ETSI TECHNICAL REPORT

ETR 175

February 1995

Source: ETSI TC-HF

Reference: DTR/HF-01008

ICS: 33.020, 33.040.40

Key words: User procedures, multipoint, videoptelephony

Human Factors (HF); User procedures for multipoint 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) was produced by the Human Factors 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.

This ETR has been produced due to the fact that multipoint videotelephony can be regarded as a promising new telecommunication service, and that user-friendly user procedures are of vital importance for the acceptance and uptake of this service.

Blank page

1 Scope

Multipoint videotelephony allows conference videotelephone calls with participants at more than two sites.

From a systematical point of view, multipoint videotelephony can be regarded as an example of multipoint videocommunication, which again is an example of multipoint communication. This is one of the reasons why services that are not explicitly designated "videotelephony", such as multipoint videoconferencing [1)], are also considered in this ETR.

From a technological point of view, multipoint videotelephony requires suitable videophones at the users' sites and specific interconnection facilities for switching multipoint connections, which in most cases are designated an Multipoint Control Units (MCUs).

This ETR focuses on "basic" or "plain" multipoint videotelephony, employing videophones as terminals. This implies, among other things, that enhanced applications, such as those employing workstations or PCs as terminals, are outside the scope of this ETR. The same applies to Integrated Services Digital Network (ISDN) supplementary services such as call forwarding, call transfer, etc.

Although the ISDN is conceived to be the first network to offer multipoint videotelephony, most of the results of this ETR are valid for other networks as well. This applies in particular to networks that already offer multipoint videocommunication services, such as broadband networks offering multipoint videoconferencing.

Multi-point/party communication that is not primarily based on audio-visual interaction, such as text-based communication, is not covered by this ETR.

Multipoint videotelephony can be established in at least two modes in terms of video presentation and control:

- "switched video" mode, in which only one single video signal from another terminal is transmitted to each terminal, enabling each user to see just one of the other subscribers' images at a given time;

- "mixed video" mode, in which a mix of some or all of the other video signals or multiple video signals are transmitted to each terminal, enabling each user to have continuous visual presence of all or a subset of all subscribers' images (e.g. by means of a split-screen presentation).

This ETR focuses on user procedures. As a consequence, other Human Factors issues of multipoint videotelephony are not covered [2)].

User procedures that are studied within the scope of this ETR are:

- procedures for setting-up a multipoint videophone call;

- procedures for switching the video signals within the framework of the switched mode;

- procedures for controlling the "mixture" of video signals within the framework of the mixed mode.

Considering that basic multipoint videotelephony will, in the first instance, allow only for the switched mode, this ETR mainly deals with "switched multipoint videotelephony". However, user procedures for "mixed multipoint videotelephony" are taken into account as far as possible.

The user procedures should be as far as possible consistent with user procedures for point-to-point videotelephony and for multipoint videoconferencing. Moreover, requirements of People with Special Needs (PWSN) ought to be taken into consideration.

---

1) If communication between more than two people is called "conferencing", multipoint videotelephony can be regarded as a special case of multipoint videoconferencing.

2) Additional Human Factors issues of multipoint videotelephony are, inter alia, audio and video quality requirements, visual indications, such as pictograms, etc.

Charging and billing principles and procedures as well as vetting procedures, although to be studied because of their importance for the acceptance and uptake of multipoint videotelephony, are outside the scope of this ETR.

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] Carey, J. (1981): "Interaction patterns in audio teleconferencing". *Telecommunications Policy*, vol. 5, no. 4, 304-314.

[2] Short, J., Williams, E., & Christie, B. (1976): "The Social Psychology of Telecommunications". London: John Wiley & Sons.

[3] Kelly, C.W. (1982): "A low bandwidth virtual space video teleconferencing system". *GLOBECOM 82*, 966-970.

[4] Nagra, A., Temime, I.P, & Viale, E. (1984): "Multipoint Videoconferencing Using Conditional Replenishment Codecs". *Proc. ITS*, 30-35.

[5] Sabri, S. & Prasada, B. (1985): "Video Conferencing Systems". *Proc. of the IEEE*, vol. 73, no. 4, 671-688.

[6] Kenyon, N.D., White, T.A., & Reid, G.M. (1985): "Behavioral and User Needs for Teleconferencing". *Proc. of the IEEE*, vol. 73, no. 4, 689-699.

[7] Chiariglione, L. & Corgnier, L. (1984): "Videoconference terminal with multipoint capability". *CSELT Rapporti tecnici*, vol. 12, no. 1, 39-44.

[8] Sawada, K., Eguchi Y., & Kishino, F. (1985): "Multipoint video teleconferencing system". *Rev. of the Electr. Commun. Lab*, vol. 33, no. 4, 647-655.

[9] Coll, D.C., George, D.A., Strickland, L.H., Guild, P.D., & Paterson, S.A. (1975): "Multidisciplinary Applications of Communication Systems in Teleconferencing and Education". *IEEE Trans. on Communications*, vol. com-23, no. 10, 1104-1118.

[10] Birrell, J.A., & White, P.N. (1982): "Using technical intervention to behavioural advantage". *Behaviour and Information Technology*, vol. 1, no. 3, 305-320.

[11] Hoecker, D.G., Brown, E.F., Wish, M., & Geller, V.J. (1978): "A Behavioral Comparison of Communication Using Voice-Switched and "Continuous Presence" Videoconferencing Arrangements". *NTC*, vol. 3.

[12] Romahn, G., & Mühlbach, L. (1988): "Multipoint Teleconferencing - Empirical Studies with Different Variants". *Proc. of the 12th Int. Symp. on Human Factors in Telecommunications*. session A, paper 1.

[13] Klein, P. (1985): "Video Workplace Conference". *IEEE Proc. of Globecom '85*.

[14] Addeo, E.J., Gelman, A.D., & Massa, V.F. (1987): "Multi-Media Multi-Point Communications Service Capability for Broadband Networks". *IEEE Proc. of ISS*.

[15] Watabe, K., Sakata, S., Maeno, K., Fukuoka, H., & Ohmori, T. (1990): "Distributed Multiparty Desktop Conferencing System: MERMAI". CSCW 90 Proceedings, 27-38.

[16] Furner, S.M., Watkinson, N., & Clark, W. (1990): "An ISDN multipoint teleconferencing environment". IEE CSCW Colloquium 24 October 1990.

[17] Furner, S.M., Watkinson, N., Clark, W., & Frowein, H. (1991): "A multipoint teleconferencing service for the ISDN". IVth Int. Conf. on Human Computer Interaction. Stuttgart.

[18] Schindler, S., & Heidebrecht, C. (1991): "TELES.VISION - Philosophy and Technology". TELES GmbH.

[19] Mühlbach, L. (1988): "User-Friendly Videotelephone Terminals". *Proc. of the 12th Int. Symp. on Human Factors in Telecommunications*, session J, paper 1.

[20] Mühlbach, L. (1990): "Procedures for Convening Multipoint Videoconferences". *Proc. of the 13th Int. Symp. on Human Factors in Telecommunications*, 121-131.

[21] Kishino, F., Okubo, S., Hisaki, T., & Murakami, S. (1984): "Multipoint Video Teleconferencing System Using Field/Frame Time Division Multiplexing Technique". *Proc. ITS*, 37-44.

[22] Cornacchia, M. & Papa, P. (1992): "Integration of Multipoint Videotelephone Systems with PC Facilities: Video Switching, Tele-pointing and Tele-drawing". RACE project R1065 (ISSUE), deliverable 65/FUB/HFG/D5/A/005/B1.

[23] Gili, J., Chacón, J.C., Ayuso, P., Rodriguez, I., & Jánez, L. (1991): "Icons in multipoint videotelephony: an international study".

[24] ETS 300 511: "European digital cellular telecommunications system (Phase 2); Man-Machine Interface of the Mobile Station".

[25] ITU-T Recommendation F.720 (1992): "Videotelephony services - General".

[26] ITU-T Recommendation F.721 (1992): "Videotelephony teleservice for ISDN".

[27] ITU-T Recommendation F.730 (1992): "Videoconference service - General".

[28] ITU-T Recommendation H.231 (1993): "Multipoint control units for audiovisual systems using digital channels up to 2 Mbit/s".

[29] ITU-T Recommendation H.243 (1993): "Procedures for establishing communication between three or more audiovisual terminals using digital channels up to 2 Mbit/s".

[30] ITU-T Recommendation H.230 (1993): "Frame-synchronous control and indication signals for audiovisual systems".

[31] CEPT Recommendation T/N 31-03 (1988): "Multipoint International Videoconference System".

[32] ETF News (1992): "The official magazine of the European Teleconferencing Federation", Issue 1.

3 State of the art

This clause is a brief review of Human Factors studies on user procedures for multipoint videocommunication and of the work that has been done or is under way on this issue within standardization bodies, such as ETSI and the ITU-TS.

3.1 Introduction

Multipoint telecommunication has been available for a long time in the form of telephone conferencing. Human Factors studies on telephone conferencing dealt with basic interaction issues of pure-audio telecommunications as well as with the impact of various technical parameters on the acceptance by the user. On the whole, a comparatively strict conversational discipline seems to be required in the course of telephone conferencing. This is particularly true in the case of larger group sizes and in the case of participants unacquainted with each other (see "Interactions patterns in audio teleconfering" by J. Carey [1]). Speaker identification sometimes causes difficulties; various technical means have been developed in order to support the conferees in coping with this problem (e.g. "The Social Psychology of Telecommunications" by Short et al., 1976 [2]). That this form of multipoint telecommunication has nevertheless hardly been taken up, is supposed to be due to poor Human Factors. Examples are given in that the set-up procedures for telephone conferencing demand too much effort by the user, that it is difficult to work with documents in pure audio-conferences and that it is not possible to see the interlocutors (e.g. Short et al., 1976 [2]). Specifically of the latter, one can only "request the floor" by verbal interjection.

At least since the early 1980s, multipoint videoconferencing has been discussed internationally (Kelly, 1982 [3]; Nagra et al. [4], 1984; Sabri & Prasada, 1985 [5]; Kenyon et al., 1985 [6]). Experimental systems and demonstrators have been developed (e.g. Chiariglione & Corgnier, 1984 [7]; Sawada et al., 1985 [8]). First services have been available and first Human Factors studies have been carried out. Besides socio-psychological issues (Coll et al., 1975 [9]; Birrell & White, 1982 [10]), these studies focused mainly on issues concerning the visual presentation of the conferees located at different sites. Among other things, it became apparent that systems offering a continuous visual presence of all participants have several advantages as compared with systems that allow the presentation only of the images of one of the conferees at a given time (Hoecker et al., 1978 [11]; Romahn & Mühlbach, 1988 [12]).

For accomplishing continuous visual presence, several versions of a split-screen configuration have been discussed. Among these the following two can be considered the most convincing ones for an up to five-point conference:

a) The screen is split-up into four equal sized windows, displaying up to four remote participants in medium reduced size as compared with a full-screen presentation (e.g. Klein, 1985 [13]).

b) The screen is split-up into one large window and several small windows, displaying one participant in full size and the others in small size (see figure 1 below, from Romahn & Mühlbach, 1988 [12]).

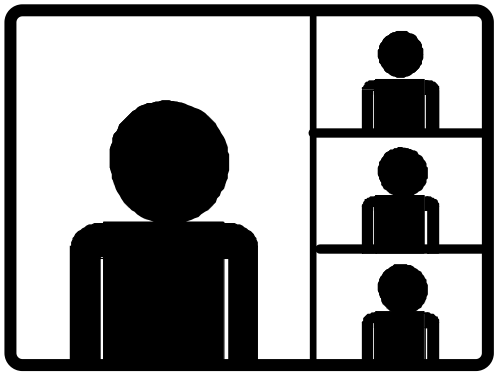

Figure 1: Split-screen presentation utilizing one large and several small windows

Because the video mixing function for split-screen techniques is considered to be a complex process from a technical point of view, procedures for the switched mode have been discussed since the time when the first systems emerged (e.g. Nagra et al., 1984 [4]). Most of the proposals date back to experimentations by AT&T with a voice-controlled video switch in the 1970s. According to the switching criterion developed by AT&T, the choice of video is automatic on the basis of the dominant sound level, such that the present speaker receives the picture of the previous speaker, while all other terminals receive the picture of the present speaker (cf. e.g. Short et al., 1976 [2]).

Since the late 1980s, the idea of integrating multipoint videocommunication and multimedia document processing has emerged (e.g. Addeo et al., 1987 [14]). Workstation and PC based prototypes for holding conferences by interchanging information through voice, video, and multimedia documents employing windowing systems were developed (e.g. Watabe et al., 1990 [15]), mainly in order to support CSCW1 **3)** applications. One example is MIAS **4)** (cf. e.g. Furner et al., 1990 [16]; 1991 [17]). Another example is the TELES.VISION system, a PC-integrated videoconferencing system for up to eight subscribers (Schindler, Heidebrecht, 1991 [18]).

The possibility of holding multipoint videotelephone conferences during which the participants can stay at their places of work, which is obvious when employing videophones as terminals, can be considered particularly attractive (Klein, 1985 [13]). Among other reasons, this is due to the fact that efforts for reserving and reaching videoconferencing studios are not necessary. But above all, these "workstation" or "desktop conferences" enable the participants to have immediate access to supplementary information, which is stored in personal filing systems or on a PC. With regard to the group of participants, multipoint videotelephony allows flexible management. Thus, for example, experts can temporarily join a multipoint call in order to deal with particular problems, or sub-conferences can be held for confidential matters.

As far as set-up procedures for establishing multipoint calls are concerned, the following can be found in the literature:

- "add-on": the initiator of a multipoint call sets up a link to a conference server and dials in the other participants;

- "meet-me": the participants of a multipoint call dial a conference server and are switched in if they have been specified in advance;

- "auto-originate": at a pre-programmed commencing time, the participants of a multipoint call are notified by a conference server.

Switching the video signals within the framework of switched multipoint videotelephony can be controlled in various ways:

a) voice controlled (i.e. on the basis of the dominant sound level);

b) chairperson controlled;

c) controlled by each user.

Whereas a) is done automatically, b) and c) are considered to be accomplished manually. In the case of a) and b), it is usually envisaged to broadcast the video signal of one of the terminals to all other terminals **5)**; c) can be implemented in either a broadcast or a non-broadcast mode. Adopting the non-broadcast or "autonomous" user controlled mode enables each user to select autonomously the video signal he/she wants to be displayed on his/her terminal, without influencing what is to be displayed on the other terminals **6)**.

---

3) CSCW stands for "Computer-Supported Co-operative Work".
4) Multipoint Interactive Audiovisual System
5) In the case of a voice-controlled video switch it is usually envisaged to broadcast the video signal of the current speaker to all other terminals, while the current speaker receives the image of the previous speaker.
6) If MCUs are connected in tandem the user can choose only from a limited sub set of pictures.

3.2 Human factors research

Without claiming to be exhaustive, some of the Human Factors studies whose outcomes can be considered important regarding user procedures for multipoint videotelephony are reported in the following.

Heinrich-Hertz-Institut

Heinrich-Hertz-Institut (HHI) has dealt with Human Factors aspects of videotelephony and multipoint videoconferencing for several years (e.g. Mühlbach, 1988 [19]; Romahn & Mühlbach, 1988 [12]; Mühlbach, 1990 [20]). The main objective of the work has been to develop recommendations on services and terminals, based on usability tests. The technical approach consists in allowing test subjects, who have to carry out realistic communication tasks, try out and assess laboratory system prototypes. Adopting this approach, three Human Factors studies on multipoint videocommunication have been carried out:

A first study concerned issues such as how to realize multipoint videocommunication systems in terms of technology, whether successful and satisfactory communication is feasible by means of such systems and what kind of user procedures for controlling the video signals should be implemented (Romahn & Mühlbach, 1988 [12]). Within the framework of a Human Factors experiment, several versions of a laboratory multipoint videoconferencing system were compared with each other. Among other things, these versions differed in terms of user procedures for controlling the video image.

By means of a laboratory five-point videoconferencing system, which allowed continuous visual presence of all participants utilizing a split-screen technique with one large window and three small windows on each terminal (see subclause 3.1, figure 1), different procedures for controlling the mix of the video signals were implemented and tested:

1) "Direct access" where each remote site was represented by a specific key which enabled the user to move the desired image into the large window by pressing the corresponding key;

2) "Sequential access" where the subscriber images were rotated clockwise through each of the display windows by pressing a "rotation key " several times until the desired image appeared in the large window.

Whereas the direct access as well as the sequential access allowed for autonomous control (i.e. each user was able to control his/her individual mix of video signals without influencing the displays of the other users), a chairperson mode, which was also evaluated as being within the scope of the Human Factors experiment, enabled the appointed chairperson of the conference to control the image to be displayed within the large windows of all subscriber terminals.

A second Human Factors study within the framework of multipoint teleconferencing was aimed at establishing the supplemental benefit of moving video images as compared with still video images and with pure-audio (Romahn & Mühlbach, 1988 [12]). Test subjects used and assessed one of the five versions of a multipoint laboratory system. Pure audio conferences, still picture conferences and various types of videoconferences were tested. Whereas two versions of video conferences adopted the switched mode, a third version allowed for the continuous visual presence of all participants.

The third HHI study on multipoint matters was aimed at developing user-friendly set-up procedures for establishing multipoint calls and for establishing sub-conference calls within the framework of multipoint videocommunication. In addition to that, it was studied how these procedures can be supported by visual indications appearing on the screens of the subscribers' terminals (Mühlbach, 1990 [20]).

Two different set-up procedures were tested:

The "dynamic add-on procedure" allowed for a dynamic insertion of participants during the course of a multipoint call, starting from a point-to-point connection. As each subscriber had the means of "adding-on" additional participants, the role of the "convenor" was not fixed. Adding participants was accomplished by first establishing a "side-call" to a new party during which the existing call was placed on hold. When the new subscriber agreed to join the conference call he/she was inserted.

The "auto-originate procedure" allowed for automatically establishing a multipoint call, pre-programmable by the user acting as the convenor.

A sub-conference call was established in a way that on the one hand the participants of the sub-conference and on the other hand the remaining participants of a multipoint call were interconnected separately via audio and video.

Considering the results of the above mentioned studies conducted by HHI the following can be said with regard to user procedures.

If a switched mode is adopted within the course of a multipoint videoconference or a multipoint videotelephone call, the selection of the video signal to be displayed should be controllable autonomously by each user. As compared with a chairperson controlled video switch, an autonomous switching by individual users has the advantage that each user can have a look at the participant in whose non-verbal signals he/she is personally interested, independently of which participant is considered important by a chairperson and independently of who is speaking. The will of the person who watches a display is respected when the receiving side can control the party to be displayed (Kishino et al., 1984 [21]). An autonomous switching by individual users can be controlled manually. As a result of the HHI studies, both a "direct access procedure" as well as a "sequential access procedure" have proved to be feasible in terms of user-friendliness. If showing and discussing documents captured by a document camera is in the focus of a multipoint call, a user procedure for broadcasting the video signal can be desirable. According to HHI's outcomes, such a possibility should not be technically restricted to a formally appointed chairperson but should be provided to each user.

With regard to set-up procedures for multipoint calls, the HHI studies have revealed that both, a set-up procedure in which, starting from a point-to-point connection, the dynamic insertion of additional participants is possible by direct dialling (dynamic add-on), as well as a set-up procedure in which a pre-programmable automatic convening is made available to the user (auto-originate) are desirable from a Human Factors point of view.

RACE

Within the scope of the RACE project R1065 (ISSUE) Human Factors studies were carried out on video switching modes for multipoint videotelephony (Cornacchia & Papa, 1992 [22]). Adopting an autonomous video switching mode, two control procedures using a mouse were examined in a laboratory experiment: picture icons versus a pull-down menu. In this experiment three workstations placed in three different rooms were linked together in a network that emulated broadband capabilities. In the picture icon condition test subjects had to control the video switching by clicking one of the presented picture icons, which were video signals (e.g. portraits and documents of other participants) "frozen" into icons. In the pull-down menu condition the video switching was performed by selecting the desired video signal from a menu bar. The results indicated that the picture icon switching control mode was strongly preferred to the pull-down menu mode.

Another study within the scope of the ISSUE project focused on icons for multipoint videotelephony (Gili et al., 1991 [23]). The main objective of the study was to test and to evaluate the suitability of pictograms for function keys and status indications. The evaluation was performed by means of a questionnaire with the help of a video tape. Recommendations were reported regarding, among other things, preferred icons for function keys such as "rotate" and "multipoint mode".

The RACE project R2025 (MIMIS) addresses multipoint multimedia communication. One of the project's major objectives is to develop a protocol infrastructure for this service, to be implemented on a demonstrator. The development of user procedures is considered to be an integral part of the project. MIMIS intends to implement two modes of video switching on its demonstrator: an automatic selection of the conferees that are to be displayed (e.g. a voice-controlled video switch) and a manual mode that allows users to make their own selection. It is intended to evaluate these procedures during the project [7)].

---

7) Personal communication from Mr. Coolegem (project manager of R2025). Mr. Coolegem mentioned also that it might be possible to run a comparison of the voice-controlled video switch versus the manual switch during 1993, provided that the procurement of the necessary equipment is successful.

In addition to that, multipoint applications are addressed by other RACE projects as well (e.g. R2008: EuroBridge).

3.3 Standardization activities

ETSI

Inquiries within ETSI revealed that, although a demand for multipoint videotelephony is taken for granted, activities on this issue are in the embryonic stages. For instance, not even a Service Description for multipoint videotelephony exists. According to the current state of information, STC HF1 seems to be the only ETSI Committee dealing explicitly with multipoint videotelephony.

TE10 (which was the AVM) created a work item, now dropped in favour of a TE4 work item, on "Multipoint for audiovisual services". The aim of this work item was to identify technical problems in multipoint configurations for audio visual services with priority to conversational services.

ETS 300 511 [24] "Man-Machine Interface of the Mobile Station (MS) (GSM 02.30)" was adopted in 1994. Although multipoint videotelephony is not considered within the scope of this ETS some interesting proposals on handling multiparty services can be found. So, for instance, it is proposed that a multiparty conversation should be established by first establishing calls to two parties (with one call active and the other on hold) and then bringing the three parties together. To add another remote party, the same procedure applies.

CCITT [8)]

During the study period 1989-1992, CCITT Study Group I approved, among other things, ITU-T Recommendations F.720 [25], F.721 [26] and F.730 [27].

According to ITU-T Recommendation F.720 [25], videotelephony conferencing is regarded as a supplementary service within videotelephony services. The procedures and functions for videotelephone conference calls are characterized as being for further study. Listed as principle applications of videotelephony are - besides face-to-face dialogues - "dialogue including interactive viewing of documents", "communication between hearing and speech impaired persons using the sign language" etc..

The purpose of ITU-T Recommendation F.730 [27] is to define and describe the general features and attributes of videoconference services. Within this Recommendation "the Videotelephone teleservice" for conferencing is mentioned, but ITU-T Recommendation F.730 [27] itself does not deal with this service. However, it is recommended that a videoconference service should be able to intercommunicate with videotelephony services. The definition of the videoconference service covers also more than two separate locations, i.e. multipoint connections are part of the general service description. According to the description, all terminals in a multipoint videoconference are connected via at least one Multipoint Control Unit (MCU). Some MCUs may use split-screen techniques to allow the continuous and simultaneous perception of the images of all participants at each terminal. Other MCUs may operate according to the switched mode, where the decision about the source signal is controlled by a chairperson or on the basis of the dominant sound level. With regard to conference management procedures, the unconducted mode, in which none of the connected terminals have priority, is considered the basic procedure. The conducted mode or chairman mode allows to rule the videoconference with procedures like "asking for the floor", "giving the floor", etc.. Pre-booking is considered the normal procedure of call establishment for the videoconference service. However, videoconference calls may additionally be established by direct dialling if this facility is offered by the service provider (see ITU-T Recommendation F.730 [27]).

Study Group I has decided to treat multipoint videotelephony calls as a new work item in the next study period [9)].

Within Study Group XV (question 4/XV), (ITU-T Recommendation H.231 [28] and ITU-T Recommendation H.243 [29]) have been adopted. Issues of multimedia multi-point connections and of a generic numbering

---

8) now ITU-TS
9) Personal communication from Mr. Matsumoto (Special Rapporteur for audiovisual services within ITU-T Study Group I)

scheme that can be used in multipoint conferencing systems are included. Question 10/VIII (in Study Group VIII) deals with a protocol stack (T.120 series) for multipoint multimedia calls, applicable to real-time audiovisual services including videotelephony.

With regard to user procedures for switching the video signals within the framework of a switched multipoint videotelephony call, the description of the Video Processor Unit of an MCU seems to be most interesting. So, according to ITU-T Recommendation H.231 [28], the video processor can operate in both, a "switched" as well as a "mixed" mode. However, it is indicated that since the video mixing function is a complex process, the alternative of video switching may be preferred. The choice of video may be automatic, such that the present speaker receives the picture of the previous speaker, while all other terminals receive the picture of the present speaker. However, ITU-T Recommendation H.231 [28] contains the possibility of overriding the automatic mechanism, e.g. for controlling the video switching by a chairperson.

ITU-T Recommendation H.230 [30] defines a number of control and indication signals, including those used in multipoint communication. According to ITU-T Recommendation H.230 [30], it is technically feasible to broadcast a video signal (e.g. to transmit the picture of a chairperson or to hold a picture source during the transmission of graphics) within the course of a multipoint conference.

CEPT

CEPT (Sub-Working Group NA3) has already dealt with multipoint issues in the 1980s. It prepared a recommendation on "Multipoint International Videoconference System", comprising, inter alia, statements on how to switch images within the course of a multipoint videoconference (CEPT T/N 31-03 [31]). CEPT took up the concept of a voice-controlled video switch as a "minimum working mode". Several manual overrides of this mode are identified. Among these are:

- each location has the choice among the channels without affecting the displays of other locations;

- complete chairman control;

- "visualisation forcing" (one location may force the MCU to send its video signal to the other locations, i.e. "broadcasting" the video signal).

3.4 Systems in use

There have been commercial multipoint services since the middle of the 1980s; now there are quite a lot of products (e.g. BT/GPT).

British Petroleum (BP) was among the first private users of videoconferencing, installing a point-to-point service in the UK in the early 1980s. In 1987 multipoint conferencing, enabling a number of sites to link into a single video meeting, was installed (ETF News, 1992 [32]).

Rank Xerox is using a Multipoint Control Unit in order to connect up to 14 sites simultaneously via its permanent videoconferencing network Xerox Team Vision. Each site can utilize voice activated switching to allow the site "talking" to be seen by others (ETF News, 1992 [32]).

3.5 Résumé

User procedures for setting-up a multipoint videophone call

Until now, only a small number of Human Factors studies have been carried out on user procedures for setting-up a multipoint videophone or videoconferencing call. These studies revealed that both "dynamic add-on" procedures as well as "auto-originate" procedures are desirable from a Human Factors point of view. Dynamic add-on procedures allow for a dynamic insertion of participants during the course of a multipoint call, starting from a point-to-point connection. Auto-originate procedures allow of automatically setting-up a multipoint call, pre-programmable by the user who is acting as the convenor. It is interesting that procedures for dynamic add-on are under discussion also with regard to multiparty services of mobile stations.

User procedures for switching the video signals

The following user procedures for switching the video signals in the case where a switched mode is adopted are under study or in use:

a) voice controlled (i.e. on the basis of the dominant sound level);

b) chairperson controlled;

c) controlled by each user.

Whereas the voice controlled video switch seems to be widely acknowledged as a basic or fall back mode by standardization bodies, Human Factors studies very often point out its disadvantages and recommend an autonomous mode, in which each user is able to select the video image he/she prefers. Apart from reliability problems it is mentioned that a voice-controlled video switch does not sufficiently meet Human Factors requirements in that, for instance, during the course of a conversation non-verbal signals from the "non-speakers" can be equal or even more important for a listener than those of the current speaker. When showing documents by means of a document camera, a voice-controlled switch can cause problems when an interlocutor asks a question referring to the displayed document and (due to the voice signal) another picture is switched on. Another argument against a voice-controlled video switch is, that this procedure may be difficult to be used by deaf videophone users using sign-language. However, it needs to be borne in mind that for the time being a controlled Human Factors experiment in which a voice-controlled video switch was compared with other switching modes, such as autonomous modes does not exist.

User procedures for controlling the mixture of video signals

User procedures for controlling the mix of the video images in case where such a control is possible or required when adopting a split-screen technique have been studied within the framework of some Human Factors experiments. Results indicate that it is possible to adopt to a far extent the same principles as for user procedures for switching the video signals within the framework of the switched mode. So, for example, moving the image of a particular participant into a specific window can be accomplished autonomously by each user or by adopting a chairperson mode.

As a final statement, regrettably, there are, for the time being, neither reports of Human Factors studies nor recommendations drafted by standardization bodies that deal explicitly with requirements of People with Special Needs with regard to user procedures for multipoint videotelephony.

4 Preliminary recommendations on user procedures

This clause contains preliminary recommendations on user procedures for:

- setting-up multipoint videophone calls;

- switching the video signals within the framework of switched multipoint videotelephony;

- controlling the mixture of video signals within the framework of mixed multipoint videotelephony.

These recommendations are based on the state of the art review (see clause 3) and are reflecting the state of discussion within ETSI TC-HF at the time of writing. In particular, with regard to the recommendations on user procedures for switching the video signals, it needs to be borne in mind that for the time being no Human Factors experiment exists in which a voice-controlled video switch has been compared with a user-controlled autonomous switching mode.

4.1 Procedures for setting-up multipoint videophone calls

Basic procedure

The basic procedure for setting-up multipoint videophone calls by direct-dialling is the following ("dynamic add-on procedure").

Starting from a point-to-point connection, each subscriber has the means of adding-on additional participants. The same applies to any n-point-connection (n > 2), which allows for a dynamic insertion of participants during the course of a multipoint call.

This principle implies that the role of the "convenor" of a multipoint videophone call is not fixed.

Adding additional participants is accomplished by first setting-up a "side-call" to a new subscriber during which the actual (point-to-point or multipoint) call is placed on hold. When the new subscriber has agreed to join the conference call he/she is inserted after the "acting convenor" confirms this to the MCU.

Additional procedure

In addition to the basic procedure, the following procedure ("auto-originate procedure") is desirable.

A subscriber - acting as the convenor - indicates to an MCU that he/she is intending to set-up a multipoint videophone call. Then the convenor inputs the videophone numbers of the desired participants of the conference call and the commencement time (which can be a pre-booking date or "immediately"). The input can be accomplished, for example, by filling in a form that is generated by the MCU on the screen of the convenor's terminal. At the commencement time the desired participants are rung up by the MCU and are inserted into the conference call as soon as they answer the (video-)phone. Connecting the desired subscribers automatically (i.e. without waiting for them to answer) should not be possible. In order to give the subscribers who have not answered the alerting signal in time, the chance to join the conference call later or a call-waiting service should be provided.

The adding-on of additional subscribers who have not been specified in advance is accomplished by adopting the basic procedure (see above).

4.2 Procedures for switching the video signals

4.2.1 Basic working mode

Within the framework of switched multipoint videotelephony the basic working mode is the following (non-automatic) one.

Each user is able to autonomously select the video image he/she prefers to be displayed on his/her screen (without influencing what is to be displayed on the other terminals).

User procedures for the autonomous switching mode

Autonomous switching according to the basic working mode can be implemented by adopting various user procedures.

Selection of the preferred video image can be accomplished **manually** by

a) direct/random access procedures where each remote site is represented by a specific instance (e.g. push-button or icon) on the terminal enabling the user to get the desired image by activating (e.g. pressing or clicking) the corresponding instance;

b) sequential access where the subscriber images are switched though sequentially by activating a specific instance (e.g. pressing a push-button or clicking an icon) on the terminal several times until the desired image appears on the screen.

In addition to manual procedures it is conceivable to control the autonomous video switching by spoken keywords (i.e. **speech input**) or other means.

When the user's choice is not available (which can be the case if MCUs are connected in tandem) appropriate feedback messages should be displayed.

4.2.2 Additional modes

The following additional modes are optional; in various applications they can be considered desirable.

- a voice-controlled video switch where the choice of video is automatic, such that the present speaker receives the picture of the previous speaker, while all other terminals receive the picture of the present speaker **10)**;

- broadcasting, i.e. transmitting the video signal of one of the terminals to all other terminals simultaneously **11)**.

The additional modes are not considered override modes in the sense of switching off the basic working mode. So, for instance, in cases where a voice-controlled mode is adopted, each user should have the chance to select another video signal than that of the current speaker.

4.3 Procedures for controlling the mixture of video signals

User procedures for controlling the mix of the video images in cases where such a control is possible or required **12)** should adopt as far as possible the same principles as for user procedures for switching the video signals within the framework of the switched mode. This implies the following recommendation with regard to a basic working mode for mixed multipoint videotelephony:

Each user is able to select autonomously the video image he/she prefers to be displayed in a specific window on his/her screen (i.e. without influencing what is to be displayed on the other terminals).

The switching can be implemented by adopting the same user procedures as being recommended for the switched mode (see subclause 4.2).

In addition to that, the user should be able to switch on and off a kind of "switched mode" within the framework of mixed multipoint videotelephony, i.e. a mode where only one of the other subscribers' images is presented at a given time **13)**. User procedures for accomplishing the switching within this mode should be same as within switched videotelephony (see subclause 4.2).

5 Recommendations for further research

1) The most urgently needed research activity with regard to user procedures for multipoint videotelephony seems to be a Human Factors experiment aiming at comparing an automatic voice-controlled video switching mode with other switching modes. Such a kind of experiment should take into account various applications and user profiles (e.g. People with Special Needs).

2) Another desirable experiment is one that compares various versions of split-screen arrangements within the framework of mixed multipoint videotelephony in terms of pros and cons (e.g. a "quarter-split" arrangement consisting of four fixed equal-sized windows vs. versions utilising windows of different size in flexible arrangements).

---

10) The voice-controlled video switching mode may be suitable in situations where
a) it is not necessary to show documents
b) non-verbal signals of "non-speakers" are of minor interest to the interlocutors, and
c) users don't want to spend effort on manual switching.

11) The broadcasting mode may be suitable in particular in situations where showing and discussing documents captured by a document camera is in the focus of a multipoint call. For selecting the video signal to be broadcast, various control procedures are conceivable (cf. clause 3). However, results of Human Factors studies indicate that selection should not be technically restricted to a formally appointed chairperson but should be provided to each user.

12) Procedures for controlling the mix of the video signals may be required in cases where, for instance, the screen is split-up into windows of different sizes and/or where it is desirable to move subscriber images from one window to another. Procedures for controlling the mix of the video signals may not be required in cases where the screen for displaying the subscriber images is split-up into four equal sized windows, allowing of presenting up to four images in fixed positions.

13) Despite the fact that from a Human Factors point of view continuous visual presence has proved to by desirable in multipoint videotelephony, there might occur situations or periods in which the user prefers to use the entire screen for presenting a particular image (e.g. to enhance the legibility of a document or to avoid being disturbed by other images).

3) For enlarging the recommendations in hand it might be desirable to study user procedures for setting-up and holding sub-conferences within the course of a multipoint videophone call.

4) Research activities that should be joint activities of TC-HF and other ETSI TCs (e.g. Terminal Equipment (TE)) should study some technical issues that turned out to be important when finalising recommendations on user procedures for multipoint videotelephony with regard to ISDN. These issues are, among other things, how multipoint videotelephony with continuous visual presence can be realized via ISDN and to investigate the effects of various (automatic or user-controlled) modes for switching video signals on coding mechanisms.

History

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

ISBN 2-7437-0045-9
Dépôt légal : Février 1995