INTERNATIONAL TELECOMMUNICATION UNION

ITU-T F.901

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(03/93)

OPERATIONS AND QUALITY OF SERVICE

HUMAN FACTORS – GENERAL (NOT SPECIFIC TO TELEPHONE SERVICE)

USABILITY EVALUATION OF TELECOMMUNICATION SERVICES

ITU-T Recommendation F.901

(Previously "CCITT Recommendation")

FOREWORD

The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the International Telecommunication Union. The 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, established the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

ITU-T Recommendation F.901 was prepared by the ITU-T Study Group I (1988-1993) and was approved by the WTSC (Helsinki, March 1-12, 1993).

NOTES

1 As a consequence of a reform process within the International Telecommunication Union (ITU), the CCITT ceased to exist as of 28 February 1993. In its place, the ITU Telecommunication Standardization Sector (ITU-T) was created as of 1 March 1993. Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.

In order not to delay publication of this Recommendation, no change has been made in the text to references containing the acronyms "CCITT, CCIR or IFRB" or their associated entities such as Plenary Assembly, Secretariat, etc. Future editions of this Recommendation will contain the proper terminology related to the new ITU structure.

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

Recommendation F.901

USABILITY EVALUATION OF TELECOMMUNICATION SERVICES

(Helsinki, 1993)

1 Introduction

This Recommendation defines guidelines for measuring the usability of human-to-system interfaces in telecommunications.

Usability is widely recognized as one of the components of the Quality of Service. It plays a pivotal role in determining the success of new telecommunication services.

2 Definitions

The concept of usability and its components has been defined within ISO and ETSI. A specific interpretation of this definition for the human-to-system interface is as follows:

usability: A concept comprising the effectiveness, efficiency and satisfaction with which specified users can achieve specified system tasks in a particular environment;

effectiveness: Measures of the accuracy and completeness of the system tasks achieved;

efficiency: Measures of the accuracy and completeness of system tasks relative to the resources (e.g. time, human effort) used to achieve the specific system tasks;

satisfaction: Measures of the comfort and acceptability of the system to its users and other people affected by its use.

3 Evaluation aspects

The current literature and laboratory reports show that there is a world-wide commonality on the experimental procedures for usability evaluation. Both the techniques of the Protocol Analysis[1)] and controlled experiments have been used to evaluate the human-to-system interface. These techniques include many of the following considerations:

1) decide which usability dimensions are to be measured;
2) explore methods by which these can be assessed within the system-specific framework;
3) set explicit, quantifiable usability criteria (i.e. reference values);
4) design experimental tasks by which to assess the degree to which usability criteria have been met by the system;
5) let a sample of experimental subjects complete the set tasks. The subjects can be tested one at a time or together, dependent on experimental design. Subjects should be given instructional material as demanded by the tasks;
6) analyse the data statistically;
7) modify those areas of the system in which the usability criteria were not met, relying on anecdotal observations of where and when subjects encountered cognitive stumbling blocks;

---

1) Protocol Analysis: Standard evaluation method in cognitive psychology.

8) re-test as under 5) above, using different subjects belonging to the same class but the same tasks as before in the modified system;

9) upon completion of the new experiment, analyse data:

   i) by themselves as before, using the same types of analyses, and

   ii) comparing data from both experiments;

10) iterate through steps 7) to 9) until usability criteria are clearly met by the system.

3.1 Methods and measures

Since usability is a multi-disciplinary concept, no single, universally "correct" usability evaluation method exists. The choice of the most suitable methods for data capture among the many available (such as data log, observation, questionnaires, interviews and so forth) depends on the system under investigation and the usability dimensions deemed important for that system.

In a public telecommunication terminal, for example, effectiveness would seem to be the most important dimension of usability. People should be able to approach the telecommunication terminal, make use of the service and pay for the service without any fuss. In the case of a public telecommunication terminal, learnability should not be an issue, as one may assume that people will not use a public telecommunication terminal which they cannot readily use.

The use of both performance and attitude measures is recommended because they focus on different aspects of usability.

Performance measures, sometimes referred to as objective measures, include performance time, error rate and number of failures to complete the task, whilst attitude (or subjective) measures capture user opinions.

Checks should be made to ensure that there are no human factors deficiencies (such as poor lighting).

3.2 Criteria

The usability criteria are quantified reference values set operationally for each usability component. The data captured is compared against the criteria to assess whether these have been reached.

3.3 Tasks

Experimental tasks are to be identified by task analysis and appropriate scenarios made up. The scenarios should cover the most important system functions and be typical of the tasks that potential users will perform in an up-and-running system.

3.4 Users

Experimental subjects should be representative of the potential users. A suitable number of subjects should participate in the experiment, so as to guarantee the statistical reliability of the results.

3.5 Analysis of data[2)]

Experimental data is processed to check how well the usability criteria have been met. Problems are identified and solutions proposed.

4 Example

An application of the general procedure to a specific instance is provided in the Annex A.

---

2) In certain situations a statistically significant sample is not necessary to bring to light a design problem.

Annex A

(to Recommendation F.901)

Usability evaluation of ISDN videotelephone terminals

(This annex forms are integral part of this Recommendation)

The following is an example of an approach to evaluating the usability of the ISDN videophone service access procedures according to this Recommendation.

This example contains specific numeric values and specific procedures for illustrative purposes. In an actual evaluation, these values and procedures and other particular details of the test would of course have to be adjusted to suit the situation being tested.

A.1 The user interface

The ISDN videophones are here referred to as terminals consisting of a basic ISDN telephone set and a colour video display (typically 9″ to 12″).

The telephone set consists of a numeric keypad, a number of additional keys and a display which provides the user with information (status of the line, dialled number, charging, etc.) and prompts.

Dialling may be performed according to the "overlap" and "en-bloc" procedure. The "overlap" procedure encompasses the following steps: going off-hook, waiting for dial tone, dialling, changing mode from audio to audiovisual communication. The "en-bloc" procedure consists of dialling, going off-hook and changing mode from audio to audiovisual communication.

The following functions are provided in addition to dialling: terminal set-up, re-dial, supplementary services, suspend/resume, dialled number editing, mute, hands-free, video pause, self-view, mode change control. They are accessed by means of dedicated keys located near the numeric keypad.

The self-view facility may be provided through a separate monitor.

Instructions of use may be provided in written form or verbally by the personnel who installs the equipment.

A.2 Users

The videophones have been designed for the general public.

A.3 Environment

The videophones for home or office applications are designed for an environment with the following viewing conditions:

- environmental lighting: $\geq$ 400 lux;
- viewing distance: 60 to 120 cm.

A.4 Tasks

When using the videophone the user can set up videophone calls with either the "en-bloc" or "overlap" procedure, edit dialled numbers in the "en-bloc" procedure, change mode (from audio to audiovisual and vice versa), display his/her own picture (unless continuously provided on a separate screen), inhibit the transmission of the outgoing picture, release videophone calls.

A.5 Usability measures

To evaluate the videophone usability both user performance and user attitudes are to be measured. The following data are captured: number of failures to complete a task, error rate, task completion time and satisfaction of user procedures.

A.6 Test design

To check whether the usability criteria are met by the videophone user interface, laboratory tests are to be performed.

In the following an example is provided of how to plan such an experiment for testing the most important aspects of videotelephony.

A.6.1 The subjects

At least thirty representative users drawn as a random sample (the higher the number, the more statistically reliable the results) take part in the experiment.

A.6.2 The environment

The videophone terminal is located in premises with 400 lux lighting and is used at 60 cm viewing distance.

A.6.3 The tasks

In each experimental session a given subject performs the following experimental tasks:

- Task 1 consists of setting up a videophone call with the "overlap" procedure. The videophone number to be dialled is a 10-digit one. This task has to be repeated five times.
- Task 2 consists of setting up a videophone call with the "en-bloc" procedure. The videophone number to be dialled is a 10-digit one. This task has to be repeated seven times.
- Task 3 consists of editing one digit in the middle of a 10-digit telephone number dialled with the "en-bloc" procedure. This task has to be repeated five times.

To assess the effect of learnability a subject will typically return for a number of sessions spread over a period of time.

The experimental conditions are administered in a different random order to each subject.

A.6.4 The usability criteria

They are as follows:

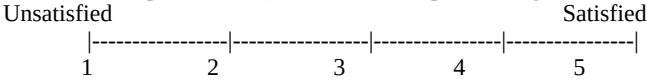
A.6.4.1 Task 1

1) 90% of the subjects should be able to set up the five videophone calls within a session using the "overlap" procedure without committing any errors[3] or failing to complete a task in the last three attempts;
2) the average performance time of the last (correct) attempt should not exceed 60 seconds;
3) the MOS[4] of satisfaction should be higher than 3 and the distribution should not be significantly bimodal.

---

3) Errors are deviations from the correct procedure (excluded misdialling).

4) MOS (Mean Opinion Score) measured on 5-point rating scales.



A.6.4.2 Task 2

1) 90% of the subjects should be able to set up the seven videophone calls within a session using the “en-bloc” procedure without committing any errors or failing to complete a task in the last three attempts;

2) the average performance time of the last (correct) attempt should not exceed 60 seconds;

3) the MOS of satisfaction should be higher than three and the distribution should not be significantly bimodal.

A.6.4.3 Task 3

1) 90% of the subjects should be able to complete the five wrong digit correction attempts within a session without committing any errors or failing to complete a task in the last three attempts;

2) the average performance time for editing the wrong digit should not exceed 30 seconds;

3) the MOS of satisfaction should be higher than three and the distribution should not be significantly bimodal.

A.6.5 The experimental procedure

All the users receive written instructions on how to use the equipment and a description of the tasks to be performed. Training may be provided if the service provider's introduction foresees so. Then the subjects carry out the assigned tasks the specified number of times.

At each attempt the accuracy (number of errors or failures) and speed (performance time) are recorded. No time-out for task completion is set. In case of failure to complete task, subjects proceed to the next attempt.

Possible questions on the experiment may be answered by the experimenter, but not ones on the tasks and the use of the terminal.

At the end of each task the subjects are asked to express their “satisfaction” assessment on a 5-point rating scale.