IIF-WT-066R4

Working Text on

TECHNICAL REPORT – VALIDATION PROCESS FOR IPTV PERCEPTUAL QUALITY MEASUREMENTS

Alliance for Telecommunications Industry Solutions

Approved Month DD, YYYY

Abstract

Abstract text here.

FOREWORD

The Alliance for Telecommunication Industry Solutions (ATIS) serves the public through improved understanding between carriers, customers, and manufacturers. The IPTV Interoperability Forum (IIF) enables the interoperability, interconnection, and implementation of IPTV systems/services by developing ATIS standards and facilitating related technical activities. This forum will place an emphasis on North American and ATIS Member Company needs in coordination with other regional and international standards development organizations.

Suggestions for improvement of this document are welcome. They should be sent to the Alliance for Telecommunications Industry Solutions, IIF Secretariat, 1200 G Street NW, Suite 500, Washington, DC 20005.

At the time it approved this document, IIF, which is responsible for the development of this standard, had the following members:

[COMMITTEE LIST]

Al Morton	AT&T
Robert Streijl	AT&T
David Hands	British Telecom
Paul Higgs	Ericsson
John Williams	JDSU
Richard Chernock	Triveni Digital, an LG company
Richard Brand	Nortel
Kishan Shenoi	Symmetricom
Stefan Winkler	Symmetricom
Paul McMenamin	Telchemy
Ken Kerpez	Telcordia Technologies
John Colombo	Verizon

The **ATIS IIF QoSM** committee was responsible for the development of this document.

NOTICE OF DISCLAIMER & LIMITATION OF LIABILITY

TABLE OF CONTENTS

1. SCOPE, PURPOSE, & APPLICATION 1

1.1. SCOPE 1
1.2. PURPOSE 1
1.3. APPLICATION 2

2. REFERENCES 3

2.1 NORMATIVE REFERENCES 3
2.2 INFORMATIVE REFERENCES 3

3. DEFINITIONS, ACRONYMS, & ABBREVIATIONS 4

3.1. DEFINITIONS 4
3.2. ACRONYMS & ABBREVIATIONS 4

4. REVIEW OF TEST PROCESSES 6

4.1. INTRODUCTION 6
4.2. PROCESS FOR QOE MODELS – DIFFERENCES AND CHARACTERISTICS 6
4.3. PROPOSED ALTERNATIVE PROCESS 8

5. THE PARTICIPANTS 8

5.1. ROLE OF MODEL DEVELOPER 9
5.2. ROLE OF INDEPENDENT TEST LABORATORY 10
5.3. ROLE OF MODEL USER 10
5.4. ROLE OF 3RD PARTY ORGANIZATION 11
5.5. ROLE OF ATIS IIF QOSM COMMITTEE 11

6. FORMALIZING MODEL VALIDATION 12

6.1. ON-DEMAND TESTING 12
6.2. OUTLINE OF THE TESTING PROCESS 12
6.3. MODEL CATEGORIES 12
6.4. PERFORMANCE THRESHOLDS AND MODEL CLASSES 13
6.5. MODEL CROSS-CALIBRATION 13
6.6. SUBMISSION OF MODELS 14

7. CONTENT LIBRARY 14

7.1. COMPONENTS OF THE CONTENT LIBRARY 14
7.2. GENRES AND APPLICATIONS 15
7.3. SOURCE SEQUENCES 15
7.4. PROCESSED SEQUENCES 16
7.5. SERVICE SCENARIOS 16
7.6. PUBLICLY AVAILABLE DOCUMENTATION 17
7.7. PUBLICATION OF RESULTS 17

8. OPERATIONALIZING THE PROCESS 19

8.1. GENERAL 19
8.2. ITL FUNCTIONS 19
8.3. POSSIBLE COMPENSATION ARRANGEMENTS 20

APPENDIX A – OVERVIEW OF INDUSTRY STANDARDS AND CERTIFICATION EFFORTS (INFORMATIVE).................... 29

TABLE OF FIGURES

FIGURE 1: THE PARTICIPANTS 9
FIGURE 2: OBJECTIVE MODEL VALIDATION 12
FIGURE 3: SUBJECTIVE TEST PROCESS 15
FIGURE 4: TEST PROCESS OPERATION 19

TABLE OF TABLES

TABLE 1: TEST PROCESS PARTICIPANTS 9

Working Text on –

Technical Report — Validation Process for IPTV Perceptual Quality Measurements

1. Scope, Purpose, & Application

1.1. Scope

This document discusses a proposed test process for IPTV Perceptual Quality Measurements (PQM). ATIS IIF distinguishes a test process from a test plan (see [2]). The following topics relative to the proposed test process are included in this document:

- On-demand testing versus algorithm standardization.
- Content library creation.
- Publication of results.
- Operational process models.
- Operationalizing the process.
- The role of Alliance for Telecommunications Industry Solutions (ATIS) IPTV Interoperability Forum (IIF).

1.2. Purpose

The number of service providers that are rolling out IPTV services continues to increase. In order to support IPTV operations, the need for measurements that can provide insights into the customer's perception of the quality of the IPTV content is apparent.

There is a general concern that PQM solutions available today and the test process that is used to validate these solutions may not be adequate to support the needs of IPTV service providers. Some of the solutions available may be adequate for limited purposes, such as basic reproduction quality checks (e.g. Peak Signal to Noise Ratio (PSNR)) or for measuring basic transmission performance, such as traditional network Quality of Service (QoS) measurements.

The general problem appears to be fourfold:

1) Different organizations create their own PQM test plans and solutions. As a consequence there may be cases where comparison of two models is not meaningful.
2) Lack of comprehensive requirements for a toolset for establishing, measuring, and monitoring Quality of Experience (QoE) for IPTV.
3) Lack of clarity in the application of measurement methods at various points in the IPTV distribution/delivery chain.
4) The current process for testing/validating methods (algorithms) is limited in scope and can be time-consuming.

ATIS IIF addresses each of these four problems in the following documents:

- ATIS-0800025 [2], *Test Plan for Evaluation of Quality Models for IPTV Services*, addresses the first point. This document provides a universal test plan and by standardizing the test plan, it encourages industry developments where multiple organizations could develop QoE models, and these models can be validated based on the same basic test plan. A basic test plan allows for "delta" documents to be developed for extensions that adopt the same fundamentals as this basic test plan but go into more detail for particular types of QoE models and applications.
- ATIS-0800031 [3] , *IPTV QoE Measurement Recommendations and Framework*, addresses points 2 and 3. The purpose of this document is to recommend a variety of IPTV QoE measurements that predict customer experience, to describe the various types of measurements (e.g. parametric, and bit-stream approaches, etc.), their inputs and outputs, and also includes measurement points where such measurements could be most useful.
- The present document, *Validation Process for IPTV Perceptual Quality Measurements*, addresses the fourth point. Today, standards groups combine the test "process" and test "plan" activities with the eventual goal of a standardized PQM solution. ATIS IIF separates these two processes. The test plan is described in [2], and this document describes a recommended test process. This separation provides greater flexibility.

This document describes the currently followed industry standards test process and indicates the shortcomings and then suggests solutions for these weaknesses. The proposed process is flexible and is believed to address market needs better than the current process.

1.3. Application

The formalization of a test process is discussed in this document at a high level. The ATIS IIF Quality of Service Metrics (QoSM) Committee has identified a streamlined process for validating the quality of proposed measurement solutions. A primary intent of the proposal in this technical report is to stimulate discussion that introduces improvements and advances the state of the art in this important aspect of the IPTV industry. ATIS IIF QoSM solicits comments in an effort to construct a coordinated industry consensus position for introducing a standardized test process for validating measurements, algorithms and the related aspects needed to support such efforts. Further development of the concepts proposed here is expected and will be guided by the comments received and the support received.

2. REFERENCES

The following standards contain provisions which, through reference in this text, constitute provisions of this ATIS Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this ATIS Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below.

2.1 Normative References

[1] ATIS-0800008, *QoS metrics for Linear/Broadcast IPTV*, December 2007. [1]

[2] ATIS-0800025, *Test Plan for Evaluation of Quality Models for IPTV Services*, work in progress.[1]

[3] ATIS-0800031, *IPTV QoE Measurement Recommendations and Framework*, work in progress. [1]

2.2 Informative References

[4] ITU-T Recommendation P.564, *Conformance testing for narrowband voice over IP transmission quality assessment models*, November 2007.

[5] 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*, February 2001.

[6] CableLabs ®, *Certification and Wave Requirements and Guidelines, Certification Wave 68-70*, May 2009, Revision 31.

[7] ISO/IEC 17025:2005, *General requirements for the competence of testing and calibration laboratories,* 2005, Edition 2.

[8] ISO/IEC 17011, *Conformity assessment - General requirements for accreditation bodies accrediting conformity assessment bodies,* 2004, Edition 1.

[9] ISO/IEC Guide 65:1996, *General requirements for bodies operating product certification systems,* 1996, Edition 1.

[10] ITU-T Recommendation P.563, *Single-ended method for objective speech quality assessment in narrow-band telephony applications*, May 2004.

[11] ITU-T Recommendation P.861, *Objective quality measurement of telephone-band (300-3400 Hz) speech codecs,* February 1998.

[12] 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*, February 2001.

[13] ITU-T Recommendation P.862.1, *Mapping function for transforming P.862 raw result scores to MOS-LQO*, November 2003.

[14] ATIS Technical Report T1.TR.72-2001, *Methodological framework for specifying accuracy and cross-calibration of video quality metrics*, October 2001.

---

[1] This document is available from the Alliance for Telecommunications Industry Solutions, 1200 G Street N.W., Suite 500, Washington, DC 20005. <http://www.atis.org>

[15] ITU-T Recommendation J.149, *Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM)*, March 2004.

3. DEFINITIONS, ACRONYMS, & ABBREVIATIONS

3.1. Definitions

3.1.1 Accreditation: [From ISO:] "Accreditation is the procedure by which an authoritative body [the accreditor] gives formal recognition that a body or person is competent to carry out specific tasks". An accreditor can accredit Independent Test Labs (ITL). This is independent verification that ITLs are competent to perform the activities for which they are accredited. An example of an accreditation agency the International Accreditation Forum (IAF).

3.1.2 Certification body/authority: An independent party that was not involved in the creation of a product. Optional: This independent party may have taken part in the creation of the requirements. A certification authority may be accredited by an accreditor. A certification authority also issues a "certification" when a product performs according to requirements.

3.1.3 Independent Test Laboratory (ITL): Laboratories have subjective test responsibilities as well as the responsibility to compare a model's performance with the appropriate subjective test content. An ITL cannot be or become a model developer (to prevent conflict of interest).

3.1.4 Model Developer: Creators/developers of algorithms that predict video, audio-video or other media quality.

3.1.5 Model User: Measurement Equipment vendors, Service providers, International Telecommunication Union – Telecommunication Standardization (ITU-T), ATIS IIF, etc.

3.1.6 Perceptual Quality Measurements (PQM): As specified by ATIS IIF, PQM involves any objective quality algorithm (the quality measurement solution), hereafter referred to as "model", capable of predicting subjective measurements to be used in (ATIS IIF defined) IPTV-based applications. The scope of PQM includes video, audio, combinations of audio and video, as well as additional content, such as textual and graphical elements (possibly as overlay) as part of the customer experience. ATIS IIF's position is that all PQM methods must be validated and their performance reported to the Model User (see 3.1.4, below, and section 5.3) who determines the level of acceptability.

3.2. Acronyms & Abbreviations

AAC	Advanced Audio Coding
AC3	Audio Coding
ANSI	American National Standards Institute
ARQ	Automatic Repeat reQuest
ATIS	Alliance for Telecommunications Industry Solutions
ATP	Acceptance Test Plan
AVC	Advanced Video Coding
CHILA	CableCARD-Host Interface License Agreement
CMTS	Cable Modem Termination System
CPU	Central Processing Unit
FEC	Forward Error Correction

FLOPS	FLoating point Operations Per Second
HD	High Definition
HRC	Hypothetical Reference Circuit
IA	Implementation Agreement
IAF	International Accreditation Forum
IEC	International Electrotechnical Commission
IIF	IPTV Interoperability Forum
ITL	Independent Test Laboratory
ISO	International Standards Organization
ITU-T	International Telecommunications Union, Telecommunications Standardization Sector
MIPS	Million Instructions Per Second
MPEG	Moving Pictures Experts Group
MOS	Mean Opinion Score
MOS-LQO	Mean Opinion Score – Listening Quality Objective
MOSp	MOS – predicted
MSF	MultiService Forum
NAMS	Non-intrusive parametric model for Assessment of performance of Multimedia Streaming
NGN	Next Generation Network
Non-CPE	Non-Consumer Premises Equipment
OCAP	OpenCable Application Platform
OLQA	Objective Listening Quality Assessment
OS	Operating System
PAMS	Perceptual Analysis Measurement System
PESQ	Perceptual Evaluation of Speech Quality
PQM	Perceptual Quality Measurements
PSNR	Peak Signal to Noise Ratio
PSQM	Perceptual Speech Quality Measure
PVS	Processed Video Sequence
PVSa	Processed Video Sequence annotated
QCIF	Quarter Common Intermediate Format
QoE	Quality of Experience
QoS	Quality of Service
QoSM	Quality of Service Metrics (committee in ATIS IIF)
RAM	Random Access Memory
RMSE	Root Mean Square Error
RTCP	Real Time Control Protocol
SD	Standard Definition
SDO	Standards Delivery Organization
SIP	Session Initiation Protocol
SRC	SouRCe, relating to original video before transmission and impairment
UDCP	Unidirectional Digital Cable Products
UDP	User Datagram Protocol

VGA	Video Graphics Array
VoD	Video on Demand
VQEG	Video Quality Experts Group
VoIP	Voice over IP
VSF	Video Services Forum

4. REVIEW OF TEST PROCESSES

4.1. Introduction

The typical industry specifications/standards process is (in order of appearance):

(1) to create a standards specification/implementation agreement

(2) for vendors to implement the specification

(3) (optional) to perform independent validation of the implementation, possibly with certification, to see if the vendor's product conforms to the specification.

This process is followed by a number of industry forums and standards organizations, including MultiService Forum (MSF), International Standards Organization (ISO), and CableLabs (see Appendix A).

Certification provides a level of comfort to the purchasers of products that the product performs according to specification (standards compliance) and has been proven to interwork with other vendor's products (interoperability). The certification is meaningful when standards are to be supported by two or more vendors. Most notably, and especially in the case of telecommunications, this is beneficial in multi-vendor environments where the interoperability of communication interfaces or protocols is involved.

4.2. Process for QoE Models – Differences and Characteristics

While there are similarities, the process for QoE models is different than the interoperability and/or compliance process (described in section 4.1).

First of all, there is no strict interoperability requirement between vendor products for QoE models; interoperability testing is not meaningful for QoE models. Instead, a certain level of accuracy is expected from a QoE model, and as a result, accuracy is used as the main criterion in QoE model testing to test whether on its own it results in a close correlation with the equivalent subjective test results. Furthermore, models that have shown a close correlation with subjective tests can form the basis of industry standards that document the results as well as the QoE model algorithms. This does not occur for a process where a vendor implements a product, demonstrates interworking and is possibly certified for it.

Secondly, the standards "process" followed (by, for example, the ITU-T and VQEG), as observed with the standardization of models for VoIP and digital video quality models, is slightly different. We can see the following characteristics:

- Competition – one or more winners emerge from a test event.
- Collaboration in part (e.g., ITU-T Rec. P.862 [5]) or in whole (see the component model in the Non-intrusive parametric model for Assessment of performance of Multimedia Streaming (P.NAMS)).

- Conformance testing (e.g., ITU-T Rec. P.564 [4]).

The interoperability/compliance process does not have these characteristics. After the specification of requirements, the implementation (underlying algorithms) of a vendor's product and how the product competes in the market place is a private matter. Certification in the interoperability/compliance process is for that vendor's product alone, not for a collection of products as is the case for QoE models according to the interoperability/compliance standards process.

The process currently followed for validating QoE models has various characteristics (in no particular order of appearance):

- QoE is subjective by nature and is difficult to capture in detailed requirements. Requirements documents for QoE models are relatively high-level compared to requirements documents related to, for example, transmission gear, for which traditional interoperability/compliance processes are suitable. QoE models are not usually specified in terms of the input and output parameters of the models and where and how these models may be used.
- The number of vendors that choose to openly test their QoE models is (at the time that this report was created) very small compared to the vendors certifying their equipment following traditional interoperability/compliance processes.
- Validating QoE models requires the acquisition of suitable multimedia content. To date such source content is selected upon the "judgment of experts" (see section 7.3 for more detail) . Furthermore, once test material has been made available to model proponents it cannot be re-used in future validation tests, requiring the selection and preparation of new content that must be used in a new round of subjective testing.
- Given the current approaches (competition, collaboration, etc.), QoE models can only be considered if they are available at a specific cut-off date, because the validation is scheduled by consensus with multiple models from multiple developers being evaluated in the same exercise.
- For the competitive and collaborative approaches, model developers must reveal the details of their algorithms if they want to be included in the standard. They must also agree to the intellectual property disclosure conditions imposed by the standards body.
- Model developers are sometimes involved in the preparation of PVS or in conducting subjective experiments due to ITL budget and time constraints, which is not ideal for an independent evaluation.
- The process for validating QoE models is very lengthy. From start to finish, a process can take a number of years. This time is spend on the creation of a unique test plan for the particular type of test (and also indicates who does what and many other process related aspects), collecting source material, impairing the source material, performing subjective tests, performing objective tests, and analysis of results.
- Successful QoE models are often documented in a standards document (i.e. a fourth step based on the three steps in section 4.1), such as an ITU-T recommendation. (This does not happen with the traditional conformance testing process). This leads to the possibility that successful QoE models can be licensed to other vendors who wish to implement the standard.
- Once a standard has been defined and approved it is very difficult to change, as the database cannot be re-used. Consequently the validation process must be repeated with the associated elapsed time to completion. This means that standardized models can quickly

become outdated, and there is no process for the models or the standards to be updated in a prompt fashion. This situation is acceptable only for applications that are somewhat static in nature.

4.3. Proposed Alternative Process

The following points summarize the main benefits of the proposed process for QoE model validation. The various items and issues are discussed in more detail in the remainder of this technical report.

1. The most critical component of the entire process is the validation and that there is an "open" process for approving QoE models. This open process allows for rapid, reliable, and inexpensive validation of QoE models. In such a process, anyone (e.g. industrial or academic or governmental organizations) who creates a model will have access to this process. With the appropriate set of policies in place, such a process is expected to result in a much more dynamic market place for quality models than the current standards-based process.
2. The inner details of an algorithm are de-emphasized. This can be done only if there is a well articulated definition of the type of model, and its input and output parameters. (Note that ATIS IIF is working on this in ATIS-0800031 [3]).
3. When the details of the algorithm are not subject to standardization, the model developer can make routine updates to algorithms to improve performance. Improved algorithms can be (re-) validated easily and quickly. The proposed process thus promotes innovation, improves the time-to-market for solutions, and reduces development costs.
4. The proposed process allows vendors to keep their intellectual property intact. Validating the results of the algorithm serves to inform the user (e.g. IPTV Service Provider or Test Equipment Manufacturer) that the algorithm performs well without necessarily revealing the inner workings of the algorithm.
5. The proposed process allows for development and deployment of models that are "fit-for-purpose" by permitting trade-offs between model accuracy and model complexity. The choice of the specific trade-off lies with the model user and not the standardization body.
6. Creation of reporting templates provides a standard format for vendors to report the performance of their model. This standard report template will help the industry by simplifying interpretation of the performance characteristics of different vendors' models. Note that ATIS-0800025 [2] specifies reports.

5. THE PARTICIPANTS

The different entities, generically referred to as "participants" and their roles are explained in more detail below.

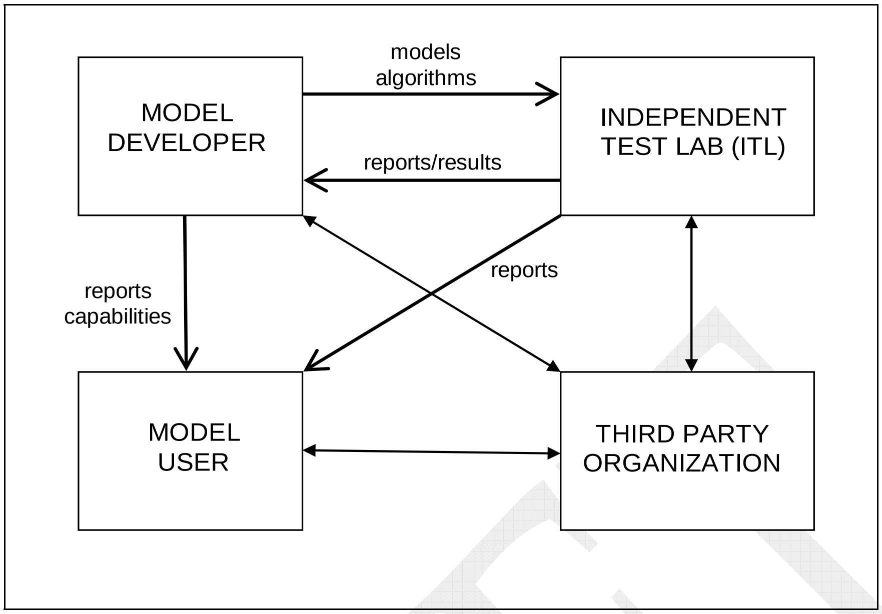

Figure 1: The Participants

The relationship between ITL and Model User indicates that test reports may be provided by the ITL directly to the Model User. However, it should be noted in the process that this exchange requires the permission of the Model Developer. It is usually required that the report be distributed unaltered and in its entirety.

Description of the participants:

Table 1: Test Process Participants

Actor	Definition
Model Developer	Creator/developer of algorithms that predict video, audio-video or other media quality, *a.k.a.* proponents. Models may be developed in academia, by equipment vendors, or in collaborative consortia.
Independent Test Laboratory (ITL)	Lab that conducts subjective tests and compares a model's performance with the appropriate subjective test results.
Model User	IPTV service providers, IPTV solution makers, Measurement Equipment vendors, System Integrators, Network Providers, etc.
3rd Party Organization	Organization that facilitates interaction between the different actors. This is an optional component of the process.

5.1. *Role of Model Developer*

The model developer is responsible for:

- Initiating a validation request for their model(s).
- Verifying the correct functionality of the model using suitable test vectors. The test vectors are mutually agreed to by Model Developer and ITL.
- Providing a working software or hardware implementation of the model to the ITL.
- Paying the validation fees.

- Checking the detailed and summary reports received from the ITL for accuracy and completeness.
- Approving the publication of the summary reports (if satisfied with the results).

5.2. Role of Independent Test Laboratory

The ITL has two main tasks:

1. Preparing and maintaining an annotated database of test material.
   The database has to be prepared before any validation can begin. The database may then be updated or expanded as needed.
2. Conducting model validation and reporting.
   This has to be done on-demand for every model that is tested.

The task of preparing and maintaining the database involves:

- Obtaining and editing source content.
- Processing source content to create test clips.
- Conducting subjective tests on test material.

Service providers and other parties that are not model developers may contribute to one or more of these steps.

The task of model validation involves:

- Receiving developer's model and verifying its correct functionality.
- Running the model on the test database.
- Analyzing model prediction performance.
- Writing detailed and summary reports on the model's performance.

5.3. Role of Model User

The Model User is an entity that deploys complete or partial solutions for providing IPTV services.

Examples include:

- Content providers that deploy quality monitoring solutions at the point of hand-off to the service provider.
- Service providers that deploy quality monitoring solutions at the hand-off with the content provider and at the hand-off to the end-user of the IPTV service and at various intermediate points to ensure high quality of experience/service.
- Measurement Equipment vendors that provide equipment that includes predicted-QoE assessment.
- Video processing equipment manufacturers who need to optimize their products, and make sure it delivers the quality that their customers expect.

In one special case the model user may take on the role of an ITL. Specifically the model user may have a private data base. In this case there may or may not be collaboration between model user and model developer. Some possible scenarios for this:

- Model user develops a data base of PVSa. Model developer shares the model with the model user. The model user runs the QoE model through a test plan (e.g. ATIS-0800025 [2]). The results are then known to the model user who may or may not share the results with the model developer.
- Model user develops a data base of PVSa and makes it available to the model developer. Model developer runs their QoE model through a test plan. The results are then shared with the model user. Implicit here is that the model user trusts the model developer and the model developer provides the model user suitable assurance that the model has not been trained for this data base in particular.

5.4. Role of 3rd Party Organization

An (optional) 3rd party organization may facilitate interaction between model users, model developers, and independent test labs. A third party can generally be considered a champion for QoE models and that its role is to promote all aspects of QoE measurement. Representing a single point of contact between model developers and test labs, this organization could:

- Collect fees from model developers.
- Distribute collected fees to the test labs according to work performed.
- Facilitate the acquisition of suitable content.

Furthermore, this organization could:

- Provide access to summary reports for model users (e.g. web site, list of validated models, etc.).
- Provide access to publicly available documentation about the test material (see above).
- Facilitate the accreditation of ITLs.

5.5. Role of ATIS IIF QoSM Committee

ATIS IIF has created a test plan for PQM validation, see ATIS-0800025 [2], which currently focuses on video and audio-video tests. It is expected that in a future version of this document the scope will be expanded to include other aspects of PQM. Nevertheless, this document is expected to be the standard test plan for PQM for ATIS-IIF-defined IPTV and could thus be input to a PQM test process as discussed in this document.

Additionally, ATIS IIF is in the process of defining PQM requirements, more specifically called *IPTV QoE Measurement Recommendations and Framework* [3]. This document is expected to provide guidance on where and what types of PQM tools should be used in IPTV deployments. This document is expected to be followed by specific solutions. These PQM solutions may be developed internally within ATIS IIF, or adopted from external activities so long as they are within the scope of ATIS IIF. Any PQM tool requirements could be used as part of the validation process to see if the proposed model is in compliance with the standard and/or fit for purpose.

6. FORMALIZING MODEL VALIDATION

While it is recognized that the design of models can be very complex, guidance in this area does not necessarily mean a path to validation results and standardization alone. At any time, certain models of a particular kind may be recommended, but this should not be perceived to mean no new development can take place. ATIS IIF is interested in a highly stimulated and energetic environment that actively promotes constant model improvements. ATIS IIF believes that on-demand model testing and validation is an integral part of such an environment.

6.1. On-Demand Testing

Instead of staging a test of various candidate models of a particular kind, a model process that supports on demand testing at any time will result a shortened test period. This elapsed time for testing/validation is expected to be in the order of weeks, rather than years.

The rationale is that while the standards industry currently achieves some very exciting results, ATIS IIF hopes that a continued improvement in the accuracy of the proposed models continues to be a goal for all model developers. In addition, the type of model proposed most likely will evolve over time. Both these developments may impact what has already been standardized. Naturally, on demand testing can only work if there is a complete and secret content library available (described later).

6.2. Outline of the Testing Process

Any model validation should be executed as depicted in the following figure. A model would be applied to a pre-selected set of PVS, and the results of the objective test would be compared with PVSa from the subjective tests. Validation would then imply a close correlation between the results.

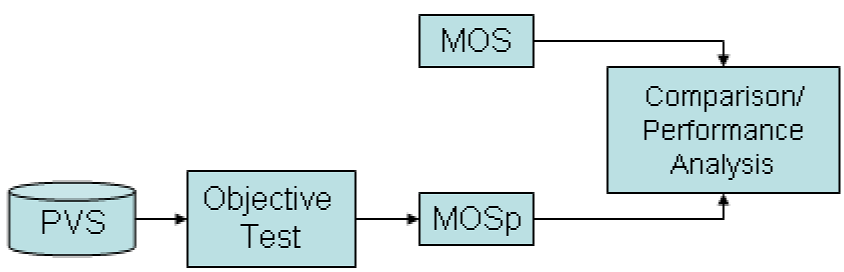

Figure 2: Objective Model Validation

6.3. Model Categories

Since different model categories can have different uses and applications, validation can be specific to certain model categories, depending on what is included in the testing database and requested by the model developer. Useful categories could include video format (Standard Definition (SD)/High Definition (HD)), video vs. video-audio, etc. As an example, a model would be validated for "HD video-only quality measurement." However, the number of different categories should be kept to a minimum in order to avoid proliferation of models with too narrow an application scope. The number and type of models should follow certain requirements, such as those specified in ATIS-0800031 [3].

The ITL is expected to be the "keeper/maintainer" of the model categories. That is, an ITL may choose to specialize in certain areas. See section 8.

6.4. Performance Thresholds and Model Classes

It is also proposed to define thresholds for one or more of the model evaluation criteria described in the IPTV test plan [2] that can be used to decide whether a given model predicts subjective Mean Opinion Score (MOS) data sufficiently well. Using the correlation of a model's MOS predicted (MOSp) with subjective MOS as an example, the criterion could be, "the Pearson linear correlation coefficient has to be at least 80%". Combinations of different evaluation criteria together with measures of statistical significance may be needed to make the threshold comparison meaningful. The underlying test plan used must identify the evaluation criteria used. Examples of evaluation criteria (see [2, 3] for detailed definitions) include:

- Pearson correlation coefficient.
- Root Mean Square Error (RMSE).
- Outlier Ratio.

To refine the simple pass/fail criteria for validation, different threshold levels for the above-mentioned criteria can be used to distinguish different levels or classes of models. Care will have to be taken that the threshold levels and ranges defined are relevant to the category or scenario in question and fit for purpose. Furthermore, as above, combinations of different evaluation criteria together with measures of statistical significance may be needed to make the class distinctions meaningful.

Again using correlation coefficient only as a simplified example, the following classes could be defined (CC = correlation coefficient) for a specific model category:

CC > 90%	Class A
CC > 85%	Class B
CC > 80%	Class C
CC > 75%	Class D

It is intended that the definition of different classes as above will encourage competition between vendors and improvement of models over time. It will also allow vendors to market and price models based on prediction performance. Suitable thresholds for classification are for further study.

From an operational standpoint the complexity of the model is also relevant.

6.5. Model Cross-Calibration

To compare PQMs and PQM results from different Model Developers, especially as multiple different solutions could be used in an operational environment, there is a need to translate (or cross-calibrate) the output of one model with that of another. This cross-calibration is performed according to the method described in [14, 15].

Cross-calibration is addressed by a transformation to a common scale through the annotated PVS database. The transformation is typically a linear or nonlinear fitting function that maps the MOSp model outputs to the subjective MOS. Computing this fitting function for a model is part of the validation and will be done by the ITL; the function and its coefficients will also be given in the summary report.

Once two PQMs are transformed to the common MOS scale, the transformation from model A to model B is composed of the transformation from model A to the common scale (MOSp-A to MOS), followed by the transformation from the common scale to model B (MOS to MOSp-B, i.e. the inverse transformation of model B).

If such an inverse transformation does not exist then there may be inconsistencies in the comparison of models A and B.

While cross-calibration of two or more models does not mean one model can be substituted error-free for another, it provides for a common scale and makes measurements using different models more easily comparable.

Editor's note: The purpose of this section may be revisited.

6.6. Submission of Models

Models may be submitted as software or as hardware implementations. It is the responsibility of model developers to ensure that the version of the model delivered to ITL works correctly. Model developers should provide full installation instructions and a user guide or, where necessary, send a representative to correctly install the model at the ITL premises.

Model developers should provide processing requirements to provide a guide to the complexity of the model. Typical metrics and information are:

- Operating System (OS) dependencies
- Central Processing Unit (CPU) load (e.g. million FLoating point Operations Per Second (FLOPS), Million Instructions Per Second (MIPS), etc.
- Random Access Memory (RAM) peak usage
- Disk space required

7. Content Library

7.1. Components of the Content Library

The content library will be comprised of three classes of sequences. First are the baseline SouRCe video/audio-video sequences (SRC). The second class of sequences are generated from the SRC sequences by the application of suitable Hypothetical Reference Circuits (HRCs) to create the Processed Video Sequences (PVSs). The third class is also referred to as "test vectors" and comprises a collection of PVSs that could be a subset of the second class but is likely to have other sequences as well. In order to validate models, a library of processed video and audio-video sequences (PVS) will be created and held by the ITL.

This content library will be annotated with subjective scores (PVSa = PVS plus associated MOS). The subjective test process is depicted in the following figure. The PVSs used in the subjective test process are the second class mentioned above.

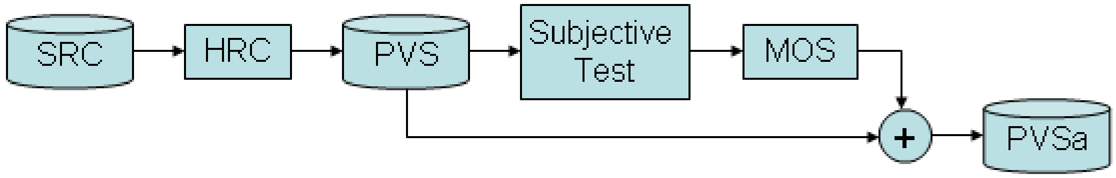

Figure 3: Subjective Test Process

Subjective scores will be obtained in line with the appropriate standardised subjective test procedures. All source (SRC) and processed (PVS) sequences will remain secret, known only to the ITL.

A small subset of sequences, known as test vectors, and not to be used for validation purposes, will be created by the ITL. These test vectors can be shared with model developers to verify that models will run correctly at the ITL's premises.

7.2. Genres and Applications

The library of test content should be representative of different content genres, for example, movies, sports, animation/cartoon, music videos, and documentaries. Within each genre, content must be representative. For example, sports content may include football, tennis, gymnastics, golf and pool.

Content should be correctly formatted for different application scenarios and should be designed with the desired model categories in mind (cf. Section 6.3). Examples of target applications include, but are not limited to:

- Linear TV delivered on a fixed-line broadband network (full frame-rate, various resolutions (HD – ¾ SD), progressive and interlace scan)
- TV delivered on a cellular wireless network (variable frame-rates, reduced resolutions (VGA - QCIF), progressive scan)
- Video on Demand.
- Others to be defined.

As indicated in section 6.3, different models may be tested on differently prepared content. In other words, a model used for assessment of HD linear TV would not be tested on prepared SD source material.

7.3. Source Sequences

The ITL will have access to a large selection of high quality source sequences (SRC). A small set of lower quality source sequences should be created (e.g. containing analog artifacts) and used to produce a special case set of processed test sequences (PVS). Source sequences will generally be available only to the ITL and will remain secret from model developers.

One primary reason for keeping the source content confidential is to respect the intellectual property rights of the original developer of the content, such as a movie studio. Some test content might not be secret. For example, camera or equipment manufacturers may have test video sequences appropriate for use in a validation test which they may be willing to release to the ITL at no cost with the proviso that the first frame includes their company name ascribing them as the source of the material.

In order to keep the test content secret, the ITL works directly with content providers who provide useful content to be used for the testing of models. This interaction is continuous, as new types of content and content transport are developed over time.

Acquisition of source sequences must be such as to have a suitable variety of characteristics. In most cases followed currently the main criterion used is "judgement of experts" where the choice of material to be included in the data base is made by those skilled in the art. ATIS IIF recommends that the selection of source content follow a more objective methodology. Specifically it should be possible to characterize content according to some well-defined properties (e.g., reporting the average of spatial and temporal components, color components, luminance values that could include mean / range, scene cuts, merges, pans; to state just a few examples). Grouping content according to such pre-defined characteristics will produce a more robust set of test content. In addition, there might be some subjective testing by experts in the field in order to confirm that the selected characteristics cover the full range of qualities that may occur in 'natural' scenarios. Suitable description of the objective properties is for further study.

7.4. *Processed Sequences*

Content should be processed using a range of representative coding and transmission methods appropriate to the application being targeted (HRC). For example, processing of linear TV scenario content may use:

- H.264/Advanced Video Coding (AVC), VC-1, Moving Pictures Experts Group (MPEG)-2 video codecs; Audio Coding (AC)3, Advanced Audio Coding (AAC)+ audio codecs
- User Datagram Protocol (UDP) packet loss
- Decoders using different error concealment methods
- Forward Error Correction (FEC), Automatic Repeat reQuest (ARQ), etc.

The ITL should capture both the bitstreams and the uncompressed versions of processed sequences. Care must be taken when storing the files. For example, since the same bitstream may be decoded using different decoders resulting in different subjective scores, the ITL must ensure the bitstreams can be matched to the correct uncompressed file. It may be desirable for bitstreams to be captured both with and without encryption.

7.5. *Service Scenarios*

Processed content may belong to different service scenarios. Content should be labelled correctly in order for service scenarios to be easily identified. Example service scenarios include:

- Linear SD IPTV
- Linear HD IPTV
- HD Video on Demand (VoD)
- Linear IPTV, maximum throughput = 6 Mbit/s, SD, General Entertainment
- Linear IPTV, maximum throughput = 20 Mbit/s, HD, General Entertainment
- Cellular TV, maximum throughput = 384 kbit/s, QCIF, Sports

This will enable model developers to submit models for validation against content datasets that are associated with one or more specific service scenarios.

For each service scenario, the ITL should process a minimum of 50 different source sequences (SRC) using a representative set of codecs and transmission errors conditions. The processed sequences along with the associated subjective scores (PVSa), will form the basis of a test library against which models will be validated. Prior to model validation, selection of test scenarios should be agreed between the ITL and the model developer. Where a model developer wishes to disclose a model's performance, performance must be reported against each service scenario in full, unaltered form. Model developers may choose to withhold performance details for one or more service scenarios.

7.6. Publicly Available Documentation

A publicly available document providing a written description of the test content will be produced by the ITL. This written record of test content should provide a detailed description of the video and, where appropriate, audio component of each test sequence. The information that is relevant includes amount of motion, levels of detail, variety of objects present in the video. The detail provided on any audio should include information regarding whether it is direct speech, whether it is commentary, soundtrack details, background noise, and so on. In addition to the written documentation detailing scene content, thumbnails (single frame taken from a video sequence) from a representative sample of test sequences should be publicly available.

In addition to a written description of source content, the method used to process sequences should be documented. This should include details of how sequences were processed (e.g. pre-processing applied to source, coding scheme and implementation used, method and process for introducing any transmissions errors, decoder, video capture method and so on).

7.7. Publication of Results

On demand testing is initiated by the model developer. The ITL conducts the validation tests and generates the appropriate reports.

There are two types of reports that the ITL generates:

1. Detailed. This is generally shared only with the model developer. The model developer may share it with others, or authorize the ITL to do so. For the agreed upon category (service scenario) the report will include the MOS and MOSp for each PVSa, in the format specified in the test plan.
2. Summary. This is what would normally be shared with model users either by the model developer or the ITL once the results are released by the model developer.

A summary report is important. While the detailed report may contain data that cannot or should not be shared with outsiders, the summary report should be accessible to a bigger audience, not just the specific model developer and/or model user. Perhaps an ITL makes the summary reports available for a fee to outsiders. To make models comparable, the summary report format should also have a clear and well-defined structure that is followed by all ITLs for all models. The summary report should represent a true summary of the detailed report. As a special service the ITL could provide results of a test done on a subset of the PVSs. If this is done, the report should clearly state the properties of the subset and also that the test represents validation over a subset. Under no circumstances can PVSs be excluded based on poor model performance.

The summary report content includes:

1. Reference to test plan, category/service scenario/application tested, PVSa database, and the number of PVS that were used in the validation test.
2. Prediction performance of the model for the set of PVSs in terms of evaluation criteria and corresponding certification/class, as well as a suggested transformation from MOSp to MOS in the form of a fitting function and its coefficients.. A scatter plot of MOSp versus MOS should also be included.
3. Identification of the following:
   a. ITL name
   b. Model developer organization, model identifier and version number model developer
   c. Testing round. This identifies whether the ITL has tested the model before, including instances where the results were kept confidential.
   d. Outline of model inputs.
   e. Computational complexity (if a uniformly applicable way of describing computational complexity is unavailable, each ITL should provide a description of how it determines computational complexity).

The format of a "typical" summary report may be as follows:

Summary Report

Testing lab: XYZ

Model developer: ABC Corp.

Model: DEFG Version 1.0 (Software model)

Scenario: SD

Application: Linear fixed-line IPTV

Testing round: 4

Number of PVSs: 110

Prediction performance:

Correlation: 85% (0.85)

RMSE: 1.7

Outlier ratio: 0.02

Accuracy class: B

Transformation function: MOS = f(MOSp, a, b, c, d); a=15.7, b=846, c=0.669, d=5.21

Computational complexity:

The minimum, average, and maximum run times for the model were 2s, 2.6s, 2.8s, respectively. This was performed on an XXX Workstation with a YYY processor rated at 2 GHz. The platform had 100 Mb of core memory and used a Linux operating system.

8. OPERATIONALIZING THE PROCESS

8.1. General

The following picture depicts an operational process.

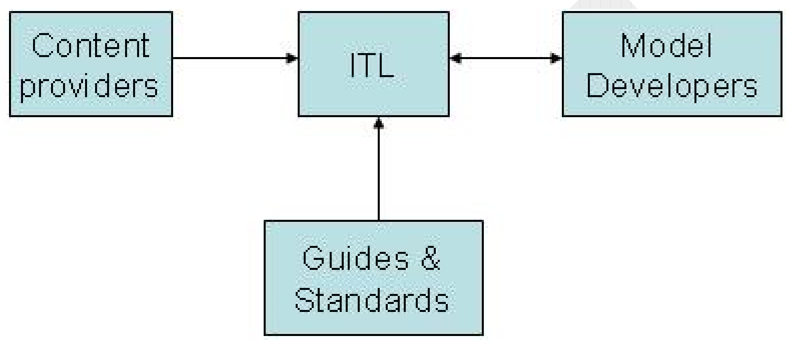

Figure 4: Test Process Operation

8.2. ITL Functions

The ITL may function not only to test models to be used in IPTV environments. This is a scope matter that is left for further discussion. At a minimum, it MUST have in its scope the IPTV industry.

The ITL operates based on guides and standards. These guides and standards are developed in standards bodies such as ATIS IIF and the ITU-T. Example documents include:

- How the ITL operates, i.e. their test process, e.g. as described in this document. There will be one test process. This test process is transparent, and may be developed/amended over time, as needed, by those who have work with the ITL (perhaps the ITL community as a whole will keep periodic meetings). Given that this document is written by ATIS IIF, the standards organizations will have a certain role to play in this.
- Depending on the scope of the ITL, it advertises which models it will validate. The types of models an ITL validates are specified in standards documents [2].
- Thresholds of model performance should also be documented and be kept public. Who is responsible for creating the thresholds is for further discussion, but this could again be transparent and thus be a document that is developed by consent of the various involved parties.

Model developers will submit their models to the ITL and have their models evaluated on-demand per the process described in this document.

Not shown in the picture above is what happens if a model performs according to threshold levels. This is for further discussion. There are several possibilities. Examples includes that the model could be "certified" or be submitted to a standards body for standardization.

8.3. *Possible Compensation Arrangements*

In the aforementioned process example compensation arrangements may be as follows:

- The ITL and the content providers have an agreement. Note, there can be multiple providers of content, and the ITL community may function as one unit and consist of one or more actual labs that are "qualified" to be an ITL or to be part of the ITL group.
    - The ITL pays the content provider a certain fee for the use of certain content. The ITL will recuperate their costs when they receive individual models from model developers for testing
    - The ITL does not pay the content providers; instead, the content providers "profit" by a certain percentage from the fees received by the model developers.
    - etc….
- Model developers pay a fee each time they would like to have a model validated. For this compensation, the ITL performs the test and provides test results back to the Model developer and, possibly, to Model Users and the 3rd Party Organization..
- To jumpstart the process and help the ITL with the initial expenses of content acquisition, PVS creation, and subjective testing, interested model developers could make a pre-payment of the necessary amount to fund these activities. In return, they would receive preferential treatment, such as a certain number of free model validations once the database is in place. Other model developers who did not contribute to this fund would be charged the regular fee.

APPENDIX A – OVERVIEW OF INDUSTRY STANDARDS AND CERTIFICATION EFFORTS (INFORMATIVE)

This appendix provides an overview of standardization and certification practices in the following areas:

- International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) Standardization of Voice over IP (VoIP) Speech Quality Measurement Models
- Certification in CableLabs
- International Standards Organization (ISO) Standardization and Certification
- MultiService Forum (MSF) Certification

NOTE -- This overview is not meant to be exhaustive. It is meant to provide illustrative examples of certification processes in other parts of the industry.

A.1 Standardization of Voice over IP - Speech Quality Measurement Models

There are a number of ITU-T recommendations related to the measurement of speech quality. Examples include:

- P.563 [10], a non-intrusive method called "Single-ended method for objective speech quality assessment in narrow-band telephony applications".
- P.862 [12], an intrusive method called "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs".

For the purposes of this appendix, the creation of the P.862 recommendation is of primary interest.

P.861 [11], Perceptual Speech Quality Measure (PSQM), is the precursor of P.862. P.861 was deprecated when PESQ was created. PESQ is a means of estimating listening speech quality by using reference and degraded speech samples. It cannot be used for the assessment of talking quality or interaction quality.

PESQ was developed by KPN Research, the Netherlands and British Telecommunications, by combining the two advanced speech quality measures Perceptual Speech Quality Measure+ (PSQM+) and Perceptual Analysis Measurement System (PAMS). PESQ builds on the PSQM and PAMS algorithms by adding additional processing steps to account for signal-level differences and the identification of errors associated with packet loss.

PESQ gives a numerical rating to voice quality. As such, it masks the underlying network problems by producing a particular voice-quality score. To determine the reasons for a score, the network statistics associated with a voice-quality test must be analyzed. Statistical events of interest are jitter, delay, packet loss, and duplicate packets. Analyses of these factors allow network operators to correlate voice-quality issues with busy-hour occurrence or to determine whether a media gateway involved in the call is the source of the voice-quality issues.

P.862.1 [13] specifies a conversion from the PESQ output R-factor to a Mean Opinion Score – Listening Quality Objective (MOS-LQO).

It should be noted that PESQ has a number of limitations. The ITU-T continues to work on improvements. One such effort is the Objective Listening Quality Assessment (P.OLQA).

A.2 Cablelabs Certification

Cable Television Laboratories, Inc. (CableLabs) creates specifications, Compliance Test Plans, and performs certifications. CableLabs performs regular testing in what are called “Waves”. http://www.cablelabs.com/certqual/ provides an overview of what is involved in certification. The following is maintained:

- Certification requirements (for each or a subset of waves, i.e. there could be different requirements for each or a subset of waves).
- rules of engagement for vendors
- how certification is determined and what happens if a product is not certified after a test
- when tests are conducted including fees.
- Maintenance of a list of certified products (http://www.cablelabs.com/certqual/lists/)

From Certification and Wave Requirements and Guidelines, Certification Wave 68-70 [6]:

“The certification process defined by CableLabs has been developed to provide cable operators, retail distributors, and consumers confidence that Certified products interoperate with products made by other manufacturers, and that the integrity and security of the cable operators’ network is maintained.”

“Non-Consumer Premises Equipment (non-CPE) or cable network element devices (e.g., CableCARDs, [Cable Modem Termination System] CMTS) are determined to be “Qualified” rather than Certified because they are only purchased by cable operators and not the consuming public at retail. The word “Certification” means “Qualification” for purposes of these non-CPE network products. Similarly, certain test tools and other devices that are purchased only by cable operators may be submitted to CableLabs for testing. These devices are “Verified for Interoperability.

A separate process of "Verification" is performed for so-called "Plug & Play" devices; also called Unidirectional Digital Cable Products, or UDCPs. Please see the UDCP area on CableLabs' website for more information.

"CableLabs Certified®", "CableLabs Qualified™" or “Verified for Interoperability” means that the device has passed a series of tests for compliance with the applicable Specification, and has thus demonstrated interoperable functionality with other CableLabs certified devices. Manufacturers are encouraged to add additional functions and features to their devices to meet the requirements of other industry specifications and standards. However, "CableLabs Certified®", "CableLabs Qualified™" or “Verified for Interoperability” should not be understood as an endorsement of these other attributes, or that the device is certified to such other specifications or standards. “

On Self Certification

“If a manufacturer was successful in certifying a uni-directional OpenCable terminal device, as described in an Issued version of the OpenCable Core Functional Requirements specification, prior to December 31, 2004, then that manufacturer may use the paper submission process to “self certify” new models of one-way OpenCable terminals after that date, at no charge to manufacturer. Such self-certification shall be effective only if the manufacturer:

- Fully tests each new terminal host device against the last version of the applicable Requirements Checklist and [Acceptance Test Plan] ATP issued by CableLabs in 2004, validating that each model conforms to all specifications and requirements for unidirectional terminal devices;

- Follows this paper submission process, and completes the Self-Certification Form available as a web-based application form at https://www.cablelabs.com/certwave/ for each model and returns it to CableLabs, along with a copy"

"Licensees who have demonstrated that they are consistently capable of obtaining Certification of OpenCable Host Devices may apply for Self Certification status. Subject to the conditions of this section, successful Certification of three (3) unique Host Devices in three (3) separate CableLabs Certification Waves within a two (2) year period, together with no Certification failures or breaches of the applicable agreements ([CableCARD-Host Interface License Agreement] CHILA, [OpenCable Application Platform] OCAP, tru2way over such two (2) year period, shall serve as prima facia evidence of such capability. After successful Certification in such third Certification Wave, Vendor may apply to the Certification Board for Self-Certification status by written request. Once granted, the Certification Board may revoke Self-Certification status for any material breach of the applicable Agreements. The Self-Certification election is optional under the sole discretion of the Vendor, who may notwithstanding the acquiring of the Self-Certification status continue to use CableLabs Certification. For Self Certification of an OpenCable Host Device, see Self Certification Submission Form available in the online-application form: http://www.cablelabs.com/certwave/"

A.3 ISO Standardization and Conformance Assessment activities

http://www.iso.org/iso/resources/conformity_assessment.htm

"Conformity assessment is the name given to the processes that are used to demonstrate that a product (tangible) or a service or a management system or body meets specified requirements. These requirements are contained in ISO/[International Electrotechnical Commission] IEC standards and guides.

The processes that need to be followed to be able to demonstrate that they meet the requirements are also contained in ISO/IEC standards and guides."

The use of ISO/IEC standards in conformity assessment procedures allows for harmonization throughout the world and this, in turn, not only facilitates international trade between countries but also facilitates trade within countries by giving the purchaser of the product or service confidence that it meets the requirements.

Conformity assessment can cover one or more of the following activities:

- **Testing** of a product/service to determine if it complies or performs in accordance with the specified requirements.
- **Inspecting** the manufacturing process of a product to ensure that it is manufactured in a safe manner and according to regulations (e.g. fire extinguishers).
- **Implementing** a management system to ensure that products/services are produced or delivered by an organization in a consistent manner and meet customers' expectations.

Conformity assessment provides benefits to everyone in the supply and demand chain. This includes the consumer, manufacturer and the supplier. It also includes regulators who are responsible for ensuring the health and safety of the general public. Conformity assessment not only provides confidence to consumers and purchasers but it also facilitates the free flow of goods and services between national boundaries.

The general requirements for laboratories or other organizations to be considered competent to carry out testing calibration and sampling are specified in the joint International Standard ISO/IEC 17025:2005 *General requirements for the competence of testing and calibration laboratories* [7].

Some conformity assessment bodies may wish to distinguish themselves from their competitors by having an impartial evaluation of their competence by an accreditation body based on internationally-recognized criteria. These criteria are contained in ISO/IEC 17011 *Conformity assessment - General requirements for accreditation bodies accrediting conformity assessment bodies* [8]. These conformity assessment bodies are then said to be accredited.

ISO/IEC Guide 65:1996 *General requirements for bodies operating product certification systems* [9] can be used in combination with a number of related product standards and guides to demonstrate that a product complies with specified criteria. There are other standards within this family of standards that give guidance of the various types of product certification schemes which can be used.

A.4 MultiService Forum (MSF) Certification

http://www.msforum.org/techinfo/certification.shtml

In 2007, the MSF launched their Next Generation Network (NGN) certification program. The program is designed to proof multi-vendor practical open architecture solutions for Next Generation Networks (NGN). The goal is to address key issues necessary to deliver multi-vendor open architecture solutions. MSF chose Iometrix as the host test lab.

The MSF key deliverables are validated, commercially viable Implementation Agreements (IAs). To date, the MSF has created a long list of Implementation Agreements (see http://www.msforum.org/techinfo/approved.shtml). Most of these IAs are related to protocols such as [Real Time Control Protocol] RTCP, [Session Initiation Protocol] SIP, Diameter, etc. Each MSF architecture version has its own new set of IAs.

The certification program is open to MSF members. The certification process is not publicly available on the MSF website. Once certified a product can carry the MSF certified logo:

MSF certified

A.5 International Accreditation Forum (IAF)

http://www.iaf.nu/

“The International Accreditation Forum, Inc. (IAF) is the world association of Conformity Assessment Accreditation Bodies and other bodies interested in conformity assessment in the fields of management systems, products, services, personnel and other similar programs of conformity assessment. Its primary function is to develop a single worldwide program of conformity assessment which reduces risk for business and its customers by assuring them that accredited certificates may be relied upon. Accreditation assures users of the competence and impartiality of the body accredited. IAF members accredit certification or registration bodies that issue certificates attesting that an organization's management, products or personnel comply with a specified standard (called conformity assessment).” For example, American National Standards Institute (ANSI) is a member.