INTERNATIONAL TELECOMMUNICATION UNION

ITU-T J.148

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(05/2003)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Measurement of the quality of service

Requirements for an objective perceptual multimedia quality model

ITU-T Recommendation J.148

ITU-T J-SERIES RECOMMENDATIONS

CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

General Recommendations	J.1–J.9
General specifications for analogue sound-programme transmission	J.10–J.19
Performance characteristics of analogue sound-programme circuits	J.20–J.29
Equipment and lines used for analogue sound-programme circuits	J.30–J.39
Digital encoders for analogue sound-programme signals	J.40–J.49
Digital transmission of sound-programme signals	J.50–J.59
Circuits for analogue television transmission	J.60–J.69
Analogue television transmission over metallic lines and interconnection with radio-relay links	J.70–J.79
Digital transmission of television signals	J.80–J.89
Ancillary digital services for television transmission	J.90–J.99
Operational requirements and methods for television transmission	J.100–J.109
Interactive systems for digital television distribution	J.110–J.129
Transport of MPEG-2 signals on packetized networks	J.130–J.139
Measurement of the quality of service	**J.140–J.149**
Digital television distribution through local subscriber networks	J.150–J.159
IPCablecom	J.160–J.179
Miscellaneous	J.180–J.199
Application for Interactive Digital Television	J.200–J.209

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Recommendation J.148

Requirements for an objective perceptual multimedia quality model

Summary

This Recommendation details the requirements for the development of an objective multimedia perceptual quality model. The requirements are set out for an auditory-visual model. The current requirements detail the form of the model, the focus for the multimedia modelling component, and the nature of the output necessary for the model to operate as a valuable assessment tool.

Source

ITU-T Recommendation J.148 was prepared by ITU-T Study Group 9 (2001-2004) and approved under the WTSA Resolution 1 procedure on 14 May 2003.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

Introduction

At present, perceptual models for the objective measurement of audio quality and video quality are at a reasonably advanced stage. Existing ITU-T Recommendations are in place for narrowband voice audio quality measurement (ITU-T Rec. P.862) and for the measurement of high quality audio (ITU-R Rec. BS.1387-1). An intermediate subjective quality assessment method is under consideration by ITU-R Working Party 6Q. Continuing research within the Video Quality Experts Group is directed towards providing additional input to the ITU on objective video quality measurement models. Although work has been reported indicating some basic rules underpinning human perception of multimedia quality (see ITU-T Rec. P.911), the development of an overall measure of auditory-visual quality has been neglected. This Recommendation is aimed at defining some of the requirements needed to develop a multimedia perceptual model.

Multimedia is defined as the combination of multiple forms of media such as audio, video, text, graphics, fax, and telephony in the communication of information (ITU-T Rec. Q.1702). For the purpose of this Recommendation, multimedia is confined to any service in which two primary information sources are transmitted. One information source is a single or composite audio signal, the second information source considered is a video or composite image signal. The audio information includes, but is not limited to, mono, stereo, spatial or some other composite form of audio signal. The video or composite image includes, but is not limited to, fixed and variable frame rate services as well as natural, synthetic and natural-synthetic pictures. The signal may be transmitted using any type of transport mechanism. Current ITU-T Recommendations will form the basis for the development of the multimedia model. Relevant Recommendations are ITU-T Recs P.910, P.911, P.920 and P.931.

ITU-T Recommendation J.148

Requirements for an objective perceptual multimedia quality model

1 Scope

The initial goal of the multimedia model is to produce an objective measurement of quality for auditory-visual services. The primary use of the model is to measure the quality of limited bandwidth services (e.g., services delivered at or below 2 Mbit/s). It is considered that limited bandwidth represents a more critical level of service, particularly with respect to the range of quality associated with the audio component of any multimedia service.

It should be noted that the multimedia model should also be applicable to higher bandwidth services, for example in defining any positive influences of high quality audio on perceptions of video/composite image quality. Such cross-modal influences can be used to modify outputs from video/composite image models defined to measure, for example, the picture quality of broadcast services.

The multimedia model should be a flexible tool capable of providing feedback on individual qualities as well as combined quality.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

2.1 Normative references

None.

2.2 Informative references

- ITU-R Recommendation BS.1387-1 (2001), *Method for objective measurements of perceived audio quality*.
- ITU-R Recommendation BT.1359-1 (1998), *Relative timing of sound and vision for broadcasting*.
- ITU-T Recommendation J.144 (2001), *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference*.
- ITU-T Recommendation P.862 (2001), *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*.
- ITU-T Recommendation P.910 (1999), *Subjective video quality assessment methods for multimedia applications*.
- ITU-T Recommendation P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications*.
- ITU-T Recommendation P.920 (2000), *Interactive test methods for audiovisual communications*.

– ITU-T Recommendation P.931 (1998), *Multimedia communications delay, synchronization and frame rate measurement*.

– ITU-T Recommendation Q.1702 (2002), *Long-term vision of network aspects for systems beyond IMT-2000*.

3 Terminology

This Recommendation defines the following terms:

3.1 multimedia: The combination of multiple forms of media such as audio, video, text, graphics, fax, and telephony in the communication of information (ITU-T Rec. Q.1702).

3.2 Aq: Objective measurement of audio quality.

3.3 Vq: Objective measurement of video quality.

3.4 Aq(Vq): Objective measurement of audio quality, accounting for the influence of video quality.

3.5 Vq(Aq): Objective measurement of video quality, accounting for the influence of audio quality.

4 Requirements for a multimedia model

Figure 1 depicts the basic form of the multimedia model. The audio and video/composite image models that are under development provide inputs to the multimedia model. The focus of research is therefore to define the form of the multimedia integration function. The integration function applies specific rules to the information provided by the audio and video/composite image inputs. The form of these rules will be based on data derived from subjective quality experiments. The aim is to produce a set of integration rules that enable the multimedia model to accurately predict human quality perception of systems and services under test. Therefore the validity of the model must be shown by comparing the performance of the model against quality ratings obtained from subjective tests for a range of test materials.

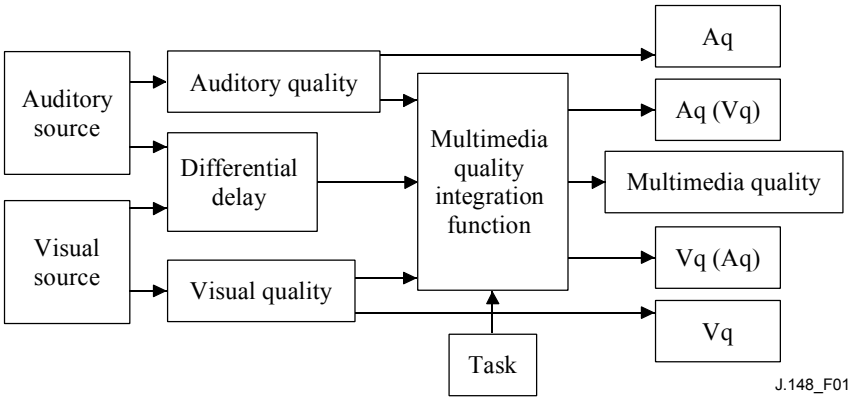

Figure 1/J.148 – Basic components of a multimedia model

The multimedia model contains three primary input modules. These modules provide regular predictions of audio quality and of video/composite image quality for some multimedia service, and an indication of the differential delay between the audio and video/composite image sources (ITU-R Rec. BT.1359-1). Standardized models will be used as the quality inputs to the multimedia model. There is ongoing research in developing perceptual quality models for audio and for

video/composite images with both full reference and reduced/no reference capabilities. A fourth input allows the model to accommodate any task-dependent influences that may impact on quality perceptions. The role of the task input module is to represent the degree of interactivity associated with the multimedia service. The multimedia model will be based on generic rules that capture human perceptions of auditory-visual quality.

The multimedia model should be applicable to the measurement:

i) of services where a reference is available; and

ii) for services in which no reference information is present.

Note that the multimedia model represents the interactions and perceptual consequences of different quality patterns present in an auditory-visual service. This information needs to be obtained from the individual models, so it is the task of the individual input modules to provide the information for the multimedia model to operate effectively. The means by which the individual models extract this information is not a concern of the multimedia model. Clearly it is important for the multimedia modelling researchers to maintain contact with the individual audio and video/composite image modelling groups to ensure that appropriate information is available for input to the multimedia model.

The following requirements follow from the preceding discussion.

1) the core inputs to the multimedia model are the audio model (e.g., ITU-T Rec. P.862) and video/composite image model (e.g., ITU-T Rec. J.144) and a measure of differential auditory-visual delay (ITU-R Rec. BT.1359-1);

2) the task of the multimedia model is to integrate together the qualities from the audio and video/composite image models;

3) the multimedia integration function should be generic, capable of handling inputs from individual models regardless of whether a reference signal is available;

4) a discussion forum with individual model developers needs to be in place so that the inputs from the audio and video/composite image models is appropriate to the needs of the multimedia model;

5) the output from the multimedia model is a prediction of multimedia quality representative of human perception.

5 Multimedia integration function

The multimedia integration function will accommodate human perceptual and cognitive processes active in the formation of quality judgments of auditory-visual services. Data obtained from subjective quality tests will be used to define the integration function. The role of the integration function is to accept inputs from the audio and video/composite image models, apply some predefined rules to the incoming data to produce a multimedia quality prediction.

The integration function will take account of human perceptual responses to multimedia services. In particular, the integration function will apply rules that represent elementary perceptual processes present in subjective quality assessment of multimedia. The multimedia integration function provides five outputs. The primary output is a predicted measure of overall, auditory-visual quality. Four subsidiary outputs provide predictions of perceptual quality for the audio (denoted Aq), video (denoted Vq), audio taking account of any influence the video/composite image may have (denoted Aa(Vq) in Figure 1); and for video/composite image perceptual quality taking account of any influence the accompanying audio quality may have (denoted Vq(Aq) in Figure 1). A number of issues need to be addressed in order for these rules to be defined. For example:

- cross-modal influences, such as the interactions between differing quality levels in different modalities;

- effects of cross-modal error frequency;
- auditory-visual error synchrony;
- cross-modal masking effects.

In addition to the perceptual factors acting on human perceptions of multimedia quality, cognitive factors must also be considered and built into the model. Perhaps the most critical of these is cross-modal attention. The selective and automatic attentional processes can have an important impact on the degree to which specific degradations influence quality opinions. Other cognitive factors, such as task, prior experience and knowledge should also be addressed. Additional requirements include:

1) the integration function contains the combination rules that are used to predict multimedia quality;
2) the integration function takes account of both perceptual and cognitive factors influencing multimedia quality assessment.

The output of the model should be expressed in terms of a standardized quality scale. Some transformation function should be built into the model to enable different scaling of outputs.

1) the multimedia model should output quality predictions of multimedia quality together with information regarding:
    i) the perceived quality of the video/composite image as mediated by the corresponding audio quality (denoted by Aq(Vq) in Figure 1); and
    ii) the perceived quality of the audio as mediated by the corresponding video/composite image quality (denoted by Aq(Vq) in Figure 1);
2) the model should be capable of providing quality predictions on different standard scales.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure and Internet protocol aspects
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2003