



INTERNATIONAL TELECOMMUNICATION UNION

**ITU-T**

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

**P.920**

(05/2000)

SERIES P: TELEPHONE TRANSMISSION QUALITY,  
TELEPHONE INSTALLATIONS, LOCAL LINE  
NETWORKS

Audiovisual quality in multimedia services

---

**Interactive test methods for audiovisual  
communications**

ITU-T Recommendation P.920

(Formerly CCITT Recommendation)

---

ITU-T P-SERIES RECOMMENDATIONS

**TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS**

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Subscribers' lines and sets	Series	P.30 P.300
Transmission standards	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 quality	Series	P.80 P.800
<b>Audiovisual quality in multimedia services</b>	<b>Series</b>	<b>P.900</b>

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

## **ITU-T Recommendation P.920**

### **Interactive test methods for audiovisual communications**

#### **Summary**

This ITU-T Recommendation is intended to define interactive evaluation methods for quantifying the impact of terminal and communication link performance on point-to-point or multipoint audiovisual communications. This methodology is based upon conversation opinion tests, and can be considered to be an extension of the methods defined in Annex A/P.800.

#### **Source**

ITU-T Recommendation P.920 was prepared by ITU-T Study Group 12 (1997-2000) and approved under the WTSC Resolution 1 procedure on 18 May 2000.

## 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 Conference (WTSC), 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 WTSC 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.

## 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.

© ITU 2001

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from ITU.

## CONTENTS

	<b>Page</b>
1 Scope.....	1
2 References.....	1
3 Experimental design .....	1
3.1 Basic approach and factors to be investigated .....	1
3.2 Stimuli used in conversational tests.....	1
3.3 Test conditions and experimental design.....	2
3.4 Subjects.....	3
3.5 Subject training and reference connections .....	3
3.6 Ambient room and equipment characteristics.....	3
4 Solicitation of opinions.....	4
Appendix I – Examples of tasks and stimuli for conversation.....	5
I.1 Stimuli for conversation.....	5
I.2 Tasks to evaluate the effects of speech delay on communication quality.....	5
I.3 Tasks to evaluate the effects of audiovisual delay and/or transmission errors on communication quality.....	6
I.4 Task to evaluate the synchronization between audio and video signals .....	6
Appendix II – Protocols for the stimuli for conversation.....	6
II.1 Protocol for the Name-Guessing task .....	6
II.2 Protocol for the Story-Comparison task .....	7
II.3 Protocol for the Picture-Comparison task.....	7
II.4 Protocol for the building blocks task .....	7
II.5 Protocol for the object-description task .....	8
Appendix III – Test condition questionnaire.....	9
Appendix IV – Exit questions .....	10
Appendix V – Bibliography .....	11

## **Introduction**

The audiovisual interactive test methods described in this ITU-T Recommendation are intended for quantifying the impact of terminal and communication link performance, that may affect the ability to conduct an interactive audiovisual communication.

The efficacy of these tests strongly depends on the ability to reproduce in laboratory environments the conditions that are very close to the real situations. In this regard, particular care must be taken in choosing the tasks proposed to the subjects. In general, those tasks used in conversation tests for telephony assessment are not suited for audiovisual assessment because they often distract the subject's attention from the video screen. Therefore new tasks have been developed following the criteria illustrated in this ITU-T Recommendation.

Substantial work has been done in this area, although all aspects of audiovisual quality are not yet completely understood.

This ITU-T Recommendation reflects the current status of research on interactive audiovisual testing.

As progress on this work continues, understanding of these interactive test methods will no doubt improve. As new knowledge is attained, this ITU-T Recommendation will be revised.

# ITU-T Recommendation P.920

## Interactive test methods for audiovisual communications

### 1 Scope

This ITU-T Recommendation is intended to define interactive evaluation methods for quantifying the impact of coding artifacts, transmission delay and transmission impairments (e.g. packet loss, cell loss, digital channel errors) on point-to-point or multipoint audiovisual communications. This methodology is based upon conversation opinion tests, and can be considered to be an extension of methods defined in Annex A/P.800.

This ITU-T Recommendation does not cover topics that are already described in other P-series ITU-T Recommendations such as the objective measurements of the link.

### 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; all 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.

- ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.
- ITU-R Recommendation BT.812 (1992), *Subjective assessment of the quality of alphanumeric and graphic pictures in Teletext and similar services*.
- ITU-T Recommendation P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications*.
- IEC Publication 60651 (1979), *Sound level meters*.

### 3 Experimental design

#### 3.1 Basic approach and factors to be investigated

In order to quantify the impact of factors, such as transmission delay, that may affect the ability to conduct an interactive communication, the approach proposed in this ITU-T Recommendation is based on an active talker conversation assessment. Further, since it is necessary to express these opinions using a rating system, several single-stimulus rating scales are proposed.

#### 3.2 Stimuli used in conversational tests

In general, in conversational opinion tests it is desired to minimize the artificiality of the environment. However, at the same time, it is necessary to invoke some method to stimulate interactive communication utilizing the conditions which are being evaluated. In telephony assessments, it is common to use a set of photographs, or some other form of printed material, to achieve this objective. In audiovisual terminal performance assessments, however, such mechanisms are likely to distract a participant's attention from the video screen, thus possibly leading to an unnatural mode of communication for this type of terminal.

For general applications, the following guidelines are provided for designing task-based tests:

- the task should be designed such that, during their conversation, the subjects primarily maintain their attention on the audiovisual terminal;
- the task must have sufficient face value, that is, it must resemble real-life audiovisual communication to a sufficient degree. In particular, it is preferable that the task be performed by two subjects and not by one subject and an experimental leader;
- the task must yield reproducible quantitative results that represent adequate measures of communication efficiency. When time delays are involved, time measures should be among the results.

A wide range of subjects, including elderly and hearing-impaired subjects, should be able to perform the task.

It is preferable that the task is, in itself, sufficiently rewarding for the subjects. This has several advantages: the subjects learn the task faster and they are less susceptible to fatigue and loss of motivation.

From past experiments, it has been found that lively audiovisual conversations can be stimulated if the participants in such a test know each other. Subsequently, the provision of written material can be used as a secondary, rather than primary, source of stimulation. Thus, unlike telephony, familiarity between pairs of conversing participants is highly desirable, if not essential.

It is recognized, however, that for specific applications, the conversational tasks may have to be modified to take into account the services that the system under test is intended to provide. In order to permit meaningful measurements to be made of the factors being investigated, it is recommended that in such cases the conversational tasks be structured so as to represent the applications of interest, particularly as regards:

- a) the rate of information exchange; and
- b) the degree of audio and video signal utilization.

For example, to account for the attributes in the first category, tasks could range from predominately one-way communication, to free-conversation, to a rapid exchange of information, be it via video, audio or both signals. Similarly, to test attributes in the second category, tasks could range from the subjects working on a hard-copy document in front of them (minimum use of video information) to reading sign language over the video link (maximum use of video information). The actual tasks should combine attributes from both categories. These guidelines have been applied to develop the tasks illustrated in Appendix I, and the protocols for the tasks are detailed in Appendix II.

### **3.3 Test conditions and experimental design**

In general, at least one transmission impairment factor or test condition is likely to be evaluated in a test, in addition to a baseline (reference) condition where the impact of such factor is minimum (when using the reference condition, this should not be identified as such to the participants). However, because conversational tests are time-consuming, the total number of conditions ought to be reasonably constrained in order to minimize participant fatigue and maximize experimental accuracy. This requirement should be balanced against the need to ensure that the duration of each conversation/condition is at least five minutes long.

As with conversational tests (audio communications), a Latin or Greco-Latin square may be found to be a suitable experimental design for this purpose. In such case, the square's rows may be associated with the test participants and the square's columns with the order in which the conditions in the test are being presented.

Other treatments may also be appropriate depending on the factors being investigated. For example, past experiments have appeared to indicate that there may be an interaction between the audiovisual communication path quality and the perception of the impact of transmission delay. Consequently, it

may be preferable to apply two treatments using a Latin-Latin square design, so that the letters of the first alphabet are associated with different values of transmission delay and the letters of the second alphabet are associated with different image/voice coding rates.

Of course, other experimental designs including replicated block designs and Youden square designs may be suitable and could be left up to the experimenter to select in order to meet specific cost and accuracy objectives in view of the number of conditions of interest.

Also any possible effects related to the order in which the tasks are performed must be taken into account.

### **3.4 Subjects**

At least 16 subjects should participate in a test, the exact number will be dictated by the experimental design and the accuracy required to the results. These subjects should be non-expert, and they should not be directly involved with either audio and/or video technology as part of their normal work.

Nevertheless, in the early phases of the development of audiovisual communications systems and in pilot experiments carried out before a larger test, small groups of experts (4-8) or other critical subjects can provide indicative results with sufficient reliability.

### **3.5 Subject training and reference connections**

Before starting the experiment, a scenario of the intended application of the system under test should be given to the subjects. The range and type of impairments should be shown in a preliminary phase. During this phase, a first level of personal introduction may thus be allowed to take place over the communication link at the worst (or best) experimental condition, while further discussion pertinent to the tasks expected of the participants can be subsequently permitted at the best (or worst) experimental condition.

Again, as with the main test, the particulars of the conditions should not be revealed to the test participants.

### **3.6 Ambient room and equipment characteristics**

Table 1 lists typical viewing and listening conditions as used in audiovisual quality assessment. The actual parameter settings used in the assessment should be specified. For the comparison of test results, all viewing and listening conditions must be fixed and equal over laboratories for the same kind of tests.

Both the size and the type of monitor used should be appropriate for the application under investigation. When sequences are presented through a PC-based system, the characteristics of the display and audio transducers must be specified, e.g. dot pitch and of the monitor, type of video display card used, characteristics of either handsets, headphones or loudspeakers, etc.

In particular, in case of loudspeaker presentation, the number and positions of the loudspeakers relative to the image should be reported.

Operational parameters, such as signal level, for the test sequences shall match those of the alignment signal used to verify the viewing and listening conditions. Any operational adjustments performed so that source or processed sequences meet this requirement should be reported.

Synchronization between audio and video signal should be measured according to ITU-T Recommendation P.911 and reported.

**Table 1/P.920 – Typical viewing and listening conditions  
as used in audiovisual quality assessment**

Parameter	Setting
Room size (Note 7)	Specify L × W × H
Viewing distance (Note 5)	1-8 H
Peak luminance of the screen	100-200 cd/m <sup>2</sup>
Ratio of luminance of inactive screen to peak luminance	≤ 0.05
Ratio of the luminance of the screen, when displaying only black level in a completely dark room, to that corresponding to peak white	≤ 0.1
Ratio of luminance of background behind picture monitor to peak luminance of picture (Note 1)	≤ 0.2
Chromaticity of background (Note 6)	D <sub>65</sub>
Background room illumination (Note 1)	≤ 20 lux
Background noise level (Note 2)	≤ 30 dBA
Listening level (Note 3)	~ 80 dBA
Reverberation time (Note 4)	< 500 ms, ∀f > 150 Hz
<p>NOTE 1 – This value indicates a setting allowing maximum detectability of distortions, for some applications higher values are allowed or they are determined by the application.</p> <p>NOTE 2 – If background noise levels used in the application are significantly higher, Hoth noise should be used for environments like office. The Hoth noise should be generated in a room with low background noise levels (≤ 30 dBA) and be measured in the acoustical domain. If for some special type of application another type of background noise is used, the spectral power density and dBA level should be specified.</p> <p>NOTE 3 – This value indicates a setting allowing maximum detectability of distortions, for some applications lower or higher values are allowed. Level setting is measured using the maximum value over the audio sequence using IEC Publication 60651 fast averaging. When loudspeakers are used, the sound level might be adjusted according to the viewing distance.</p> <p>NOTE 4 – This value is only relevant for loudspeaker presentation, larger reverberation times generally lead to a decrease in detectability of distortions.</p> <p>NOTE 5 – For a given screen height, it is likely that the viewing distance preferred by the subjects increases when visual quality is degraded. Concerning this point, the preferred viewing distance should be predetermined for qualification tests. Viewing distance in general depends on the applications. Viewing distance should be defined taking into account not only the screen size, but also type of applications, type of screen and the goal of the experiment.</p> <p>NOTE 6 – For PC monitors, the chromaticity of background may be adapted to the chromaticity of the monitor.</p> <p>NOTE 7 – Room size is important only for loudspeaker presentation.</p>	

#### **4 Solicitation of opinions**

As with telephone conversational assessments, each participant should also be separately solicited for his, or her, opinion after the completion of each condition. It is preferable to structure in advance the type of questioning and minimize the number of questions posed after each condition and to minimize uncontrolled variations. A possible test condition questionnaire that could be used for this purpose can be found in Appendix III.

Several category judgement scales can be used to evaluate the audiovisual terminal performance. The sensitivity of these scales to various transmission impairment factors may be different.

Examples of scales that may be used for this purpose are:

- Overall Audiovisual Quality, Video Quality, and Audio Quality are generally assessed using the categories: Excellent, Good, Fair, Poor, or Bad.
- Effort Needed to Interrupt is generally assessed using the categories: No Effort, Minor Effort, Moderate Effort, Considerable Effort, or Extreme Effort.
- Communications Difficulty and Acceptability of Communication are generally assessed using a binary choice: Yes or No.

Although selection of any particular scale may be subject to the goals of an individual experiment, it is important that these scales (and their associated wording, or translation) be used consistently by different laboratories.

If the use of audiovisual terminals is novel for most participants, it is recommended that "Exit Questions" are presented after the last test condition has been rated. Such questions should attempt to capture any other factors that may have been inadvertently omitted from the experiment. A sample of such questions is given in Appendix IV.

## APPENDIX I

### Examples of tasks and stimuli for conversation

#### I.1 Stimuli for conversation

The following tasks differ from each other in the degree to which free conversation can occur. The protocols for the tasks are described in more detail in Appendix II.

Task 1: The Name-Guessing task. The name-guessing task is a question-answer game performed according to a fixed protocol. This results in a very restricted conversation.

Task 2: The Story-Comparison task. In the story-comparison task, subjects have to discover a number of differences between two versions of a story. They are allowed unrestricted conversation. Prior to the test, both subjects have to read and memorize a short story.

Task 3: The Picture-Comparison task. In this task, the subjects have to memorize a picture and subsequently determine whether they were given identical or different pictures. Conversation is not restricted.

#### I.2 Tasks to evaluate the effects of speech delay on communication quality

In the following tasks the talk spurt increases from task 1) to task 6), whereas the conversation switching rate decreases.

- 1) take turns in counting;
- 2) take turns reading random numbers aloud as quickly as possible;
- 3) take turns verifying random numbers aloud as quickly as possible;
- 4) words with missing letters are completed with letters supplied by the other talker;
- 5) take turns verifying city names as quickly as possible;
- 6) determine the shape of a figure described verbally;
- 7) free conversation.

The previous tasks [with exception of task 1) and task 7)] cannot be used for audiovisual quality evaluations because most of them require the subjects to concentrate their attention on a sheet of paper and not on the screen.

### **I.3 Tasks to evaluate the effects of audiovisual delay and/or transmission errors on communication quality**

The following tasks should draw the attention of the assessors to the video signal.

The protocols for the tasks 1) and 2) are described in more detail in Appendix II.

- 1) one of the subjects shows and describes a plastic building block and the other one is required to reproduce it;
- 2 each subject in a pair is given a detailed object to describe to their partner;
- 3) one of the two subjects shows some exercises for the alleviation of neck pain and the other one is required to reproduce them.

### **I.4 Task to evaluate the synchronization between audio and video signals**

The following task is intended for drawing the attention of the assessors to the synchronization between audio and video signals:

- One person claps while the other checks the synchronization between the movement and the sound.

The previous task is not suitable for assessing lip synchronization.

Additional tasks aimed at evaluating the synchronization between the two signals are for further studies.

## **APPENDIX II**

### **Protocols for the stimuli for conversation**

#### **II.1 Protocol for the Name-Guessing task**

The following protocol for the Name-Guessing task can be used:

"The first task, the name-guessing task, is a question-answer game performed according to a fixed protocol. In this task, one of the subjects receives three pieces of information: first, either the word "Brand" or the word "Person", indicating whether it is a brand name or a name of a (well-known) person that should be guessed; second, in the case of a brand, a description of the product, and in the case of a person, his or her profession; third, the name to be guessed. Thus, the subject would, for example, be presented with the text:

- Brand;
- Cigarettes;
- Camel."

"In guessing the name of the brand or person, the second subject has to adhere to the following protocol: the first question should be "Is it a brand or a person?"; the second, in the case of a brand, "What is the product?", and, in the case of a person "What is the profession?"; subsequently, one guess can be made; when the guess is false, consecutive letters of the name should be asked for and a guess can each time be made; this continues until the name is guessed correctly or until the entire name has been spelled out. The subjects being interrogated are allowed to consult their text during the conversation. Most subjects do not find this necessary, however, because they had no trouble retaining the limited amount of information contained in it.

After the name is guessed, the total time required and the number of letters that are requested are scored. Several names, varying in length and difficulty, should be guessed in a given experimental condition. A linear regression analysis can then be applied to the data so that estimates are obtained of the time required for a direct guess (zero letters suggested) and the extra time per letter."

## **II.2 Protocol for the Story-Comparison task**

The following protocol for the Story-Comparison task can be used:

"In the second task, subjects have to discover a number of differences between two versions of a story. They are allowed unrestricted conversation. Prior to the actual test, both subjects have to read and memorize a short story. They are given stories that are essentially the same but that contain a certain number of small but distinct differences. (For example stories of approximately 200 words, containing six differences.) In addition, they receive a list of questions about the story, which they have to answer for themselves. The questions are intended to improve retention of the story. After the memorization period, the subjects have to start a conversation with the aim of discovering as quickly as possible all differences between the two stories. They are not allowed to consult the text during the conversation. The subjects know how many differences there are and they get feedback informing them when they have detected a difference and how many differences they have detected thus far. The conversation continues until all differences are detected or until no differences have been detected for a certain time."

"The starting time of the conversation and the times at which detection of a difference occurs are registered. In the analysis, the time interval between the start and the detection of the first difference, and the intervals between subsequent detections are determined. These intervals are then ordered according to their duration. This is done because the time interval, i.e. the time required for the detection of the next difference, depends both on the strategy used by the subject pair and on where the differences are located in the text. Reordering the intervals reduces the variance introduced by these factors. In addition, it allows evaluation of the effect being investigated as a function of the duration of the interval."

## **II.3 Protocol for the Picture-Comparison task**

The following protocol for the Picture-Comparison task can be used:

"In this task, the subjects have to memorize a picture and subsequently determine whether they were given identical or different pictures. (Pictures of various subjects, such as landscapes, buildings and urban sites, can be used.) Conversation is not restricted. The subjects are not allowed to look at the picture during the conversation.

The results that are scored are the total time required by the subjects to reach a decision and whether or not the decision is correct. As subjects normally need more time when the pictures are identical than when they are different, the results for both cases have to be separated. Though this task is somewhat similar to the story-comparison task, there is a difference in the degree to which free conversation will occur. This is because, in comparing stories, the subjects tend to borrow phrases from the text, known to both of them, whereas the task of comparing pictures forces them to use their own wordings."

## **II.4 Protocol for the building blocks task**

The following protocol for the building blocks task can be used:

"One subject was given a bag of multicolour, interlocking construction blocks. The other subject was given a completed figure made from an identical set of blocks. The first subject's goal was to build the same figure with help from the other subject, and verify its correctness.

The blocks chosen were designed for pre-school children (5 years old or less), with a minimum size of 3cm × 3cm × 2cm. Some example figures were birds, trucks, and arches, typically composed of ten or fewer pieces. Completed figures were easily held in one hand, leaving the other hand free to build or point out features. The resolution of the video system was sufficient to see the shape of the figures, but the borders between blocks of the same colour were not easily visible. Again, use of self-

view was prohibited. The verification suggested in the instructions required both subjects to rely on the video channel, as well as the audio.

The instructions to subjects had the same features as those of the "Describe an Object" task. Examples follow:

**BUILD an OBJECT from DESCRIPTION – INSTRUCTIONS:**

Your partner has been given an object to describe to you. You must build a similar object from loose blocks and your partner's instructions. Your partner will display the object to you and point out its features to help you build it.

When you have completed the call, please answer the following questions.

**DESCRIBE AN OBJECT TO BE BUILT – INSTRUCTIONS:**

You have been given an object to describe to your test partner. Your partner must build a similar object from loose blocks and your instructions. You should hold the object in front of you and point out its features to help your partner. Confirm that your partner understands the description of each feature.

When you have completed the call, please answer the following questions."

## **II.5 Protocol for the object-description task**

The following protocol for the object-description task can be used:

"Each subject in a pair was given a detailed object to describe to their partner. Partners changed between the active listening role, and the describing role at the halfway point of the call. The touch screens in the boothcarrels gave instructions to place calls, and indicated the halfway point and end of the call with audible indicators and messages on the screen.

The objects easily fit in one hand, so that the subject in the describing role could hold the object and point to a specific feature while narrating. The objects chosen had large features with different colours that could easily be seen on compressed video, as well as detailed features that the video system could not reproduce. Also, there was no self-view available to the subjects, so the listener would often need to direct the describer to hold the object higher, closer, etc., to improve their view. This combination of features helped to balance the reliance on the video and audio channels to convey the descriptions, and maintained good interactions throughout the call.

The instructions to subjects were organised as follows. The subjects received their object for each call in a plastic bag. A card attached to the bag gave a short label for the task, and detailed instructions. After conducting the task a few times with different objects, the test subjects only needed to refer to the task label to understand their roles (but instructions were always included).

**DESCRIBE FIRST, THEN LISTEN – INSTRUCTIONS:**

You have been given an object to describe to your test partner. You may hold it in front of you and point out its features as you describe them. Confirm that your partner understands the description of each feature.

Next, your test partner will describe a different object. You may ask questions about the object during the description process.

When you have completed the exchange, please answer the following questions.

**LISTEN FIRST, THEN DESCRIBE – INSTRUCTIONS:**

First, your test partner will describe an object. You may ask questions about the object during the description process.

Next, you have been given a different object to describe to your test partner. You may hold it in front of you and point out its features as you describe them. Confirm that your partner understands the description of each feature.

When you have completed the exchange, please answer the following questions."

### APPENDIX III

#### **Test condition questionnaire**

Questionnaire to be used after each test condition:

Administrator: I will ask you a few questions (again) relative to your opinion of the connection over which you just conversed. Are you ready?

- 1) How would you rate the overall audiovisual quality? (Circle one)  
Excellent    Good    Fair    Poor    Bad
- 2) How would you rate the video quality of the connection? (Circle one)  
Excellent    Good    Fair    Poor    Bad
- 3) How would you rate the audio quality of the connection? (Circle one)  
Excellent    Good    Fair    Poor    Bad
- 4) How would you judge the effort needed to interrupt the other party (or parties)? (Circle one)  
No effort    Minor effort    Moderate effort    Considerable effort    Extreme effort
- 5) Did you have any difficulty during the connection? (Circle one)  
Yes    No
- 6) Was the connection acceptable? (Circle one)  
Yes    No

Administrator: Thank you for your answers. We will now go ahead and re-establish our connection in a few minutes so that we can continue the experiment. As soon as the connection is set up, I shall let you know that you can go ahead and converse with the other party.

APPENDIX IV

**Exit questions**

Additional questions to be used at the end of the test:

Administrator: Since we are at the end of the test, I would like to ask you some supplementary questions. Are you ready?

Could you tell me, in order of importance, which of the following factors would you consider sufficiently important that you would like to see improved. (Evaluator will read a number of factors here.)

7a \_\_\_\_\_ 7b \_\_\_\_\_ 7c \_\_\_\_\_ 7d \_\_\_\_\_

Do you have any other comments relative to the entire test you wish to make known to us?

---

---

---

---

---

Thank you for your cooperation in taking part in the test. I shall be over in a few minutes to wrap up the session and let you out of the building.

## APPENDIX V

### **Bibliography**

- [1] KIRK (R.E.): Experimental Design, *Procedures for the Behavioural Sciences*, 2nd Edition, Brooks/Cole Publishing Co., California 1982.
- [2] BRONKHORST (A.W.) (J.A.) VERHAVE: The effect of audio-video desynchronization on communication efficiency in videotelephony, *Report IZF 1992 C-35* Study for PTT Telecom Netherlands, TNO Institute for Perception 1992.
- [3] HOTH (D.F.): Room noise spectra at subscribers' telephone locations, *J.A.S.A.*, Vol. 12, pp. 99-504, April 1941.

## 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	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
<b>Series P</b>	<b>Telephone transmission quality, telephone installations, local line networks</b>
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