ETSI TECHNICAL REPORT

ETR 198

October 1995

Source: ETSI TC-HF

Reference: DTR/HF-01022

ICS: 33.020

Key words: ISDN, user control procedures, videotelephony

Human Factors (HF); User trials of user control procedures for Integrated Services Digital Network (ISDN) videotelephony

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 the application of ETSs or I-ETSs, or which is immature and not yet suitable for formal adoption as an ETS or an I-ETS.

The intended users of this ETR are given in table 1 below:

Table 1: Intended users and potential benefits

	User	ETR used for	Potential benefit
1	Videotelephony service and terminal designers and providers	Development and qualification of videotelephony user control procedures	Increased usability through harmonized and supportive procedures
2	User groups	To identify problems within videotelephony user control procedures	Increased awareness by user groups of user requirements for videotelephony services
3	ETSI Technical Committees	Development of videotelephony standards	Improved usability of services by ensuring consideration of user needs
4	TC-HF	Development of videotelephony user control procedures	Improved usability through consistency and coherence of recommendations

Introduction

This ETSI Technical Report (ETR) contains a discussion and recommendations based on the results of user trials of recommended user control procedures in Integrated Services Digital Network (ISDN) videotelephony. Four separate trials were conducted by:

- British Telecom Research Laboratories, UK (BTRL);
- PTT Research, Netherlands (NL-PTT);
- Philips Kommunikation Industrie, Germany (PKI);
- Centro Studi e Laboratori Telecomunicazioni, Italy (CSELT).

The user procedures themselves are discussed as a background to the final recommendations. A discussion is also included about controls and indications, because they are an essential part of procedures and will be the subject of recommendations which are to be included in "Human Factors in Videotelephony" currently being developed within TC-HF.

One of the tasks was to define a set of generic user control procedures for access to ISDN teleservices and validate them for at least one particular area, e.g. videotelephony, by carrying out usability trials.

The ultimate aim of the work was to ensure the provision of consistent user procedures that can be applied to European-wide ISDN telecommunications services. This ETR supports activities where user procedures are under study for a variety of telecommunications services, including videotelephony, supplementary services, Universal Personal Telecommunications (UPT), and other phone-based interfaces. The aim of defining and validating generic user procedures and methods for their development and description was to avoid the risk of recommending procedures suitable for individual services that may nevertheless be inconsistent with others.

Blank page

1 Scope

This ETR presents and discusses the combined results from four pan-European experiments in videotelephony. These were designed to evaluate a set of user control procedures for videotelephony, as a specific example of generic procedures in ISDN telecommunications as recommended in ETR 170 [1].

2 References

This ETR incorporates by dated and undated reference, provisions from other publications. These references are cited at the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this ETR only when incorporated in it by amendment or revision. For undated references the latest edition of the publication referred to applies.

[1] ETR 170 (1995): "Human Factors (HF): Generic user control procedures for telecommunication terminals and services".

[2] ETR 095 (1993): "Human Factors (HF): Guide for usability evaluations of telecommunications systems and services".

[3] Anderson, D.M.M. et al (1993): "Empirical evaluation of user procedures in videotelephony". 14th International Symposium on Human Factors in Telecommunications, Darmstadt.

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of this ETR, the following definitions apply:

effectiveness: The number and types of errors made in the execution of each task. In some cases errors may be compared against pre-determined criteria, and they may be analysed to determine the cause of error.

efficiency: A measure of human resources taken to complete each task, as indicated by time to completion. Comparison may also be made with any known criteria for existing tasks in telephony, e.g. dialling.

learnability: As assessed by comparing time and error rates for each task across successive trials.

satisfaction: As measured by using questionnaires and interviews, of the subjective acceptability of a given procedure to the user. Data may also be used to identify possible areas for improvement.

usability: The effectiveness, efficiency and satisfaction with which specified users can achieve specified goals (tasks) in a particular environment. In telecommunications usability should also include the concepts of learnability and flexibility; and reference to the interaction of more than one user (the A and B parties) with each other and with the telecommunications system, see ETR 095 [2].

3.2 Symbols

For the purposes of this ETR, the following symbol applies:

✓ Affirmative

3.3 Abbreviations

For the purposes of this ETR, the following abbreviations apply:

A Party	The originating party in a telecommunications call
B Party	The receiving party in a telecommunications call
BTRL	British Telecom Research Laboratories, UK
CSELT	Centro Studi e Laboratori Telecomunicazioni, Italy
CT-V	Convert a Telephone call to a Videotelephone call
CT-Vacc	Accept a Conversion of a Telephone call to a Videotelephone call
CUSI	Computer User Satisfactory Inventory
CV-T	Convert Videotelephone call to Telephone call
DC	Document Camera
DCin	Document Camera in (On)
DCout	Document Camera out (Off)
HF-HS	Handsfree to Handset conversion
HF-LS	Handsfree to Loudspeech conversion
HFend	Handsfree end, terminate a call in Handsfree mode
HS-HF	Handset to Handsfree conversion
HS-LS	Handset to Loudspeech conversion
HSend	Handset end, terminate a call from Handset mode
ISDN	Integrated Service Digital Network
LED	Light Emitting Diode
LSend	Loudspeech end, terminate a call in Loudspeech mode
MHF	Make Handsfree call
MHS	Make Handset call
MT	Make a Telephone call
MTHF	Make a Telephone call in Handsfree mode
MTHS	Make a Telephone call in Handset mode
MV	Make a Videotelephone call
MVHF	Make a Videotelephone call in Handsfree mode
MVHS	Make a Videotelephone call in Handset mode
NL-PTT	Netherlands Post Telegraphy and Telephony Research
PKI	Philips Kommunikations Industrie
Q	Question (as in subjective questionnaires)
RACE	Research and development in Advanced Communications technologies in Europe
RHF	Receive a call in Handsfree mode
RHS	Receive a call in Handset mode
RT	Receive a Telephone call
RTHS	Receive Telephone call in Handset mode
RTVPetsi	Receive a videotelephone call with Video Pause on (i.e. audio only) - ETSI TC-HF suggested procedure
RV	Receive a Videotelephone call
RV-A	Receive a Videotelephone call in Audio only
RV-T	Receive a Videotelephone call in Telephone mode (i.e. audio only)
RVHF	Receive Videotelephone call in Handsfree mode
RVHS	Receive Videotelephone call in Handset mode
SB	Selfview Before a call
SD	Selfview during videotelephone call
SDT	Selfview During Telephone call
Soff	Selfview off
Son	Selfview on
VP	Video Pause
VPoff	Video Pause off
VPon	Video Pause on
VPS	Video Pause in selfview
VPSin	Video Pause in selfview in (i.e. switch into Video Pause whilst in selfview mode)
VPSout	Video Pause in selfview out (switch out of Video Pause whilst in selfview mode)

4 Videotelephone user trials

4.1 Background

At the time when Human Factors (HF) Technical Committee was first set-up in 1988, one of the first tasks was to investigate and recommend user control procedures for videotelephony in response to a request from Network Aspects (NA) Technical Committee development of the human factors recommendations and guidelines included consideration of recommendations from NA and other technical committees in ETSI, CCITT and CEPT documents and the general human factors literature.

The result was a report from HF to NA Technical Committee, which aimed to provide guidelines for videotelephone terminal design, and which included a set of user procedures specific for videotelephony. These aimed to be as usable as possible by being simple, rational and consistent with normal telephony, especially for new, video-related functions. No recommendations were made which were terminal specific, nor any recommendations about controls and displays or messages.

The generic control procedures for videotelephony included the following functions:

a) call set-up functions (on/off hook, dialling);

b) control of service mode (video/audio, one/two B channels);

c) video control functions:

   1) incoming video indication;
   2) video pause;
   3) selfview;

d) switching between audio modes (hands-free and handset).

The first and last are truly telephone generic as they do not only concern videotelephones, but are increasingly important as features proliferate in current telephone sets.

The following subclauses summarise the methodology and results of the four individual user trials.

4.2 User trials

This ETR tests experimentally the usability of the proposed videotelephony procedures. By inference, this also tests the reference models for generic user procedures recommended in ETR 170 [1]. This has been achieved by carrying out experiments in four European countries, using different videotelephone implementations but one basic experimental design. The advantages of taking this approach were:

- by incorporating the user procedures in a number of different videotelephone implementations, procedures can be evaluated to see whether they are generic, or specific enough. At the same time, problem areas can be identified which might exist in the procedures;

- by using one basic experimental design for all experiments, the results from all four experiments can be compared with one another, thus increasing the amount of data available as a basis for drawing conclusions about human factors aspects of user procedures for videotelephony and telecommunications in general.

4.2.1 Videotelephony procedures tested

The user control procedures are listed below. It should be noted that in developing these procedures, it was agreed that there should be a choice of service modes (limited to 5 or 6) and that one service mode should be selectable as the default. The possible service modes include: audio 3,1 kHz, audio 7 kHz, audio-visual 1B (audio 3,1 kHz), audio-visual 2B (audio 3,1 kHz), audio-visual 2B (audio 7 kHz). Also, there are a number of alternative methods for signalling audio-visual 2B, and these may be additional service modes.

To enable specific videotelephone procedures to be developed, it has been assumed that one of the audio-visual service modes would be the default state that the product is a videotelephone, and therefore, users expect that calls made are videotelephone calls, with the possibility of making ordinary telephone calls if required (default video).

However, videotelephony can be required a natural accessory to the telephone and users may expect to make telephone calls as normal and add in the video element when required (default telephone). This conforms with the view that privacy is respected, and that incoming calls, at least, should always be telephony.

These two options were tested within the series of experiments, due to the different defaults chosen by different administrations and manufacturers. The specific default conditions used were:

- BTRL : videotelephony, all calls out and in;
- CSELT : telephony, all calls out and in;
- PKI : telephony, all calls out and in;
- NL-PTT : mixed, videotelephony out, telephony in.

The user control procedures proposed, based on a videotelephone default, are given in table 2.

Table 2: Proposed videotelephone user control procedures

Function	User control procedure
Make video call (default video)	Go off-hook, dial, wait for connection
Make telephone call (default video)	Go off-hook, select telephone mode, dial, wait for connection
Receive video call (default video)	"Ring signal", go off-hook
Receive telephone call (default video)	"Ring signal", go off-hook
Receive video call in audio (default video)	"Ring signal", select telephone mode, go off-hook
Receive video call, do not send picture	"Ring signal", select video pause on, go off-hook
Upgrade from telephone to video	"During telephone call", select video/telephone mode change
Downgrade from video to telephone	"During video call", select video/telephone mode change
Accept upgrade from telephone to video	No provision
Accept downgrade from video to telephone	No provision
Use of selfview (in any state)	Select/deselect selfview
Use of video pause (in any state)	Select/deselect video pause
Handset to handsfree	Select handsfree and replace handset
Handsfree to handset	Select handset (handsfree off, automatic)

The relative effects of these different defaults are discussed in the results presented in subclauses 4.3 and 4.4.

4.2.2 Experimental method

The broad outline for the experimental method was first discussed and the ideas and proposals were then developed into a practical form by BTRL, who were first to run an experiment. The other trials adopted, in principal, the BTRL methods, subject to variations in local implementation.

Some general requirements were laid down for the experiment:

- to use actual videotelephones where possible, or representative simulations of realistic videotelephone interfaces;

- to use every subject as their own experimental control by getting them to complete all procedures;

- to collect objective data, time and errors for each element of a procedure or task and to collect subjective data which relates to the equipment or procedures as a whole;

- to use as many subjects as possible, covering both sexes, and a wide range of age and computing experience;

- to match the procedures used and data collected in the different experiments in order to maximise the validity in comparing the data;

- to repeat the experimental tasks so that learning curves could be compared.

The BTRL team proposed general criteria for usability that the procedures should be easy to use, easy to learn and consistent with user expectations (cf. normal telephony). These could be assessed by measures of effectiveness, efficiency, satisfaction and learnability, as defined in subclause 3.1.

4.2.2.1 Data requirements

To generate qualitative and quantitative data to support these requirements, all subjects were tested on all implicit tasks in the procedures. Rather than asking each subject directly to perform each task, it was decided to encourage greater involvement and motivation from them by presenting a number of scenarios, indirectly invoking the required procedure and actions. For example, subjects may be asked to imagine they are about to call an old friend, and may want to check their appearance before making the call. The subject should then activate selfview.

The time taken to complete each task and any errors were noted. In order to analyse any errors or problems arising, subjects were recorded on video.

4.2.2.2 Scenarios

Three scenarios were prepared by BTRL. The first involved asking a number of friends to a dinner party, including showing a map. The second was an attempt to make an appointment at the hairdresser, including showing a specimen hairstyle. The third involved making a complaint to the local authority, regarding roadworks, also including a sketch.

Each scenario required the subjects to make and answer a number of telephone and videotelephone calls. Each call included one or more of the procedures to be tested. Each subject then completed four trials. One scenario per trial, with the first scenario being repeated. The time taken to complete each procedural task and any mistakes made provided the data for subsequent analysis. Repetition of tasks from scenario to scenario allowed learning to be assessed over the four trials.

For each national experiment, the scenarios were translated and adapted as required by cultural, language, network or administrations' needs.

The specific procedures that each experiment included in its scenarios and tested are shown in table 3.

Table 3: Control functions included in the videotelephone user control procedures tested

	BTRL	CSELT	NL-PTT	PKI
Make telephone call	✓	✓	✓	✓
Make videotelephone call	✓		✓	✓
Receive telephone call		✓	✓	✓
Receive videotelephone call	✓		✓	✓
Receive video call in audio only (change service)	manual		auto	auto
Receive video call in audio only (use video pause)	✓		✓	
Upgrade telephone to video call		✓	✓	✓
Downgrade video to telephone call			✓	✓
Accept an upgrade to video		✓	✓	✓
Accept a downgrade to telephone				
Reject an upgrade to video		✓		
Reject a downgrade to telephone				
Select selfview, before a video call	✓	*	✓	✓
Select selfview, during video call	✓		✓	✓
Select selfview, during a telephone call				✓
Deselect selfview			✓	
Select video pause, during video call	✓	✓	✓	✓
Select video pause, during selfview (during a video call)				✓
Deselect video pause			✓	in selfview
Select second camera				✓
Deselect second camera				✓
Convert handset to handsfree	✓		✓	✓
Convert handset to loudspeech				✓
Convert handsfree to handset			✓	✓
Convert handsfree to loudspeech				✓
End call in handset			✓	✓
End call in handsfree			✓	✓
End call in loudspeech				✓

(continued)

Table 3 (concluded): Control functions included in the videotelephone user control procedures tested

NOTE: The CSELT equipment was set up to reflect the preferred procedure, which requires that a telephone call is established first and then a conversion to video may be requested by the "A" party which has to be accepted by the "B" party before the second channel or mode change is invoked by the network. CSELT also used two workstations working back to back to emulate the interface so that both the "A" and "B" parties were subjects. Using the workstation also required the user to use a mouse to activate the "keys" and to "lift the handset".
CSELT also introduced some methodological differences into the experimental method, which significantly reduced the number of times each procedure was completed and the time taken for each procedure. Consequently, the overall number of errors were significantly reduced.
The PKI implementation used a PKI commercial videotelephone, so consequently the procedures were fixed by the implementation, which meant that PKI could not sensibly test the procedure recommended for "Receiving a Video call in Audio only using Video Pause". The NL-PTT did include this procedure, even though they had a telephone default on incoming calls, and not surprisingly it registered the greatest number of errors across all the NL-PTT's procedures.

4.2.3 The four experiments

Four different national experiments were conducted as shown in table 4.

Table 4: Summary of the four experiments conducted

Member	Country	Type of videotelephone	Subjects	Date
BTRL	UK	Macintosh simulation	33	July 1991
PKI	Germany	Philips	30	Nov. 1991
PTT Research	Netherlands	RACE	36	Dec. 1991
CSELT	Italy	SUN simulation	22	Feb. 1991

4.2.3.1 Interfaces

Some idea of the differences in implementation of hardware can be judged by the interfaces, illustrated in figures 1 to 4, It should be noted that BTRL employed a touch screen, while CSELT used a mouse. The effects these had on the results is discussed in their respective reports. Only the PKI trial utilized real ISDN videotelephones, connected through an actual switch.

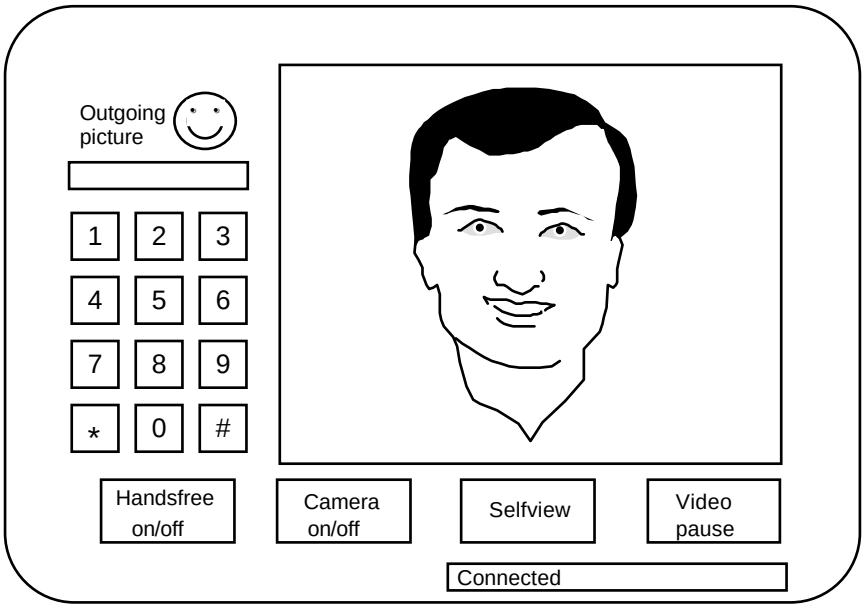

Figure 1: BTRL, Macintosh simulated videotelephone interface

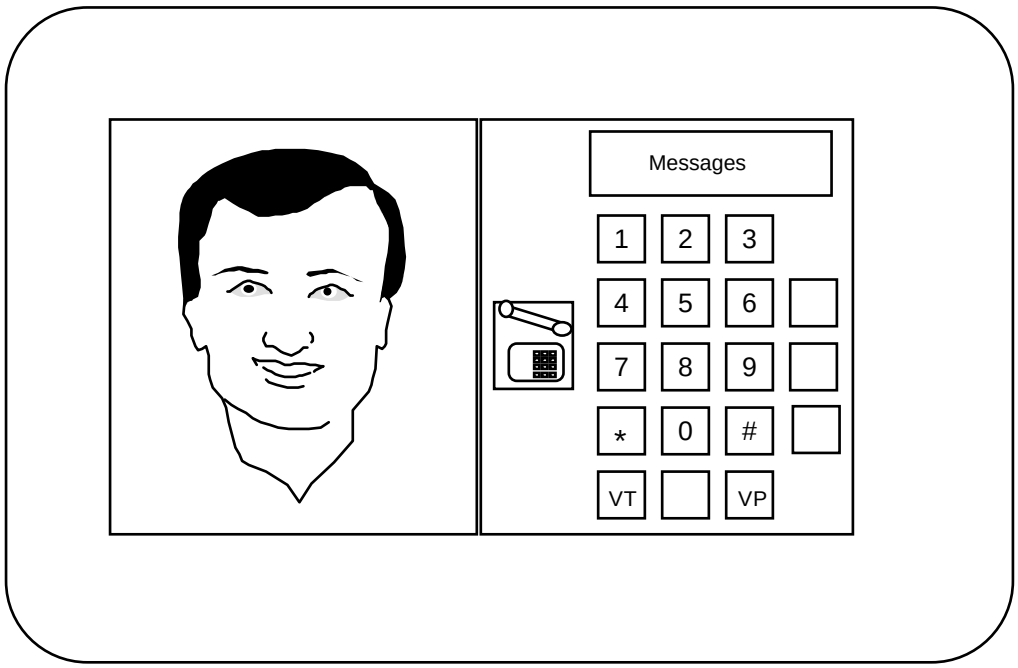

Figure 2: CSELT, Sun workstation simulated videotelephone interface

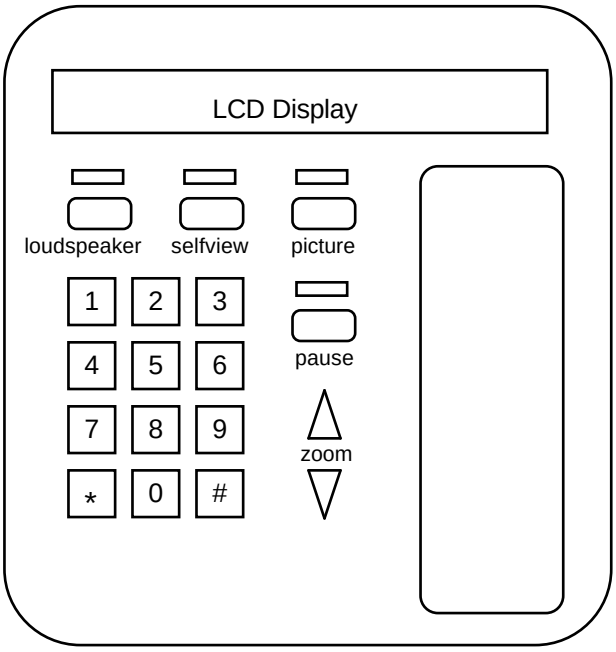

Figure 3: NL-PTT Research, RACE videotelephone/telephone interface

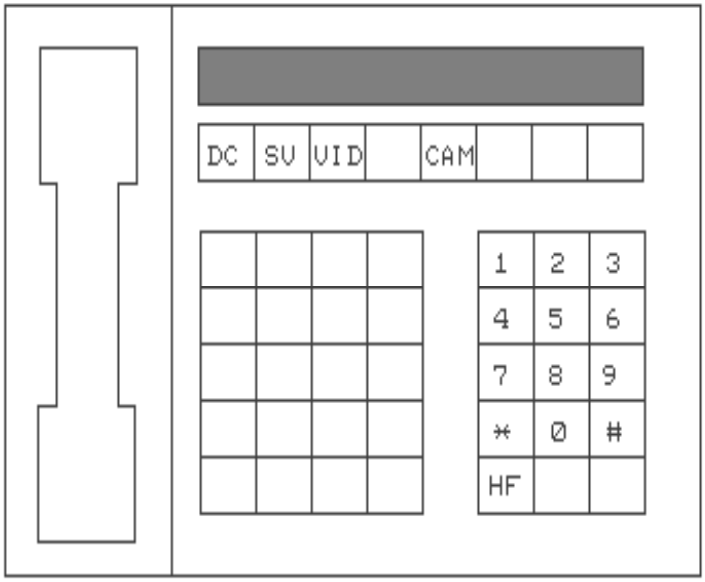

Figure 4: PKI, videotelephone/telephone interface

Unfortunately, all the experiments did not take place using real equipment. However, the consistency of the resulting data suggests that this was not an important factor. The common elements were the procedures and tasks required from the subjects and the scenarios embodying them.

4.2.3.2 Experimental procedure

All except CSELT employed an experimenter, who acted as the second party in the calls, and who also did a certain amount of "prompting" for the subject when they got into difficulties. CSELT used subjects "face-to-face", in pairs, without the experimenter being actively involved.

The experimental procedure was similar in each case. The subject was briefed and sat in front of a videotelephone or the simulation. The subjects were then expected to proceed through all four scenarios,

each containing 12 procedures in about eight calls, both incoming and outgoing. Presentation order of the scenarios was rotated.

In the case of BTRL and PKI, the experimenter read aloud the next procedure for each call. Errors were either self-corrected or the experimenter "prompted" the subject. In the case of the NL-PTT, the experimenter gave the whole procedure to the subject, who made the best he could of it. The CSELT experimenter, in the event of a subject making an error, drew their attention to the help sheets. These minor differences in experimental procedure are referred to in relation to the overall error scores later.

4.2.3.3 Questionnaire

The questionnaire for the collection of subjective data adopted by BTRL was used in translated form by all participants. It is fully described and illustrated in some of the national reports. It was administered after the scenarios had been completed by each subject, followed by a post-trial interview. Overall data from the questionnaires and interviews is referred to in subclause 4.3.2.

4.2.3.4 Subjects age and experience

There appeared to be a good distribution of attributes amongst the subjects used in all the experiments. An approximate distribution of age and sex is given in table 5.

Table 5: Summary distribution of subjects age and sex in the four experiments

Expt	< 25	25 - 35	35 - 45	45 - 55	> 55	Total	Male	Female
BTRL	(All between 18-60+, 16 between 35-50)					31	17	14
CSELT	11	3	(8 between 36-50+)			22	11	11
NL-PTT	17	10	4	2	3	36	15	21
PKI	1(4)	5(2)	4(4)	2(3)	2(0)	27	14	13

The total figures given in table 5 represent those whose data were included in the analysis (compared with the table 4 in subclause 4.2.3). BTRL "lost" two subjects through equipment malfunction, and PKI used two for a pilot and one for a base-line test.

Computer experience ranged from none for nearly 50 % of the BTRL sample to nearly 40 % with high experience in the PKI sample. The usual categories used were: None, Little or Low (some word processing perhaps), Medium (a few software packages), High (a programmer) are shown in table 6.

Table 6: Summary of subjects computer experience

Expt	None	Low	Medium	High	Total
BTRL	14	11	5	3	33
CSELT	10	7	3	2	22
NL-PTT	14	(14 low-medium, 8 medium-high)			36
PKI	4	6	7	12	30

None of the subjects, in any of the experiments, had any experience in videotelephony. However, there is sufficient variation between the groups to suggest that the subjects for each experiment may not come from a homogenous group. Nor was it possible in any single experiment to balance the subjects in terms of age, experience and sex. Consequently, the standard position is taken that this imbalance may impart an effect on the results obtained but that any bias is random in effect. Some effort to look at these aspects was made, but no consistent effect was perceived within the data available except in the PKI data where there was a tendency for less computer experienced subjects to make conservative responses to the subjective questions and for more computer experienced males to make more errors. Full details of each experiment are given in the respective reports. The subclauses that follow compare and contrast the results and findings in relation with the original aim of validation of the procedures, both specifically and generically.

4.3 Experimental results - general

The full results are given for each experiment in each national report, these include discussions of the results in terms of the procedures studied. This subclause presents a comparison of the overall results from each experiment. The data collected for each of the four experiments is shown in table 7.

Table 7: Summary of data collected in four experiments

	BTRL	CSELT	NL-PTT	PKI
Objective				
Elapsed time per procedure element	✓	✓	✓	✓
Errors per procedure, raw score	✓	✓	✓	✓
Errors per procedure, percentages	✓	✓	✓	✓
Error analysis or confusion matrix	✓		✓	✓
Subjective				
Videotelephone in general	✓	✓**(4Q)**	✓	✓
Procedure by procedure	✓**(6Q)**	✓**(1Q)**	✓**(4Q)**	✓**(2Q)**
Videotelephone at home			✓	✓
Who sees what				✓
Computer User Satisfaction Inventory (CUSI)				✓

NOTE 1: The timing data was collected from the completion of the instruction or previous control action to the selection of the last control element in the procedure. The time taken for dialling was excluded. CSELT were unable to use the timing data as their subjects were able to read and reread the instructions, and their interface required mouse operation. Both of which would have obscured any differences in time caused by the procedure.

NOTE 2: The error data was collected as raw scores, excluding dialling errors, and converted to percentage scores. The percentage was based on the total number of errors for a procedure, divided by the total number of occurrences of a procedure. For example if a procedure happened three times in a scenario, with four scenarios and sixteen subjects, the total number of occurrences is obviously 3 x 4 x 16 = 192.

NOTE 3: The error analysis is a simple confusion table or matrix, which shows the raw score or percentage error made for each procedure analysed for the controls actually activated. There is obviously the potential to reflect both procedure and implementation effects within this analysis.

NOTE 4: The subjective questions relating to videotelephone, in general, were developed as a set by BTRL. Both PKI and the NL-PTT used the same question set with the exclusion of one question which related to the instructional material. CSELT selected four questions from the set of ten/eleven for inclusion in their overall question set.

NOTE 5: The subjective data relating to the individual procedures varied between the experiments. BTRL, for example, asked a profile of six questions, selected from the general eleven, for each of seven procedures. Whereas the NL-PTT used two basic questions together with their converse across twelve procedures, PKI used two basic questions on thirteen procedures and CSELT used one question on each of four procedures.

NOTE 6: The NL-PTT and PKI had the advantage of conducting their experiments later and also looked at perceptions of acceptability of the videotelephone with reference to the home. The purpose behind this was to get a perspective on the issue of privacy.

NOTE 7: PKI also used two other question sets. Who sees what? To examine the subjects understanding of what was being seen by the other party during different procedures. Also, CUSI, a standard validated satisfaction inventory which could be compared with the responses to the BTRL questionnaire and could be used to compare satisfaction of the videotelephone with other products.

4.3.1 Comparison of objective data

The objective data collected from the four experiments can be compared graphically by the four graphs, shown in figures 5 to 8. These have been prepared by averaging the mean times and errors per procedure, from the learning curves shown in the individual reports, and presenting these on a set of graphs using the same axis.

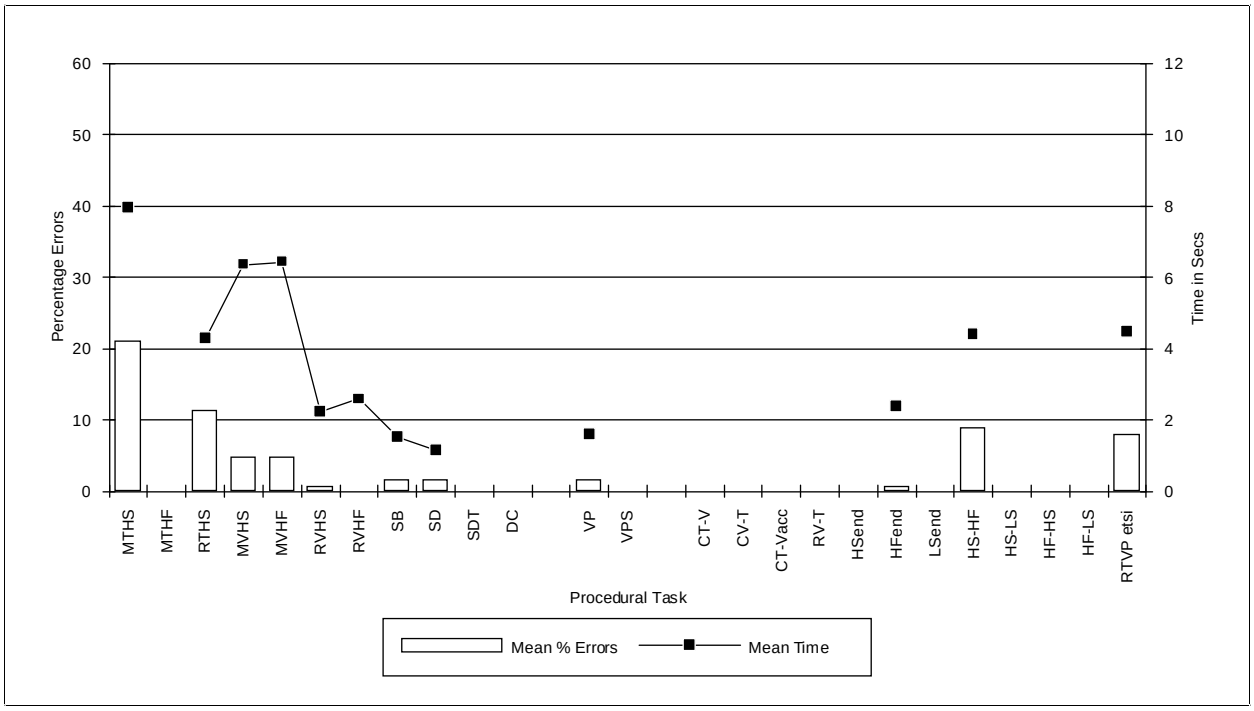

Figure 5: BTRL data: mean percentage errors and mean time across all trials for each task

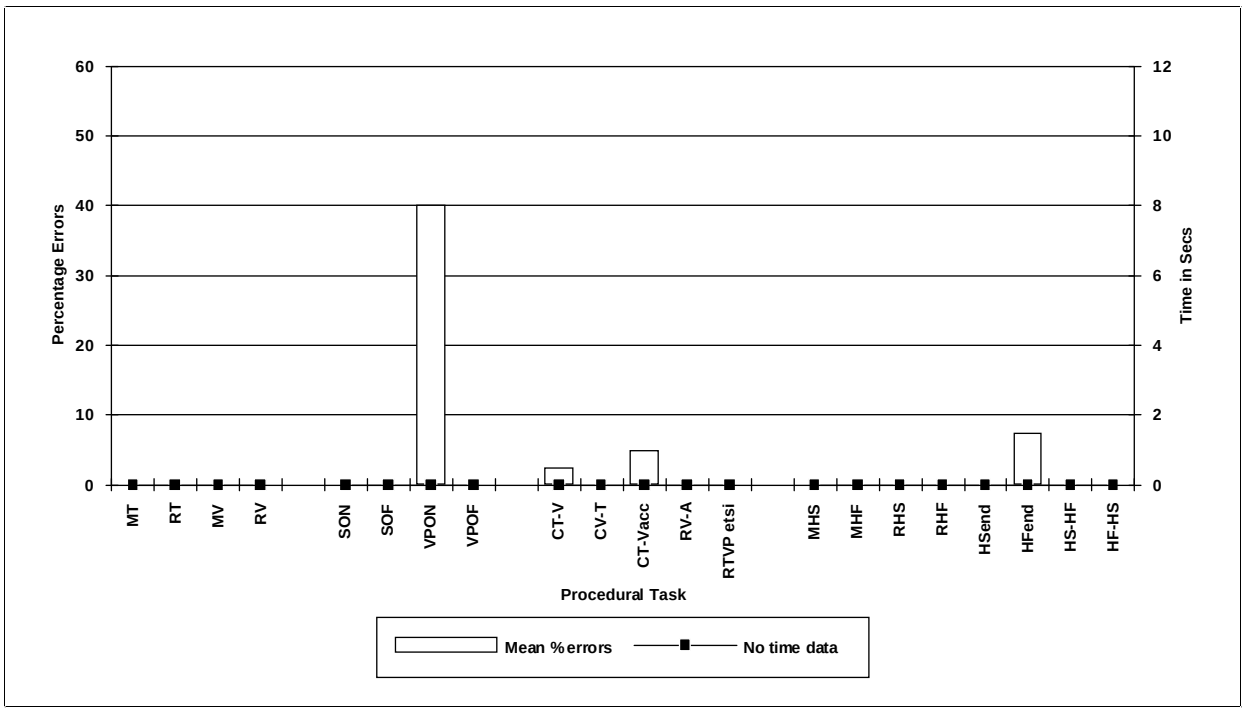

Figure 6: CSELT data: mean percentage error across all trials for each task

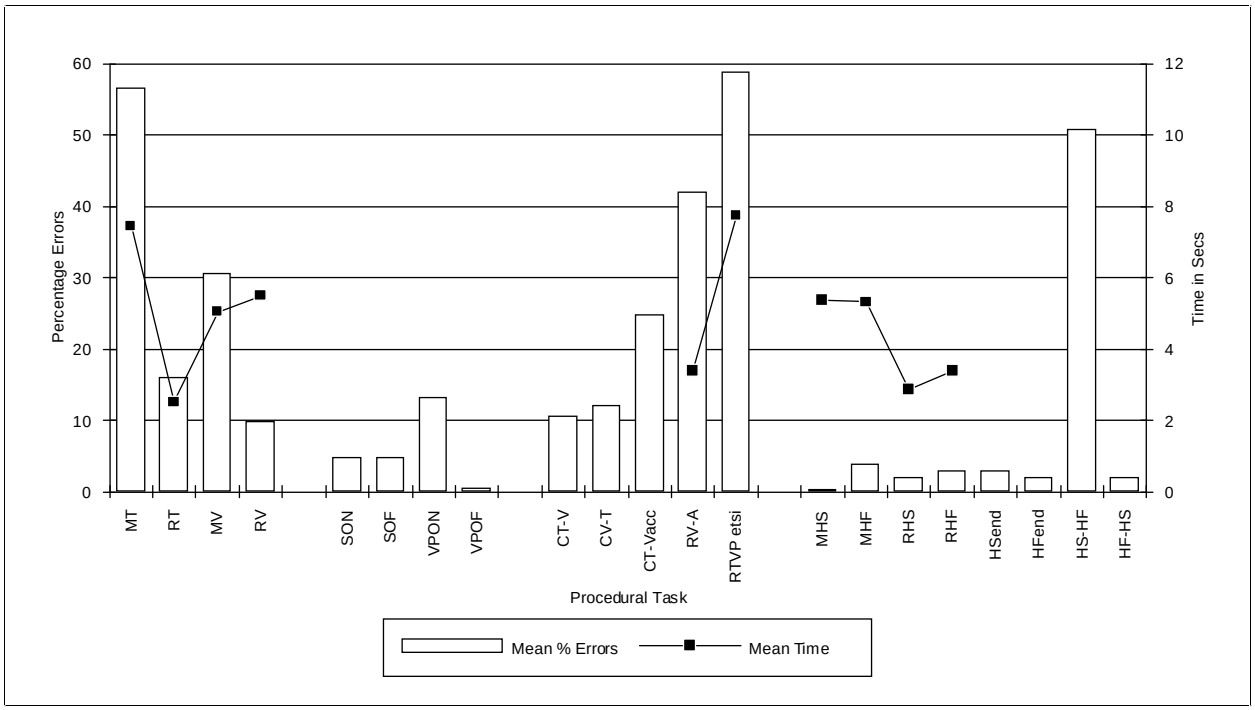

Figure 7: NL-PTT data: mean percentage error and mean time across all trials for each task

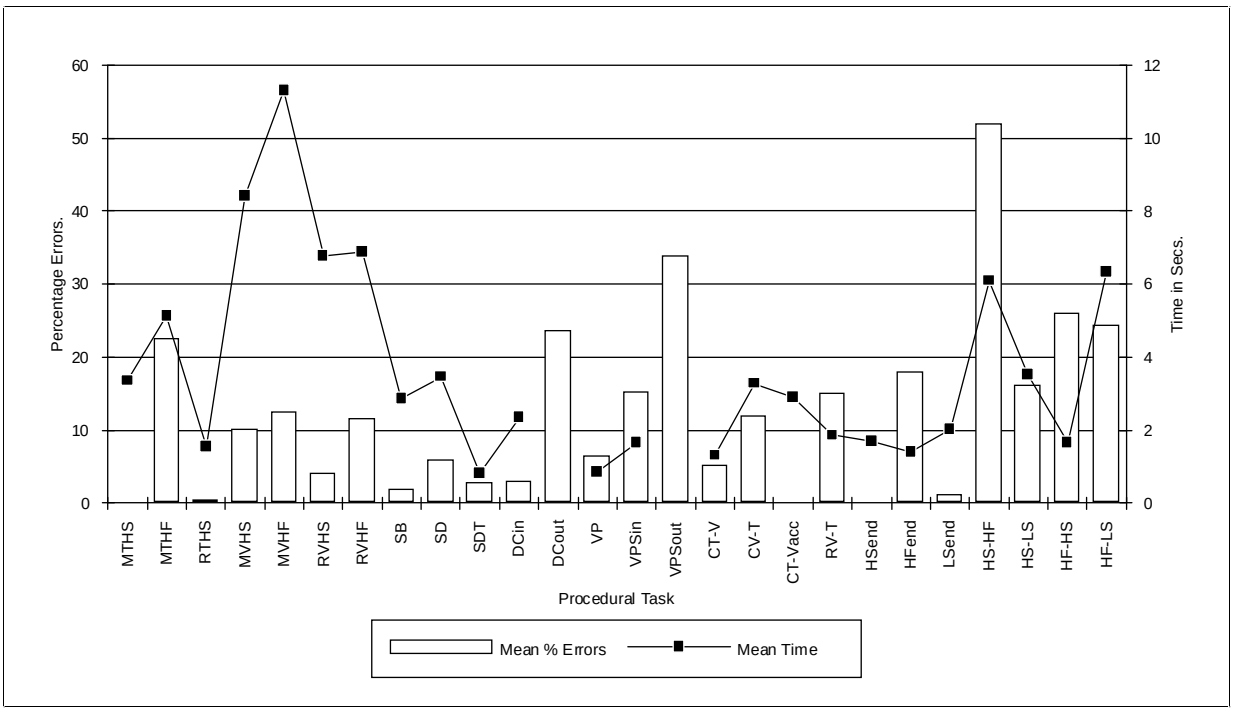

Figure 8: PKI data: mean percentage errors and mean time across all trials for each task

4.3.1.1 Implementation effects

There is no real significance in joining the mean times for each procedure to draw a graph line, except to emphasise a difference in the error histogram. What is interesting and significant is the degree of variation between the different sets of results. The errors for making a basic telephone call range from 0 % for CSELT to more than 50 % for the NL-PTT. Similarly, the time taken to make a video call ranges from 5 seconds for the NL-PTT to as much as 11 seconds for handsfree video calls for PKI. Both of these broad ranges are caused by the different implementations and specifically by the choice of default service mode. This is considered in detail later. The similarities between these graphs can be found, for example, where the peaks and troughs of errors and the times occur in each graph. The call set up procedures, some of the handset/handsfree controls and perhaps receiving video as audio, show the peaks; while the basics of selfview, video pause and second camera access show the troughs. However, PKI showed a variance to this trend by asking users to perform the more complex task of selecting video pause (camera off) in selfview and then to recover normal picture. Getting into this nested procedure caused some problems, but to get out again the subjects were looking for a single key rather than the correct two key action. A similar effect was seen as PKI subjects tried to get back to the main camera. The fact that this problem was not shared by the NL-PTT, who also asked people to get out of a procedure, is most probably due to the NL-PTT implementation, which included LED indicators to show when those controls were selected.

Another effect of an implementation could be the generally low level of errors within the BTRL results. This may be due to the simplified procedure built on a touch screen and with instantaneous response, and also to the feedback on which controls was active and which had just been activated (the key "greyed" out, if it was not available; and reversed its contrast, if it was activated). The effect may also be due to the simpler set of calls within each scenario. Because, in order to get the maximum data, both PKI and the NL-PTT increased the number of procedures tested in each call from BTRL's average of 1,57 per call (33 procedures in 21 calls in 3 scenarios) to 3,87 (PKI).

The final effect from implementation is the data available from CSELT. Because the procedure invoked required the user to set up a telephone call and then convert it to a video call, there were fewer procedures that could be tested. In addition, this implementation made selfview available at all times through a picture in a picture window on the main screen, which again restricted the procedures available to test and, for the user severely limited, the number of controls they had to consider. At the same time, the procedure used by CSELT was also simplified. Each subject only made or received 8 calls (2,25 procedures per call) against the 28 made by BTRL's subjects and the 32 made by the NL-PTT's.

4.3.1.2 Learning effects

A second way of looking graphically at the objective data is to consider the learning effect. For BTRL, the NL-PTT and PKI each subject completed four scenarios, the first three were different in detail, but the fourth always replicated the first for each subject. Therefore, learning can be considered by comparing time and error data across each procedure between first and fourth scenarios (or trial). These are shown graphically in figures 9 to 11. The obvious impression is that the BTRL graph had fewer errors and lower times than other PKI or the NL-PTT. This is due to the three factors: a simpler more supportive implementation, less procedures tested (therefore no data points) and no switch needed to make the connection (hence the lower times).

Aside from the visual impact, there is a very significant consistency across almost all the procedures in all the experiments: both the mean errors and mean times have been reduced between trial one and trial four.

The exceptions (in errors only) include:

- receive video in audio (NL-PTT);
- make telephone call handsfree, Downgrade Video to Telephony, and Receiving Video in audio (PKI).

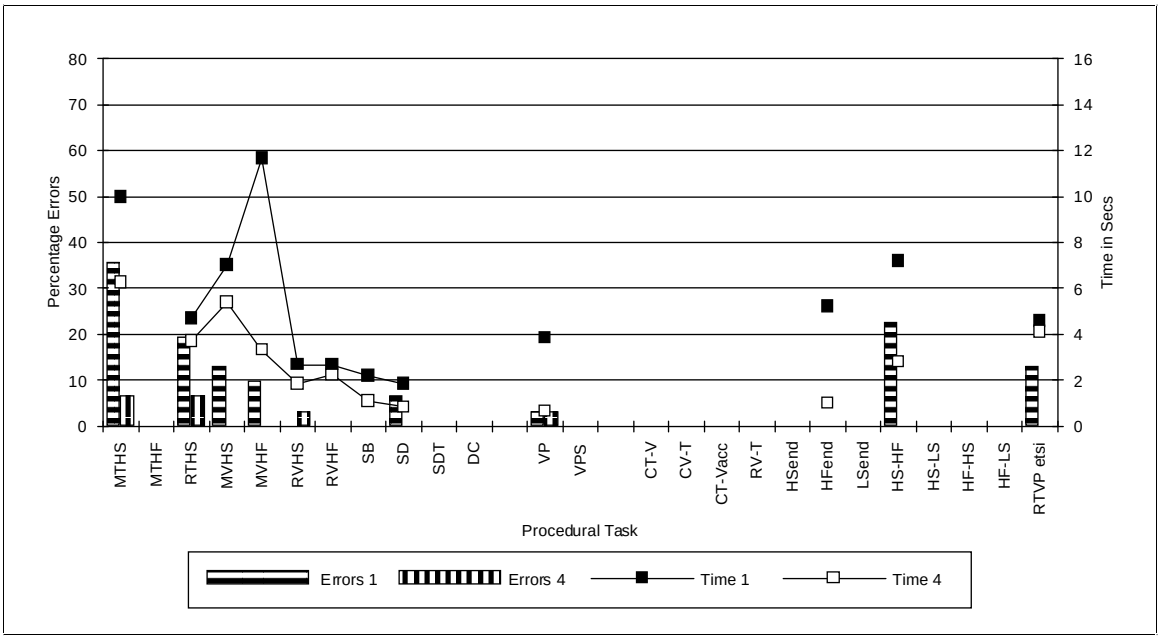

Figure 9: BTRL learning data: mean percentage errors and mean time for first and fourth trial for each task

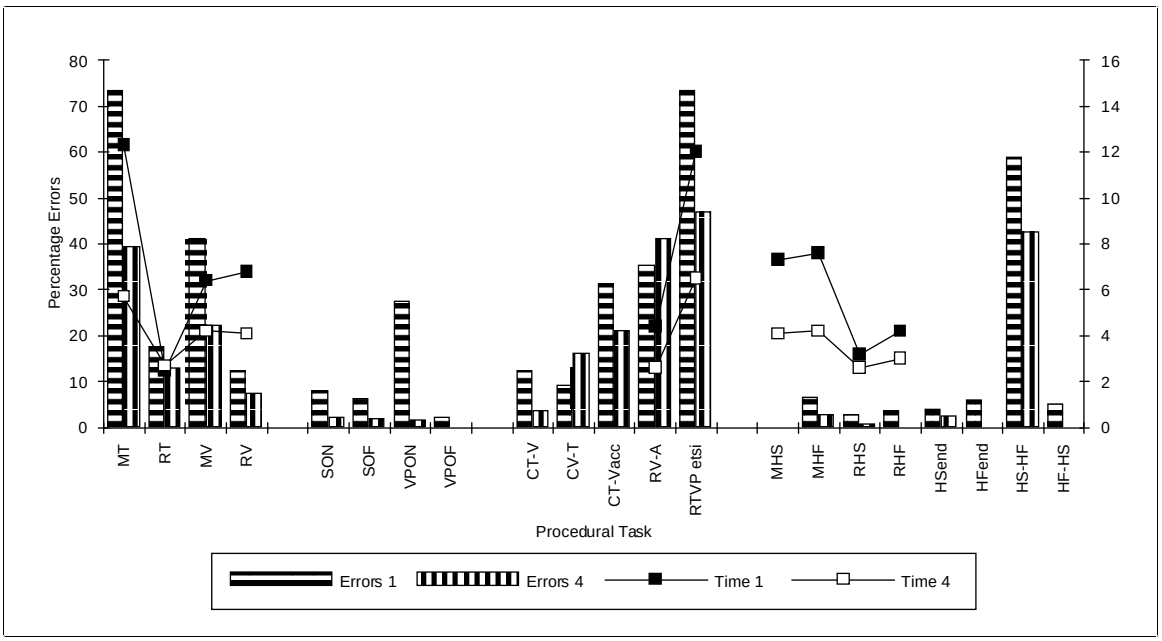

Figure 10: NL-PTT learning data: mean percentage error and mean time for 1st and 4th trial for each task

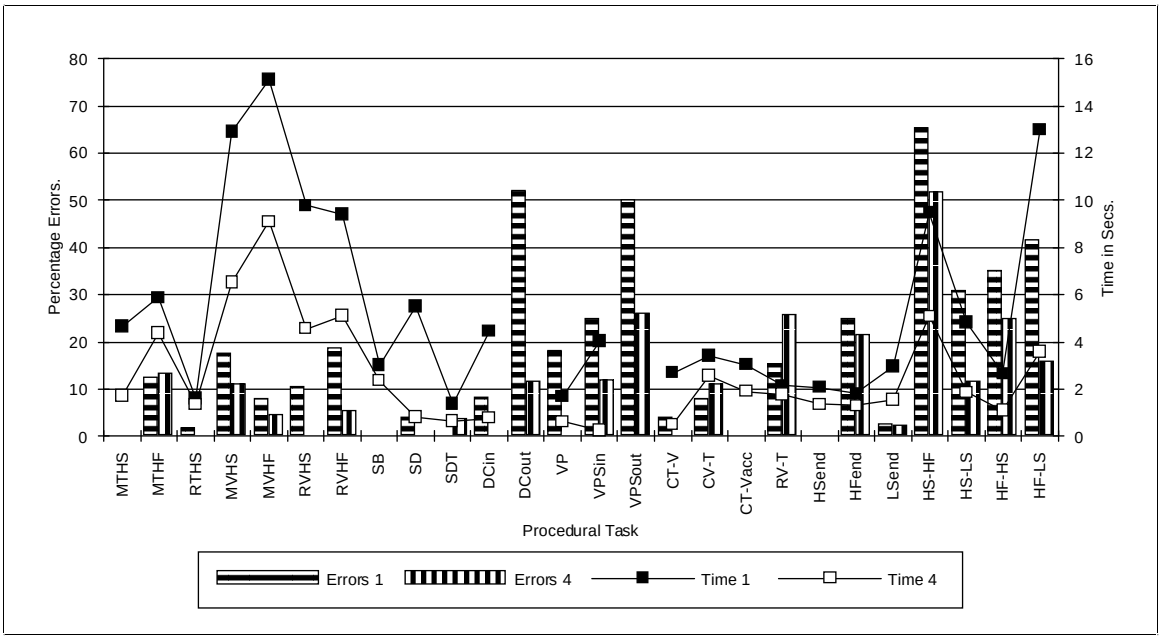

Figure 11: PKI learning data: mean percentage errors and mean time for first and fourth trial for each task

4.3.2 Comparison of subjective data

The subjective data from the questionnaires collected from the four experiments can be compared graphically within the three bar charts, shown in figures 12 to 14. These have been compared by normalising the mean results obtained by each experiment. Comparison can be made possible on three aspects. First, "videotelephony in general" and second, across the procedures for two questions: "Is the procedure useful?" and "Is the procedure too complicated?".

Comparing the mean responses to the "videotelephone in general" questions shows both considerable agreement and disagreement. There is agreement between the experiments when the question can be related to videotelephony generally, for example, "The videotelephone could be very useful", or "I enjoyed using the videotelephone". However, there is disagreement when the question can be related more to

specific implementations, for example, "The videotelephone is more complicated than it needs to be" or "I did not feel in control of the videotelephone".

The comparison also showed that there is no consistent trend of responses for any one videotelephone or implementation, although this statement has been tempered by the knowledge that the subjects across the four experiments may not be considered a homogenous group.

Comparing the mean responses, procedure by procedure, to the two questions "Useful?" and "Too complicated?", also shows considerable agreement and disagreement. The agreement stems from the subjects response to the question "Being able to use (specified) procedure is very useful?". Consistently, subjects, for example, rated telephony higher than videotelephony, and receive video in audio higher than downgrade to telephony. In fact, except for the procedure selfview, all the responses for each procedure are within 0,5 points on the scale. For some reason, the NL-PTT subjects rated selfview particularly low. The agreement turns to disagreement between the experiments when the questions become "Procedure (specified) is more complicated than it needs to be?". Within most of the procedures there is a much higher degree of variation between the means, than in the "Useful" bar chart (see figure 13). "Handset to Handsfree" is obviously too complicated at PKI, and "Receive video as telephone" is criticised by the NL-PTT subjects.

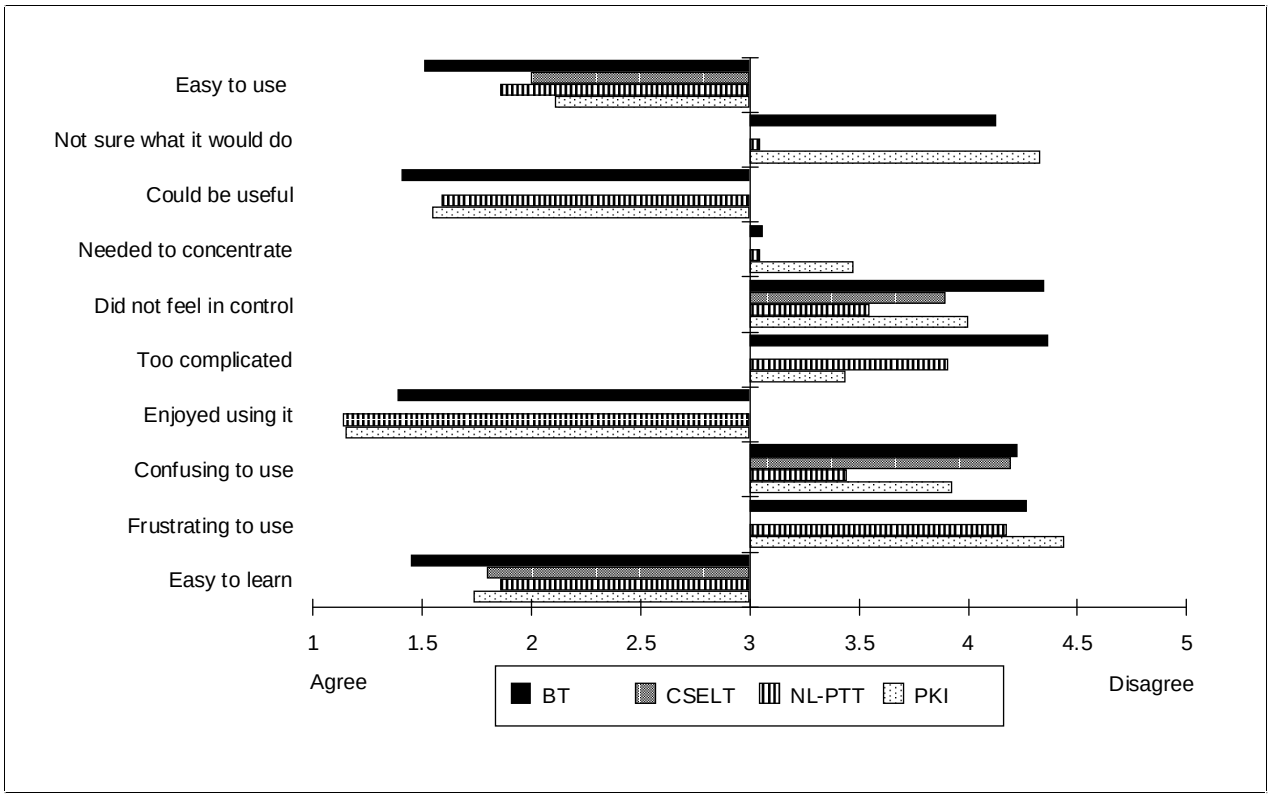

Figure 12: Comparison of mean responses across four experiments to the "Videophone in General" questions

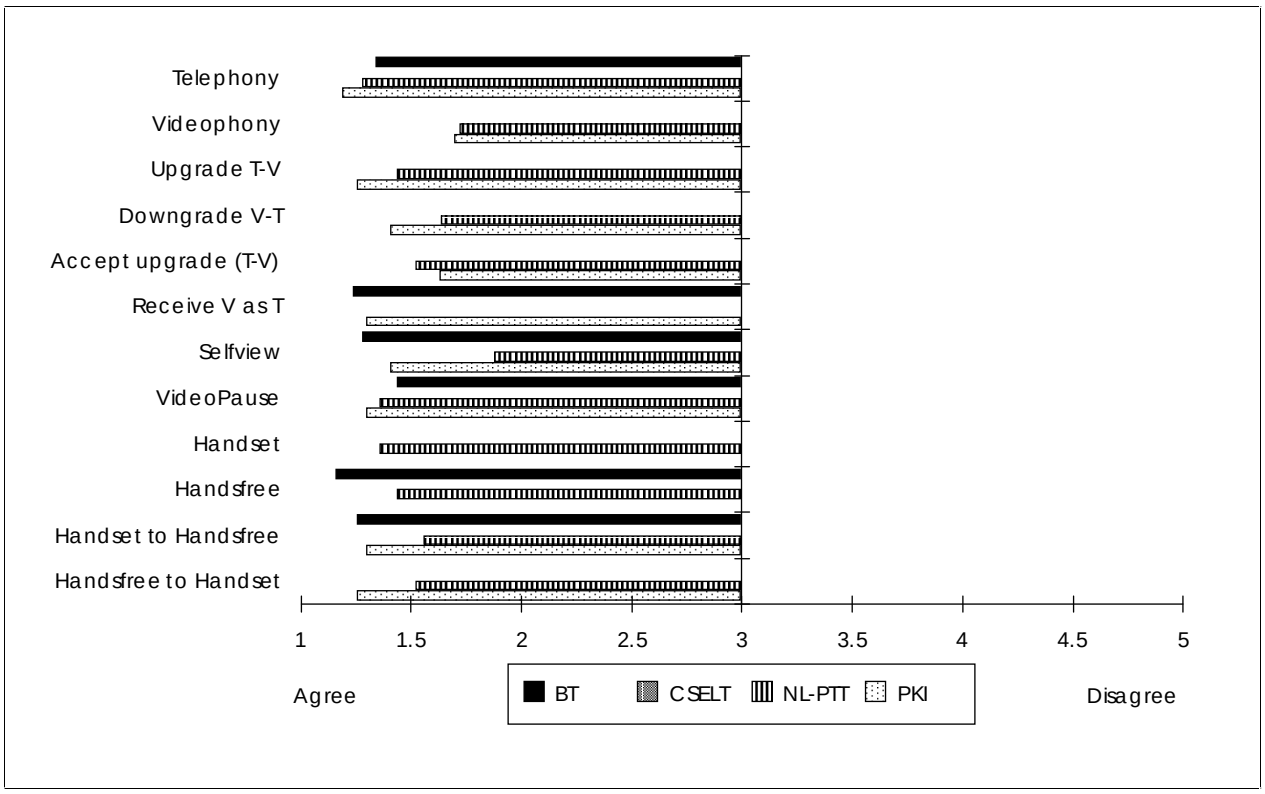

Figure 13: Comparison of subjective mean responses to "Useful" question, procedure by procedure

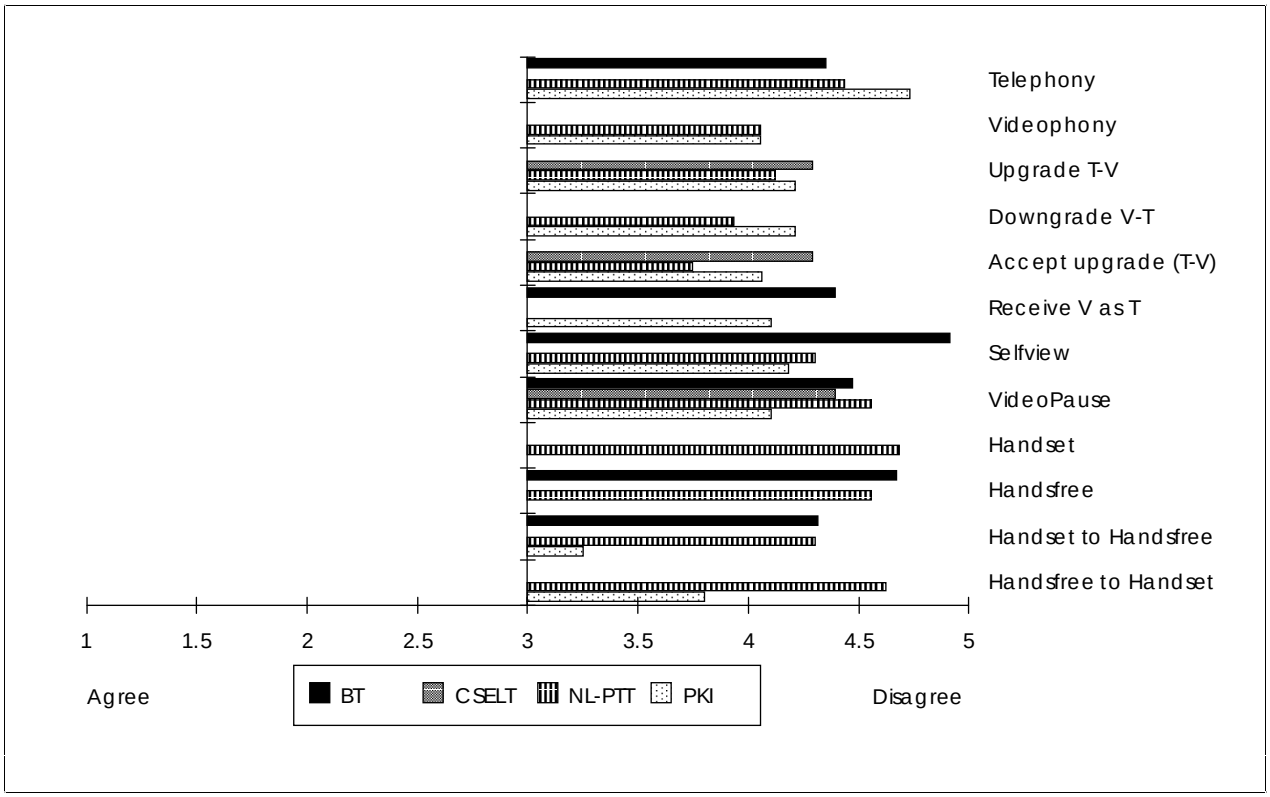

Figure 14: Comparison of subjective mean responses to "Too Complicated" question, procedure by procedure

However, there is some agreement, as can be seen between PKI and the NL-PTT for making video calls, or between BTRL and the NL-PTT to switch between Handset and Handsfree. However, what is interesting is that even when subjects were having extreme difficulty, as in PKI with Handsfree conversions

and the NL-PTT with Receiving video in audio, subjects were very reluctant to agree with the statement. They reflected the general trend to blame themselves rather than the procedure. In consequence, any low score on the disagreement side should be viewed by the system developers with concern.

4.4 Experimental results - specific

This subclause looks at the results from the four experiments in order to consider three key issues:

a) does the selected default service mode affect the users performance? (see subclause 4.4.1);

b) is there evidence to support or discount the concept of reciprocity? (see subclause 4.4.2);

c) is there evidence to indicate how specific videotelephone functions, e.g. selfview, video pause, etc. should work? (see subclause 4.4.3).

4.4.1 Default service mode versus. user performance

In the four experiments, the defaults used in each equipment were:

BTRL videotelephony default;
CSELT telephony default;
PKI telephony default;
NL-PTT mixed defaults, videotelephony out - telephony in.

Because the time data collected was in part affected by the response time of the equipment used, particularly in call set up, the comparison of user performance in relation to the specific default is restricted to error rates.

Table 8 was prepared from the raw score error data quoted in each of three reports BTRL, the NL-PTT and PKI, and converting to percentage errors with careful reference to the number of occurrences of each procedure in each scenario, the number of scenarios completed and the numbers of subjects. This normalising of the error data was done to ensure there were no differences in the basis for the percentage error calculations. The CSELT data was not included, because of the very different procedures used to conduct the experiment.

Table 8: Comparison of error scores between different default conditions

	Making a call	**Receiving a call**
Basic handset telephony		
Default telephony	**0 %**	**0,46 %**
Default videotelephony	**21 %**	**9,6 %**
Mixed defaults	**51 %**	**14 %**
Basic handset videotelephony		
Default telephony	**9,7 %**	**3,7 %**
Default videotelephony	**4,8 %**	**0,9 %**
Mixed defaults	**15,3 %**	**5,5 %**

Clearly then, to make and receive basic handset telephone calls, it is an advantage to have the default set on telephony. Similarly, to make and receive basic handset videotelephone calls, it is an advantage to have the default set as videotelephony. Equally clearly, there is a penalty for making and receiving either type of call if the default is mixed (videotelephony when making, telephony when receiving).

This is clarified if the table is re-drawn as table 9, to show the errors for default calls versus. non-default calls.

Table 9: Comparison of error scores between default and non-default calls

	Making a call	Receiving a call
Default call		
Default telephony	**0 %**	**0,46 %**
Default videotelephony	**4,8 %**	**0,9 %**
Mixed defaults	**15,3 %**	**14 %**
Non default call		
Default telephony	**9,7 %**	**3,7 %**
Default videotelephony	**21 %**	**9,6 %**
Mixed defaults	**51 %**	**5,5 %**

As indicated by table 9, it is clear that there is a penalty for making non-default calls, whichever is the default service mode set. There is also an additional penalty if the defaults are mixed (in an attempt to get the best of both worlds). There is also a minor advantage if the default condition is telephony rather than videotelephony (9,7 % versus. 21 % on making calls and 3,7 % versus. 9,6 % on receiving calls).

In compiling these data, the errors were averaged across all trials. However, it is also possible to look at the learning aspect and compare the differential error rates on trials one and four. If, for example, a learning based usability criteria is set for videotelephony, it could be as follows:

"A representative sample of users making and receiving videotelephone calls should achieve 80 % success on the first trial and 95 % success on the fourth trial. The same sample should also achieve 90 % success on the first trial and 100 % success on the fourth trial for making and receiving telephone calls on the same equipment".

Looking back at figures 9 to 11, the error rates on the first and the fourth trials of figure 9 which represent a videotelephone default, for MVHS, MVHF, RVHS, RVHF all meet the videotelephony element, but the error rates for MTHS and RTHS fail to meet the telephony element of the criteria. Whereas for the telephony default shown in figure 11, the error rates for first and fourth trials for MVHF, RVHS and RVHF all meet the videotelephone element and MTHS and RTHS both meet the telephony element. Unfortunately, MVHS and MTHF just fail to meet their respective target figures. Table 10, which shows the mixed default data confirms that this solution would be unlikely to meet the chosen criteria.

Again, the conclusion is that there is a penalty whichever default is chosen, but that the penalty is slightly less severe if the default is telephony.

Further analysis shown in table 10 to 13 was also carried out using the normalised percentage error scores to examine whether the chosen default service mode had an effect on the other videotelephone facilities.

Converting between videotelephony and telephony

Table 10: Comparison of error scores when converting between videotelephony and telephony

	Upgrade T-V	Downgrade V-T	Accept T-V
Default telephony	**5,5 %**	**13,9 %**	**0 %**
Default videotelephony	**not tested**		
Mixed defaults	**5,5 %**	**5,5 %**	**10,4 %**
NOTE: Unfortunately, BTRL did not include this procedure, however is it interesting that even the mixed default condition did not elicit high errors. Over half the downgrade errors for default telephony were due to the subjects choosing to turn the camera off (video pause) as a way to select telephony.			

Conclusion: There is no evidence for a difference between default modes.

Use of selfview

Table 11: Comparison of errors scores within the use of selfview

	Before making a call	During videotelephony	During telephony	Deselecting
Default telephony	**0 %**	**7,4 %**	**3,7 %**	**-**
Default videotelephony	**1,6 %**	**1,6 %**	**-**	**-**
Mixed defaults	**2 % (before and during video and telephony)**			**2 %**
NOTE: The variation in errors between the default telephony and the mixed default, tends to suggest that any difference is more due to implementation than the default setting. The higher figure is probably due to testing video pause nested within selfview (i.e. selfview on - video pause on).				

Conclusion: No evidence for an effect due to default mode.

Use of video pause (camera off)

Table 12: Comparison of errors scores within the use of video pause (camera off)

	Select	Select in selfview	Deselect	Deselect in selfview
Default telephony	**5,5 %**	**7,4 %**	**-**	**27,8 %**
Default videotelephony	**1,6 %**	**-**	**-**	**-**
Mixed default	**5,9 %**	**-**	**0,26 %**	
NOTE: The high error rates in default telephony occurred in the combined task (select selfview, select video pause, return to normal view).				

Conclusion: There is no evidence for a difference between default modes.

Receiving a video call as a telephone call (audio only)

Table 13: Comparison of errors scores when receiving a video call as a telephone call (audio only)

	Service mode change	Using video pause
Default telephony	**18,5 %**	**-**
Default videotelephony	**11,3 %**	**8 %**
Mixed default	**38 %**	**54 %**
NOTE: For default telephony and the mixed default equipment this should have been a low error procedure. All the subjects had to do was: lift the handset (or press handsfree). However, it seems they were looking to press another key. This could be an experimentally induced error. Almost all the other procedures required the subjects to select one of the available keys. The 54 % error rate for using video pause on the mixed default equipment is due to this being a very complex and unnecessary task. The user had to accept the call first as a video call and then opt to select video pause.		

Conclusion: There is some evidence that there is an advantage in the default mode being videotelephony, and a lot of evidence against the use of mixed defaults.

Consideration of default mode

There is then a lot of evidence that choice of default mode does impact users performance. As could be expected, the impact is most significant on call set up, but perhaps, surprisingly, also on receiving video calls in audio mode. The conclusions are:

a) there is a penalty for choosing mixed defaults (e.g. videotelephony outgoing calls, telephony incoming calls);

b) there is a penalty in making and receiving non-default calls, the penalty is significantly smaller across all trials if the default is telephony;

c) there is probably no difference in making and receiving default calls, as long as there is a clear default service mode established;

d) there could be a penalty in receiving video calls as telephone calls with a telephone default.

4.4.2 Reciprocity versus non-reciprocity

Unlike the default conditions, no distinction was made between the experimental equipment for reciprocity. All provided a non-reciprocal service. The concept of reciprocity is derived from a simple model of the visual world because, in normal circumstances, if party "A" can see party "B" then party "B" can see party "A".

The impact of a principal of reciprocity within videotelephony would cover several areas see tables 14 to 16).

Table 14: Reciprocity during call set-up

Terminal "A"	Terminal "B" reciprocal	Terminal "B" non-reciprocal
Idle	Idle	Idle
Call "B"	Ringing - caller ID	Ringing - caller ID
Call "B"	Answer - caller ID	Answer - party A
Call "B"	Accept - send pic	Accept - send pic
Party "B"	Party "A"	Party "A"

Table 15: Reciprocity during selfview

	Terminal "A"	Terminal "B" reciprocal	Terminal "B" non-reciprocal
During call	Live "B"	Live "A"	Live "A"
Select SV	Live "A"	Not Live "A"	Live "A"
De-select SV	Live "B"	Live "A"	Live "A"

Table 16: Reciprocity during video pause

	Terminal "A"	Terminal "B" reciprocal	Terminal "A"	Terminal "B" non-reciprocal
During call	Live "B"	Live "A"	Live "B"	Live "A"
Select VP	Not Live "B"	Not Live "A"	Live "B"	Not Live "A"
Deselect VP	Live "B"	Live "A"	Live "B"	Live "A"

From this brief analysis, it can be seen that the principal of reciprocity forces a change on both screens in order to meet the requirements of one of the users. From the experiments, it was seen that error rates increased substantially if the equipment did not behave as the users were expecting. For example, the error rates on call set up in mixed defaults equipment, or the error rates in the PKI data when the users tried to exit a video pause inside a selfview by a single action (select video). Significantly, users had sufficient problems understanding their own videotelephone and what happened in different states. Again,

as the PKI data showed in the "who sees what?" questions, the functions causing greatest confusion over what the "B" party could see were, basic telephony, and selfview, but even then the predominant response was for the correct one according to the non-reciprocal model.

In conclusion, there was no clear evidence available from the experiment on the effect of reciprocity, but there is some circumstantial evidence that people are not expecting the principal to be adhered to. For example, no subject questioned this aspect during the debriefing and the "who sees what?" data from PKI confirmed that during camera off (video pause) the majority of subjects correctly reported that the "B" party gets a blank screen, and during selfview the "B" party still gets the "A" party live.

4.4.3 The working of specific functions

4.4.3.1 Video pause or camera off

Within the set of experiments there were two interpretations of how a video pause function should work. The two interpretations can be considered as **camera off** (the current camera is switched off or covered, which prevents any further image being captured and sent) or **stop picture** (the current camera continues working, but somewhere the transmission of the picture to the "B" party is stopped). The different effect of these two can be seen if the nested procedure of video pause during selfview is considered (see tables 17 and 18).

Table 17: What may happen if "video pause" equals "camera off"

	Terminal "A"	**Terminal "B"**
Video call	Live "B"	Live "A"
Selfview on	Live "A"	Live "A"
Video pause on	Blank screen	Blank screen + message
Video pause off	Live "A"	Live "A"
Selfview off	Live "B"	Live "A"

Table 18: What may happen if "video pause" equals "stop picture"

	Terminal A	**Terminal B**
Video call	Live "B"	Live "A"
Selfview on	Live "A"	Live "A"
Video pause on	Live "A" (+ message?)	Blank screen + message
Video pause off	Live "A"	Live "A"
Selfview off	Live "B"	Live "A"

Within the four experiments, the function was implemented as follows:

- PKI: Camera off;
- CSELT: Stop picture;
- BTRL: Stop picture/camera off;
- NL-PTT: Stop picture.

Unfortunately, only PKI tested the nested procedure so there is no objective data evidence. Subjectively, if camera off is not acceptable then the PKI responses for the "Too complicated" question should be lower than those for the other three. This is the case shown in table 18, but this could be an affect caused by testing the nested procedure. Therefore, there is no real evidence for either implementation. The low error score from BTRL for video pause on, and those for the NL-PTT video pause on and off may simply be a function of better feedback. The BTRL touch key went into reverse video when it was operated, and the NL-PTT had a key with a LED.

4.4.3.2 Selfview

Within the four experiments, there were three implementations of selfview:

- CSELT: had a picture in picture, with a small selfview image available at all times;

- BTRL and the NL-PTT: had an exchange image solution with selfview occupying the whole screen, selfview made no impact on the "B" party;

- PKI: had an exchange image solution, which also included the second camera. Therefore, if document camera was on and selfview was on, Party "A" and "B" saw the document image.

There was no objective data within the trials which looked at this aspect, and there is no distinction within the subjective data which can assess the value, for example, of picture in picture. The differences that are present within the subjective data, implies that the NL-PTT selfview is less useful than PKI or BTRL and that, perhaps, the NL-PTT and PKI selfview is too complicated, in respect of the BTRL implementation. This may be an aspect of feedback of keys status or actuation. BTRL's key went into reverse video, PKI's key had no similar indication (LED on, etc.).

Subsequently, it was suggested that selfview should be available at all times, and should not be affected by video pause, or even perhaps by selection of camera. This issue was not resolved and the data does not offer any guidance.

4.4.3.3 Answering video calls in telephone only

This is a particular requirement of videotelephone systems to ensure the user can maintain privacy when answering an incoming video call. It is a highly rated facility, and is frequently one of the prime reasons for implementing the telephony default. The experiments had different defaults and different implementations which enabled this specific aspect to be examined.

The error rates from the different experiments are shown in table 19.

Table 19: Receiving a video call in telephone only

	BTRL	**CSELT**	**NL-PTT**	**PKI**
Telephony default	**-**	**17 %**	**38,%**	**18,5 %**
Videotelephony default:				
Service change	**11,3 %**	**-**	**-**	**-**
Video pause	**8 %**	**-**	**-**	**-**

A quick analysis of table 19 seems to give advantage in having the videotelephony default, particularly as the telephone default meant that all the user had to do was answer the call in the normal way. Privacy was automatically guaranteed.

Therefore the types of errors, which are summed up in table 20, should be considered in more detail:

- in the normal NL-PTT experiment, 64 % of the 38 % of errors (i.e. 24 %) were due to the subject pressing the video key, which was next to a flashing indicator (LED) which was there to highlight that the incoming call was a video call. Pressing the key switched the LED into a steady state and accepted the call as a video call. The other 36 % of subjects who made an error selected video pause, this may seem to guarantee there is no picture sent, but in reality had no effect. These errors now look to be distinctly implementation and experimentally induced;

- in the PKI experiment, 60 % of the 18,5 % of errors (11,1 %) were due to subjects selecting video pause, another 30 % selected video on. Again the errors seem to be essentially experiment induced. In all other procedures, except make and receive a telephone call, the subject was cued to press one or more of the five marked keys. Therefore even when it was not necessary some of the subjects conditioned response was to press something; and presumably something which is not selfview, handsfree, or document camera;

- in the CSELT experiment, the subjects errors haven't been analysed in this way, but the expectation is that the 17 % were induced by the experimental set-up to press something;

- in the BTRL experiment, the induced behaviour was in fact the correct one. Therefore, the only errors occurred in deciding which of the two most likely keys, labelled video pause and camera off, was correct in each circumstance. In the calls where they should have selected camera off to change the service to telephony before answering, 57 % of the 11.3 % errors (i.e. 6,4 %) selected video pause, the other 43 % simply forgot and answered the call as a video call. In the other case, where they should have selected video pause before answering, 90 % of the 8 % of errors selected camera off. Therefore in both cases the majority of the subjects who failed, failed safe.

Table 20: Analysis of fatal and safe errors in receiving an incoming videotelephone call as a telephone call, i.e. in audio only

		Total errors	Fatal errors	Safe errors
NL-PTT		**38 %**	**24 %**	**14 %**
PKI		**18,5 %**	**5,5 %**	**13 %**
BTRL	Service change	**11,3 %**	**4,9 %**	**6,4 %**
	Video pause	**8 %**	**0,8 %**	**7,2 %**

The evidence in table 20 seems to suggest that if a videotelephony default is used the users should be encouraged to use video pause as the method for answering a video call as a telephone call. It also suggests that the telephony default may present a problem for some users with this procedure. It certainly confirms that if there is a LED or other indication with the key to accept a video call, then it may be unacceptable to flash this to gain attention. There is probably too strong a stereotype for a significant number of people to cancel the flashing light, for this type of implementation to be acceptable. It again underlines the need for sensitive design and the value of user testing.

4.5 Conclusions from the trial results

The conclusions from the results of the trials are:

- four experiments were successfully conducted in four different countries to evaluate videotelephony user control procedures. The four experiments used four different implementations of the videotelephone user control procedures as initially agreed. Two experiments were based on simulated videotelephone terminals, one on a prototype Research and development in Advanced Communications technologies in Europe (RACE) terminal and the last one on a commercial videotelephone;

- the four experiments were conducted with a sufficient degree of consistency within the experimental design to enable comparisons to be drawn from the data;

- the data collected were valid measures of usability which covered the four dimensions of efficiency, effectiveness, satisfaction and learnability; which also gave valuable insight into the procedures and implementations of user interfaces for videotelephony;

- the agreed user procedures do not restrict user interface design, but do provide a consistent base line which can ensure a minimum level of usability;

- specific implementations can adversely and positively effect user performance. In particular, mixed service mode defaults and some flashing indications should be avoided;

- both telephony and videotelephony as the service mode default have a user performance related penalty associated with them. From the evidence collected there was no real distinction between the two options. What was clear, was that mixed default modes (e.g. Videotelephony out and telephony out) which may appear to offer the best solution have significant user performance penalties associated;

- the principle of reciprocity was not possible to test within the experiments, and may present significant confusion to the user if it was implemented rigorously across all videotelephone functions;

- that the functioning of some essential videotelephone facilities, notably video pause, selfview and second camera can be significantly different depending upon the implementation. For example, the function label can have a particular effect on the users expectation of how the function should work, video pause might be implemented as camera off or stop picture or freeze picture. Unfortunately, there was insufficient data from the experiments to clarify the best options within these functions;

- that the user control procedures defined for videotelephony were incomplete. In particular, no provision is made to enable the telephony default, and the need to allow change of service mode within the call set-up and the incoming call procedures. This criticism has been addressed in HF: User control procedures in basic call, point to point connections, for ISDN videotelephony" currently being developed within HF Technical Committee.

5 General conclusions and recommendations

5.1 General conclusions

The general conclusions are:

- the videotelephone experiments have tested an elaboration of the generic procedures specifically applied to videotelephony. The results of the testing have given valuable insights into the requirements of users and the opportunities for maximising usability within user control procedures. The results confirm the validity of the generic concept, the general rules and the generic procedures recommended within ETR 170 [1];

- the results of the videotelephone trials broadly confirm the requirements and recommendations which are included within "Human Factors in videotelephony" under development in TC-HF. It also highlighted some user control procedures for videotelephony that were omitted from the earlier document;

- the results of the videotelephone trials confirm that the definition of user control procedures as described in HF: User control procedures in basic call, point to point connections, for ISDN videotelephony" currently being developed within HF Technical Committee, do not significantly restrict the implementation of different user interfaces.

5.2 General recommendations

The following general recommendations are made:

- the documentation of user control procedures for videotelephony should be revised to accommodate the missing elements identified in subclause 4.5;

- consideration should be given to including a recommendation on default service mode options in HF: User control procedures in basic call, point to point connections, for ISDN videotelephony" currently being developed within HF Technical Committee;

- the issues relating to the functionality of video pause and the interaction of video pause, selfview, second camera, and perhaps other functions should be urgently addressed and documented in.

History

Document history	
October 1995	First Edition
January 1996	Converted into Adobe Acrobat Portable Document Format (PDF)

ISBN 2-7437-0271-0
Dépôt légal : Octobre 1995