred from the number of interruptions and the quality subjective perception of the sound when PDT was varied by the NistNet emulator, keeping BW without any limitation. PLR and PDV were also set to zero. PDT was varied in the range from 0 ms (QoS = 100%) thru 4500 ms (QoS = 0), where the sound was not recognized anymore due to the excess of the number of interruptions. Fig. 8 also shows a possible fitting by an exponential curve (dashed line), and a linear approximation that is used by the restoration algorithm (dotted line). This latter approximation is better explained later.
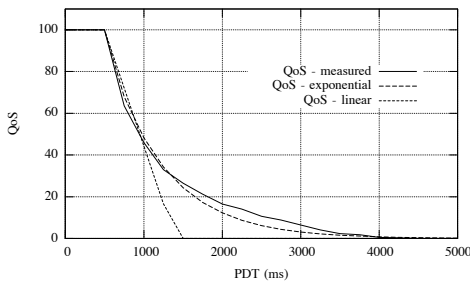
Fig. 8.   QoS dependence on PDT for the audio stream scenario; BW without any limitations and PDV and PLR set to zero

Fig. 9 shows QoS level curves as function of PDT and THRU. These curves show that the PDT severely limits the maximum achievable QoS. For example, for a PDT = 800 ms, QoS is limited to 50%, no matter how much network bandwidth is available. This is a very important conclusion because the available bandwidth (ABW) is not dependent only on the link speed, but is a function of the interfering network load, that is not always under complete control of the service provider.

Fig. 10 is another way of illustrating the dependence of QoS on PDT for six different values of BW in the range from
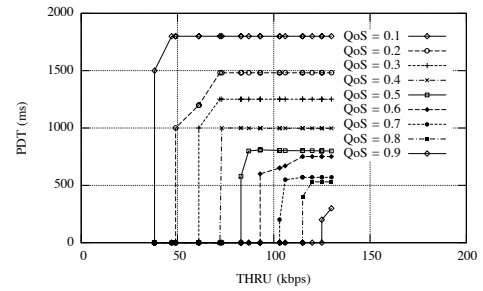

Fig. 9.   QoS level curves as function of PDT and THRU

28 to 136 kbps. It can be observed that for BW = 48 kbps, QoS is constant and equal to 20% for $0 \leq PDT \leq 1500$ ms. However, for BW = 136 kbps, QoS maintains a level of 100% for $0 \leq PDT \leq 300$ ms. In other words, when designing a QoS restoration algorithm one should consider what is better, in terms of subjective perception of the end-user:

- obtain the best possible QoS, that may vary very much if the network transmission conditions are continuously changing due to the interfering load;
- keep a moderate QoS, that may be stable even under a wide range of transmission conditions variations due to the interfering load.

It is also important to realize that all curves converge to a limiting boundary, beyond which it is not possible to work.
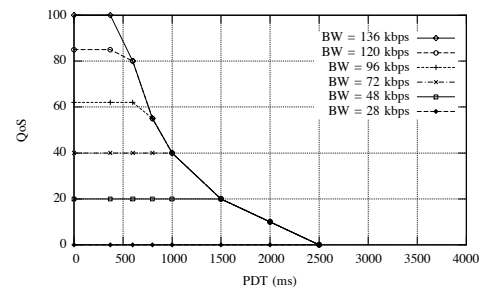

Fig. 10.   Dependence of QoS on PDT for six different values of BW in the range from 28 to 136 kbps

The FFmpeg allows to code the audio stream at different rates. In this experiment, coding rates between 32 kbps and 256 kbps were used; 128 kbps was considered the reference for QoS = 100%. Fig. 11 shows the experimental results. The leftmost curve corresponds to a code rate of 32 kbps and the rightmost curve corresponds a code rate of 256 kbps.

It can be seen that for coding rates lower than 128 kbps, the QoS improves faster for smaller increments of THRU, but the maximal achieved values are less than 100%. For example, for THRU = 40 kbps, an audio stream coded at 32 kbps achieved QoS = 90%. The same audio stream coded at 128 kbps, for the same THRU, achieved a very low QoS equal to 12%.

It can also be observed that for coding rates higher than 128 kbps, QoS = 100% demands higher values of THRU, maintaining the linear relationship between QoS and THRU. Thus, an audio stream coded at 256 kbps needs a THRU of 272 kbps to achieve QoS = 100%, twice the value needed for a 128 kbps coding rate.
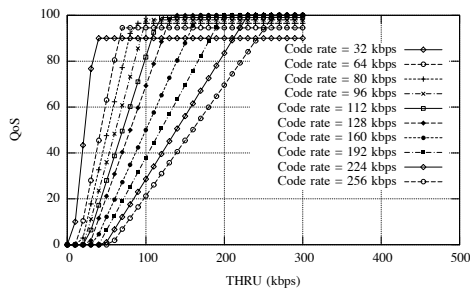
Fig. 11. Dependence of QoS on THRU for different coding rates of the audio stream

E. Interactive Audio: Scenario Characterization and QoS Metrics Assignment

As already mentioned, the SJPhone has been chosen for the interactive audio characterization. The tests with it provided the following data that define the boundary conditions for the QoS metrics assignment for this scenario:

- transport protocol: UDP;
- RTP and RTSP protocols were not implemented;
- signaling protocol: SIP;
- audio coding: G.723.1.

For the QoS quantification in the interactive audio scenario the number of interruptions (cuts) in the received audio and the initial delay to hear the audio after the speaker has begun to transmit were considered by the average listener as the main parameters that determine the quality.

The tests were performed with the speaker seeing the listener, but in a real environment, where there isn't such visualization, the speaker may interpret this initial delay as a loss of connection or as a listener's misunderstanding and start to repeat when the answer is up to arrive, provoking an undesirable overlap.

Fig. 12 shows the initial delay to start audio reception as function of BW. It can be seen that for $BW \geq 29.6$ kbps the initial delay is zero and the audio reception is continuous. When BW is decreased to 29.1 kbps, the delay increases to 530 ms and the audio perception is not confortable. For $BW = 29.0$ kbps the delay is equal to 840 ms and the sound begins to be cut. For $BW = 28.8$ kbps the delay is equal to 3 s and the sound has too many interruptions. For $BW = 28.0$ kbps the delay reaches 10 s and the voice is not understandable anymore.
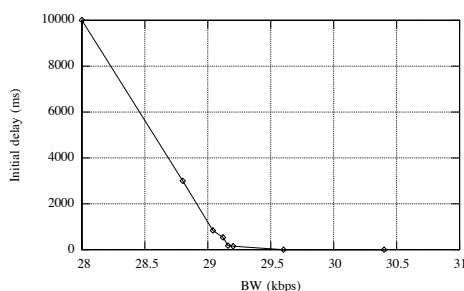

Fig. 12. Initial delay for audio reception as function of BW

TABLE VII
METRICS ASSIGNMENT CHARACTERIZING THE DEPENDENCE OF
INTERACTIVE AUDIO QoS ON BW

BW (kbps)	Initial delay (ms)	QoS (%)	Subjective Perception
28.0	10000	0	Initial delay extremely long and unconfortable. Sound with many cuts and not understandable.
28.8	3000	20	Initial delay extremely long and unconfortable. Sound with many cuts.
29.0	840	30	Initial delay still long and unconfortable. Sound with short cuts.
29.1	530	60	Initial delay starts to bother. Sound without cuts.
29.3	150	90	Short initial delay. Normal sound.
29.6	0	100	Sound without interruptions and nice to be heard.

Fig. 13 shows that there is a linear relationship between THRU and BW up to $BW = 29.0$ kbps. From this value on, BW is not a limiting factor anymore.
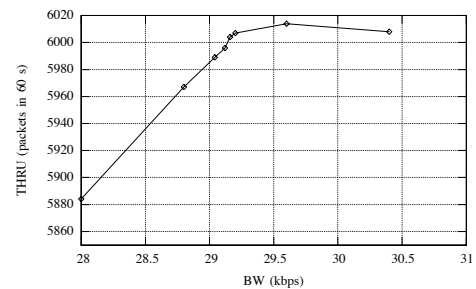

Fig. 13. THRU as function of BW

Table VII introduces the metrics assignment characterizing the dependence of interactive audio QoS on BW. Fig. 14 shows the QoS dependence on BW.
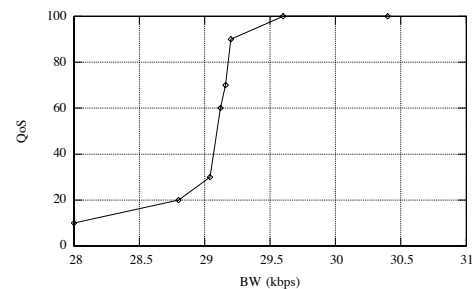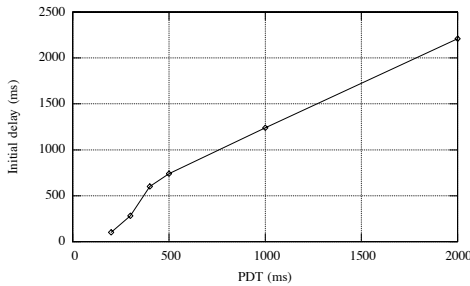

Fig. 14. QoS as function of BW

The characterization of the interactive audio QoS dependence on PDT has also shown that the initial delay for hearing the sound was the most important parameter that provokes discomfort on the listener. In this scenario, clippings in sound reception were not perceived. Fig. 15 shows the initial delay as function of the PDT imposed by the network emulator. It

TABLE VIII

METRICS ASSIGNMENT CHARACTERIZING THE DEPENDENCE OF INTERACTIVE AUDIO QoS ON PDT

PDT (ms)	Initial delay (ms)	Queue size (bytes)	QoS (%)	Subjective Perception
2000	2210	100	0	Initial delay extremely long and unconfortable. The queue in the router is very long when compared with the minimum registered value. After the initial delay, the sound is normal.
1000	1240	60	20	Initial delay extremely long and unconfortable. After the initial delay, the sound is normal.
500	740	26	50	Initial delay still long and unconfortable. After the initial delay, the sound is normal.
400	600	20	70	Initial delay starts to bother.
300	280	12	90	Short initial delay. After the initial delay, the sound is normal.
200	100	10	100	The sound is normal after an almost unnoticeable initial delay.

Fig. 16. QoS as function of PDT

Fig. 17. QoS as function of the router's queue size

can be observed that for an imposed PDT = 200 ms the initial delay perceived till the reception begins is perceived as null and there are no sound interruptions. For PDT = 2000 ms the initial delay is very uncomfortable, even if the sound remains continuous afterwards. Tests with PDT values much higher than those accepted by telephone systems were performed because they may occur under heavy traffic loads that may arise in cases of disaster or social disturbance.

Fig. 15. Initial delay as function of PDT

Table VIII introduces the metrics assignment characterizing the dependence of interactive audio QoS on PDT and on the router queue size. Fig. 16 shows the QoS dependence on PDT and Fig. 17 shows the QoS dependence on the router queue size. The router queue size may be an interesting parameter to be monitored to start a QoS restoration algorithm. It may be observed that Figs. 16 and 17 present almost the same QoS behavior as function of the forementioned parameters.

Table IX introduces the metrics assignment characterizing the simultaneous dependence of interactive audio QoS on PDT and PDV and the QoS behavior is shown in Fig. 18. It has been observed that the PDV has a much bigger impact on QoS
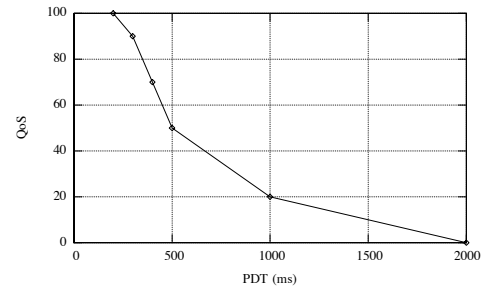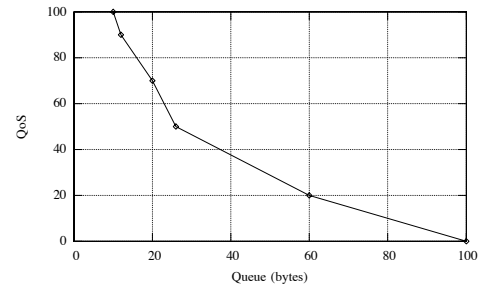
than the PDT, even for small PDV values. For PDV $\geq$ 50 ms the speach is not understandable for any value of PDT. As PDT increases, for a constant PDV value, the QoS improves.

F. Video Stream: Scenario Characterization and QoS Metrics Assignment

As already mentioned, the WMS has been chosen for the video stream characterization. The adopted configuration was: 640 x 480 pixels screen format, 4:3 aspect ratio and 25 frames per second. The video stream had 4 minutes and 33 seconds duration and all tests were 60 seconds long. The tests with it provided the following data that define the boundary conditions for the QoS metrics assignment for this scenario:

TABLE IX

METRICS ASSIGNMENT CHARACTERIZING THE SIMULTANEOUS DEPENDENCE OF INTERACTIVE AUDIO QoS ON PDT AND PDV

| PDT (ms)	PDV (ms)	Initial delay (ms)	QoS (%)	Subjective Perception
100	10	200	95	Short distortions
100	20	200	70	Understandable with many cuts
100	40	200	30	Many words are not understandable
100	50	200	0	Completely non-understandable
200	10	330	95	Short distortions
200	20	330	75	Understandable with a few cuts
200	40	330	35	Several words are not understandable
200	50	330	0	Completely non-understandable
400	10	440	95	Short distortions
400	20	440	80	Understandable with a few cuts
400	40	440	40	Few words are not understandable
400	50	440	0	Completely non-understandable
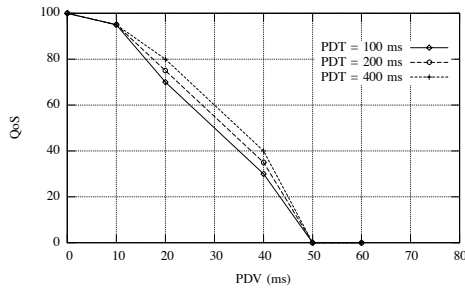
Fig. 18.   QoS simultaneous dependence on PDT and PDV

TABLE X

METRICS ASSIGNMENT CHARACTERIZING THE DEPENDENCE OF VIDEO STREAM QoS ON THRU

THRU (kbps)	QoS (%)	Subjective Perception
1344	0	Frame-by-frame reproduction and completely unconfortable.
1360	70	Too many instabilities and cuts in the video reproduction.
1380	80	Short instabilities
1460	95	Almost normal.
1536	100	Normal.

- transport protocol: UDP and MMS over UDP;
- real time protocol: RTSP;
- video coding: proprietary format with .wmv extension;

For the QoS quantification in the video stream scenario the intermitence in the received video and the initial delay to see the video after the receiver has asked for a transmission were considered by the average watcher as the main parameters that determine the quality. Other parameters made available by the NistNet, router queue and BW increases, were also considered in the QoS metrics assignment.

Table X introduces the metrics assignment characterizing the dependence of video stream QoS on THRU, for 1.5 Mbps coding rate. For other coding rates, similar criteria were used. Fig. 19 shows the video stream QoS dependence on THRU. It is important to realize that for 768 kbps and 1.0 Mbps the maximum achieved QoS was less than 1 due to a perceived color degradation.
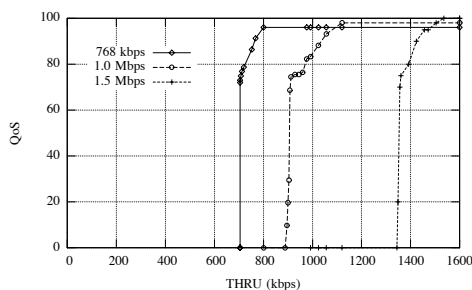

Fig. 19.   QoS as function of THRU for different coding rates

The video stream QoS dependence on PDT has also been characterized and the results are shown in Fig. 20 for 768

kbps coding rate. This characterization was made for three conditions concerning the available bandwidth: no limitation (curve designated as free), 800 and 784 kbps. The PDT range was far beyond the usual values we expect to face in normal conditions, but that may be the case for exceptional situations of catastrophes and social comotion.
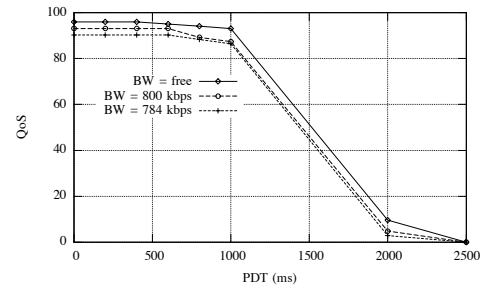

Fig. 20.   QoS as function of PDT for 768 kbps coding rate and three different available bandwidths conditions

Fig. 21 shows the QoS level curves for simultaneous variations of THRU and PDT, and 768 kbps coding rate.
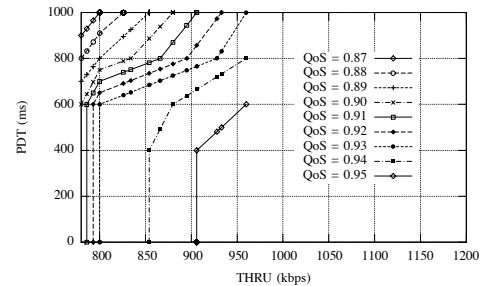

Fig. 21.   QoS level curves for simultaneous variations of THRU and PDT, and 768 kbps coding rate

The QoS dependence on the simultaneous variations of PDV and PDT has also been characterized, without imposing any restriction on the available bandwidth. Fig. 22 shows the QoS level curves for simultaneous variations of PDV and PDT, and 768 kbps coding rate. For all tests PDV is always less than PDT. For small PDV and PDT values, the QoS remains at its maximum level.
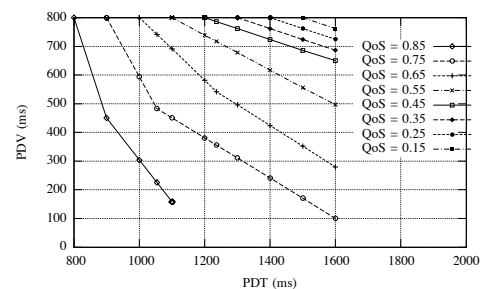

Fig. 22.   QoS level curves for simultaneous variations of PDV and PDT, and 768 kbps coding rate

V. QoS MODEL VALIDATION

In Section IV has been presented the QoS metrics assignment and characterization for the audio stream scenario. Fig.

7 shows the dependence of QoS on THRU, what should correspond to the function $t(\text{THRU})$. Fig. 8 shows the dependence of QoS on PDT, what should correspond to the function $f(\text{THRU})$. Fig. 9 shows the dependence of QoS when THRU and PDT are simultaneously changed. It was experimentally verified that the dependence of QoS on PDV was not important, from the user subjective perception for the audio stream scenario. Due to limitations of the experimental setup available, it was not possible to act on PLR (PLR = 0 for all experiments). However, for normal transmission conditions, i. e., low PLR values, this parameter has almost no impact on QoS for audio stream reproduction.

According to the former comments, equation (4) can be simplified to the form given in (6):

$$\text{QoS}_{\text{service}} = [\alpha f(\text{PDT}) + \gamma t(\text{THRU})] \times f(\text{PDT}) \times t(\text{THRU}). \tag{6}$$

In order to verify if the proposed model is an adequate representation of the experimentally observed QoS behavior, it is first necessary to determine the values of V1 and V2. The V1 values are easy to be characterized because it is not difficult to note when the audio stream reception starts to present to some degradation, not being perfect anymore. So, these values have been obtained directly from the experimental observation. On the other hand, the V2 values determination is not easy at all. When the QoS values are very low, it is very difficult to establish an objective criterium for the beginning of the QoS = 0 region. For this reason, it was decided to impose that the linear approximation of the QoS transition region should pass through the experimental QoS = 0.5 point. Taken this into consideration, the V2 parameter is calculated from the obtained straight line and its crossing with the horizontal axis. The V1 and V2 values for the $f(\text{PDT})$ and $t(\text{THRU})$ functions are the following:

- $f(\text{PDT})$: V1 = 500 ms, V2 = 1400 ms;
- $t(\text{THRU})$: V1 = 128 kbps, V2 = 24 kbps.

The model proposed in (6) has been evaluated for different values of the α and γ coefficients. Figs. 23, 24 and 25 illustrate the results.
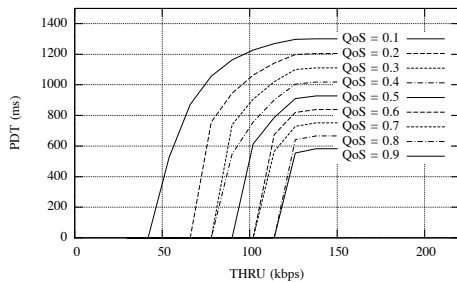

Fig. 23.   QoS level curves modeled by (6) for $\alpha = 0.1$ and $\gamma = 0.9$

Observing Figs. 23, 24 and 25, the following remarks can be made:

- the proposed model exhibits a behavior that is similar to the experimental results;
- for QoS ≥ 0.5, the adjusted curves follow quite well the experimental behavior;
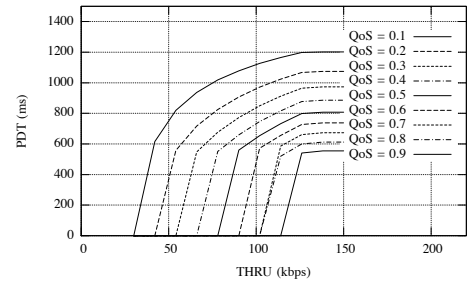

Fig. 24.   QoS level curves modeled by (6) for $\alpha = 0.7$ and $\gamma = 0.3$
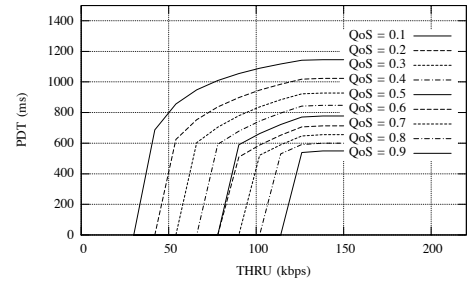

Fig. 25.   QoS level curves modeled by (6) for $\alpha = 0.9$ and $\gamma = 0.1$

- looking at the curve for which QoS = 0.5, the point corresponding to PDT = 0 varies very much depending on the specific values of α and γ, i. e., depending on the relative weight that is given to PDT or THRU;
- the values of $\alpha = 0.7$ and $\gamma = 0.3$ are those that best mimic the experimental behavior. Fig. 26 compares the experimental and modeled curves for QoS = 0.5 and QoS = 0.8.
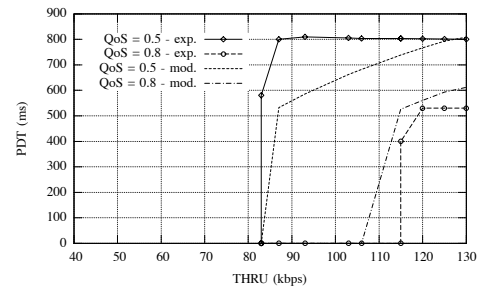

Fig. 26.   QoS level curves: experimental *vs.* modeled (6) for QoS = 0.5 and QoS = 0.8 ($\alpha = 0.7$ and $\gamma = 0.3$)

For the interactive audio scenario the appropriate model is given by (7):

$$\text{QoS}_{\text{service}} = [\alpha f(\text{PDT}) + \beta g(\text{PDV})] \times f(\text{PDT}) \times g(\text{PDV}). \tag{7}$$

The V1 and V2 values for the $f(\text{PDT})$ and $g(\text{PDV})$ functions are the following:

- $f(\text{PDT})$: V1 = 200 ms, V2 = 2000 ms;
- $g(\text{PDV})$: V1 = 10 ms, V2 = 50 ms.

Figs. 27, 28 and 29 show the QoS experimental and modeled behavior for PDT equal to 100, 200 and 400 ms, respectively.

It can be seen that for all practical purposes, the modeled behavior tracks reasonably well the experimental one. The best results are obtained for small values of PDT.
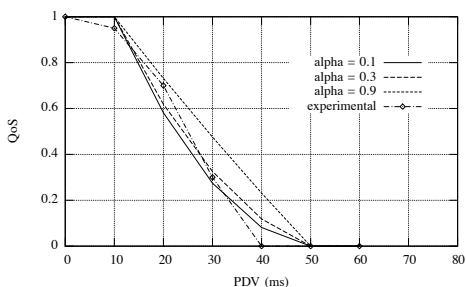

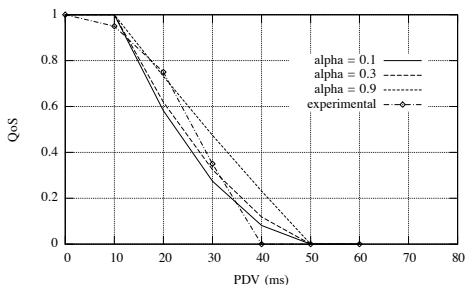Fig. 27. Experimental and modeled QoS behavior for PDT = 100 ms

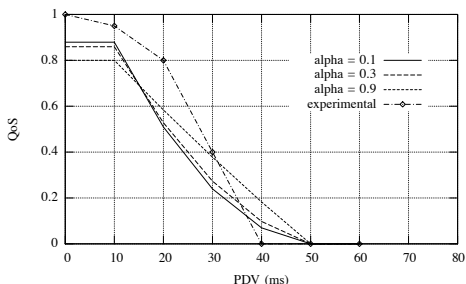Fig. 28. Experimental and modeled QoS behavior for PDT = 200 ms

Fig. 29. Experimental and modeled QoS behavior for PDT = 400 ms

For the video stream scenario the appropriate model is given by (8):

$$
\begin{aligned}
\mathrm{QoS}_{\mathrm{service}} = {} & [\alpha f(\mathrm{PDT}) + \beta g(\mathrm{PDV}) + \gamma t(\mathrm{THRU})] \\
& \times f(\mathrm{PDT}) \times g(\mathrm{PDV}) \times t(\mathrm{THRU}).
\end{aligned} \tag{8}
$$

The V1 and V2 values for the $f(\mathrm{PDT})$, $g(\mathrm{PDV})$ and $t(\mathrm{THRU})$ functions are the following:

- $f(\mathrm{PDT})$: V1 = 900 ms, V2 = 2500 ms;
- $g(\mathrm{PDV})$: V1 = 300 ms, V2 = 700 ms;
- $t(\mathrm{THRU})$: V1 = 785 kbps, V2 = 754 kbps.

Fig. 30 shows the QoS level curves modeled for $\alpha = 0.7$, $\beta = 0.0$ and $\gamma = 0.3$. It can be seen that for all practical purposes, the modeled behavior tracks reasonably well the experimental one. It is worth remarking that the experimental results show a positive slope trend of PDT as THRU increases, while the modeled behavior presents a saturation of PDT. This

happens because the video server employs a dejitter buffer that is not being modeled.
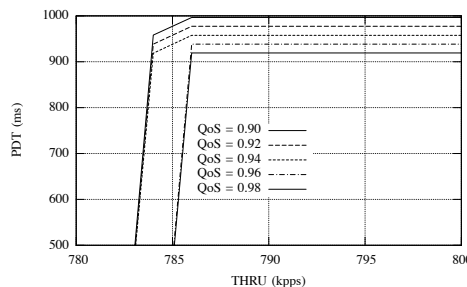

Fig. 30. Modeled QoS behavior for video stream considering $\alpha = 0.7$, $\beta = 0.0$ and $\gamma = 0.3$

VI. A QoS Restoration Algorithm

A possible QoS restoration algorithm flowgraph is depicted in Fig. 31. The general structure of this algorithm can be used for QoS restoration of any kind of service, but the description will focus on the audio stream case. Each box in the figure is identified by a number on its bottom right corner. Boxes 5, 9 and 10 are decision boxes. The algorithm works as follows:

- Box 1 is the *Initial State* where the service's QoS is QoS1;
- Box 2 continuously monitors the networks parameters PDT, PDV, THRU and PLR, and forwards them to Box 3;
- Box 3 evaluates the functions $f(\mathrm{PDT})$, $g(\mathrm{PDV})$, $t(\mathrm{THRU})$ and $w(\mathrm{PLR})$. In fact, in the audio stream case, only the functions $f(.)$ and $t(.)$ are used;
- Box 4 evaluates the new QoS, i. e., QoS2 using the α and γ obtained experimentally;
- Box 5 compares QoS1 and QoS2. If QoS2 $\geq$ QoS1 there is nothing to be done and the algorithm goes back to Box 2. Otherwise, a restoration must be attempted and the algorithm proceeds to Box 6;
- Box 6 implements the change of a controllable variable to improve the QoS. In this example, only the audio stream coding is being used. The algorithm searches a table to find a better coding;
- Box 7, using the new coding parameters, re-evaluates $f(.)$ and $g(.)$;
- Box 8 evaluates the new possible QoS, i. e., QoS3 if the new coding is employed;
- Box 9 compares QoS3 and QoS2. If QoS3 $\geq$ QoS2 the new coding is accepted and the algorithm proceeds to Box 11. Otherwise, the new coding is not accepted and the algorithm proceeds to Box 10;
- Box 10 checks if the coding table end has been reached. In the affirmative case, QoS cannot be restored and the algorithm goes back to Box 2 (QoS = QoS2 < QoS1). Otherwise, the algorithm goes back to Box 6 to try another coding;
- Box 11 restores the QoS: QoS = QoS3 > QoS2. It is most likely that QoS3 < QoS1. The algorithm returns to the *Initial State*.
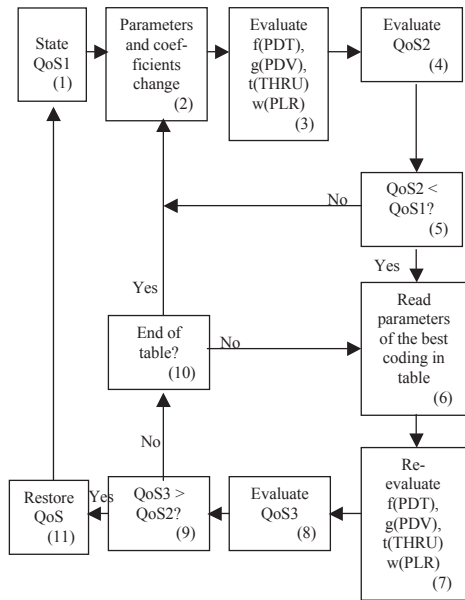
Fig. 31. A possible QoS restoration algorithm

VII. CONCLUSIONS AND FUTURE WORK

In this paper it was proposed a conceptual model to quantify the QoS variation as a result of changes in THRU, PDT, PDV and PDT, that enabled a QoS restoration algorithm to be devised. This algorithm acts only on the controllable transmission variables.

The QoS variation was experimentally quantified. This quantification is not simple, because it represents the translation of a subjective perception into an objective numerical value. Besides, during the metrics assignment it was necessary to associate the QoS value to measurable values of the trasmission system parameters. This being so, an extensive characterization work has been performed. The transmission parameters were changed in a controlled way by the NistNet emulator. Three scenarios were characterized, namely:

- audio stream transmission;
- video stream transmission;
- interactive audio transmission.

The experimental environment setup was very important to provide repetitive and reliable results. The audio and video servers software was also critical, because it has been observed that an interaction between them and the NistNet may occur, leading to strange behavior or interrupting completely the transmission. So far, a satisfactory explanation for this interaction has not been found. Nevertheless, the environment has been setup to minimize the influence of the used programs on the characterization results.

It is also not possible to claim that the obtained results have statistical validity. This would require to rely on a big population of users to grade the QoS and the experimental environment should be better controlled. However, it is licit to state that a more detailed work could lead to minor numerical variations that would not invalidate the proposed QoS restoration method.

The QoS variation has been modeled by a piecewise linear function characterized by two parameters: V1 and V2. This model has been validated by the experimental charaterization, for all three scenarios. It has been verified that the model reproduces with excellent approximation the experimental results in the QoS transition region of most practical interest, i. e., $0.5 < QoS \leq 1$.

At last, as an example, a possible QoS restoration algorithm was shown for the audio stream scenario, using the coding rate as the control variable.

According to the best of our knowledge, the proposal of this work is quite original. The references show that most of researchers adopt ON-OFF approaches, i. e., either there is QoS or there is not. The published results in terms of soft QoS are still preliminary.

On the other hand, the results presented in this paper are still seminal, creating new research opportunities. Thus, some of possible future works are:

- perform a new characterization of the QoS dependence on transmission parameters substituting the NistNet by an interfering traffic generator;
- practical implementation of the QoS restoration algorithm an its deployment in the test environment;
- identification of other transmission control variables and, consequently, the development of a new restoration algorithm;
- extension of the ideas developed for $QoS_{service}$ to $QoS_{session}$, to optimize the multimedia traffic transmission;
- characterization of other scenarios, e. g., interactive video transmission and wireless networks;
- evaluate the efficacy of the ideas presented in this work in networks that provide some form of controlling the QoS, e. g., DiffServ and MPLS.

REFERENCES

[1] P. K. Sharma, J. Loyall, R. E. Schantz, J. Ye, P. Manghwani, M. Gillen, and G. T. Heineman, "Managing end-to-end qos in distributed embedded applications," *IEEE Internet Computing*, vol. 10, no. 3, pp. 16 – 23, may 2006.

[2] V. Bhat, M. parashar, H. Liu, M. Khandekar, N. Kandasamy, and S. Abdelwahed, "Enabling self-managing applications using model-based online control strategies," in *IEEE International Conference on Autonomic Computing, ICAC 06*, june 2006, pp. 15 – 24.

[3] M. F. Moreno, S. Colcher, and L. F. G. Soares, "Adaptable resource management based on the virtual resource tree model," in *International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies*, april 2006, pp. 27 – 27.

[4] A. Lopes, F. Amaro, G. Elias, G. Lemos, and M. F. Magalhaes, "Qos specification and management in a middleware for distributed multimedia systems," in *20th International Conference on Advanced Information Networking and Applications, AINA 2006*, april 2006, pp. 959 – 964.

[5] Y. Yang and M. H. Williams, "Handling dynamic qos requirements in a pervasive system," in *International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies*, april 2006, pp. 11 – 11.

[6] Y. Bai and M. R. Ito, "A study for providing better quality of service to voip users," in *20th International Conference on Advanced Information Networking and Applications, AINA 2006*, april 2006, pp. 799 – 804.

[7] D. Picovici, A. Raja, and C. Flanagan, "Real-time non-intrusive voip evaluation using second generation network processor," in *2006 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006 Proceedings*, may 2006, pp. 1208 – 1211.

[8] G. Casale, P. Cremonesi, G. Serazzi, and S. Zanero, "Performance issues in video streaming environments," in *2005 Workshop on Techniques, Methodologies and Tools for Performance Evaluation of Complex Systems*, september 2005, pp. 3 – 14.

[9] W. C. Hardy, *VOIP Service Quality: measuring and evaluating packet-switched voice*. U.S.: McGraw-Hill Networking, 2003.

[10] "Itu-t recommendation p.800, methods for subjective determination of transmission quality, 1996."

[11] "Itu-t recommendation p.862, perceptual evaluation of speech quality (pesq): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001."

[12] "Itu-t recommendation p.861, objective quality measurement of telephone-band (300 - 3400 hz) speech codecs, 1996."

[13] "Itu-t recommendation g.113, transmission impairments due to speech processing, 2001."

[14] "Itu-t recommendation g.107, the e-model, a computational model for use in transmission planning, 2005."

[15] N. Thio and S. Karunasekera, "Automatic measurement of a qos metric for web service recommendation," *Proceedings of the 2005 Australian Conference on Software Engineering*, pp. 202–211, 2005.

[16] C. C. Wust and et al., "Qos control strategies for high-quality video processing," *Real-Time Systems*, vol. 30, may 2005.

[17] F. I. V. Martin, "Modelling an adaptive-rate video-streaming service using markov-rewards models," *QSHINE-2004 First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, october 2004.

[18] "Nistnet network emulator - version 2.0.12," 2003, available in: URL:http://snad.ncsl.nist.gov/itg/nistnet; accessed on 03/21/2003.

[19] "Ethereal - network protocol analyser," 2003, available in: URL:http://www.ethereal.com; accessed on 03/22/2003.

[20] "Winamp audio and video player," 2003, available in: URL:http://www.winamp.com/support/help/50/Winamp5.htm; accessed on 06/01/2003.

[21] "Sjphone voip client software," 2004, available in: URL:http://www.sjlabs.com; accessed on 10/05/2004.

[22] "Ffmpeg audio and video server," 2004, available in: URL:http://sourceforge.net; accessed on 10/05/2004.

[23] "Windows media server - audio and video server," 2004, available in: URL:http://www.microsoft.com; accessed on 06/03/2004.

Carlos Rebolledo Barra graduated in electrical engineering from the Chile State Technical University (UTE today Usach) in 1970. He received the MSc and PhD from the São Paulo University, Polytechnic School (EPUSP) degrees in 1999 and 2005 respectively. His research interests are in the area of wired and wireless networks, quality of service (QoS) and remote learning.

José Roberto Amazonas graduated in electrical engineering from the Escola Politécnica of the University of São Paulo (Epusp), Brazil, in 1979. He received the MSc, PhD and postdoctoral degrees from Epusp in 1983, 1988 and 1996, respectively.

He is associate professor of the Telecommunications and Control Engineering Department at Epusp, where he is in charge of optical communications and high-speed communications networks education and research. He held various positions in universities in Brazil and Europe. He has also led research in partnership with several brazilian, european and north-american companies.

His research interests are in the area of optical communications, wired and wireless networks, quality of service (QoS) and remote learning.