EUROPEAN TELECOMMUNICATION STANDARD

DRAFT
pr ETS 300 807
August 1996

Source: ETSI TC-TE

Reference: DE/TE-04043

ICS: 33.080

Key words: audio, ISDN, terminal, video

Integrated Services Digital Network (ISDN); Conference Services; Conference terminal video and audio characteristics

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 draft European Telecommunication Standard (ETS) has been produced by the Terminal Equipment (TE) Technical Committee of the European Telecommunications Standards Institute (ETSI), and is now submitted for the Public Enquiry phase of the ETSI standards approval procedure.

Proposed transposition dates	
Date of latest announcement of this ETS (doa):	3 months after ETSI publication
Date of latest publication of new National Standard or endorsement of this ETS (dop/e):	6 months after doa
Date of withdrawal of any conflicting National Standard (dow):	6 months after doa

Blank page

1 Scope

This European Telecommunication Standard (ETS) specifies the audio characteristics of terminals designed to support the audiographic conference teleservice as specified in draft prETS 300 675 [1]. The same audio requirements of this ETS are also applicable to terminals supporting the Integrated Services Digital Network (ISDN) videoconference teleservice.

This ETS does not specify the terminal procedures, both with respect to in-band signalling and to ISDN signalling on the D channel. Also the procedures and protocols for data exchange and conference control are outside the scope of this ETS.

2 Normative References

This ETS incorporates by dated and undated reference, provisions from other publications. These normative 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 ETS only when incorporated in it by amendment or revision. For undated references the latest edition of the publication referred to applies.

[1] draft prETS 300 675: "Integrated Services Digital Network (ISDN); Audiographic conference teleservice; Service description".

[2] I-ETS 300 245-2 (1996): "Integrated Services Digital Network (ISDN); Technical Characteristics of Telephony Terminals, Part 2: PCM A-law, handset telephony".

[3] I-ETS 300 245-3 (1995): "Integrated Services Digital Network (ISDN); Technical Characteristics of Telephony Terminals, Part 3: Pulse Code Modulation (PCM) A-law, Loudspeaking and Handsfree telephony".

[4] I-ETS 300 245-5 (1996): "Integrated Services Digital Network (ISDN); Technical Characteristics of Telephony Terminals, Part 5: Wideband (7 kHz) Handset Telephony".

[5] I-ETS 300 245-6 (1996): "Integrated Services Digital Network (ISDN); Technical Characteristics of Telephony Terminals, Part 6: Wideband (7 kHz) Loudspeaking and Handsfree telephony".

[6] I-ETS 300 245-8 (1996): "Integrated Services Digital Network (ISDN); Technical Characteristics of Telephony Terminals, Part 8: Speech transmission characteristics when using Low Delay Code-Excited Linear Prediction (LD-CELP) coding at 16 kbit/s".

[7] I-ETS 300 302-1 (1996): "Integrated Services Digital Network (ISDN); Videotelephony teleservice: Part 1: Electroacoustic characteristics for 3,1 kHz bandwidth handset terminals".

[8] I-ETS 300 302-2 (1995): "Integrated Services Digital Network (ISDN) Videotelephony teleservice: Part 1: Electroacoustic characteristics for 3,1 kHz bandwidth loudspeaking and handsfree terminals".

[9] I-ETS 300 302-3 (1996): "Integrated Services Digital Network (ISDN); Videotelephony teleservice: Part 3 Audio aspects-Wideband and Handset".

[10] I-ETS 300 144 (1996): "Integrated Services Digital Network (ISDN); Audiovisual services, Frame structure for a 64 to 1 920 kbit/s channel and associated syntax for in-band signalling".

[11] I-ETS 300 143 (1994): "Integrated Services Digital Network (ISDN); Audiovisual services, In-band signalling procedures for audiovisual terminals using digital channels up to 2 048 kbit/s".

[12] CCITT Recommendation G.711 (1988): "Pulse code modulation (PCM) of voice frequencies".

[13] CCITT Recommendation G.722 (1988): "7 kHz audio coding within 64 kbit/s".

[14] CCITT Recommendation G.728 (1992): "Coding of speech at 16 kbit/s using low-delay code-excited linear prediction".

[15] CCITT Recommendation G.725 (1988): "System aspects for the use of the 7 kHz audio codec within 64 kbit/s".

[16] CCITT Recommendation P.57 (1992): "Artificial Ears".

[17] ITU-T Recommendation P.10 (1993): "Vocabulary of terms on telephone transmission quality and telephone sets".

[18] ITU-T Recommendation G.701 (1993): "Vocabulary of digital transmission and multiplexing and pulse code modulation (PCM) terms".

[19] ITU-T Recommendation P.51 (1993): "Artificial Mouths".

[20] ITU-T Recommendation P.79 (1993):"Calculation of loudness ratings for telephone sets".

[21] ITU-T Recommendation P.64 (1993): "Determination of sensitivity/frequency characteristics of local telephone systems".

[22] CCITT Recommendation P.76 (1988): "Determination of loudness rating; fundamental principles".

[23] TBR 3 (1995): "Integrated Services Digital Network (ISDN); Attachment requirements for terminal equipment to connect to an ISDN using ISDN basic access".

[24] ISO 3 (1973): "Preferred numbers-Series of preferred numbers".

[25] ITU-T Recommendation P.34 (1993): "Transmission characteristics of hands free telephones".

[26] ITU-T Recommendation G.122 (1993): "Influence of national systems on stability talker echo in international connections".

[27] IEC Publication 651 (1979): "Sound level meters".

[28] ITU-T Recommendation P.31 (1993): "Transmission characteristics for digital telephones".

[29] ITU-T Recommendation P.50 (1993): "Artificial Voices".

[30] CCITT Recommendation G.101 (1988): "The transmission plan".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of this ETS, the definitions provided in the referenced standards and the following definitions apply:

Acoustic Reference Level (ARL): Acoustic level which gives -10 dBm0 at the digital interface.

audiographic terminal: Terminal supporting the audiographic teleconference service.

Hands Free Reference Point (HFRP): A point located on the axis of the Artificial Mouth, at 50 cm from the lip ring, where the level calibration is made in free field. It corresponds to the measurement point n.11, as defined in ITU-T Recommendation P.51 [19].

reference sphere: Sphere of radius 1 metre where the anechoic conditions of the acoustic testing environment are verified.

lip synchronization delay: The delay introduced in the sending and receiving audio paths in order to align the audio signals with the moving pictures respectively transmitted and received by the terminal.

digital interface: For the purposes of this ETS, the digital interface refers to the B channels available at the coincident S and T reference points at an ISDN basic access.

3.2 Abbreviations

For the purposes of this ETS, the abbreviations used in ITU-T Recommendations G.701 [18], P.10 [17], P.51 [19], P.57 [16], P.64 [21], P.76 [22] and P.79 [20] and the following abbreviations apply:

ARL	Acoustic Reference Level
ERP	Ear Reference Point
F_r	Correction factor for receiving measurements (annex A, subclause A.1.1.2.2)
F_s	Correction factor for sending measurements (annex A, subclause A.1.1.2.1)
F_{tcl}	Correction factor for terminal coupling loss measurements (annex A, subclause A.1.1.2.3)
HFRP	HandsFree Reference Point
ISDN	Integrated Services Digital Network
LD-CELP	Low Delay-Code Excited Linear Prediction
MRP	Mouth Reference Point
PCM	Pulse Code Modulation
PSTN	Public Switched Telephone Network
rms	root mean square
RLR	Receiving Loudness Rating
SB-ADPCM	Sub Band-Adaptive Differential Pulse Code Modulation
S/D	Signal to Distortion
SLR	Sending Loudness Rating
TCL	Terminal Coupling Loss
TCLw	Weighted Terminal Coupling Loss
TEUT	Telephone Equipment Under Test

4 System description

For the purposes of this ETS a teleconference terminal can be represented by the functional block diagram shown in figure 1.

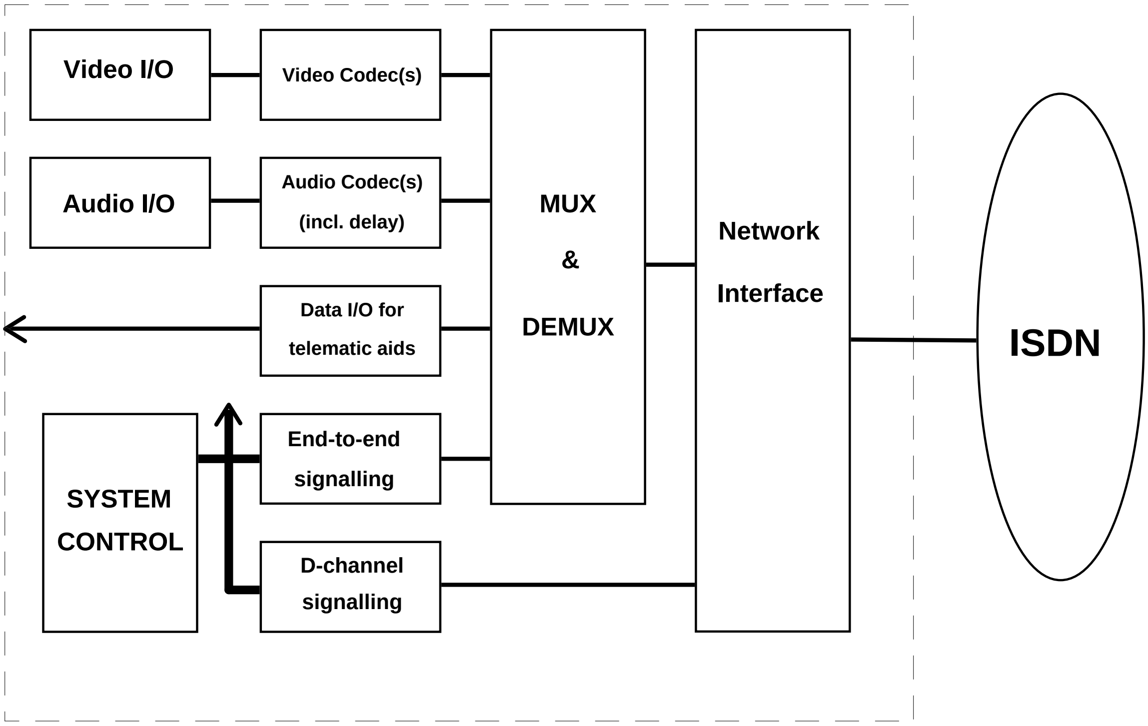

Figure 1: Functional block diagram of a teleconference terminal

The following units compose a teleconference terminal:

- video I/O-includes camera(s), monitor(s) and video processing units providing functions like split screen, picture superposition (e.g. for combining the pictures generated by telewriting and still pictures) and text generation (e.g. for supporting conference management and text processing features);

- audio I/O-inclusive of handset/headset and handsfree facilities, both for single users and for multiple users;

- data I/O-used for interfacing telematic aids to the teleconference terminal (e.g. external PCs, fax machines, telewriting facilities, electronic blackboards, etc.);

- system control-carrying out the following functions: Network access through user-to-network signalling, end-to-end in band signalling, control of audio and video codecs, control of data I/O facilities, human interface with the user;

  NOTE 1: End-to-end signalling is defined in I-ETS 300 143 [11] and I-ETS 300 144 [10].

- audio codec(s)-performing CCITT Recommendations G711, G722 and G728 coding according to in band and D-channel signalling information;

- video codec(s)-performing moving and still picture coding according to the terminal characteristics;

- mux/demux unit-multiplexes transmitted video, audio, data and control signals into a single bit stream and demultiplexes the received bit stream, as defined in I-ETS 300 144 [10];

- network interface-makes the necessary adaptation between the network and the terminal according to the user-network interface requirements.

NOTE 2: Teleconference terminals using up to 2 B channels can use the ISDN Basic Access, while terminals designed for using more then 2 B channels can either be equipped with more Basic Access interfaces or with a Primary Rate Interface.

The teleconference terminals can be connected through the ISDN both to other terminals, either of the same type or of different types, or to Multipoint Conference Units. The behaviour of the terminals for both type of connections shall be the same.

4.1 Audio facilities

The teleconference terminals are intended in principle for supporting communications between groups of users. However, single-user terminals compliant with this ETS can also be designed for allowing a single-user access to audiographic teleconferencing in order to permit, for instance, single users to enter multipoint conferences or to connect end-to-end to multi-user terminals.

The following range of audio facilities can be provided by teleconference terminals:

- handset (single user);
- headset (single user);
- handsfree (single user);
- handsfree (multiple users).

Some implementations can provide audio facilities to groups of users by supplying each user with a personal hands-free unit. The same requirements applicable to single user handsfree terminals apply to this type of multiple user terminal.

4.2 Audio encoding

Audiographic terminals shall be able to operate both narrow band, CCITT Recommendations G.711 [12] Pulse Code Modulation (PCM) and G.728 [14], and wideband CCITT Recommendation G.722 [13] coding. The default mode for narrow band PCM coding is A-law, however µ-law coding shall also be implemented.

4.2.1 CCITT Recommendation G.711 encoding

4.2.1.1 A-law

At the beginning of a call operation, mode 0F (CCITT Recommendation G.711 [12]) shall be used. The default encoding shall be A-law.

When in mode 0U the requirements of I-ETS 300 245-2 [2] and I-ETS 300 245-3 [3] shall be met for handset and handsfree operations respectively.

4.2.1.2 µ-law

If information is available to the terminal, either by configuration or by user input, as to whether the destination is within a µ-law region, then this encoding law shall be used after the reception of the ALERTING message or, if the ALERTING message is not received, the CONNECT message, or inband signalling, as described in I-ETS 300 245-5 [4] shall be initiated. The information shall be encoded using the µ-law at 64 kbit/s as defined in CCITT Recommendation G.711 [12].

It is the responsibility of the calling terminal to ensure that the correct encoding law is used. If no indication on the coding law has been received during the D-channel signalling sequence or during the in-band signalling sequence, the calling terminal shall use the default coding law while monitoring the statistics of the incoming signal. In order to determine whether the incoming signal was encoded by A-law or µ-law PCM, the algorithm described in appendix 1 to CCITT Recommendation G.725 [15] shall be used. Compliance with the CCITT Recommendation G.725 [15] algorithm implementation shall be checked by the test described in annex A, subclause A.2.7.

For terminals also supporting handset operations, conformance to the µ-law coding requirements shall be checked in the handset mode, as specified in I-ETS 300 245-2 [2].

For terminals not supporting the handset mode the coding requirements shall be checked by using the test methods specified in I-ETS 300 245-3 [3], with the following amendments:

- the test signal generator and analyser shall use µ-law encoding/decoding;
- only sensitivity/frequency, loudness rating, harmonic distortion and noise requirements shall be verified;
- the sending noise shall meet the requirement described in ITU-T Recommendation P.31 [28], Section 4.

4.2.2 CCITT Recommendation G.722 encoding

When operating in modes 1, 2 or 3 the signal shall be encoded as specified in CCITT Recommendation G.722 [13].

The lower sub-band shall be encoded using 6 bits independent of operational mode. When operating in modes 2 or 3, the least significant bit or the two least significant bits shall be used for the auxiliary data channel, see CCITT Recommendation G.722 [13], Section 1.3.

4.2.3 CCITT Recommendation G.728 encoding

When audio coding at 16 kbit/s, the Low Delay-Code Excited Linear Prediction (LD-CELP) speech coding algorithm as defined in CCITT Recommendation G.728 [14] shall be used.

4.3 Audio decoding

4.3.1 CCITT Recommendation G.711 decoding

4.3.1.1 A-law

At the beginning of a call, the operating mode 0F (CCITT Recommendation G.711 [12]) shall be used. The default decoding law shall be A-law.

When in mode 0U the requirements of I-ETS 300 245-2 [2] and I-ETS 300 245-3 [3] shall be met for handset and handsfree operations respectively.

4.3.1.2 µ-law

When a terminal has detected, by D-channel signalling, in-band signalling or by other means, that the incoming bit stream is µ-law encoded, the decoding algorithm shall fulfil the characteristics described in CCITT Recommendation G.711 [12].

For terminals also supporting handset operations, the conformance to the µ-law decoding requirements shall be checked in the handset mode as specified in I-ETS 300 245-2 [2]. For terminals not supporting the handset mode the coding requirements shall be checked by using the test methods specified in I-ETS 300 245-3 [3], with the first two amendments specified in subclause 4.2.1.2 above.

4.3.2 CCITT Recommendation G.722 decoding

The same requirements specified in I-ETS 300 245-5 [4] for wideband handset terminals with respect to the operating mode selection and to fallback procedures to 3,1 kHz telephony and to Public Switched Telephone Network (PSTN) are applicable.

4.3.3 CCITT Recommendation G.728 decoding

When audio coding at 16 kbit/s, the LD-CELP speech decoding algorithm as defined in CCITT Recommendation G.728 [14] shall be used.

> NOTE: The quality provided by CCITT Recommendation G.728 speech coding is not recommended for high quality hands free operations.

4.4 Relative level

The digital interface is defined as a 0 dBr point according to CCITT Recommendation G.101 [30].

5 Speech transmission characteristics

The audiographic terminals shall be able to interwork with PSTN and ISDN telephones, as well as with ISDN videophones and wide band telephones. To this purpose, the audio standards applicable to these devices are also applicable to audiographic terminals as appropriate and specified below.

5.1 Handset mode

The requirements and measurement methods of I-ETS 300 245-2 [2] and I-ETS 300 245-5 [4] apply, respectively for narrow band and wide band operations at an audio bit rate of 64 kbit/s.

For narrow band operations at 16 kbit/s I-ETS 300 245-8 [6] applies.

If lip-synchronization delay is introduced, the measurement method specified by I-ETS 300 302-1 [7] applies.

5.2 Headset mode

In principle the requirements for headset operations shall be based on the requirements for handset operations, with the following allowances:

- the suitable positioning of the microphone with respect to the Mouth Reference Point (MRP) position shall be defined;
- an appropriate Artificial Ear shall be used and the measurement results shall be referred to the Ear Reference Point (ERP) position as specified in ITU-T Recommendation P.57 [16]

> NOTE: The subject is for further study.

5.3 Hands-free mode (single user)

The requirements and measurement methods of I-ETS 300 245-3 [3] and I-ETS 300 245-6 [5] apply, respectively for narrow band and wide band operations at an audio bit rate of 64 kbit/s.

For narrow band PCM coding at 56 kbit/s and LD-CELP coding at 16 kbit/s, I-ETS 300 302-2 [8] applies.

For wide band operations at 56 kbit/s and 48 kbit/s, I-ETS 300 302-3 [9] applies.

5.4 Hands-free mode (multiple users)

Multiple users audiographic terminals can be designed for guaranteed optimum audio performance when installed in "typical" conference rooms (i.e., in "live" environments with acoustic reverberation effects and background noise). To this purpose, suitable signal processing techniques can be used for eliminating reverberation in the transduced signals and/or suppressing room noise. Performance under real operating conditions can then only be evaluated in the actual installation environments, whose characteristics can not be specified in a general way.

This ETS is not intended to cover all the installation conditions of multiple user audiographic terminals, but specifies the audio performances under reference echo-free conditions. In case the specified testing environment is not compatible with the technical characteristics of specific terminals, the manufacturers are given the option to state alternative, more suitable acoustic testing environments.

A testing procedure intended to guarantee the correct audio alignment of actually installed terminals is provided in annex C.

5.4.1 Receiving volume control and sensitivity adjustments

Due to the wide range of operating conditions, both the sending and receiving sensitivity of multiple-user handsfree terminals can generally be adjusted at the installation, based on the room characteristics and operational distance of the transducers. These adjustment controls shall not be accessible to normal users. An additional user-accessible receiving volume control (referred to in the following as "receiving volume control") can also be provided in order to allow the user to establish the optimum listening conditions on the basis of room noise, number of participants and other variable factors.

5.4.1.1 Sending sensitivity adjustment

The sending sensitivity adjustment is intended to allow for varying the positioning of the microphone(s) with respect to the user(s), according to the installation environment constraints. The manufacturer shall state the regulation range of this control and the microphone-speaker operating distance ($\mathbf{d_s}$) for which the terminal submitted to the tests has been regulated.

5.4.1.2 Receiving sensitivity adjustment

The receiving sensitivity adjustment is intended to allow for varying the positioning of the loudspeaker(s) with respect to the user(s) according to the installation environment constraints. The manufacturer shall state the regulation range of this control and the loudspeaker-listener operating distance ($\mathbf{d_r}$) for which the terminal actually submitted to the tests has been regulated.

5.4.1.3 Receiving volume control

Unless stated otherwise, the compliance requirements refer to the maximum position (maximum sensitivity) of the receiving volume control (when manually operated).

The minimum range of the receiving volume control (when manually operated) shall be 15 dB.

The operation of the receiving volume control shall not affect the sending sensitivity. The only user-accessible control allowed for sending, is the "muting" function of the hands-free microphone.

5.4.2 Sensitivity-frequency response

5.4.2.1 Sending

The sending sensitivity-frequency response (from MRP to digital interface) shall fall within the mask specified in table 1.

The mask ordinates are in relative dB units; compliance shall be checked by vertically shifting the mask with respect to the sending characteristic of the terminal under test.

Table 1: Sending sensitivity/frequency mask

Center Frequency [Hz]	Upper limit [dB]	Lower limit [dB]
100	4	-∞
125	4	-7
160	4	-5,5
200	4	-4
250	4	-4
315	4	-4
400	4	-4
500	4	-4
630	4	-4
800	4	-4
1 000	4	-4
1 250	4,6	-4
1 600	5,2	-4
2 000	5,9	-4
2 500	6,5	-4
3 150	7,1	-4
4 000	7,7	-4
5 000	8,4	-4
6 300	9	-7
8 000	9	-∞

NOTE: Under ideal conditions, the speech quality can be improved by transmitting frequencies lower than 125 Hz. However, under normal operating conditions, extending the frequency range to lower frequencies can cause excessive transmission of unwanted noise. Furthermore, a rising sensitivity-frequency response is recommended.

Compliance shall be checked by the tests described in annex A, subclause A.2.1.1.

5.4.2.2 Receiving

The receiving sensitivity-frequency response (from the digital interface to the measurement point) shall fall within the mask specified in table 2.

The mask ordinates are in relative dB units; compliance shall be checked by vertically shifting the mask with respect to the receiving characteristic of the terminal under test.

Table 2: Receiving sensitivity/frequency mask

Center Frequency [Hz]	Upper limit [dB]	Lower limit [dB]
100	8	-∝
125	8	-∝
160	8	-15
200	8	-9
250	8	-6
315	7	-6
400	6	-6
500	6	-6
630	6	-6
800	6	-6
1 000	6	-6
1 250	6	-6
1 600	6	-6
2 000	6	-6
2 500	6	-6
3 150	6	-6
4 000	6	-6
5 000	6	-6
6 300	6	-9
8 000	6	-∝

Compliance shall be checked by the tests described in annex A, subclause A.2.1.2.

5.4.3 Loudness rating

5.4.3.1 Sending

The nominal value of Sending Loudness Rating (SLR) shall be:

$$SLR = (+\ 12 - F_s)\ dB$$

where F_s is as defined in annex A, subclause A.1.1.2.1.

A manufacturing tolerance of ±3 dB is allowed.

Compliance shall be checked by measurement of the SLR as described in annex A, subclause A.2.2.1.

5.4.3.2 Receiving

The nominal value of Receiving Loudness Rating (RLR) shall be:

$$RLR = (+\ 6 - F_r)\ dB$$

where F_r is as defined in annex A, subclause A.1.1.2.2.

A manufacturing tolerance of ±3 dB is allowed.

The RLR_{min} measured with the manual volume control at the maximum position shall be:

$$RLR_{min} = (-\ 4 - F_r)\ dB$$

A manufacturing tolerance of ±3 dB is allowed.

The RLR measured with the volume control at its minimum position shall be by 15 dB to 30 dB quieter (higher) than RLR_{min}.

The nominal value of RLR shall be met (within its tolerance) for at least one setting of the volume control (when manually operated).

For sets only equipped with automatic (receiving) gain control, the RLR measured with an input signal of -15 dBm0 shall be higher by 10 dB to 15 dB than the RLR measured with an input signal of -30 dBm0. The nominal RLR shall be included in the measured range. The RLR measured with an input signal of -30 dBm0 shall be:

$$RLR\ (@\text{-}30\ dBm0) = (-\ 4\text{-}F_r)\ dB.$$

A manufacturing tolerance of ±3 dB is allowed.

Compliance shall be checked by measurement of the RLR as described in annex A, subclause A.2.2.2 with the volume control set as specified.

5.4.4 Terminal coupling loss

5.4.4.1 TCL_w

The Weighted Terminal Coupling Loss (TCL_w), measured from the digital input to the digital output shall be at least 35 dB.

NOTE: The TCL_w requirement specified here refers to stationary single-talk conditions. Reduced performances may occur during the transient switching conditions of the echo control devices and under double-talk conditions. Further information on this subject is available in table 3 of I-ETS 300 245-3 [3].

Compliance shall be checked by the tests described in annex A, subclause A.2.3.1.

5.4.4.2 Stability loss

The attenuation from the digital input to the digital output shall be at least 6 dB at all frequencies in the range from 100 Hz to 8 kHz.

Compliance shall be checked by the test described in annex A, subclause A.2.3.2.

5.4.5 Distortion

The distortion requirement is specified in terms of the total distortion evaluated in the frequency band from 100 Hz to 7 kHz (harmonic and quantizing) measured with input signals of 300 Hz, 1 kHz and 6 kHz.

According to CCITT Recommendation G.722 [13], where the nominal 1 kHz frequency is indicated, the actual measurement frequency shall be 1 020 Hz (+2 Hz/ -7 Hz).

5.4.5.1 Sending

The sending Signal to Distortion (S/D) ratio is the ratio of the signal power of the measurement tone to the distortion power at the digital output. The S/D ratio shall be above the limits given in table 3.

Limits for the intermediate levels of the 1 kHz measurement are found by drawing straight lines between the breaking points in the table on a linear (dB signal level)-linear (dB ratio) scale.

Measurements at 300 Hz and 6 kHz shall only be carried out at -3 dB rel. Acoustic Reference Level (ARL).

Table 3: Limits for signal to total distortion ratio

Tone input level [dB rel.ARL]	300 Hz [dB]	1 kHz [dB]	6 kHz [dB]
10		24,5	
3		35,0	
-3	29,0	35,0	29,0
-11		35,0	
-18		35,0	
-40		15,0	

Compliance shall be checked by the tests described in annex A, subclause A.2.4.1

5.4.5.2 Receiving

The receiving S/D ratio is the ratio of the signal power of the measurement tone to the distortion power at the measurement point.

The S/D ratio shall be above the limits given in table 4. Limits for the intermediate levels of the 1 kHz measurement are found by drawing straight lines between the breaking points in the table on a linear (dB signal level)-linear (dB ratio) scale.

Measurements at 300 Hz and 6 kHz shall only be carried out at 0 dBm0.

Table 4: Limits for signal to total distortion ratio

Tone input level [dBm0]	300 Hz [dB]	1 kHz [dB]	6 kHz [dB]
8		24,5	
0	29,0	30,5	29,0
-7		35,0	
-13		35,0	
-23		35,0	
-30		30,0	
-40		20,0	
-50		10,0	

Compliance shall be checked by the tests described in annex A, subclause A.2.4.2

5.4.6 Out-of-band signals

5.4.6.1 Discrimination against out-of-band input signals (sending)

With any sine-wave signal above 9 kHz up to 15 kHz applied at the HandsFree Reference Point (HFRP) at a level of -28,7-F_S dBPa, the level of the image frequencies produced at the digital interface shall be below a reference level obtained at 1 kHz (-28,7-F_S dBPa at the HFRP) by at least 25 dB.

Compliance shall be checked by the tests described in annex A, subclause A.2.5.1

5.4.6.2 Spurious out-of-band (receiving)

With a digitally simulated sine wave signal in the frequency range of 100 to 7 000 Hz and at a level of 0 dBm0 applied at the digital interface, the selectively measured level of spurious out-of-band image signals in the frequency range of 9 kHz to 16 kHz shall be lower than the in-band acoustic level produced by an input signal at 1 kHz set at the level specified in table 5.

Table 5: Discrimination levels-receiving

Image signal frequency	Equivalent input signal level
9 kHz	-50 dBm0
16 kHz	-60 dBm0

The limit at intermediate frequencies lies on a straight line drawn between the given values on a logarithmic (frequency)-linear (dB) scale.

Compliance shall be checked by the tests described in annex A, subclause A.2.5.2.

5.4.7 Noise

5.4.7.1 Sending

The A-weighted noise produced by the apparatus in the sending direction shall not exceed -62 dBm0(A).

Compliance shall be checked by the tests described in annex A, subclause A.2.6.1.

5.4.7.2 Receiving

The A-weighted noise produced by the apparatus in the receiving direction at the measurement point shall not exceed -49 dBPa(A)+F_r.

Compliance shall be checked by the tests described in annex A, subclause A.2.6.2.

5.4.8 Delay

The total delay of the sending plus the receiving parts shall be less than 26 ms.

NOTE: The specified maximum delay allows for the 4 ms delay in the CCITT Recommendation G.722 [4] codec plus the transmission delay in the air paths between the users and the transducers, and the additional delay due to speech processing algorithms implementing echo control functions.

Compliance shall be demonstrated by a suppliers declaration based on measurements conducted according to one of the alternative principles described in annex B, or by an alternative and documented approach. The declaration shall take into account the maximum distances ($d_{s\ max}$ and $d_{r\ max}$) allowed between the participants and the transducers.

Annex A (Normative): Test methods

A.1 General conditions for testing

A.1.1 Testing environment

A.1.1.1 Test room

The acoustic measurement environment shall be free field (anechoic) down to the lowest frequency of the 1/3 octave band centred at 125 Hz.

Satisfactory free field conditions are deemed to exist where errors due to the departure from ideal conditions do not exceed the limits shown in table A.1, inside a one meter radius sphere (referred to later on as "reference sphere").

Table A.1

1/3 Octave band centre frequency [Hz]	Allowable departure [dB]
< 300	±1,5
300 to 5 000	±1
> 6 300	±1,5

The test source used for the verification of free field conditions shall be an Artificial Mouth compliant with ITU-T Recommendation P.51 [19] and the level of the signal used shall be -20 dBPa at the HFRP. A wideband noise signal shall be used and third octave spectrum measurements shall be carried out at the measurement points. Measurements shall be made along the three orthogonal axes through the centre of the sphere (x, y horizontal, z vertical). By default, the x axis shall be used for measurement purposes, unless the ITU-T Recommendation P.34 [25] measurement set-up is used. The main axis of the Artificial Mouth shall coincide with the testing axes and the lip plane of the Mouth shall be tangential to the sphere, at 500 mm from its centre. Measurement points along each axis, taken from the lip plane of the Artificial Mouth, shall be at distances of 315 mm, 400 mm, 500 mm, 630 mm, 800 mm and 1 000 mm.

The broadband level of the background room noise shall not exceed -70 dBPa(A). Furthermore, the octave band noise level shall not exceed the limits given in table A.2:

Table A.2: Noise level limits

Octave centre frequency [Hz]	Octave band pressure level [dBPa]
63	-45
125	-60
250	-65
500	-65
1 k	-65
2 k	-65
4 k	-65
8 k	-65
16 k	-65

A.1.1.2 Testing arrangements

The acoustic measurement conditions are free-field, however for specific designs, it is up to the manufacturer to specify an alternative (reverberant) testing environment. In this case the measurement conditions shall be stated in the test report.

A.1.1.2.1 Sending

Two different testing conditions shall be used, according to the type of microphone(s) equipping the audiographic terminal:

- microphones placed on the table surface in front of one or more speakers:
- microphone(s) placed in other parts of the room.

Microphones intended for being placed on the table:

- The same arrangement as specified in I-ETS 300 245-6 [5] shall be used, placing each microphone in turn at point B (figure A.1). The other microphones shall be removed from the acoustic field of the Artificial Mouth.

Microphones intended for being placed elsewhere in the room:

- The acoustic pick-up port of the microphone shall be placed at the centre of the reference sphere and the Artificial Mouth shall be placed with the lip plane tangential to the sphere. The reference axis of the Mouth shall coincide with the x axis of the sphere, while the microphone shall be oriented as specified by the manufacturer. In case of microphones intended to pick up the sounds coming from multiple directions (either omnidirectional or steered devices), the extreme incidence angles on the horizontal plane stated by the manufacturer shall be tested, as well as an intermediate incidence direction. The testing angles actually used for measurements shall be stated in the test report.

In order to take into account the difference between the reference test positioning and the actual microphone-speaker operating distance (d_S) for which the terminal is adjusted, the following correction factor F_S is defined:

$$F_S \text{ (dB)} = 20 \text{ Log } (d_S/0{,}5) \qquad (d_S \text{ in meters})$$

For sending noise measurements the sound source shall be muted and all microphones shall be placed in the testing environment.

NOTE: No method is specified for testing handsfree terminals equipped with lapel-mounted (lavalier) microphones.

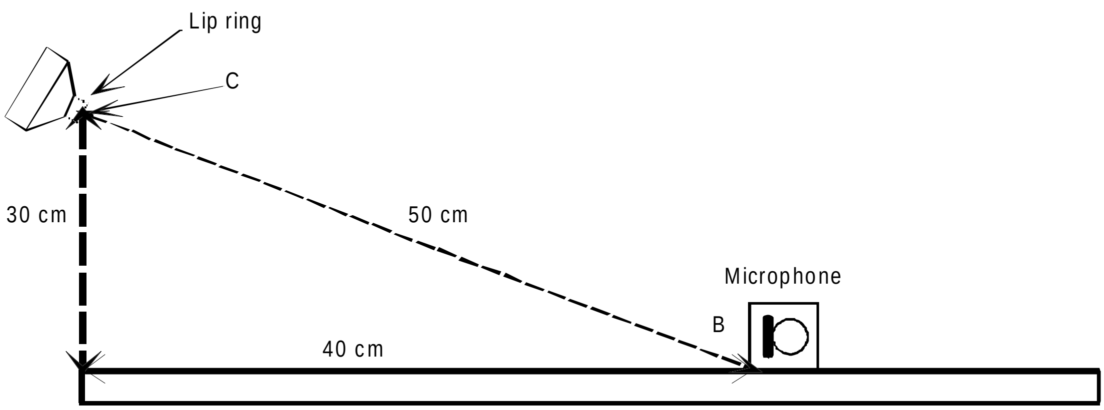

Figure A.1: Measurement set-up for table place microphones

A.1.1.2.2 Receiving

Each loudspeaker shall be placed in turn at the centre of the reference sphere, with its main axis coincident with the x-axis and its outer surface containing the z-axis. The measurement microphone shall be placed on the x-axis, its acoustic centre being at the intersection with the reference sphere.

In order to take into account the difference between the reference test positioning and the actual loudspeaker-listener operating distance (d_r) for which the terminal is adjusted, the following correction factor F_r is defined:

$$F_r \text{ (dB)} = 20 \text{ Log } (d_r/0,5) \qquad (d_r \text{ in metres})$$

A.1.1.2.3 Terminal coupling loss

For TCL and stability measurements the more efficient loudspeaker shall be placed at the centre of the reference sphere, with its main axis coincident with the x-axis and its outer surface containing the z-axis. The more efficient microphone shall be placed on the x-axis, its acoustic pick-up port being at the intersection with the reference sphere and its maximum efficiency pick-up direction parallel to the x-axis. For surface effect microphones, a suitable horizontal positioning plane shall be used with the minimum surface guaranteeing the proper working of the microphone. In any case, the positioning surface shall not enter the reference sphere by more than 100 mm.

In order to take into account the difference between the reference test positioning and the actual loudspeaker-to-microphone operating distance for which the terminal is adjusted, the following correction factor F_{tcl} is defined:

$$F_{tcl} \text{ (dB)} = 20 \text{ Log } (d_{min}/0,5)$$

d_{min} (in metres) is the minimum loudspeaker-to-microphone distance specified by the manufacturer. If no specification is provided, a default d_{min} = 2 m value shall be assumed and stated in the test report.

A.1.2 Test equipment interface

The interface of the test equipment connected to the terminal under test shall be capable of providing the signalling and supervision necessary for the terminal to be working in all test modes. The connection of the test equipment to the terminal under test at the S and T coincident reference point shall be in accordance with TBR 3 [23].

A.1.3 Test equipment requirements

A.1.3.1 Electroacoustic equipment

Artificial Mouth: The Artificial Mouth shall conform to ITU-T Recommendation P.51 [11].

Sound level meter: The sound level measurement equipment shall conform to IEC Publication 651 [27], type 1.

A.1.3.2 Test signals and spectrum measurements

A.1.3.2.1 Standard test signals

Unless differently specified (see subclause A.1.3.2.2), the test signal shall be either sinusoidal or pink noise, as specified for the different measurements. The pink noise shall be band limited to the frequency range 100 Hz to 8 kHz, with a band pass filter with at least 24 dB/oct slopes and 25 dB out-of-band attenuation. The third octave spectrum of the electrically generated pink noise shall be equalized to within ±1 dB, while the acoustically generated pink noise shall be equalized at the MRP within ±3 dB. The crest factor of the (continuous) pink noise signal shall be indicated in the test report.

An on/off modulation (T_{ON} = 250 ms (±5 ms) and T_{OFF} = 150 ms (±5 ms)) shall be applied both for noise and for sinusoidal measurements. The excitation levels refer to the ON component of the signals.

For noise excitation, measurements shall be made by 1/3rd octave filters, at centre frequencies as specified by ISO 3 [24], in the range from 100 Hz to 8 kHz.

A.1.3.2.2 Composite source signal

In case of terminal equipment using technologies for which the test specifications in this annex are not suitable to prove conformance to this ETS, equivalent evaluation methods can be used. These can be based on the Composite Source Signal specified in annex B of I-ETS 300 245-3 [3]. All tests shall be carried out by using only one type of test signal. The test signal used shall be indicated in the test report. The alternative methods shall be validated by measuring linear terminal equipment and comparing the results with those obtained by using the switched noise/sine signals.

The results of measurements on terminal equipment incorporating adaptive automatic gain control, acoustic echo cancelling or other non-linear functions may depend on the signal used. The Composite Source Signal may be used for testing equipment incorporating Adaptive Acoustic Echo Control functions, particularly when the switched signals do not properly activate the terminal for all tests described in this annex.

A.1.3.3 Test signals levels

A.1.3.3.1 Sending

Unless specified otherwise, the test signal level shall be ($-28{,}7-F_S$) dBPa at the HFRP. The characteristics of the Artificial Mouth shall be according to ITU-T Recommendation P.51 [19]. Two different methods can be alternatively used for calibrating the Artificial Mouth.

Method 1-at the HFRP

The input signal from the Artificial Mouth is equalized under free field conditions at the HFRP, such that the spectrum is as specified in subclause A.1.3.2 and the total level in the frequency range corresponding to the third octave bands from 100 Hz to 8 kHz is ($-28{,}7-F_S$) dBPa. For calculating SLR and response characteristics the reference point is MRP. The sound pressure level at the MRP shall be calculated (by definition) by adding 24 dB to the sound pressure level at the HFRP.

Method 2-at the MRP

The signal generated by the Artificial Mouth is equalized at the MRP under free field conditions to obtain the spectrum specified in subclause A.1.3.2, at a level of -4,7 dBPa in the frequency range corresponding to the third octave bands from 100 Hz to 8 kHz. The spectrum at the MRP is then recorded and the level is adjusted to obtain ($-28{,}7-F_S$) dBPa at the HFRP. The spectrum recorded at the MRP is used as a reference for calculating SLR and response characteristics.

A.1.3.3.2 Receiving

Unless specified otherwise, the test signal shall be -30 dBm0 when measurements with the receiving volume control at its maximum position are carried out. For measurements with the volume control at its minimum position, a test signal level of -15 dBm0 shall be used.

Terminals not equipped with a manual volume control shall be measured both with -15 dBm0 and with -30 dBm0 test signals.

A.1.3.4 Test equipment for the digital interface

Either a codec approach or a direct signal processing approach can be used. In the following the first solution is specified. The latter or other alternative approaches can be adopted, provided their equivalence with the codec approach is demonstrated.

A.1.3.4.1 Codec specifications

The reference codec and its audio parts shall comply with CCITT Recommendation G.722 [13]. The tests shall be carried out with the codec operating in mode 1.

A.1.3.4.2 Analogue interface

Measurements shall be carried out by connecting the measurement instrumentation to the test points A and B of the reference codec (see CCITT Recommendation G.722 [13], figure 2).

For compatibility with existing telephone instrumentation, 600 Ω ohm balanced electrical interfaces shall be implemented.

A.1.3.4.3 Definition of 0 dBr point

D/A conversion: A digital sequence representing the Sub Band-Adaptive Differential Pulse Code Modulation (SB-ADPCM) equivalent to an analogue sinusoidal signal whose r.m.s value is 9 dB below the maximum full-load capacity of the codec, generating a 0 dBm signal across a 600 ohm load.

A/D conversion: A 0 dBm signal generated by a 600 Ω source giving the digital sequence representing the Sub Band-Adaptive Differential Pulse Code Modulation (SB-ADPCM) equivalent to an analogue sinusoidal signal whose root mean square (r.m.s) value is 9 dB below the maximum full-load capacity of the codec.

A.1.5 Accuracy of calibrations

Unless specified otherwise, the accuracy of measurements made by test equipment shall be better than:

Item	Accuracy
Electrical signal power	±0,2 dB for levels > = -50 dBm
Electrical signal power	±0,4 dB for levels < -50 dBm
Sound pressure	±0,7 dB
Time	±5 %
Frequency	±0,2 %

Unless specified otherwise, the accuracy of the signals generated by the test equipment shall be better than:

Quantity	Accuracy
Sound Pressure level at the MRP	±3 dB (100 Hz to 8 kHz)
Sound Pressure level at the HFRP	±3 dB (100 Hz to 16 kHz)
Electrical excitation level	±0,4 dB (see note 1)
Frequency generation	±2 % (see note 2)

The measurement results shall be corrected for the measured deviations from the nominal level.

NOTE 1: Across the whole frequency range.

NOTE 2: When measuring sampled systems, it is advisable to avoid measuring at sub-multiples of the sampling frequency. There is a tolerance of ±2 % on the generated frequencies, which may be used to avoid this problem, except for 8 kHz, where only -2 % tolerance may be used.

A.2 Testing of transmission requirements

A.2.1 Sensitivity-frequency response

A.2.1.1 Sending

The measurement shall be carried out for each microphone, which shall be mounted as specified in subclause A.1.1.2.1. In case of microphones intended to pick up the sounds coming from multiple directions (either omnidirectional or steered devices), the extreme incidence angles on the horizontal plane (as stated by the manufacturer) shall be tested, as well as an intermediate incidence direction. If stated by the manufacturer, the nominal direction shall be used as intermediate. The testing angles actually used for measurements shall be stated in the test report.

The noise signal shall be generated by the Artificial Mouth at the level specified in subclause A.1.3.3.

The spectrum of the output signal shall be measured at the output interface of the reference codec.

The sending sensitivity shall be calculated as follows, according to the calibration method used (subclause A.1.3.3):

Method 1

The sending sensitivity is given by the difference between the electric spectrum and the acoustic spectrum at the MRP:

$$S_{mJ} = 20 \text{ Log Vs} - 20 \text{ Log } P_m \qquad (P_m = P_{HFRP} + 24 \text{ dB})$$

where:

20 Log Vs:	Electric spectrum;
20 Log P_m:	Acoustic spectrum at MRP.

Method 2

The sending sensitivity S_{mJ} is given by the following relationship:

$$S_{mJ} = 20 \text{ Log Vs} - 20 \text{ Log } P_m + \text{Corr} - 24 \text{ dB}.$$

where:

20 Log Vs:	Electric spectrum;
20 Log P_m:	Acoustic spectrum recorded at MRP;
Corr:	Correction factor (20 Log P_{MRP}/P_{HFRP}) of the Artificial Mouth.

A.2.1.2 Receiving

Each loudspeaker shall be placed in turn in the test room as specified in subclause A.1.1.2.2.

The noise signal generator shall be connected to the input of the reference codec.

The sensitivity at each 1/3 octave band shall be calculated by subtracting the spectrum of the electric signal from the acoustic spectrum at the measurement point.

The measurement shall be repeated at the minimum and maximum position of the (manual) volume control, changing the input level accordingly. In case of devices not provided with manual volume control, the measurement shall be repeated for excitation levels of -30 dBm0 and -15 dBm0.

A.2.2 Loudness rating

A.2.2.1 Sending Loudness Rating

The sending sensitivity shall be measured for each of the fourteen 1/3rd octave bands given in table 1 of ITU-T Recommendation P.79 [20], bands 4 to 17 (200 Hz-4 kHz). The sensitivity is expressed in terms of dB(V/Pa) and the SLR shall be calculated according to the formula 2-1 of ITU-T Recommendation P.79 [20], according to section 3 and using the sending weighting factors from table 1/P.79.

The measurement shall be carried out for each microphone, which shall be mounted as specified in subclause A.1.1.2.1. In case of microphones intended to pick up the sounds coming from multiple directions (either omnidirectional or steered devices), the extreme incidence angles on the horizontal plane (as stated by the manufacturer) shall be tested, as well as an intermediate incidence direction. If stated by the manufacturer, the nominal direction shall be used as intermediate. The testing angles actually used for measurements shall be stated in the test report.

A.2.2.2 Receiving Loudness Rating

The receiving sensitivity shall be measured for each of the fourteen 1/3rd octave bands given in table 1 of ITU-T Recommendation P.79 [20], bands 4 to 17 (200 Hz-4 kHz). The sensitivity is expressed in terms of dB(Pa/V) and the RLR shall be calculated according to the formula 2-1 of ITU-T Recommendation P.79 [20], according to section 3 and using the receiving weighting factors from table 1/P.79.

The receiving sensitivity shall not be corrected by the leakage correction (L_E) factor. The calculated RLR shall be corrected by subtracting 14 dB, according to ITU-T Recommendation P.34 [25].

For testing the volume control range, an additional test signal level of -15 dBm0 shall be used.

A.2.3 Terminal Coupling Loss

The measurements shall be carried out by coupling, as specified in subclause A.1.1.2.3, the most efficient loudspeaker with the most efficient microphone.

A.2.3.1 Weighted Terminal Coupling Loss

The test signal shall be a pink noise, with a level of ($-15-F_{tcl}$) dBm0.

The attenuation from digital input to digital output shall be measured at the 1/3rd octave frequencies given by the R10 series of preferred numbers in ISO 3 [24] for frequencies from 100 Hz to 8 000 Hz.

The TCLw shall be calculated according to the method in ITU-T Recommendation G.122 [26], annex B, clause B.4 (trapezoidal rule) on the frequency band from 100 Hz to 8 kHz. The value calculated according to ITU-T Recommendation G.122 [26] ($TCLw_{calculated}$) shall be corrected by adding the correction factor F_{tcl} and by considering the total sound power produced by all the loudspeakers:

$$TCLw = TCLw_{calculated} + F_{tcl} - 10 \text{ Log } [(n+1)/2]$$

where n is the total number of loudspeakers equipping the terminal under test.

NOTE: The formula above assumes that only one microphone is active at each time and that the sound pressure levels of signals picked up from the more distant loudspeakers is 3 dB less than the signal level generated by the better coupled loudspeaker. It is also assumed that a power summation law applies to the signals picked up by the different loudspeakers. This is not absolutely true in principle, but some decorrelation of the sound signals generated by the different loudspeakers can be assumed to result from the different transmission path lengths and from the reverberation effects occurring in actual use.

A.2.3.2 Stability loss

The test signal shall be sinusoidal, with a level of ($-15-F_{tcl}$) dBm0.

The attenuation from digital input to digital output shall be measured at 1/24th octave intervals for frequencies from 100 Hz to 8 kHz. The actually measured values shall be corrected by adding F_{tcl} and by subtracting 10 Log [(n+1)/2] (as for the TCLw measurement).

A.2.4 Distortion

A.2.4.1 Sending

The set shall be placed as specified in subclause A.1.1.2.1. A pulsed sine tone at the measurement frequency shall be generated by the mouth. The level of this signal shall be adjusted until the output of the terminal is -10 dBm0 (ON periods). The level of the signal at the MRP is then the ARL.

The test signal shall be applied at the following levels: -40 dB, -18 dB, -11 dB, -3 dB, 3 dB, 10 dB relative to ARL.

The ratio of the signal to total distortion power of the signal at the reference codec output shall be measured.

It shall be verified that, when producing the sound pressure level for testing at +10 dB rel.ARL, the distortion generated by the Artificial Mouth does not exceed 5 %. The distortion of the signal generated for testing at the other measurement levels shall not exceed 1 %.

A.2.4.2 Receiving

The set shall be placed as specified in subclause A.1.1.2.2 and with the volume control set at its nominal position (RLR = 6 dB-F_r). A pulsed sine tone at the measurement frequency is applied at the electrical input of the reference codec at the following levels: -50 dBm0, -40 dBm0, -30 dBm0, -23 dBm0, -13 dBm0, -7 dBm0, 0 dBm0, 8 dBm0.

The receiving distortion shall be calculated by normalizing the distortion spectrum according to the receiving sensitivity-frequency response. This is accomplished by subtracting from each frequency component of the distortion spectrum the difference between the receiving sensitivity at this frequency and the sensitivity at the measurement frequency.

A.2.5 Out-of-band signals

A.2.5.1 Discrimination against out-of-band input signals (sending)

The set shall be placed as specified in subclause A.1.1.2.1. For input signals at the frequencies of 9 kHz, 10 kHz, 12 kHz, 13 kHz, 14 kHz and 15 kHz, at ($-28{,}7-F_S$) dBPa at the HFRP, the level of each image frequency shall be measured at the output interface of the reference codec.

As the Artificial Mouth is only specified up to 8 kHz, the acoustic signal can be generated by a suitable alternative loudspeaker, placed in the same position. The sound pressure developed by the loudspeaker at the measurement point shall be calibrated under free field conditions.

To activate the set in the sending direction, every second measurement burst shall be substituted by an in-band burst of the same duration at 1 kHz. The correct activation shall be checked by measuring the output level of the transduced in-band bursts.

A.2.5.2 Spurious out-of-band (receiving)

The set shall be placed on the measurement table as specified in subclause A.1.1.2.2. For an input signal at the frequencies 200 Hz, 350 Hz, 500 Hz, 1 000 Hz, 2 000 Hz, 3 500 Hz, 5 000 Hz and 7 000 Hz, applied at a level of -30 dBm0 at the input port of the reference codec, the acoustic level of spurious out-of-band image signals at frequencies up to 16 kHz shall be selectively measured.

A.2.6 Noise

A.2.6.1 Sending

With the set placed as specified in subclause A.1.1.2.1, the noise level at the digital output shall be measured, at the output interface of the reference codec terminated by a 600 Ω resistor, by an apparatus including A weighting according to IEC Publication 651 [27].

A.2.6.2 Receiving

The set shall be placed as specified in subclause A.1.1.2.2. The input port of the reference codec shall be terminated by a 600 Ω resistor and the A weighted noise level measured.

A.2.7 CCITT Recommendation G.725 encoding

The test is described in CCITT Recommendation G.725 [15], appendix 1. The input signal shall be a sinusoid with a level as specified in subclause A.1.3.3.

Annex B (Informative): Test Methods for Delay measurement

B.1 Introduction

This informative annex presents two alternative methods for measuring delay. Before a single method can be specified in this ETS, the stability, accuracy and reproducibility of these methods need to be verified.

B.2 Cross-correlation method

B.2.1 Sending

The Artificial Mouth is placed as specified in subclause A.1.1.2.1 and calibrated as specified in subclause A.1.3.3.1. The modulated pink noise is generated by the Artificial Mouth and the transduced signal is picked up at the output interface of the reference codec. The cross-correlation function $\Phi_{xy}(\tau)$ between the input (Mouth excitation) and output signals is calculated in the time domain:

$$\Phi_{xy}(\tau) = \frac{1}{nT} \int_0^{nT} s_x(t) s_y(t+\tau) dt$$

where:

- $s_x(t)$: input signal (Artificial Mouth excitation);
- $s_y(t)$: measured output signal.

The measurement window (nT) needs to be chosen exactly equal to a multiple of the period (T) of the envelope modulation signal ($T = T_{ON} + T_{OFF} = 400$ ms $\pm$ 10 ms). It is recommended to integrate over at least ten periods ($n \geq 10$). The cross-correlation function needs to be calculated for the values of τ between $-T_{ON}$ (250 ms ± 5 ms) and $T/2 + T_{ON}$.

NOTE 1: The measurement method specified here assumes that the measured delay is less than T/2. If the delay exceeds this value, measurement misleading results may occur. The existence of this validity condition needs to be approximately checked in advance by alternative means (i.e. by single burst excitations) before applying this method.

To smooth the cross-correlation function and to avoid misinterpretations in case two relative maxima of a similar magnitude occur, the delay is calculated from the envelope $E(\tau)$ of the cross-correlation function $\Phi_{xy}(\tau)$. The maximum of the envelope function occurs in correspondence to the measured sending delay D_s. The envelope is calculated by the Hilbert transformation of the cross-correlation function $H\{\Phi_{xy}(\tau)\}$:

$$H\{\Phi_{xy}(\tau)\} = \int_{-T_{ON}}^{T/2+T_{ON}} \frac{\Phi_{xy}(u)}{\pi(\tau - u)} du$$

$$E(\tau) = \sqrt{[\Phi_{xy}(\tau)]^2 + \{H[\Phi_{xy}(\tau)]\}^2}$$

NOTE 2: It is advantageous to calculate the Hilbert transform by operating in the frequency domain by means of Fourier Transform techniques.

B.2.2 Receiving

The measurement microphone is placed at point C, as specified in annex A, subclause A.1.1.2.1, and calibrated as specified in annex A, subclause A.1.3.3.1. The modulated pink noise is generated at the input of the reference codec interface and the transduced signal is picked up by the measurement microphone at the measurement point. The cross-correlation function $\Phi_{xy}(\tau)$ between the input signal of the reference codec and output signals picked up by the microphone is calculated in the time domain and the receiving delay D_r is evaluated as described in annex B, subclause B.2.1.

B.2.3 Total delay

The total delay D is given by:

$$D = D_S + D_r - D_E$$

where D_E is the delay of the test equipment, which needs to be measured, by means of the same method, between the mouth excitation point and the microphone output port with the microphone positioned at the MRP.

B.3 Method based on group delay

B.3.1 Sending

The artificial mouth is placed as specified in annex A, subclause A.1.1.2.1, and calibrated as specified in annex A, subclause A.1.3.3.1. The excitation signal is generated by the Artificial Mouth and the transduced signal is picked up at the output interface of the reference codec. The weighted average group delay between the input (sound pressure @ Measurement Point) and output (voltage @ Reference Codec output) of the Telephone Equipment Under Test (TEUT) is calculated from the measured complex sending frequency response function, H_S:

$$\mathrm{Dn} = \int_{f_{\min}}^{f_{\max}} \mathrm{W}(f)\frac{-1}{2\pi}\frac{d\varphi_S(f)}{df}\,df = \frac{-1}{2\pi}\int_{f_{\min}}^{f_{\max}} \mathrm{W}(f)\frac{d\,\mathrm{Im}(\ln \mathrm{Hs}(f))}{df}\,df$$

where:

f_{min} and f_{max} are frequencies outside the passband of the TEUT;

$W(f)$ is a normalized weighting function that corresponds to a typical transmitted voice spectrum;

$\varphi_S(f)$ is the phase angle in radians associated with $H_S(f)$. The imaginary part of the complex logarithm (phase) is defined as a continuous function.

To optimize the method to produce results similar to the delays experienced for human speech, the long term power spectrum density values from ITU-T Recommendation P.50 [29], table 1, column 4, in dB(Pa/Hz) are used as a weighting function with a normalization factor of +4,7 dB and conversion to linear power spectrum density values. Considering the transmission characteristics (CCITT Recommendation G.722 [13], figure 10) no further weighting need be applied within the integration range from 100 Hz to 7 kHz.

In order to avoid spurious influence from non-linearities as well as aliasing in the time domain, signals with a continuous spectrum combined with selective analysis are to be preferred. If adequate signal to noise ratios can be established, modulated pink noise may be used for the measurement of the complex frequency response function. Otherwise a swept sine (modulated if necessary to activate any speech detector) combined with suitable tracking filter analysis is used.

When discrete frequency analysis or excitation signals with discrete spectra (sine or pseudorandom) are being used, the distance between the measurement points must be less than 1/T, where T is the widest range within which delays may possibly occur. Unless more detailed knowledge is available, the distance between the sampling frequencies should not exceed 10 Hz (equivalent to a useful delay range of 100 ms).

B.3.2 Receiving

The calibrated measuring microphone is placed at point C as specified in annex A, subclause A.1.1.2.1. The excitation signal is generated at the input interface of the reference codec and the transduced signal is picked up by the measuring microphone. The weighted average group delay, D_r, between the input (voltage @ Reference Codec input) and output (sound pressure @ Measurement Point) of the TEUT is calculated as D_s, but based on the complex receiving frequency response, $H_r(f)$.

B.3.3 Total delay

The total delay is given by:

$$D = D_s + D_r$$

Using this method based on frequency response functions, the delay of the test equipment is inherently excluded from the result, provided the same microphone measurement "chain" is used for calibrating the Artificial Mouth and for measuring the received signals

Annex C (Informative): Audio alignment and practical installation guides

C.1 Introduction

An audiographic or videoconference terminal can have one or more of the following audio arrangements:

- handset function;
- handsfree function for single user;
- handsfree function for a group of users.

The audio performances, and in particular the transmitted speech levels, for the first two arrangements do not depend very significantly on the installation environment and are consequently tightly controlled by the terminal specifications. On the contrary, the handsfree function for a group of users is normally provided with audio arrangements which widely interact with the installation environment and suitable guides should be provided for properly aligning the electroacoustic sensitivities and for checking the correct transduction of the whole frequency band.

Also the positioning of transducers in the acoustic environment can strongly influence their effective performances and suitable installation criteria should be followed in order to maximize the signal-to-noise and signal-to-reverberation ratios.

C.2 Audio alignment and sensitivity-frequency characteristics

Audio alignment is an operation carried out when the terminal is installed and is intended to guarantee the correct setting of sending and receiving level adjustments. If the line interface is used for aligning, the terminal should be set to operate CCITT Recommendation G.722 [13] coding in mode 1 (64 kbit/s).

The following equipment should be used for performing the audio alignment of audiographic terminals:

- Artificial Mouth (complying with ITU-T Recommendation P.51 [19]);
- signal generator (digital or analogue);
- sound level meter (complying with IEC Publication 651 [27]);
- electric level meter.

The acoustic test signal generated by the Artificial Mouth should comprise speech shaped noise (ITU-T Recommendation P.50 [29]) with a level of -4,7 dBPa at the MRP.

The signal generator is used for performing the receiving alignment and can either generate a CCITT Recommendation G.722 [13] encoded digital path or an analogue signal. In the second case a reference codec could be necessary if no specific alignment interface is available in the terminal. The signal for receiving alignment should be a speech shaped noise (ITU-T Recommendation P.50 [29]) with a level of -20dBm0. If necessary, an ON-OFF modulated signal can be used instead of stationary noise. In this case the level of the signal is defined as the level measured during the ON periods of the signal.

NOTE: The terminal manufacturers are encouraged to provide an analogue 0dBr interface for alignment purposes in their equipment.

If no specific alignment interface is provided, the electric level meter should either directly process the digital pattern generated by the terminal or be used in conjunction with a CCITT Recommendation G.722 [13] Reference Coder.

C.2.1 Sending

C.2.1.1 Sensitivity adjustment(s)

The Artificial Mouth should be placed on the border of the conference table, as shown in figure A.1, in correspondence of each conferee position. The conference microphones should be positioned on the conference table (or in the conference room) as in normal use.

If a sensitivity adjustment is provided for each microphone, it should be regulated in turn in order to achieve an output level of -20 dBm0 when the Mouth is placed in correspondence with the associated microphone. Otherwise, if only one sensitivity control is provided, it should be regulated in order that the average reading obtained by placing the Mouth in correspondence of all the conferee positions is equal to -20 dBm0.

C.2.1.2 In-site frequency response

The in-site measurement of the sending frequency response is defined as the difference between the octave spectra of the output signal and of the acoustic excitation signal at the MRP. In order to avoid excessive fluctuations of the frequency response, and considering that measurements are in-site, octave band analysis is recommended in the bandwidth from 125 Hz to 4 kHz (centre frequencies). It is recommended that the sum of the absolute differences between the measured values and their average should not exceed 8 dB.

C.2.2 Receiving

C.2.2.1 Sensitivity adjustment

The speech shaped noise is fed to the input of the terminal at a level of -20 dBm0. With manual volume control (if any) in the maximum position, the receiving gain should be adjusted in order to achieve a sound pressure level of at least 54 dBA at each conferee position. The measurements should be made by placing the measurement microphone at point C (figure A.1) in correspondence of each conferee position.

NOTE: Experience has shown that increased listening levels up to 65 dBA may be desirable. The maximum settable gain being limited by the terminal stability constraint.

C.2.2.2 In-site frequency response

The in-site measurement of the of the receiving frequency response is defined as the difference between the octave spectra of the output signal and of the input excitation signal at the terminal interface. In order to avoid excessive fluctuations of the frequency response, and considering that measurements are in-site, octave band analysis is recommended in the bandwidth from 125 Hz to 4 kHz (centre frequencies). It is recommended that the sum of the absolute differences between the measured values and their average should not exceed 12 dB.

C.3 Practical installation criteria

It is recommended that the acoustic treatment of audiographic teleconference and videoconference rooms be carefully designed in order to achieve best overall electro-acoustic performances.

Whilst audiographic teleconferencing may still be possible in rooms not meeting the following recommendations, the overall system performance is likely to be degraded. In particular, it is recommended that barrel effects and background noises be prevented and a suitable sound proof insulation of the room be provided.

The main parameters to be taken into account when installing audiographic teleconference systems are:

- room reverberation;
- background noise;
- sound insulation (privacy).

While the former two parameters independently affect the maximum talker-to-microphone allowable distance, the latter ensures privacy between the conference participants.

C.3.1 Maximum talker to microphone distance

C.3.1.1 Background noise level constraints

The following maximum talker-to-microphone distance d_{max} should be maintained in order to achieve a signal to noise ratio of at least 30 dB for the speech level of average talkers:

$$d_{max} \leq 52{,}5 \times 10^{-(L+10 \cdot \mathrm{Log}\ n)/20}$$

where d_{max} is in metres, L is the long-term average background noise level (in dBA) and n is the number of simultaneously open microphones.

In the case where cardioid microphones are used, the calculated distance d_{max} may be increased by 50 %.

NOTE: The above calculation of d_{max} is based on the assumption that the average talker delivers a speech level of -24 dBPa at 50 cm from the mouth.

C.3.1.2 Reverberation constraints

To avoid barrel effects, the microphones should be placed in such a way that the direct sound field of associated talkers is high enough compared to the reverberation field. This is achieved by restricting the maximum talker-to-microphone distance d_{max} to half the critical distance of the room:

$$d_{max} \leq \frac{1}{8}\sqrt{\frac{0{,}161V}{\pi T}}$$

where d_{max} is in metres, V is the room volume (m^3) and T is the reverberation time (s).

Extremely directional microphones should not be used. When directional microphones (cardioid) are used, the distance may be increased by 50 %. The reverberation time of the room can be measured, alternatively the reverberation time of geometrically simple rooms can be calculated from the absorption characteristics of walls and furniture and from room dimensions. Realistic talker-to-microphone distances for microphones placed on the conference table are normally attained if the reverberation time is less than 0,4 s in the frequency range from 100 Hz to 7 kHz (figure C.1). With a longer reverberation time it is recommended to take particular care in placing the microphones and loudspeakers and regarding the echo control performances. The relationship between the room volume, the reverberation time and the critical distance is graphically presented in figure C.1.

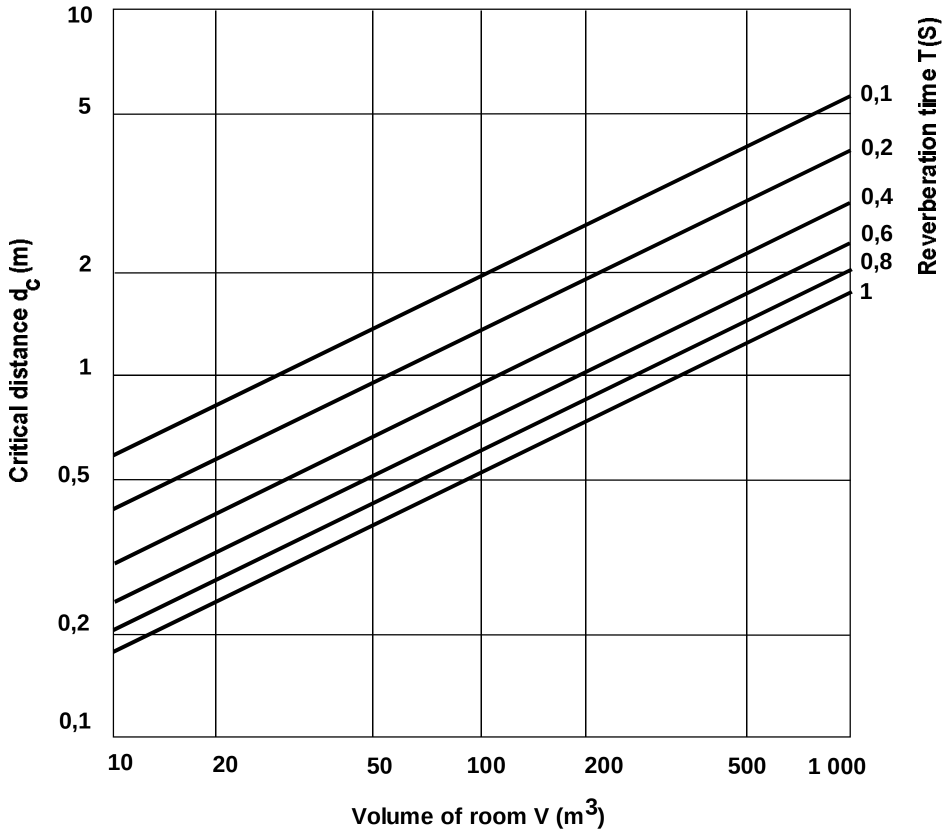

Figure C.1

C.3.1.3 Preferred maximum talker to microphone distance

The preferred maximum talker-to-microphone distance is given by the least of values obtained from formulas at clauses C.3.1.1 (room noise) and C.3.1.2 (reverberation).

C.3.2 Sound insulation

In order to insure good privacy for the conferees, the speech level transmitted to the premises around the conference room should be at least 15 dB below the background noise level in these premises. Assuming that the maximum talking and listening level in the reverberation field of the conference room is 70 dBA, and the external noise level around 40 dBA, it follows that the sound insulation between the room and the external premises should be around 45 dB.

Annex D (Informative): Bibliography

For the purposes of this I-ETS, the following informative references apply:

- ETS 300 111 (1992): "Integrated Services Digital Network (ISDN); Telephony 3.1 kHz teleservice, Service description".

- ETS 300 263 (1993): "Integrated Services Digital Network (ISDN); Telephony 7 kHz teleservice, Service description".

- ETS 300 264 (1994): "Integrated Services Digital Network (ISDN); Videotelephony teleservice, Service description".

- I-ETS 300 245-1: "Integrated Services Digital Network (ISDN); Technical characteristics of telephony terminals, Part 1-General".

- I-ETS 300 302-4: "ISDN, Videotelephony teleservice: Part 4 Audio aspects-Wideband (7 kHz) Loudspeaking and Handsfree".

History

Document history			
August 1996	Public Enquiry	PE 112:	1996-08-19 to 1996-12-13