|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION**  **STANDARDIZATION SECTOR**  STUDY PERIOD 2022-2024 | | | SG12-Cn |
| STUDY GROUP 12 |
| Original: English |
| **Question(s):** | | [19/12] | [San Mateo, 26-27 June 2023] | |
| **CONTRIBUTION** | | | | |
| **Source:** | | NTT | | |
| **Title:** | | Draft Test plan of P.obj-recog | | |
| **Contact:** | | Masanori Koike  Japan  NTT | Tel: +81-422-59-4453  Fax: +81-422-59-6364  E-mail: masanori.koike@ntt.com | |
| **Contact:** | | Kazuhisa Yamagishi  Japan  NTT | Tel: +81-422-59-4397  Fax: +81-422-59-6364  E-mail: kazuhisa.yamagishi@ntt.com | |
| **Contact:** | | Noritsugu Egi  Japan  NTT | Tel: +81-422-59-4008  Fax: +81-422-59-6364  E-mail: noritsugu.egi@ntt.com | |

|  |  |
| --- | --- |
| **Keywords:** | P.obj-recog, test plan, subject test, video-quality |
| **Abstract:** | This contribution proposes test plan for P.obj-recog. |

[List of Acronyms 2](#_Toc138188531)

[Abstract 2](#_Toc138188532)

[1 Experimental environment 3](#_Toc138188533)

[1.1 Display equipment 3](#_Toc138188534)

[2 Experimental materials 3](#_Toc138188535)

[2.1 Scene of the materials 3](#_Toc138188536)

[2.2 Visual material 3](#_Toc138188537)

[3 Subjective assessment method 4](#_Toc138188538)

[3.1 Instruction and training phase 4](#_Toc138188539)

[3.2 The procedure of object recognition task 4](#_Toc138188540)

[3.3 The structure of sessions 5](#_Toc138188541)

[3.4 Viewers 5](#_Toc138188542)

[3.5 Calculation of object recognition ratio 5](#_Toc138188543)

[4 Experimental conditions 6](#_Toc138188544)

[4.1 SRC features and conditions 6](#_Toc138188545)

[4.2 Conditions for key quality factors and parameters 6](#_Toc138188546)

[5 References 7](#_Toc138188547)

List of Acronyms

TRV Target Recognition Video

PVS Processed Video Sequence

SRC Source Reference Channel or Circuit

Abstract

This document defines the procedure for conducting subjective tests to evaluate the recognition ratio in surveillance video of autonomous driving conducted by ITU-T study group 12, Question 19 for the P.obj-recog work item.

As described in P.912 [1], the goal of test methods for TRV is to assess the ability of a viewer to recognize appropriate information in a video, regardless of the viewer’s perceived quality of the viewing experience. For object recognition for surveillance video of autonomous driving, the timed task method is appropriate for the subjective test. In the timed task method, a viewer may be asked to watch for a particular action or object to be recognized in the video clip, and when the viewer perceives that the target has appeared, a stop button can be pushed and the time taken for recognition can be recorded.

# 1 Experimental environment

## 1.1 Display equipment

Table 1 lists the display devices to be used in the test and their screen sizes and resolution ranges.

Table 1: Display devices parameters

|  |  |  |
| --- | --- | --- |
| **Display device** | **Display size** | **Resolution** |
| PC | Larger than 14 inches and smaller than 50 inches | Up to 1080p |

# 2 Experimental materials

To test many autonomous driving monitoring service scenarios and make the test process repeatable, some TRVs for this subjective test are selected.

## 2.1 Scene of the materials

The content of the scenes should be determined by the situation in which the driving is conducted, and the surveillance camera of the car took video outside. The scene should contain targets consistent with the application under study. Since multiple possible objects can be seen in the video, the object type can be different for each video. The video may contain scenes where an object is on the road and a collision is anticipated, or a person jumps out onto the road and a collision is anticipated. These situations are determined for each experiment. The purpose and criteria for recognizing objects should be clearly explained to the viewer.

## 2.2 Visual material

The video codec will be H.265, and the video resolution will be FHD (1080p). Table 2 lists the recommended video parameters.

**Table 2:** **Recommended video parameters**

|  |  |  |
| --- | --- | --- |
| **Video resolution** | **Bitrate range** | **Framerate** |
| 1080p | 200 kbps – 5Mbps | 60 fps |

# 3 Subjective assessment method

The subjective test consists of instruction, training phase, and test phase. The training phase and test phase have the object recognition task. The time for the recognition will be evaluated using the timed task method in accordance with Recommendation ITU-T P.912 [1].

## 3.1 Instruction and training phase

Before actually starting the experiment, a scenario of the intended application of the system under test should be given to the subjects. In addition, the procedure of recognition of an object and presentation of the stimuli is given in the written form. The range and type of impairments should be presented in the training phase, which may contain video sequences other than those used in the actual tests. The structure of the training phase is the same as that of the session in the test phase described in Section 3.3, and the minimum number of PVSs for training is 8. Questions about the procedure or the meaning of the instructions should be answered carefully to avoid bias and only before the start of the session.

### 3.2 The procedure of object recognition task

The test sequences are presented one at a time in the time task method. A subject may be asked to watch for objects to be recognized in the video clip. The procedure of a time task method is shown in Figure 1. Subjects watch driving videos, and when the subject perceives the target has appeared, a stop button such as a keyboard or mouse can be pressed. After that, test sequences stop, and the time taken for the object recognition is recorded. After stopping the video, a confirmation screen is displayed to check whether the subject recognizes the object correctly. If the object is recognized correctly, the subject moves on to the next video. If the subject mistakenly presses the stop button despite not recognizing an object, the subject can watch the video again. Note that the subject should only watch the video again if they mistakenly pressed the stop button and not because they did not recognize an object. After this confirmation screen, the next video will be displayed.

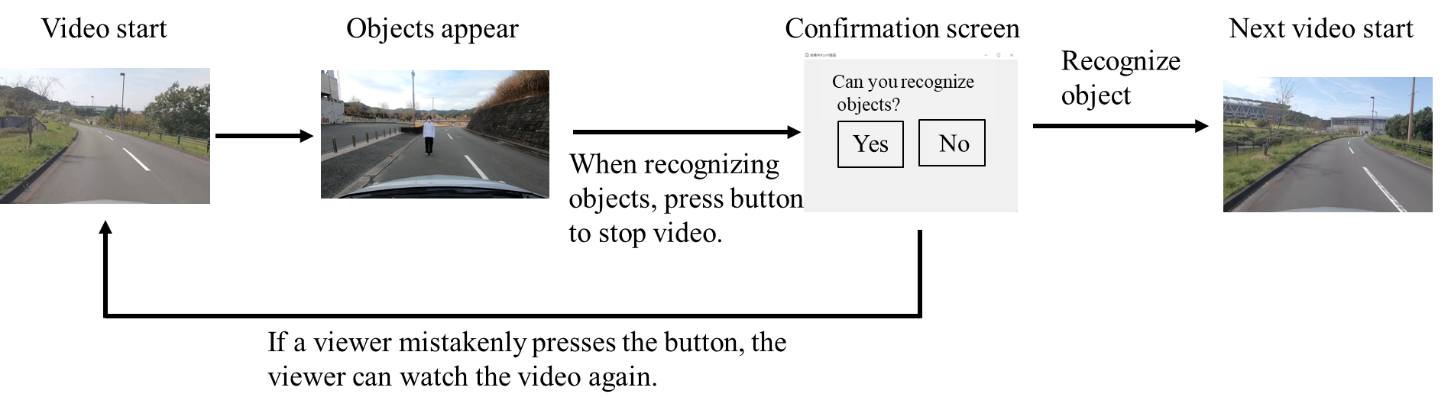


Figure 1: Procedure of object recognition task

## 3.3 The structure of sessions

The session is the time when the subjects watch PVSs and recognize objects (with no breaks). The time pattern for one session can be illustrated in Figure 2. The duration of each video is about 30 to 50 seconds. It is desirable that an object should be displayed after about 10 to 20 seconds from the time the video starts. However, the duration for an object to appear should differ among the videos because if objects are displayed at the same time from when the video starts, the viewer will soon recognize that the object appears at a certain timing. Note that the display time varies depending on the viewers and videos. The maximum number of PVSs for one session is 10, and the maximum number of sessions is 8.

The maximum length of a subjective test is 3 hours, and the maximum number of PVSs that one subject watches in the experiment is 80. In the time tasked method, one SRC should be watched only once because even if the subject watches the same SRC with different qualities, the subject may remember the object’s position.



Figure 2: Structure of the session

## 3.4 Viewers

The number of subjects in a viewing test should be the same as that specified in P.913 [3], which is mainly as described in P.912, and the minimum number of subjects is 24. The subjects should not be directly involved in multimedia quality evaluation as part of their work or be experienced as assessors.

## 3.5 Calculation of object recognition ratio

The object recognition ratio is calculated on the basis of the recognition percentage at object recognition points. To obtain the distance to the object when the subject recognizes the object, the distance between the object and the frame when the subject presses the stop button is measured. Therefore, for each PVS, the distance to the object is calculated. The object recognition ratio is defined as how many people can be recognized at what distance for each PVS. For example, if there are 20 subjects, and 12 can recognize the object 20 m away, then the recognition ratio for 20 m is 60%. In this way, the object recognition ratio is calculated for each PVS.

# 4 Experimental conditions

## 4.1 SRC features and conditions

As described in Section 2.1, SRCs contain scenes where an object is on the road and a collision is anticipated, or a person jumps out onto the road and a collision is anticipated. The recognition ratio may differ between autonomous cars drives in rural and urban areas, because the different areas have different backgrounds on the road, which changes the difficulty of encoding the video. The weather or day and night may also affect the recognition ratio. Moreover, the recognition ratio differs in accordance with object size, colors, and other features of objects. The coefficients of the object-recognition-estimation model differ from these source-related factors. Ideally, source-related factors should be tested for, but these factors have a wide range of topics, and the necessary conditions differ depending on the situation in which the system is used. Since the recognition ratio varies greatly depending on how the assumed video is used, the test is conducted by limiting to one source-related factor.

## 4.2 Conditions for key quality factors and parameters

Experimental conditions are selected from the range described in the ToR. The following sets of key quality factors that affect the object recognition ratio in autonomous driving monitoring services will be considered in the P.obj-recog, and the included parameters are shown in Table 3.

Table 3: Key quality factors and parameters

|  |  |  |
| --- | --- | --- |
| **Key quality factors** | **Parameters** | **The value corresponds to the parameters** |
| Codec-related factors | Video codec | HEVC |
| Video resolution | 1920 x 1080 |
| Video bitrate | 200 kbps – 5 Mbps |
| Video framerate | 60 fps |
| Network-related factors | Packet-loss rate | 0.2 – 10 % |
| Error concealment | Block noise, freeze |
| Car-related factors | Car velocity | 10 – 40 km/h |
| Source-related factors | Object color, size | Red, height:70 cm, width: 40 cm |
| Object background | Rural area |
| Weather | Sunny |
| Day and night | Day |

# 5 References

1. Recommendation ITU-T P.912, Subjective video-quality assessment methods for recognition tasks, 2016
2. Recommendation ITU-T P.913, Methods for the subjective assessment of video quality, audio quality, and audiovisual quality of Internet video and distribution quality of television in any environment, 2021

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_