ITU INTERNATIONAL TELECOMMUNICATION UNION

ITU-T   F.910

TELECOMMUNICATION   (02/95)
STANDARDIZATION SECTOR
OF ITU

OPERATIONS AND QUALITY OF SERVICE

HUMAN FACTORS

PROCEDURES FOR DESIGNING, EVALUATING AND SELECTING SYMBOLS, PICTOGRAMS AND ICONS

ITU-T Recommendation F.910

(Previously "CCITT Recommendation")

FOREWORD

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

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1 (Helsinki, March 1-12, 1993).

ITU-T Recommendation F.910 was prepared by ITU-T Study Group 1 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 21st February 1995.

_____

NOTE

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

CONTENTS

SUMMARY

Various national, regional and international standards organizations are responsible for the design, evaluation and standardization of symbols, pictograms and icons. In order to coordinate the work of these standards organizations, and in order to optimize the designs, it is desirable to follow a well-defined methodology. This Recommendation describes a framework for a common methodology to be used by the ITU-T. To the extent that other standards organizations can work within this framework, the development of important designs should progress more rapidly. A uniform methodology will minimize duplication of effort and will maximize the relevance of collected data.

PROCEDURES FOR DESIGNING, EVALUATING AND SELECTING SYMBOLS, PICTOGRAMS AND ICONS

(Geneva, 1994)

1   Background

Various national (e.g. ANSI, CIAJ, DIN), international-regional (e.g. ETSI) and international-worldwide (e.g. ITU-T, ISO) standards organizations are responsible for the design, evaluation and standardization of symbols, pictograms and icons. In order to coordinate the work of these standards organizations, and in order to optimize the designs, it is desirable to follow a well-defined methodology. This Recommendation describes a framework for a common methodology to be used by the ITU-T. To the extent that other standards organizations can work within this framework, the development of important designs should progress more rapidly. A uniform methodology will minimize duplication of effort and will maximize the relevance of collected data[1].

There are four high-level steps in the methodology for standardizing designs. First, a need must be determined. Second, designs must be created. Third, designs must be evaluated. Fourth, designs must be selected and approved. The steps of this process are shown in Figure 1. The remainder of this Recommendation describes the methodology in detail.

2   Methodology

2.1   Determine the need for an international symbol, pictogram or icon

Two or more countries should express the need for standardization in order to make the proposed symbol, pictogram or icon of international concern.

2.2   Design one or more alternatives

2.2.1   Actual design process

The ITU-T will not restrict suggested designs. The only requirement for suggested designs is that each must be usable in all appropriate forms (e.g. printed, engraved, embossed, screen-displayed).

2.2.2   Process for soliciting alternative designs from ITU-T Members

Suggested designs will be published in ITU-T meeting reports and/or Special Rapporteurs' reports which will include a request for alternative designs. Liaison statements will be sent to other relevant standards organizations. A reasonable time interval will be allowed for submissions before the evaluation procedures begin. The suggested reasonable time is a minimum of two months and a maximum of the interval from one Working Party meeting to the next.

2.3   Evaluation procedures

The various stages in the evaluation procedure are intended to measure initial (*a priori*) meaning and appropriateness, ease of learning, ease of remembering, and the probability of confusion with other designs. The order of procedures should be:

     a)   explain, define and demonstrate functions for each design;

     b)   evaluate contextually meaningful associations by using the matching task shown in Figure 4 and described below in 2.3.3 b);

----

[1] This Recommendation was developed with the help of ITU-T contributors who are also active in other standards organizations. Some procedures used by CIAJ and some used by ETSI (in its multiple index approach) have been adapted for this Recommendation.

c) collect subjective data;

d) if indicated by the subjective data, eliminate bad designs and create new candidate designs and then re-start the experiment;

e) evaluate memory for meaning by a recall test;

f) evaluate confusibility by a recognition-matching test.

The evaluation should be a within-subjects design such that each person provides data in stages b), c), e) and f).

The recommended procedures can be performed in one session, or can be accomplished in a series of sessions with the participants. If a small number of designs is under consideration, immediate memory for all of the designs may be perfect or nearly perfect, whereas delayed memory may show important differences in recall and/or recognition. It is advisable, therefore, for the evaluation coordinator to do a pilot experiment to determine a delay that will reveal differences in memory for the various designs. The evaluation coordinator should specify the procedural details before data are collected by the experimenters in various countries.

2.3.1 Explain, define and demonstrate the function of each design

a) *Evaluate pre-existing associations* – The matching task that is illustrated in Figure 4 and described in 2.3.3 b) is administered to determine the strength of pre-existing associations between a function's verbal description and a design. These data can be used to eliminate the worst designs and to identify the most confusing designs.

b) When possible, construct a prototype device with the test designs on it. Use this prototype to familiarize the subjects with the functions of each symbol, pictogram or icon.

c) When construction of a prototype is not practical, the functions of the symbols that will occur together should be explained in a scenario that corresponds to a scenario that would be typical when using a device.

2.3.2 Collect subjective data

– After all the to-be-rated symbols have been described (explained), present the symbols one at a time to each subject to collect the subjective opinion and rating data.

Figure 2 shows an example page for a questionnaire that can be used to collect these subjective data.

In Figure 2, questions i) and ii) are to help identify those proposed designs that exist or are perceived to be too similar to designs that exist, whereas, questions iii) through vii) ask people to judge the designs *a priori* meaning, future memorability, and graphic qualities.

If these data indicate high levels of confusion among designs or with existing designs, it may be necessary to interrupt the evaluation to generate new designs as indicated in 2.3 d). If this situation occurs, the evaluation procedure should be re-started with a new set of participants.

2.3.3 Objective data

a) *Recall test*

The data from this test will measure the ability of subjects to recall the meaning of individually seen designs. These data are of particular relevance to situations where designs are experienced individually (e.g. on signs, sometimes on screens, etc.).

The general procedure is to present a design (on-screen, projected, or on paper), then to allow time for the subject to write the meaning on an answer sheet (or, alternatively, to type the meaning if the designs are computer-presented).

Whereas, it is desirable to promote the use of consistent verbal descriptions worldwide, it is recognized that this is difficult. Therefore, care must be taken to count as correct the various verbal expressions of meaning that are essentially the same.
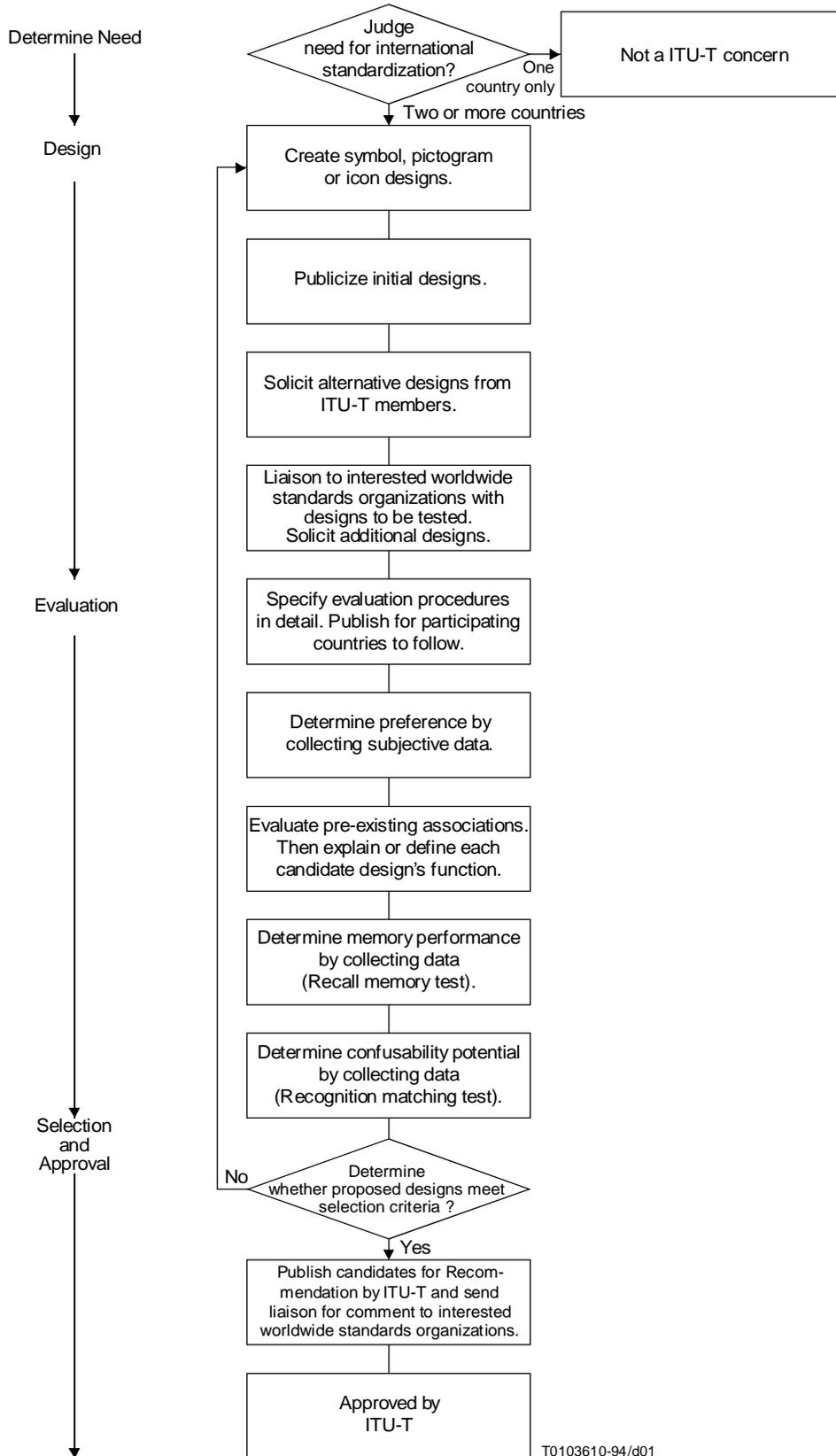
FIGURE 1/F.910

Procedure for standardizing symbol, pictograms and icon designs

Function	Description	Design
Hold	Pressing the button holds a line to interrupt communication	

I. Have you ever seen this exact symbol (or pictogram or icon) before today?
   Yes.   No.
   If yes, where? _____

II. Have you ever seen a symbol that is very similar to this one?
   Yes.   No.
   If yes, where? _____

Please answer each of the next five questions by selecting the value on the five-point rating scale. You may also explain your rating by writing a comment in the space provided.

III. Given the context in which it will be used, is the symbol's meaning easy to understand?

　　Very　　　Difficult　　　　　Easy　　Very
　　difficult　　　　　　　　　　　　　easy

IV. Do you thing that the appearance (from) of this symbol will be easy to remember?

　　Very　　　Difficult　　　　　Easy　　Very
　　difficult　　　　　　　　　　　　　easy

V. If you see this symbol in the future, do you think you will remember its meaning?

　　Very　　　Unlikely　　　　　Likely　　Very
　　unlikely　　　　　　　　　　　　　likely

VI. Are you likely to confuse this symbol with others that you have seen during this session?

　　Very　　　Likely　　　　　Unlikely　　Very
　　likely　　　　　　　　　　　　　unlikely

VII. In appearance, how attractive is this design?

　　Very　　Unattractive　　　Attractive　Very
　　unattractive　　　　　　　　　　　　attractive

Thank you very much for you cooperation.

T0103620-94/d02

FIGURE 2/F.910

Subjective data collection format

Figure 3 shows an example format for the recall test.

b) *Recognition test*

The final stage in evaluation is a recognition-matching test that will provide data to identify confusions among the designs. The general arrangement for the test items is shown by the example in Figure 4.

i) The first column gives a short label or technical term that describes a possible design.

ii) The second column provides a sentence or phrase to more fully explain the verbal label.

iii) The third column consists of an empty box wherein the subject may write an answer.

iv) The fourth column shows the whole set of to-be-matched designs.

v) Standard procedures for counteracting order effects should be used. That is, for a given respondent, the order of designs should vary across trials. The order of the functional descriptions should vary across respondents. The coordinator of an experiment should advise the experimenters on the exact details of such counterbalancing or randomizing procedures.

NOTE – The most realistic test would require subjects to interact with a prototype device that has the new symbols, pictograms or icons displayed. This type of test evaluates two things: the subject's memory for each design's meaning; the subject's comprehension of the procedures required to use the device. While it may be too expensive to perform this optional final test, it may be important in some instances.

Designs	Function

Thank you for your cooperation.   T0103630-94/d03

FIGURE 3/F.910

Recall memory test format

2.4   Selection criteria

a) *Two or more design candidates per function*

When two or more designs are evaluated for a function, selection of the best design should follow these guidelines:

– select the most preferred design (subjective rating data);

– select the design for which recall of its function was best (recall data);

– select the least confusable design (recognition data).

If a particular design does not meet all three criteria, it may be necessary to create new design candidates. Alternatively, the design that best approximates these three criteria may be selected for standardization. The design that is determined to be the best of two or more design candidates should also meet the minimum criteria for acceptance listed in 2.4 b).

b)   *One design candidate per function*

When only one design is evaluated for a function, selection criteria are more difficult. In such situations we are not asking "which design candidate is best?", but rather "is this particular design good enough?". The minimum criteria for acceptance of a design are suggested to be as follows:

–   subjective ratings should be at least neutral (the mid-point on the five-point scale), and preferably should be positive;

–   recall of the function (meaning) for a design should be at least 50 per cent correct when averaged across subjects;

–   recognition matching should at least exceed chance.

3   General parameters of experiments

a)   There are two types of evaluation studies:

In one type, only one design is evaluated for each function. Acceptance of a design under these conditions is based on comparison to arbitrary minimum preference and performance values. Suggested minimum criteria are given in 2.4 b).
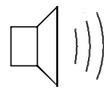
The second type of evaluation study includes two or more design candidates per function. In general, this is the preferred type of study. Winning designs should still meet the minimum criteria given in 2.4 b).

b)   It may be desirable to test a whole set of designs that will be used together against an alternative whole set of designs.

c)   The minimum number of countries to provide data should be three, and these should represent as widely different cultures and/or markets as possible. A greater number of countries should participate whenever possible.

d)   Sampling characteristics should be balanced so that the samples fully represent the user populations of interest. For example, it will be generally desirable to test equal numbers of males and females, equal numbers of users from each decade of life (up to 19, 20-29, 30-39, etc.), and it may be important to sample evenly from various educational or work experience categories (e.g. clerical, technical, labourer, homemaker, etc.). For some designs it may be important to include young children, elderly adults, or the sensory impaired.

e)   The presentation techniques for explaining, learning, rating, and memory testing can vary from one experiment to the next, but care must be taken to use techniques that control extraneous variables. Acceptable techniques include paper-and-pencil, video presentation, and computer-based presentation.

f)   The number of people to be tested in each country is to be determined by the experimenters, but it is suggested that the minimum number be 20 per category.

g)   The statistical procedures and summaries are to be determined by the experimenters.

4   Other considerations

a)   If a design is generally known and used in non-telecommunication contexts, some evaluation steps may be unnecessary to adopt the design for use in telecommunications.

b)   If new designs are to be used in context with existing symbols, it is advisable to include the existing symbols in the recognition test to measure confusability.

Technical term	Description	No.	Design
On-hook dialling	Can dial without picking up the handset		1
			2
			3
			4
			5
			6
			7
			8
			9
			10

Please write down any other observations that you have about these symbols:

Thank you for your cooperation.

T0103640-94/d04

FIGURE 4/F.910

Recognition-matching test format