ETSI TECHNICAL REPORT

ETR 070

June 1993

Source: ETSI TC-HF

Reference: DTR/HF-1010B

ICS: 33.020, 33.040.40

Key words: Human factors, Multiple Index Approach

Human Factors (HF); The Multiple Index Approach (MIA) for the evaluation of pictograms

ETSI

European Telecommunications Standards Institute

ETSI Secretariat

Postal address: F-06921 Sophia Antipolis CEDEX - FRANCE
Office address: 650 Route des Lucioles - Sophia Antipolis - Valbonne - FRANCE
X.400: c=fr, a=atlas, p=etsi, s=secretariat - **Internet:** secretariat@etsi.fr

Tel.: +33 92 94 42 00 - Fax: +33 93 65 47 16

Contents

Blank page

Foreword

This ETSI Technical Report (ETR) has been produced by the Human Factors (HF) Technical Committee of the European Telecommunications Standards Institute (ETSI).

ETRs are informative documents resulting from ETSI studies which are not appropriate for European Telecommunication Standard (ETS) or Interim European Telecommunication Standard (I-ETS) status. An ETR may be used to publish material which is either of an informative nature, relating to the use or application of ETSs or I-ETSs, or which is immature and not yet suitable for formal adoption as an ETS or I-ETS.

Blank page

1 Scope and aim of the document

This ETR describes a method for the evaluation of pictograms, the Multiple Index Approach (MIA). This method has been developed, tested and employed in the context of an ETSI (TC Human Factors) study on pictograms for basic videophone functions and the examples given in this ETR are taken from this study. This method has been found to be suitable as a general testing method for pictograms from all areas. As presented in this ETR, the method takes the form of a questionnaire test but it can be administered by other means as well (e.g. on a personal computer).

The main purpose of a pictogram evaluation study using the Multiple Index Approach is to collect data with the help of which the best suited pictograms of a number of pictogram proposals (i.e. alternative pictogram sets) can be selected for use on products, or for standardisation. Depending on the evaluator's aims, aesthetic criteria may or may not play a role. The Multiple Index Approach provides seven indices or parameters that support the evaluator in making a selection. It does not, however, provide a formula that computes one best solution - the task of weighing the importance of the various indices is left to the evaluator. In most cases, performance data (Hit rate, False alarm rate and Missing values) will be the prime criteria.

2 Introduction

2.1 Definition

In this ETR, the term "pictogram" is used for the graphical representation of a function or element of a user interface and includes both "icons" (concrete representations) and "symbols" (abstract representations).

2.2 Pictograms in user controls

Pictograms and icons are used more and more frequently in the context of the controls and indications of a large variety of devices. Recently, the trend towards pictograms has experienced an additional boost through the advent of a new generation of graphical interfaces on personal computers and workstations. In other areas, such as in telecommunications and transport, pictograms have been used for a long time, sometimes officially standardised, sometimes following quasi-standards.

Pictograms and icons have the potential of easing the use of telecommunications devices. Well designed pictograms allow the user to intuitively understand which function of a device is supposed to be represented. In many cases, pictograms require less learning time and effort than text based alternatives. They are "international" in the sense that they are not bound to a particular language and no level of literacy is required. Standardised pictograms for the functions of widely used devices allow the user to recognise the basic functions of any such device without the need for extra instruction.

Unlike in areas with well established design guidelines, e.g. menu structures in computer software, design recommendations for pictograms are somewhat vague leaving the designer a great deal of artistic freedom. Some of the resulting pictogram designs are highly ambiguous and lessen the usability of the device. This state of affairs makes it imperative that pictograms be empirically tested in order to establish whether the user does indeed associate the function, location, etc. to be represented (from this point known "referent"). If the pictogram is intended to be used on interfaces of devices for the international market, or if it is a candidate for international standardisation, the empirical testing has to be conducted in several languages in order to ascertain that the pictogram does not draw on language mediated associations that work in some languages only, thus losing the pictogram's potential benefit of being free from language and culture biases.

2.3 Rationale of an empirical testing method of pictograms

Before comparing the different methodological options available for testing pictograms, the criteria by which the pictograms are to be judged have to be made explicit, or in other words, the question of what establishes a good set of pictograms needs to be addressed (in the following, it is assumed that a set of pictograms is to be tested which represents a number of functions of a device).

A set of pictograms will optimise learning and user performance if:

- each of its elements is associated with the corresponding referent;
- none of its elements is associated with any referent other than the corresponding one;
- users feel subjectively certain in their choice of pictogram;
- users like their aesthetic appeal.

Therefore, the testing method should focus on both correct associations and on errors and it should take into account the respondent's subjective certainty. In addition, it is desirable that the method provides data on the pictograms' aesthetic appeal, a criterion which, inter alia, may be important in product design.

The most realistic evaluation approach is one that tries to represent an actual usage situation, i.e. a recognition situation in which a user with a certain intention is confronted with the controls of a device and has to make a choice as to which control will bring about the desired effect.

There are basically four ways of assessing the associativeness of pictograms.

	Display	**Task**
1	One pictogram at a time.	Name referent.
2	Set of pictograms and one referent.	Pick pictogram that represents the referent.
3	One pictogram and the list of referents.	Pick the referent that is represented by the pictogram.
4	Set of pictograms and list of referents.	To map the elements of each list.

The four options meet the criteria specified above to different degrees. The first test is one of recall rather than recognition processes. The third one, in which all referents and only one pictogram are presented at a time, is equally unsuited for the present aim since in a real-life situation, the user of a videophone will have all pictograms visually present but he/she will not necessarily have a complete cognitive representation of all functions of the terminal as defined by the referents. The same applies to Option 4 which has the additional disadvantage that it is a one-to-one mapping and that certain errors (like one pictogram being associated with two referents) do not occur.

Option 2, i.e. the test in which the complete set of pictograms is presented to the subject (as would be in the case of a real videophone call situation in which the pictograms are placed on the terminal) and only one referent is presented at a time, is the testing method on which the Multiple Index Approach is based. In addition to its greater validity, it has the advantage of allowing for all four kinds of outcomes of a signal detection situation (Hit, Miss, False Alarm and Correct Rejection) to occur thus making possible a detailed analysis of the respondents' selections.

3 Evaluating pictograms by multiple indices

3.1 The seven Multiple Index Approach indices

The Multiple Index Approach was developed on the basis of the above considerations. It enables the evaluator to collect data on seven indices on which his final selection of pictograms can be based and which are presented as follows.

1 The Hit rate
This index is the main parameter of performance and it is equivalent to the score of correct associations between the referent and pictogram.

2 The False alarm rate
The False alarm rate tells the evaluator in how many cases a pictogram has been associated with the wrong referent. Depending on the pictogram context, False alarm errors (sometime referred to as Type II or μ-errors) can be more hazardous than a Miss (incorrectly rejecting an association, Type I or α-error).

3 Missing values

The percentage of Missing values tells us in how many instances a respondent did not answer a question presumably because he/she did not know the answer. Missing values represent usage situations in which the user does not know which control to use to bring about a certain effect.

4 Subjective certainty

The Subjective certainty index indicates how certain the respondents feel in their association between a pictogram and referent. If the users of a device are extremely uncertain about the effects of the controls of a device, they may decide not to use it at all, which in turn may seriously hamper the uptake of the device.

5 Subjective suitability

In addition to making the association between pictogram and referent and to indicating how certain they are in this association, the respondents can tell us their subjective impression as to how well a pictogram represents its referent.

6 Pictogram Preference

The respondents indicate, which of the candidate pictograms for one referent represents best the referent in question. In this, we do not know which criteria (aesthetic or functional) the respondents apply.

7 Pictogram Set Preference

This index is an indicator for which pictogram set is preferred in toto mainly on aesthetic grounds.

The seven indices are collected with a questionnaire that is organised in three tests (it is, of course, possible to implement the test on computers with sufficiently high resolution screens):

- test of pictogram associativeness (Hit rate, False Alarm Rate, Missing values, Subjective certainty, and Subjective suitability). In this part of the questionnaire, one referent (name and description of a function) is presented at a time with all pictograms of one set. The respondent's task is to choose the appropriate pictogram for the function in question. In addition, Subjective certainty and suitability ratings are required for each rating;

- test of pictogram preference. Here, the respondent is asked to give preference ratings on the level of function, i.e. all candidates for one function are shown and the most suitable one is to be indicated;

- test of pictogram set preference. Preference ratings are requested on the level of sets, i.e. all pictogram sets are displayed and the preferred one is to be indicated.

The results of the test of pictogram associativeness are the main indicator for the usability of the sets to be tested. The tests of pictogram preference and of pictogram set preference are to be used mainly to verify that a pictogram set fulfils not only the associativeness criterion, but also aesthetic criteria. Furthermore, these indices can be used in cases in which there are competing sets with similar results for associativeness.

Finally, order and learning effects should be controlled by employing versions of the questionnaire with a different presentation order of the pictograms.

4 Structure of the MIA questionnaire

4.1 The introduction section

In the introduction section of the questionnaire, the purpose of the study should be made explicit and the referents should be described in detail. For the evaluation study to yield meaningful results, it is important that the referents be understood by the respondents. Ideally, the respondents are shown a model or mock-up of the device on which the pictograms will be used. Different models, each with one of the pictogram sets to be tested can then be used for the MIA tests. However, this procedure will rarely be feasible. An alternative way of familiarising the respondents with the referents is to give them a videotape presentation. In many cases, graphical representations can be used to support the description. In addition, it may be useful to present the referents with the help of a usage scenario - this is recommended particularly in the case of relatively novel devices.

An example of the introduction section for an evaluation questionnaire (in this case of pictograms for basic videophone functions) can be found in Annex A.

4.2 The test of pictogram associativeness

This test collects data on the individual pictogram's associativeness with regards to its corresponding referent and in the context of the pictogram set it belongs to. This test provides data on the Hit rate, False alarm rate and Missing Values indices. Secondly, Subjective certainty and Subjective suitability ratings are given in this subclause.

Each pictogram is tested on a separate page (i.e. if there are *r* referents and *s* pictogram sets each containing *r* candidate pictograms, there are *r* * *s* pages in this section of the questionnaire. One pictogram set is being dealt with at a time so that the resulting order of pages is

SET1 REFERENT1, SET1 REFERENT2 ... SET1 REFERENT*R*, SET2 REFERENT1 ... SET*S* REFERENT*R*.

Hit rate, False alarms, and Missing values: at the top of each page, a short description of the referent is given, followed by a representation of all pictograms of the set to be tested. The size of the pictogram representations may depend on the context in which they are intended to be employed (e.g. on keys, screens or signposts). The respondent's task is to mark the pictogram that represents the referent in question.

Subjective certainty is measured using a 5-point scale ranging from very uncertain to very certain along with the question:

How certain are you of your choice?

very uncertain [ ] [ ] [ ] [ ] [ ] very certain

Subjective suitability ratings are made using a similar 5-point scale ranging from "very badly" to "very well" along with the question "How well do you personally think does the pictogram represent Referent?".

An example of a test of pictogram associativeness of pictograms can be found in Annex A.

4.3 The Test of Pictogram Preference

In this test, the respondent is asked to give preference ratings on the level of function, i.e. there are as many pages in this section as there are referents. At the top of each page, a short description of the referent is given followed by a representation of all candidates for one function. The respondents task is to indicate the most suitable candidate answering the question: "Which pictogram do you think represents best Referent? I prefer pictogram number _____." An example of a test of pictogram preference for basic videophone functions can be found in Annex A.

4.4 The test of pictogram set preference

In the test of pictogram set preference, all pictogram sets are presented in grouped form and the respondents are asked to select the one pictogram set that they prefer. The question to be answered is "Which set of pictograms do you prefer? I prefer pictogram set number _____." An example of a test of pictogram set preference can be found in Annex A.

4.5 Control variables

It is advisable to collect some descriptive data from the respondents in order to be able to show that the sample was sufficiently representative for the expected user population in terms variables such as age, gender, and professional background. It may also be useful to collect data on technical experience and attitude towards technology. An example of the control variable section of a pictogram evaluation study can be found in Annex A.

5 Analysis

The main purpose of a pictogram evaluation study using the Multiple Index Approach is to collect data with the help of which the best suited pictograms of a number of pictogram proposals (i.e. alternative pictogram sets) can be selected for use on products, or for standardisation. Depending on the evaluator's aims, aesthetic criteria may or may not play a role. The Multiple Index Approach provides seven indices or parameters that support the evaluator in making a selection. It does not, however, provide a formula that computes one best solution - the task of weighing the importance of the various indices is left to the evaluator. In most cases, performance data (Hit rate, False alarm rate and Missing values) will be the prime criteria.

5.1 Test of order and learning effects

Primacy and recency effects may have a biasing effect on the respondents' performance and preference. The questionnaire should, therefore, be administered in at least two versions with a different pictogram order. Every respondent fills in only one version of the questionnaire. If learning and order effects are found, it may be worthwhile to re-test the pictogram material using additional item orders in order to guarantee that every pictogram set was tested in both favourable and unfavourable positions in the questionnaire.

Tests for statistically significant differences should be performed on summated, or average performance and/or preference scores to identify any order or learning effects. This is the only inferential statistical test required. For the following analyses, descriptive statistics (tables of means and tables or rank order) are sufficient.

5.2 Analysis of the test of associativeness data

5.2.1 Pictogram selection

In the test of associativeness, one referent is presented together with all the pictograms of the set to be tested in random order. The task is to mark (circle, cross out) the pictogram that represents the referent. Mean values of the three indices Hit rate, False alarm rate and Missing value, across subjects and across pictogram sets are sufficient to identify the more successful candidates - no further statistics are usually required.

The average Hit rate (i.e. percentage or absolute number of correct selections) is a prime index of pictogram associativeness. The results can be compared across pictogram sets indicating how well the different candidates did for any particular referent. Comparisons across referents point to weaknesses in the design throughout the sets or in the referent descriptions.

Tables with percentages of False alarms (instances of the selection of a pictogram in the context of a different referent) indicate how precisely the pictograms of a set differentiate between referents. It is not sufficient that all respondents identify a pictogram in the context of its appropriate referent: if the same pictogram is frequently identified wrongly in the context of another pictogram, this high False alarm rate decreases the pictogram's usability.

Missing values are a third important index for the quality of a pictogram set. No, or ambiguous, responses indicate that the respondent simply did not know the answer (or would not have known which button to push).

5.2.2 Subjective certainty

In addition to marking the pictogram that corresponds to the referent in question, the respondents are asked how certain they are in their choice ("How certain are you of your choice?" - Subjective certainty, rating scales ranging from very uncertain (1) to very certain (5)).

Average subjective certainty data indicates how the candidate pictograms compare in terms of how certain the respondents felt in making their selection.

5.2.3 Subjective suitability

The respondents are asked for a suitability rating ("How well do you personally think does the pictogram represent referent?" - Subjective suitability, rating scales ranging from very badly (1) to very well (5)). Mean Subjective suitability ratings can serve as an additional indicator of a pictogram's suitability as judged by the respondents.

5.3 Analysis of the test of pictogram preference and test of pictogram set preference data

Frequency counts or means of pictogram preference indicate the number of respondents that felt that a pictogram is better suited than any of its competitors from the other pictogram sets. It shall be kept in mind that it is not necessarily clear which criteria (aesthetic or functional) the respondents apply in the Test of pictogram preference.

This is different in the case of the test of pictogram set preference where the wording of the test question points to aesthetic criteria. Again, no statistics beyond means or frequency counts are required.

6 Making a decision based on the results

As previously pointed out, a decision about the relative importance of the collected indices has to be made by the evaluator before drawing conclusions from the data. Performance data (Hit rate, False alarm rate and Missing Values) are prime selection criteria. Of two pictograms with equal hit rates, the one with the lower False alarm and Missing values rates is the better pictogram. Of two pictograms with equal performance rates, the one with favourable results regarding the subjective parameters (Subjective certainty, Subjective suitability and preferences) is to be preferred.

The results may not always be clear-cut. It is possible that a pictogram set obtains good results for all but one or two of its members. In this case it is possible to re-design the pictograms that obtained poor results in the style that was used for the whole set. Another possibility is to borrow more successful candidates from other pictogram sets. For both solutions it is necessary to re-test the new set and compare the results obtained with the ones from the original study. More than one design and test cycle may be required until significantly better results are achieved (for a recommendation on how to test one set only see Clause 7).

7 The special case of testing one pictogram set only

If just one pictogram set is to be tested (maybe because only one set has been designed or because a previously tested and rearranged set is to be re-tested), only the test of associativeness is possible (i.e. the collection of Hit rate, False Alarm rate, Missing Values, Subjective certainty, and Subjective suitability data) because the tests of pictogram preference and pictogram set preference require selections among pictogram sets. This may not be a problem if performance data are the prime selection criterion. If preference data are required for the evaluation of the set, it may be possible to include other, possibly previously rejected sets as distracters.

8 Conclusion

The design of pictograms is a creative process. We currently do not have available pictogram design guidelines that guarantee usable pictograms. For this reason, it is important that pictograms be evaluated prior to use or standardisation.

The MIA is a tool for the empirical evaluation of alternative pictogram designs. The seven indices collected through the MIA enable the evaluator to analyse the results in terms of the parameters that are important for the application at hand, be the emphasis on associativeness or on subjective and aesthetic aspects.

Annex A (informative): Examples for the different sections of a MIA-questionnaire

The following pages show examples for:

- the introduction section of a MIA-questionnaire (pages 15-17);
- the test of associativeness section of a MIA-questionnaire (page 18);
- the test of pictogram preference section of a MIA-questionnaire (page 19);
- the test of pictogram set preference section of a MIA-questionnaire (pages 20-21); and
- the control variables section of a MIA-questionnaire (page 22).

in co-operation with
- **your institution**
- name/address
- telephone

Dear participant,

The European Telecommunications Standards Institute (ETSI), works on developing standards for telecommunications services and products in order to make them easy to use and enable the subscriber to use them abroad just as easily as at home.

This questionnaire is about a new type of telephone, namely the videophone. You do not need to know anything about videophones because what a videophone is and what it is used for will be explained on the following pages.

There are no right or wrong answers. We would therefore ask you to fill in the questionnaire page by page without browsing ahead or back.

Your answers will be treated anonymously.

This questionnaire was administered by your name of your institution on behalf of ETSI.

European Telecommunications Standards Institute
Institut européen de normes de télécommunication - Europäisches Institut für Telekommunikationsnormen
Postal address : ETSI - F - 06921 Sophia Antipolis CEDEX Tél. : + 33 92 94 42 00
Office adress : 650 Route des Lucioles - Sophia Antipolis - France Fax. : + 33 93 65 47 16
X.400: c=FR, a=ATLAS, p=ETSI, s=SECRETARIAT - Internet : secretariat@etsi.fr
Siret No 348 623 562 00017 - Association à but non lucratif enregistrée à la Sous-Préfecture de Grasse (06) No 7803/88 - NAP 742C

What is a videophone?

A videophone is a new type of telephone. Whereas on a regular telephone you can only hear the person you are talking to, on a videophone you can see them and they can see you. For this reason, a videophone is equipped with the following parts: a screen on which you can see the other person, a camera so that the other person can see you too, a handset for hearing and speaking, and a built-in microphone and loudspeaker to enable you to speak without having to hold the handset. Figure A.1 shows what a videophone should look like:

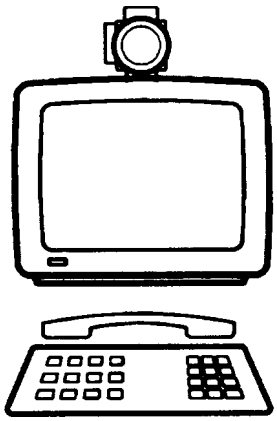


Figure A.1

Because of the additional functions, a videophone is more complex to use than a regular telephone. The following functions are possible on a videophone:

- to call someone or to receive a call;
- to switch between a telephone call (voice only) and a videophone call (voice and picture) (VIDEOPHONE/TELEPHONE Function);
- to turn the camera on and off (CAMERA ON/OFF Function);
- to make a call without using the handset = handsfree using the built-in microphone and loudspeaker (HANDSFREE Function);
- to turn the microphone in the handset or the built-in microphone on and off (MICROPHONE ON/OFF Function);
- to see oneself on the screen (SELF VIEW Function);
- to "freeze" the picture on the screen (STILL PICTURE Function);
- to transmit the picture from a ETR Camera when you wish to show to the other person a document or object (DOCUMENT CAMERA Function).

A number of switches or buttons are needed for making use of these functions of a videophone. The manufacturers of modern equipment sometimes print small pictures next to a button to remind the user of what it does. This kind of picture is called a "Pictogram" (you may know pictograms from places like train stations and airports where they are used to indicate "Information", "Exit", etc.). ETSI is studying pictograms which may in the future be placed next to the buttons of videophones, and we would like you to help us choose the best pictograms.

Pictograms for videophones

Above, we have introduced a number of special videophone functions like CAMERA ON/OFF. We have collected seven proposals for how these videophone functions can be represented as pictograms. Each of these proposals, known as pictogram families, comprises seven pictograms, one for each of the seven videophone functions. Our aim is to find out, which family is best suited for this purpose.

For this reason we would like to give you three tasks, and we have also arranged this questionnaire in three parts.

Part 1

This deals with the pictograms of Family 1. On each of seven consecutive pages, all pictograms of this family are printed. On the top of each page, one videophone function will be briefly described. Your task will be to cross out the one pictogram which you think belongs to the videophone function in question. Following this you will be asked to express, on response scales, how certain you are of your choice and how well you think the pictogram represents the function. The same tasks follow for pictogram families 2 to 7.

Part 2

We ask you to choose among all the candidate pictograms that belong to one function the one pictogram that you think represents best the function in question.

Part 3

Which of the seven pictogram families you prefer.

If you have any questions at this point, please ask the experimenter.

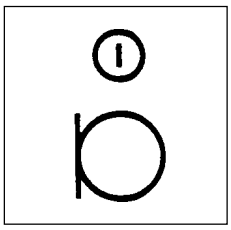
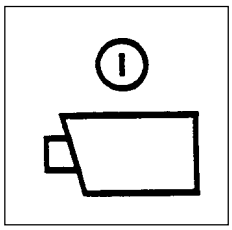
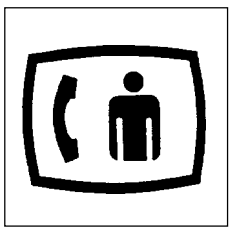
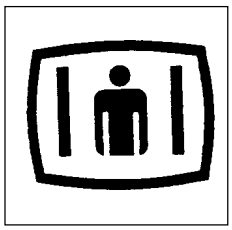
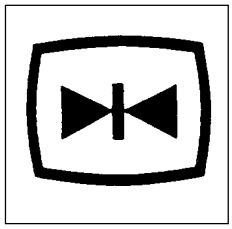
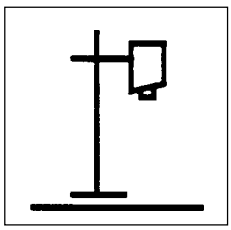
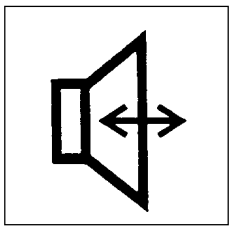
Part 1

Family 1

Function 1: VIDEOPHONE/TELEPHONE

This function allows you to switch between a telephone call (sound only) and a videophone call (picture and sound). Among other reasons, this function is important because a videophone call will be more expensive than a regular telephone call.

Please cross out the pictogram that you think represents VIDEOPHONE/TELEPHONE.

How certain are you of your choice?

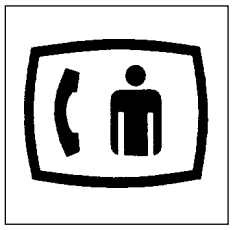
very uncertain [ ] [ ] [ ] [ ] [ ] very certain

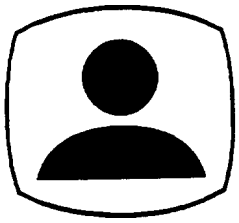
How well does the pictogram you selected represent VIDEOPHONE/TELEPHONE? The pictogram represents VIDEOPHONE/TELEPHONE?

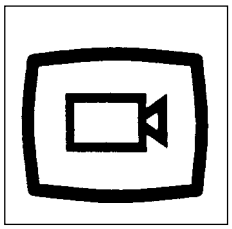
very badly [ ] [ ] [ ] [ ] [ ] very well

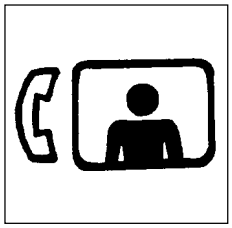
Part 2

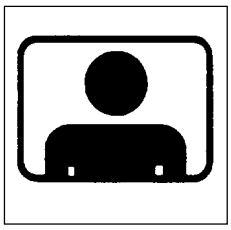
On this page you see seven pictograms that represent the VIDEOPHONE/TELEPHONE function. As you know, this function allows you to switch between a videophone call (seeing and hearing) and a telephone call (hearing only).

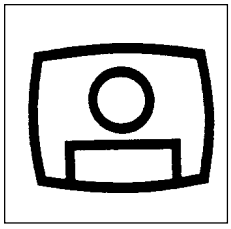
1 

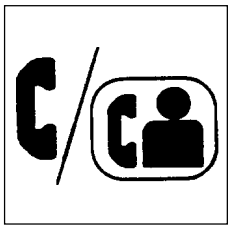
2 

3 

4 

5 

6 

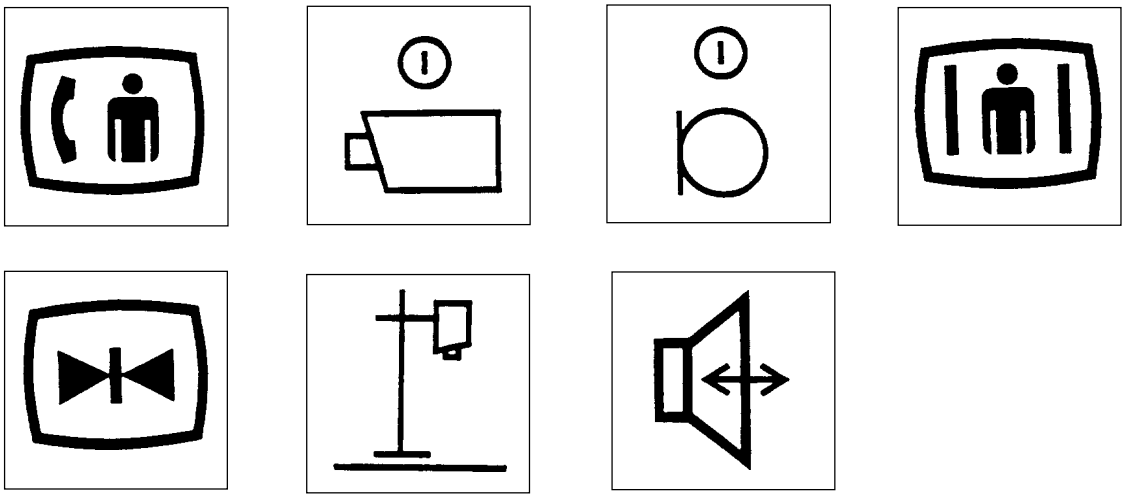
7 

Which pictogram do you think represents best the VIDEOPHONE/TELEPHONE function?

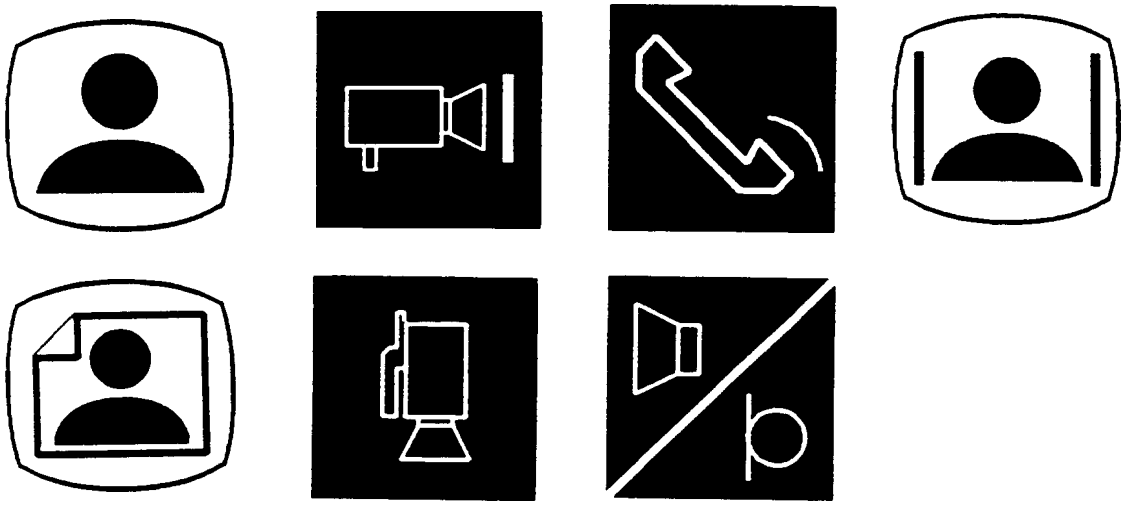
I prefer pictogram number _____

Part 3

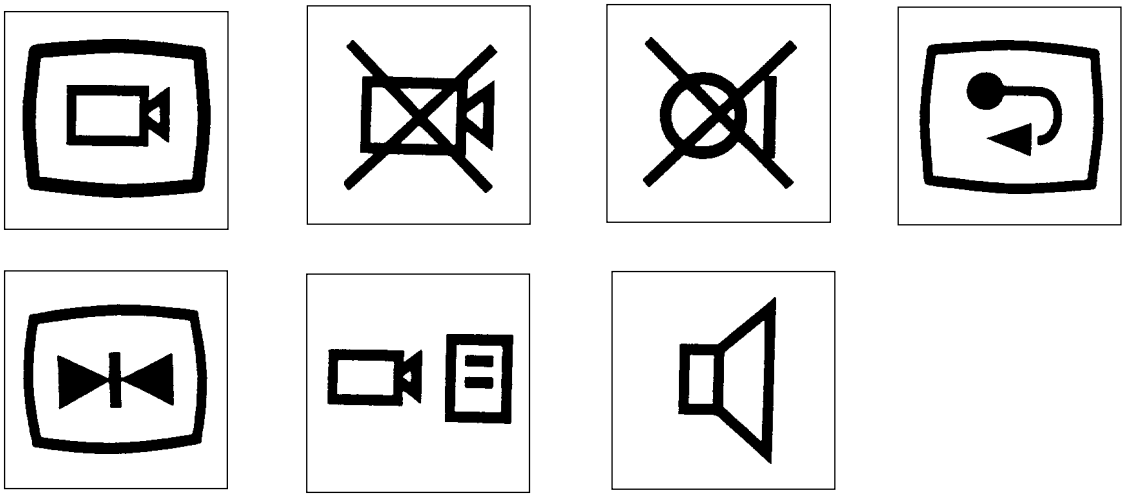
Family 1



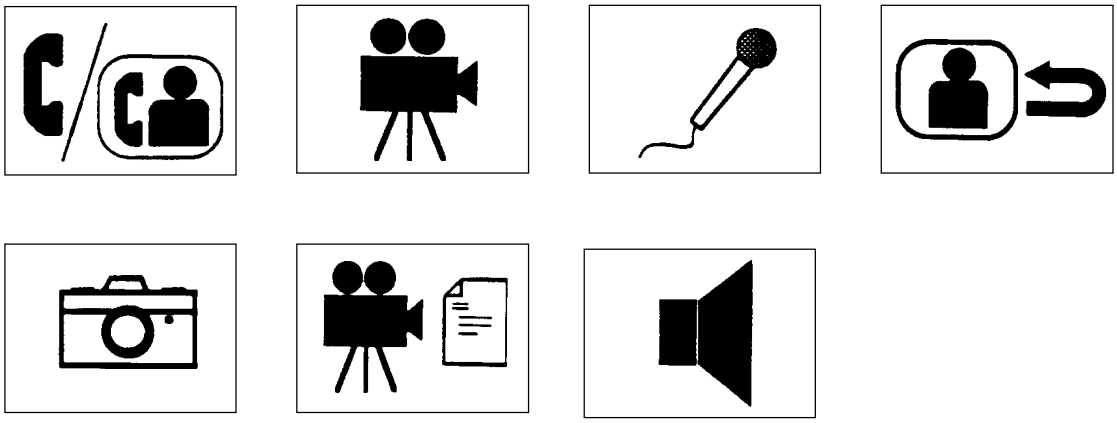
Family 2



Family 3



Family 7



Which pictogram family do you prefer? I prefer pictogram family number _____

On this last page of the questionnaire we would like to ask you some questions about yourself:

Age:

female ❑ male ❑

Profession:

Have you ever used a videophone?

❑ never ❑ once ❑ more than once

The next three questions are about your attitude towards technological progress. Please express your opinion by crossing one of the boxes of the scale.

1. The advantages of technological progress outweigh the disadvantages.

[1]	[2]	[3]	[4]	[5]
do not agree at all				totally agree

2. Progress in technology makes life a lot easier.

[1]	[2]	[3]	[4]	[5]
do not agree at all				totally agree

3. Some aspects of today's technological progress are worrying.

[1]	[2]	[3]	[4]	[5]
do not agree at all				totally agree

History

Document history	
June 1993	First Edition
February 1996	Converted into Adobe Acrobat Portable Document Format (PDF)