International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.912

(08/2008)

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Audiovisual quality in multimedia services

Subjective video quality assessment methods for recognition tasks

Recommendation ITU-T P.912

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30
		P.300
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50
		P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80
		P.800
Audiovisual quality in multimedia services	**Series**	**P.900**
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100

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

Recommendation ITU-T P.912

Subjective video quality assessment methods for recognition tasks

Summary

Recommendation ITU-T P.912 defines subjective assessment methods for evaluating the quality of one-way video used for target recognition tasks. "Target" refers to something in the video that the viewer needs to identify (e.g., a face, object, or number). Target recognition video (TRV) is video that is used as a tool in order to accomplish a specific goal through the ability to recognize specific targets of interest in a video stream. TRV can be used in various video services such as surveillance, human identification, license plate identification, telemedicine, robot control and remote monitoring and decision making.

Source

Recommendation ITU-T P.912 was approved on 13 August 2008 by ITU-T Study Group 9 (2005-2008) under Recommendation ITU-T A.8 procedure.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). 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, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

CONTENTS

Recommendation ITU-T P.912

Subjective video quality assessment methods for recognition tasks

1 Scope

This Recommendation defines subjective assessment methods for evaluating the quality of one-way video used for target recognition tasks. "Target" refers to something in the video that the viewer needs to identify (e.g., a face, object, or number). Target recognition video (TRV) is video that is used as a tool in order to accomplish a specific goal through the ability to recognize specific targets of interest in a video stream. TRV can be used in various video services such as surveillance, human identification, license plate identification, telemedicine, robot control, and remote monitoring and decision making.

This Recommendation considers three categories of target:

1) Human identification (including facial recognition).

2) Object identification.

3) Alphanumeric identification.

Each of these areas requires specific video test material that spans realistic conditions with stimuli that are carefully chosen to allow multiple scenarios to be created repeatedly with different objects of interest, in different lighting conditions, or with small changes in scene details.

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.

[ITU-T P.910] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications*.

[ITU-R BT.500-11] Recommendation ITU-R BT.500-11 (2002), *Methodology for the subjective assessment of the quality of television pictures*.

3 Terms and definitions

This Recommendation defines the following terms:

3.1 discrimination class (DC): One of four levels of visual discrimination at which the target can be analysed:

- Elements of the action – in a very broad and general sense, identification of the series of events that took place.
- Target presence – recognition/detection of the presence or absence of valid targets.
- Target characteristics – recognition of unique characteristics of the target (e.g., markings, scars, tattoos, dents, color).
- Target positive recognition – recognition of a specific instance of the target (e.g., recognition of a person, a specific object, or an exact alpha-numeric sequence).

3.2 scenario group (SG): A collection of scenes of the same basic scenario, with very slight differences between the scenes.

4 Abbreviations

This Recommendation uses the following abbreviations:

ACR Absolute Category Rating

DC Discrimination Class

SG Scenario Group

TRV Target Recognition Video

5 Source signal

Test sequences should follow the general principles stated in [ITU-T P.910] and [b-T1.801.01], which specify that scenes should be consistent with the transmission service under test, and should span the full range of spatial and temporal information. It is critical for the nature of these evaluations that the stimuli used actually reflect the true operational parameters of the conditions under which the video material is collected, and cover the entire range of scenarios possible for the application area that one is identifying. Unlike other subjective assessment methods developed for quality evaluations, this method is directed at the usefulness of the video material to complete a task and not the quality of the video itself.

6 Test methods and experimental design

For video that is used to perform a specific task, it may not be appropriate to rate the quality of the video according to a subjective scale such as absolute category rating (ACR) [ITU-T P.910]. The goal of test methods for TRV is to assess the ability of a viewer to recognize the appropriate information in the video, regardless of the viewer's perceived quality of the viewing experience. To assess the quality level of TRV, methods that reduce subjective factors and measure the ability of a participant to perform a task are useful in that they avoid ambiguity and personal preference.

The application of TRV is directly related to the ability of the user to recognize targets at increasing levels of detail. These levels are referred to as discrimination classes (DC). When determining the DC for particular scenarios, one must consider that for a set distance from the camera to the object of interest, the DC directly correlates to decreasing video resolution of the target, and therefore the object is represented by fewer cycles per degree of resolution. Fewer cycles per degree of resolution also means that the object subtends less of the information content of the video, making identification of the target more difficult.

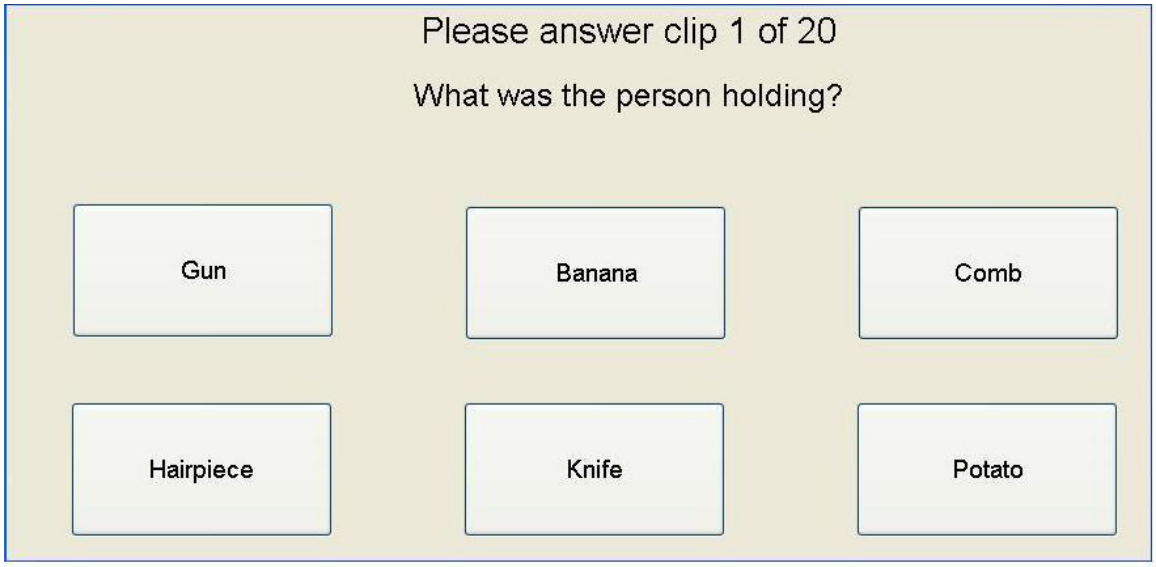
Experimental methods should consist of responding to questions relating to the content in the image or video. The parameter addressed by the question is the target to be recognized.

6.1 Multiple choice method

This method is appropriate for all DC levels and target categories (human, object and alphanumeric). For this method, the video is shown above a list of verbal labels representing the possible answers. After presenting the video, the viewers must choose the label closest to what they recognized in the clip. The use of fixed multiple choices eliminates any possible ambiguity that could arise from open questions, and allows for more accurate measurements.

The number of choices offered to the viewer will depend on the number of alternative scenes being presented. "Unsure" may be one of the listed choices.

An example of the test screen a viewer would see is shown below.



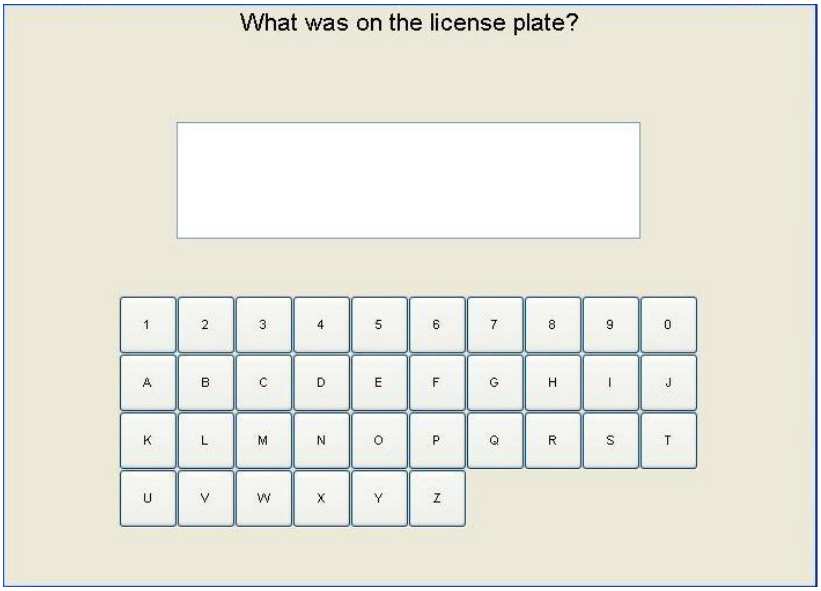
6.2 Single answer method

If there is a non-ambiguous answer to an identification question, the single answer method may be used. This method is appropriate for alphanumeric character recognition scenarios. A viewer is asked what letter(s) or number(s) was present in a specific area of the video, and the answer can be evaluated as either correct or incorrect.

Yes or No tests also fall under this method. A viewer may be asked if a certain object was present in the clip, for example. In this method, it is important to ensure that the procedure used to gather the viewers' responses is easy to understand, so that the test interface does not distract from the cognitive processing required for actual identification of the alphanumeric characters or object. Care must also be taken to avoid terminology that may differ from participant to participant.

"Unsure" may be the third possible answer.

An example of an alphanumeric single answer viewer screen is shown below.



6.3 Timed task method

A viewer may be asked to watch for a particular action or object to be recognized in the video clip. When the viewer perceives that the target has occurred, a timer button can be pushed. In the timed task, the experimenter is able to determine if the time falls within an acceptable time-frame for decision making. These time-frames will be defined by the field in which the video is used, e.g., a person responding to a riot who needs to identify if the crowd has real weapons versus a person who is chasing a car and needs to read the license plate.

6.4 Real-time vs. viewer-controlled viewing

Depending on the nature of the task, TRV test methods can be used either in real time, without the ability to freeze or rewind, or they can be used for non real-time analysis. The experiment should mimic the real world application of the video. If the intended use of the video is for analysis, the ability should be provided to the subject under test to control the playing of the test clip.

6.5 Scenes

Since TRV is generally used to perform a recognition task, the scenes should contain targets consistent with the application under study. However, because the measurements are focused on a subject's ability to identify objects and actions, the possibility that a viewer may memorize the scene content and use other visual clues to remember the identity of the target must be addressed. Therefore, an individual scene may be replaced by a set of scenes containing multiple versions, with controlled differences between the versions. This is called a scenario group (SG). For example, the scenario could be that a person walks across the field of view carrying an object. The SG would consist of multiple shots using different objects or different people. The number of scenes in a SG should be large enough so that scene memorization is unlikely. An example of three scenes from one scenario group is shown below. The scene content is almost identical except for the single change in the object being held.

The content of the scenes should be determined by experts in the application for which the video will be used. These experts should identify critical *tasks*, critical *scenes* in which these tasks are accomplished and critical *parameters* of the scenes. These parameters will be used in the design of the experiment to create the set of multiple choice answers. The scenes should be created in a way that the parameters of interest appear in the video at the resolution that would be realistically expected; that is, the parameters should occupy a realistic percentage of the field of view.

6.6 Experimental design

The experimenter should follow the guidelines outlined in [ITU-T P.910].

6.7 Reference conditions

The experimenter should follow the guidelines outlined in [ITU-T P.910].

7 Evaluation procedures

7.1 Viewing and listening conditions

The experimenter should follow the guidelines outlined in [ITU-T P.910].

7.2 Processing and playback system

The experimenter should follow the guidelines outlined in [ITU-T P.910].

7.3 Subjects

Subjects who are experts in the application field of the TRV should be used. The number of subjects should follow the recommendations of [ITU-T P.910].

7.4 Instructions to subjects and training session

The subject should be given the context of the task before the video clip is played, and told what they are looking for or trying to accomplish. If questions are to be answered about the content of the video, the questions should be posed before the video is shown, so that the viewer knows what the task is.

8 Statistical analysis and reporting of results

The statistical analysis for each method will vary slightly.

Single answer

For single answer conditions, where the answers are correct or incorrect, a statistical metric to determine if the subject is performing above the level of chance for answering correctly should be implemented. "Unsure" answers should be pooled with the incorrect answers.

Multiple choice

For multiple choice answers, the probability of an incorrect answer needs to be balanced against the ability to answer the questions correctly. The statistic metric in this situation will require an examination of the stability of the answers within and between subject performance metrics. "Unsure" answers should be pooled with the incorrect answers.

Timed task

For the timed tasks, the statistical analysis should incorporate two metrics that will in the end be correlated against each other to understand the impact of correctness versus time taken to perform the task.

The timed factor will be a straight average of time to identify the object, that will then be weighted against the correctness of the answer. For the correctness factor, the same statistical analysis for the single answer conditions will also be applied.

For all conditions a correlation and understanding of the number of cycles per degree or area subtended of the target will be taken into consideration to determine the correlation between success and cycles per degree.

For cases where there are multiple answers, a statistical validity indicator will be required.

Bibliography

[b-ITU-T P.911] Recommendation ITU-T P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications*.

[b-ITU-T P.920] Recommendation ITU-T P.920 (2000), *Interactive test methods for audiovisual communications*.

[b-T1.801.01] ANSI T1.801.01-1995*, *Digital Transport of Video Teleconferencing/Video Telephony Signals – Video Test Scenes for Subjective and Objective Performance Assessment*.
<http://engineers.ihs.com/document/abstract/POWPBBAAAAAAAAAA>

[b-Moyer] Moyer, S., Hixson, J., Edwards, T., and Krapels, K. (2004), *Identification of Small Handheld Objects for Electro-Optic/FLIR Applications*, Optical Engineering. P. 0632201-1 – 12.
<http://adsabs.harvard.edu/abs/2004SPIE.5407..116M>

[b-O'Connor] O'Connor, J., O'Kane, B., Ayscue, K., Bonzo, D., and Nystrom, B. (1998), *Recognition of human activities using handheld thermal systems*, Proceedings of the SPIE Conference on Sensor Technology for Soldier Systems, Orlando, Florida, April, SPIE Vol. 3394.
<http://adsabs.harvard.edu/abs/1998SPIE.3394...51O>

* T1 standards are maintained since November 2003 by ATIS.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
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	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Terminals and subjective and objective assessment methods**
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, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2009