INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.340

(05/2000)

SERIES P: TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Subscribers' lines and sets

Transmission characteristics and speech quality parameters of hands-free terminals

ITU-T Recommendation P.340

(Formerly CCITT Recommendation)

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Subscribers' lines and sets	**Series**	**P.30** **P.300**
Transmission standards	Series	P.40
Objective measuring apparatus	Series	P.50 P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of quality	Series	P.80 P.800
Audiovisual quality in multimedia services	Series	P.900

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Recommendation P.340

Transmission characteristics and speech quality parameters of hands-free terminals

Summary

This Recommendation defines the generic transmission characteristics of hands-free terminals, whatever the applications or technologies integrated. In addition to the parameters that ensure performance levels equivalent to these of handset terminals (sensitivity values, response curves, etc.), it defines the parameters which have a bearing on conversation quality (gain switching characteristics, double-talk performances,etc.). It also defines categories of hands-free terminals based on the terminal's double-talk capability.

This Recommendation deals with the characteristics of the signal-processing equipment (e.g. acoustic echo cancellation (AEC)) that can be incorporated in hands-free terminals.

Source

ITU-T Recommendation P.340 was revised by ITU-T Study Group 12 (1997-2000) and approved under the WTSC Resolution 1 procedure on 18 May 2000.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSC Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

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

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

Page

Introduction

The object of this Recommendation is to obtain for hands-free terminals transmission performance equivalent with handset telephones, at least with respect to send and receive loudness.

Other important features contributing to the quality of telephone calls made from hands-free terminals, as switching characteristics, duplex capability, are defined in this Recommendation.

This Recommendation covers generic characteristics and requirements that are applicable to both analogue and digital hands-free terminals. Additional requirements that are applicable strictly to digital terminals can be found in ITU-T P.342 [3] and P.341 [4].

For loudspeaking telephones (see ITU-T P.10 [14]) which do not provide full hands-free operation, the relevant parts of this Recommendation may be referred to.

Test methods appropriate for parameters defined in this Recommendation may be found in ITU-T P.502 [12].

For the use of HATS for testing, ITU-T P.581 [21] applies.

ITU-T Recommendation P.340

Transmission characteristics and speech quality parameters of hands-free terminals

1 Introduction

The object of this Recommendation is to obtain for hands-free terminals transmission performance equivalent with handset telephones, at least with respect to send and receive loudness.

Other important features contributing to the quality of telephone calls made from hands-free terminals, as switching characteristics, duplex capability, are defined in this Recommendation.

This Recommendation covers generic characteristics and requirements that are applicable to both analogue and digital hands-free terminals. Additional requirements that are applicable strictly to digital terminals can be found in ITU-T P.342 [3] and P.341 [4].

For loudspeaking telephones (see ITU-T P.10 [14]) which do not provide full hands-free operation, the relevant parts of this Recommendation may be referred to.

Test methods appropriate for parameters defined in this Recommendation may be found in ITU-T P.502 [12].

For the use of HATS for testing, ITU-T P.581 [21] applies.

2 Normative references

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision: all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

[1] ITU-T G.121 (1993), *Loudness Ratings (LRs) of national systems*.

[2] CCITT G.223 (1988), *Assumptions for the calculation of noise on hypothetical reference circuits for telephony*.

[3] ITU-T P.342 (2000), *Transmission characteristics for telephone band (300-3400 Hz) digital loudspeaking and hands-free telephony terminals*.

[4] ITU-T P.341 (1998), *Transmission characteristics for wideband (150-7000 Hz) digital hands-free telephony terminals*.

[5] ITU-T P.78 (1996), *Subjective testing method for determination of loudness ratings in accordance with Recommendation P.76*.

[6] ITU-T P.79 (1999), *Calculation of loudness ratings for telephone sets*.

[7] ITU-T P.50 (1999), *Artificial voices*.

[8] ITU-T P.51 (1996), *Artificial mouth*.

[9] ITU-T P.501 (2000), *Test signals for use in telephonometry*.

[10] ITU-T G.167 (1993), *Acoustic echo controllers*.

[11] ITU-T Handbook on Telephonometry, 1993.

[12] ITU-T P.502 (2000), *Objective test methods for speech communication systems using complex test signals*.

[13] ITU-T P.832 (2000), *Subjective performance evaluation of hands-free terminals*.

[14] ITU-T P.10 (1998), *Vocabulary of terms on telephone transmission quality and telephone sets*.

[15] ITU-T G.114 (2000), *One-way transmission time*.

[16] ITU-T G.131 (1996), *Control of talker echo*.

[17] ITU-T G.174 (1994), *Transmission performance objectives for terrestrial digital wireless systems using portable terminals to access the PSTN*.

[18] ITU-T G.122 (1993), *Influence of national systems on stability and talker echo in international connections*.

[19] ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.

[20] ITU-T G.168 (2000), *Digital network echo cancellers.*

[21] ITU-T P.581 (2000), *Use of head and torso simulator (HATS) for hands-free terminal testing.*

3 Definitions and abbreviations

The relevant definitions given in ITU-T P.10 apply along with the following:

3.1 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 11, as defined in ITU-T P.51 [8].

3.2 hands-free (telephone) set (HFT): A telephone set using a loudspeaker associated with an amplifier as a telephone receiver and which may be used without a handset.

3.3 loudspeaking (telephone) set (LST): A handset telephone using a loudspeaker associated with an amplifier as a telephone receiver.

3.4 acoustic echo canceller (AEC): A device which reduces the acoustic echo level with negligible effects on the local and distant users' speech. It is generally implemented by adaptive identification of the acoustic echo path response.

3.5 loss controller: A device which reduces the acoustic echo level by inserting variable losses on the received and/or transmitted audio signals.

3.6 non-linear processor: A device which reduces or cancels small echo signals by non-linear operation on the samples of the transmitted audio signal. A centre-clipper is a typical device of this kind.

3.7 supplementary howling control device: A device which modifies some characteristics of the transmitted and/or received signals in order to improve the stability margin of the terminal. This function is typically implemented by a harmonic processor. To prevent network disturbances, such devices should be avoided in the case of terminals likely to be used on connections including network electric echo cancellers conforming to ITU-T G.168 [20] which are not able to work properly with time-variant echo paths (e.g. frequency shift).

3.8 noise reduction system: A device which increases the signal-to-noise ratio (on the overall or a specific range of bandwidth of the signal).

This Recommendation uses the following abbreviations:

AEC Acoustic echo canceller

AGC Automatic gain control

a_H Insertion attenuation range (dB)

$a_{H,S}$	Attenuation range (dB) in the sending path
$a_{H,R}$	Attenuation range (dB) in the receiving path
Ardt	Received speech attenuation (dB) during double talk
Asdt	Sent speech attenuation (dB) during double talk
CSS	Composite Source Signal (see ITU-T P.501[9]))
Drdt	Received speech distortion during double talk
Dsdt	Sent speech distortion during double talk
EEB	Early Energy Balance
EPDn	Round trip echo path delay – network interface
ERL	Echo return loss
ERLtst	Temporally weighted echo return loss – single talk
ERLtdt	Temporally weighted echo return loss – double talk
HFT	hands-free terminal
HFRP	Hands-Free Reference Point
LOT	Listening Only Test (see ITU-T P.832 [13])
MRP	Mouth Reference Point
MOS	Mean Opinion Score (see ITU-T P.800 [19])
OLR	Overall Loudness Rating
RLR	Receiving Loudness Rating
SLR	Sending Loudness Rating
TCLwdt	Weighted terminal coupling loss – double talk
TCLwst	Weighted terminal coupling loss – single talk
TCLtst	Terminal Coupling Loss temporally weighted – single talk
TCLtdt	Terminal Coupling Loss temporally weighted – double talk
Tc	Convergence time
Tic	Initial convergence time
T_R	Build-up Time
T_{Rdt}	Build-up time – double talk
T_{Rst}	Build-up time – single talk
T_H	Hang-over Time
T_S	Switching Time
TUT	Terminal Under Test
V_{TH}	Threshold Level

4 Transmission characteristics

4.1 Sending sensitivity

The Sending LR (SLR) of an HFT should be about 5 dB higher than the SLR of the corresponding handset telephone (the actual value will depend on the type of handset used).

NOTE 1 – The average acoustic speech level from a terminal user is about 3 dB higher when using a HFT than when using a handset telephone.

The output level from the handset telephone in conversational use is about 1-2 dB lower than what is obtained in the LRGP position specified for loudness ratings measurements of handset telephones. For HFTs there is, however, no such difference.

NOTE 2 – Hands-free terminals having a sending sensitivity that complies with this Recommendation can be assumed to fulfil ITU-T G.223 [2].

Furthermore, in order to avoid excessive crosstalk from the high-level speech currents and/or inadequate received volume from low-level speech currents, care should be taken to ensure that the variation of speech currents is not substantially greater than that from handset telephones.

NOTE 3 – Necessary precautions should be taken so that the terminal user may be able to break the sending circuit if oscillations occur, or to provide suitable methods so that a device controlled by the voice may prevent oscillations.

4.2 Receiving sensitivity

The receiving sensitivity of a hands-free telephone without automatic gain control should be adjustable within a range of 15 to 30 dB. This range should span the value of the Receiving Loudness Rating (RLR) which is equal to that of the corresponding handset telephone, as well as an RLR value about 10 dB lower.

NOTE 1 – Every precaution should be taken to ensure that the increase in gain due to the volume control does not allow the overhearing of other telephone conversations due to crosstalk.

NOTE 2 – In principle, the RLR of the HFT should be equal to the RLR of the corresponding handset telephone in a quiet room. The range of room noise levels met in normal office use necessitates, however, an additional gain of at least 10 dB.

For hands-free terminals equipped with an automatic gain control for the receive level (the gain being controlled by the incoming speech voltage), loudness ratings may not be applicable. In this case, the HFT should be designed so that the listening level at the maximum of the Overall Loudness Rating (OLR) of the connection for which the HFT is intended to be used can be preset to a value that may be considered as the best compromise between the levels required for listening in quiet and noisy rooms.

NOTE 3 – The preferred listening level depends on the room noise level as well as on other external conditions. There is, furthermore, a great variance between individual listeners.

The average preferred level for listening only appears to be a sound pressure level of about −29 dBPa for −49 dBPa(A) room noise, or −24 dBPa for −39 dBPa(A) room noise. However, to obtain maximum mean opinion scores in conversation tests, listening levels of about 5 to 10 dB higher may be required.

4.3 Frequency response curves

4.3.1 Sending

Available information indicates that the optimum slope of the sending response curve when measured with the HFT on a table lies between 0 and +3 dB/octave, if the receiving response curve is flat.

Only under highly reverberant conditions may a somewhat higher pre-emphasis increase the intelligibility. Therefore, if a frequency equalization is used to compensate the cable attenuation of the analogue part of a connection, the sending curve should not rise with frequency by more than 2-3 dB/octave.

Below 300 Hz there should be a gradual roll-off. The slope may be steeper below 200 Hz.

NOTE – The interval 200-300 Hz makes a significant contribution to the naturalness of the transmitted speech and should therefore be included in the transmission band of the HFT.

Above 4000 Hz, a roll-off by at least −6 dB/octave (preferably −12 dB/octave) is appropriate for analogue terminals in order to avoid interference by crosstalk to adjacent channels in certain types of long-distance circuits.

If the analogue terminals are intended to be connected through short lines to a digital connection, it is appropriate that the sensitivity above 4 kHz be as low as possible, to avoid spurious out-of-band signals.

For (300-3400 Hz) digital terminals, information is available in ITU-T P.342 [3].

For wideband digital terminals, information is available in ITU-T P.341 [4].

4.3.2 Receiving

The receiving response curve should be substantially flat in the frequency range of 200-4000 Hz.

The requirement refers to the sound pressure in the undisturbed field at the listener's position with a setup including the table as described in clause 5.

4.4 Subjective evaluations of loudness ratings

Loudness rating should be determined in accordance with ITU-T P.78 [5].

NOTE – Some information about reference equivalents can be found in the Handbook on Telephonometry.

4.4.1 Sending

The talking level for the measurement of Sending Loudness Rating (SLR) of an HFT should normally be the same as specified for measurements on handset telephones.

It is not necessary for the talker during the test to shift between the reference microphone guard-ring and the guard-ring positioned relative to the HFT if the obstacle effect of the reference microphone can be assumed to be negligible.

Normally the specified talking level and the use of a conventional test phrase or sentence should be sufficient to ensure that a voice-switched HFT will be in the sending condition during the determination of SLR. If this is not the case, the talking level may be increased by up to 5 dB, which may be compensated in the reference system to preserve the same listening level.

If the sending sensitivity is controlled by the room noise level, the subjective measurement should be done in a quiet environment [< –59 dBPa(A)]. Further information about the HFT performance may then be estimated by repeating the sending measurements with increasing levels of room noise, up to a maximum of –34 dBPa(A).

4.4.2 Receiving

The talking level at the reference microphone for the measurement of RLR should normally be the same as specified for the measurement of handset telephones. This should normally ensure that when loudness balance is achieved between the reference system and the test system path, a signal of sufficient magnitude is present at the HFT to switch it into the receive condition.

Problems can sometimes occur when approaching the balance condition from the condition of high attenuation in the balance attenuators, when the low level input signal may fail to switch the HFT into the receiving condition. If this does occur, the talking level may be increased by up to 5 dB in order to minimize the difference in loudness.

NOTE – The listening level will thus also increase at balance, but in this case it will not be possible to correct it by changing the reference system attenuator.

Obtaining the loudness balance for the receiving condition may be facilitated by use of a loudspeaking intermediate reference system. The specification of such a system is, however, outside the scope of this Recommendation.

4.5 Objective evaluations of sensitivity and loudness ratings

4.5.1 Sensitivity measurement

Objective evaluations of loudspeaker and hands-free terminals concern:

– the sending and receiving frequency sensitivity curves measurements;

– the objective determination of loudness ratings according to the method described in ITU-T P.79 [6].

4.5.1.1 Sending sensitivity measurements

The sending response curves of a hands-free telephone is recorded at the output terminals of the telephone with the same electrical connections as for the handset telephones. The acoustical input to the telephone microphone is supplied from an artificial mouth in the position shown in Figure 3.

In such a case, the sending sensitivity of the local telephone system $\mathbf{S_{mJ}}$ is expressed as dB relative to 1 V (electrical interface)/Pa (MRP) as follows:

$$\mathbf{S_{mJ}} = 20 \log \mathbf{V_s} - 20 \log \mathbf{P_{MRP}} + \mathbf{Corr} - 24$$

where:

$\mathbf{V_s}$ is the measured voltage across the appropriate termination (unless stated otherwise, a 600 ohms termination).

$\mathbf{P_{MRP}}$ is the applied sound pressure at the MRP.

Corr is 20 log (P_{MRP}/P_{HFRP}) of the used artificial mouth.

NOTE – The value of Corr is the value given in the calibration chart of the artificial mouth (24.0 dB is the ideal value).

4.5.1.2 Receiving sensitivity measurements

The receiving sensitivity of a loudspeaker and/or hands-free telephone is expressed as follows:

$$S_{Je} = 20\log_{10}\frac{p_R}{(1/2)E_J}\, dB\ rel\ 1\ Pa/V$$

where:

p_R is the sound pressure at point C in Figure 3 and E_J is the e.m.f. in the 600 ohms source.

4.5.2 Measure and computation of loudness ratings

4.5.2.1 Sending loudness rating

The computation of the sending loudness rating may be performed according to ITU-T P.79 [6] by using the frequency sensitivity response measured between the electrical output of the set and the acoustical sound pressure at the MRP.

NOTE – Appropriate types of signals can be found in ITU-T P.501 [9].

4.5.2.2 Receiving loudness rating

Objective measurements described in 4.5.1.2 are made with a free field microphone at point C (see Figure 3).

Loudness ratings are computed following ITU-T P.79 [6].

When calculating the RLR of a hands-free or loudspeaking telephone, a correction factor of −14 dB applies and the L_e factor used in the algorithm of ITU-T P.79 [6] is equal to 0.

4.6 Switching parameters

Most loudspeaker and hands-free terminals contain voice-switched circuits whose main function is to avoid singing through acoustic feedback. Such circuits insert a loss in either the sending or receiving direction in various ways. Switching from one direction to the other occurs when a signal above a given threshold is applied from the opposite direction, or when the control circuit, taking into account the relative levels and the nature of the signals in both directions, allows the switching.

The fundamental voice-switching parameters of the switching function are defined as follows (see Figures 1 and 2):

- Threshold level V_{TH} – Minimum necessary signal level for removing insertion loss.
- Build-up time T_R – Time from the input signal going above the threshold level until the time at which the output level reaches 3 dB below complete removal of the insertion loss.
- Hang-over time T_H – Time from the input signal going below the threshold level until 3 dB of the switched loss is inserted in the output signal.
- Switching time T_S – Time from one transmission direction to the other. T_S is measured from the removal of the signal in the first direction until the level in the second direction reaches 3 dB below its final value.

By a suitable choice of parameter values, the degradation of speech quality that is introduced by voice switching can be made negligible, while an inadequate choice of parameter values, Switching Times in particular, may lead to serious clipping effects and loss of initial or final consonants in the transmitted speech.

The following values or guidance are recommended:

Threshold levels should be chosen so that switching is not interrupted by random (environmental) noise sources at either end of the call. In addition, ambient room/network noise effects on threshold should not impair performance. Ambient noise levels can be used to improve threshold performance, as talkers tend to speak louder in a noisy environment than in a quiet one.

Build-up time should be short enough so that the initial transient components of speech are not lost, but not so short that insertion loss removal would be noisy. The build-up time T_R should be less than 15 ms, preferably below 10 ms.

Hang-over time should be long enough to cover average pauses in speech so that intermittent unwanted switching does not occur before the initial talker is finished, but short enough to allow for reasonable break-in from the second talker. The hang-over time T_H is defined in Table 4.

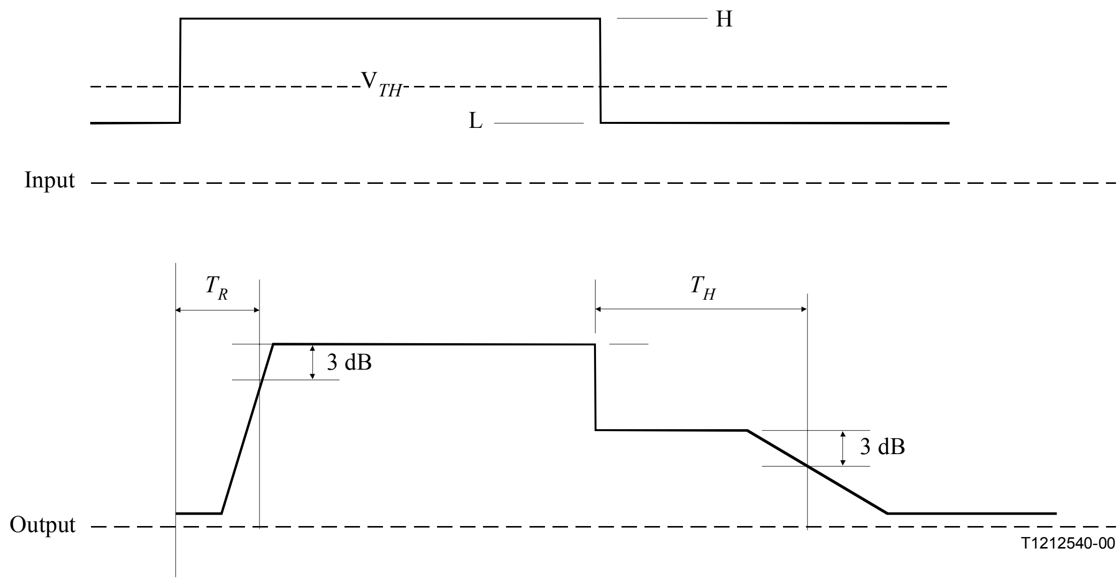
Switching time from one active state to the other should be balanced to best simulate full duplex operation. Switching time is also dependent on both build-up time and hang-over time. The switching time T_S should be approximately 100 ms.

Measurements of voice switching characteristics may be divided into those dealing with:

a) Characteristics for single talk situations, in which two parties communicate by alternating speech spurts without interrupting each other. In this case, it may be assumed that the voice switch circuit returns to an idle state before being activated by an input signal in either direction.

b) Characteristics for double talk conversation, in which both parties may interrupt each other by simultaneous talk, or where speech at one end of a connection is present simultaneously with noise at the other end.

If single talk situation is of fundamental importance, it has been shown that the behaviour of hands-free terminals in double talk situation affects seriously the overall quality.

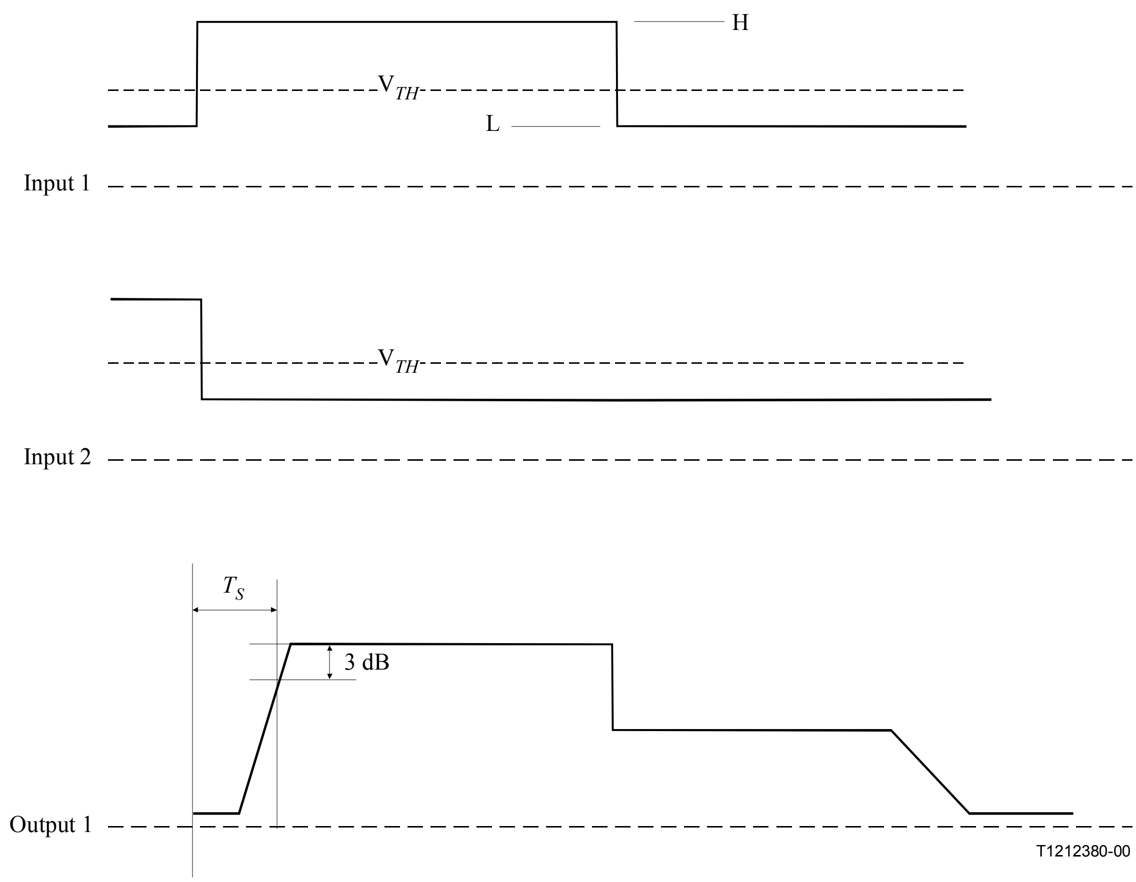


T_R Build-up Time

T_H Hang-over Time

V_{TH} Threshold Level

Figure 1/P.340



T_S Switching Time

Figure 2/P.340

5 Testing conditions

For both subjective and objective measurements, physical test arrangements as described in this clause should be used.

5.1 Test table

During the measurements, the HFT is placed on a table defined as follows:

The surface of the table should be hard (e.g. polished marine plywood or suitable hardwood), flat, rigid and horizontal to provide a sound-reflecting surface on which the HFT being tested rests. The dimensions of the table should be such that the surface area is about 1 m^2 but not less than 0.96 m^2 and the width not less than 800 mm [1].

NOTE – This arrangement should be used for all measurements, including the recording of frequency responses, although diffraction effects due to the table are likely to cause severe dips or peaks in the response curve.

5.2 Test arrangement

The physical test arrangements of one piece (single unit) HFTs for subjective and objective measurements are shown in Figure 3.

If the projections of the housing are not rectangular, the point B is positioned at the crossing of the centre line through the housing and the outline of the vertical projection of the housing.

The edge of the front of the box should be perpendicular to the line A-B.

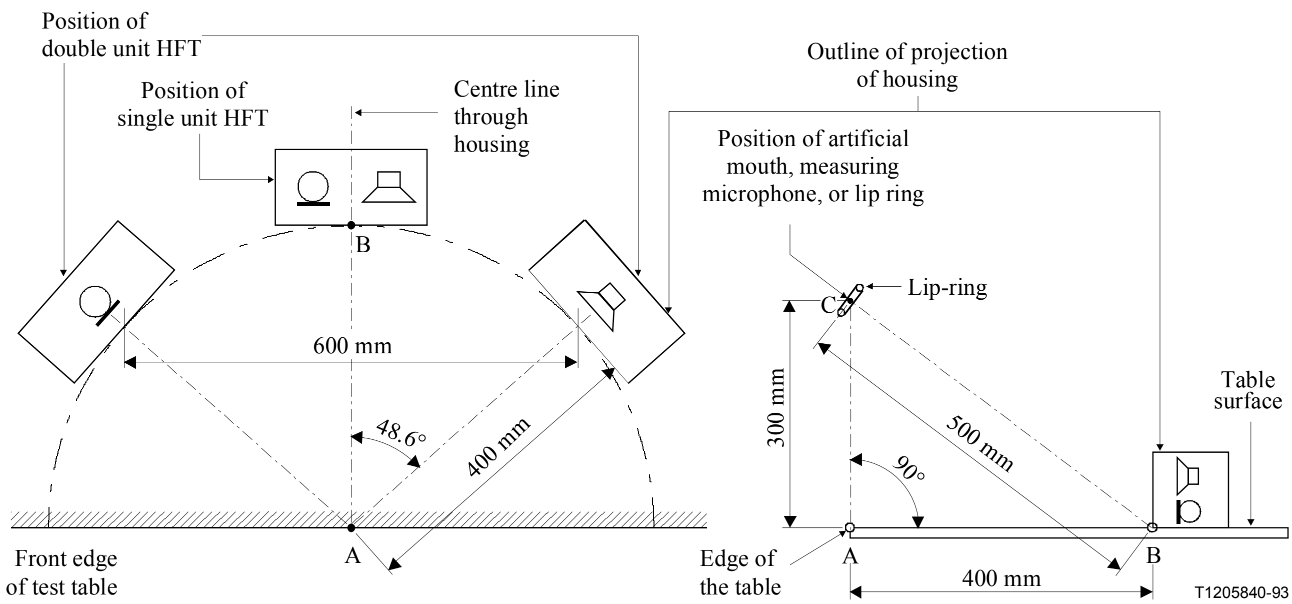


Figure 3/P.340 – Physical test arrangements for subjective and objective measurements

For more-than-one-piece HFTs, videophony and multimedia terminals, the test arrangement shall be modified to what it is stated in the instruction manual of the Terminal Under Test (TUT). The positioning of the terminal shall be referred to the point C defined and located as in Figure 3. Sending and receiving loudness ratings, sending and receiving sensitivity responses shall be tested and adjusted for this test arrangement. The axis of the artificial mouth shall be defined by manufacturer's declaration.

If there is no manufacturer's declaration, the test arrangement of Figure 3 shall be implemented.

NOTE – When a HATS is used, the test arrangement defined in ITU-T P.581 [21] applies.

5.3 Calibration of the artificial mouth

ITU-T P.51 [8] defines the tolerances that apply to the acoustic radiation of the mouth. To reduce the possible errors that could be due to these tolerances, the following procedure shall be used to calibrate the artificial mouth.

The spectrum of acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP.

The level of the acoustic signal is adjusted to −4.7 dBPa at the MRP.

The spectrum at MRP is then recorded and used as a reference.

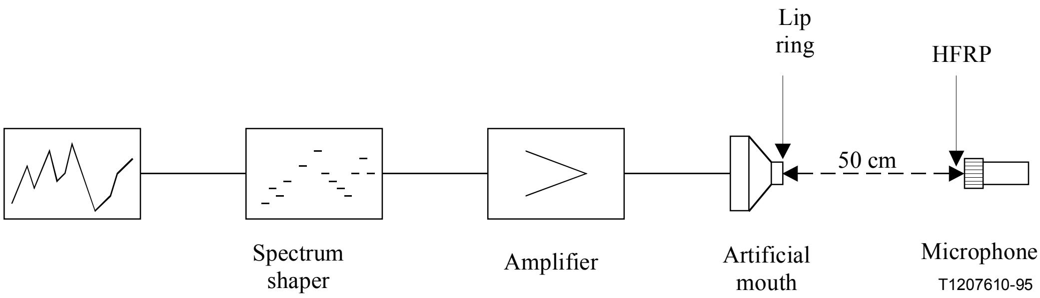


Figure 4/P.340 – Physical test arrangement for calibration of the artificial mouth

The broadband signal level then is adjusted to –28.7 dBPa at the HFRP and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in third octaves) is used as reference to determine the sending sensitivity S_{mJ} (see Figure 4).

NOTE – When a HATS is used, the calibration defined in ITU-T P.581 [21] applies.

5.4 Test environment

1) For the repeatability of the tests, the environment for most of the measurements shall be free field (anechoic) down to the lowest frequency of the 1/3 octave band centred on 200 Hz.

   Satisfactory free field conditions exist where errors, due to the departure from ideal conditions, do not exceed the values defined in Table 1, inside a sphere centred on point B (see Figure 5), with one metre radius, in absence of the table.

Table 1/P.340

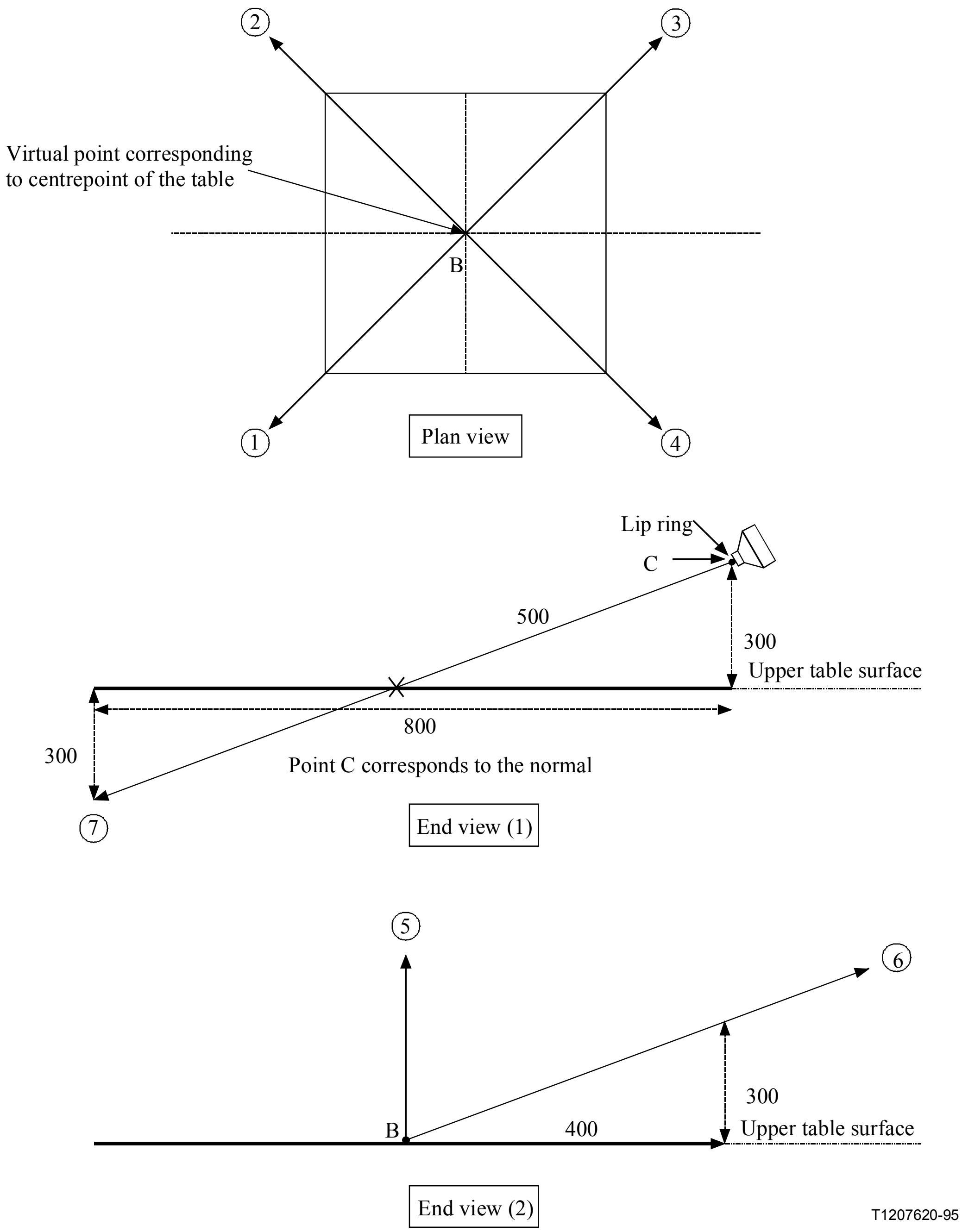
1/3 octave band centre frequency (Hz)	Allowable departure (dB)
<630	±1.5
800 to 5000	±1
>6300	±1.5

   The test signal level for verification of the free field is –20 dBPa.

   Measurements are made along the seven axes which are numbered ① to ⑦ in Figure 5, with the sound source placed at positions equivalent to B or C, as appropriate. Measurement points along each axis, taken from the front plan of the artificial mouth lip-ring are at the distances of 315 mm, 400 mm, 500 mm, 630 mm, 800 mm and 1000 mm.

   NOTE – Complementary information for test rooms is available in 10.3.1.

2) The broadband noise level shall not exceed –70 dBPa(A). The octave band noise level shall not exceed the values specified in Table 2.



Dimensions in millimetres.
Points 1, 2, 3 and 4 are in the horizontal plane normally occupied by the table surface.
Measurements of free field sound pressure are made in absence of the table.
Axes used in the determination of free field conditions for 1 m radius sphere.

Figure 5/P.340 – Determination of the free field conditions

Table 2/P.340 – Noise level

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

5.5 Test signals

The test signals shall comply with the contents of ITU-T P.501 [9].

5.6 Test signal levels

Measurement level, unless otherwise specified, corresponds to the standard level. This is –28.7 dBPa at HFRP for send, –30 dBm0 for receive.

6 Conversational quality

In the case of many of the hands-free terminals which have been hitherto available, the advantages over handsets have usually meant a sacrifice in terms of a reduction in speech transmission quality. The parameters relevant to a quality description were obtained from listening tests. The instrument-based parameters to be correlated were extracted. Hands-free terminals are not linear, time-invariant systems, and newly developed measurement engineering techniques for measuring transmission response must be additionally applied. A comprehensive definition of quality for hands-free terminals requires the adaptation or extension of current test measurements.

Subjective evaluation for hands-free terminals are defined in ITU-T P.832 [13].

The parameters "background noise transmission" and "duplex capability" are the most important auditive characteristics and are crucial in deciding the subjective impression of quality of hands-free equipment. The most important value measurable in terms of instrumentation, correlating to these auditive values, is, in the case of duplex capability, the attenuation range, and for background noise transmission the attenuation distribution at idle. These parameters provide a first indication of the product quality of hands-free equipment and allow a classification of the equipment. Three types can be included in such a classification:

1) *Full duplex capability*

   In duplex operation, attenuation of the conversation partner either does not occur or is unnoticeable. Background noise is transmitted in the send path (with or without room noise reduction).

2) *Partial duplex capability*

   In duplex operation, one or both of the speech paths is attenuated. The conversation partner and background noise in the send path are still audible.

3) *No duplex capability*

When one conversation partner is talking, the other is fully attenuated. Background noise in the send direction is not transmitted.

7 Parameters measured instrumentally

Besides the transmission parameters defined in the preceding subclauses, the following objective parameters influence significantly the overall quality of the communication using hands-free terminals.

– attenuation range, a_H;
– attenuation distribution at idle mode;
– hang-over time;
– dynamic compression;
– reverberance: impulse response, EEB;
– time-dependent frequency response;
– duplex behaviour;
– echo cancellation (see clause 10);
– quality of speech recognition.

The appropriate test procedures are defined in ITU-T P.502 [12].

7.1 Attenuation range a_H

Attenuation range is determined by the difference in sensitivity response which results when one speech path is activated and when the duplex branch is activated. An activating signal (e.g. voiced sound of CSS) is used as the receive measurement signal, immediately followed by an activating signal in the send direction. In the analysis the level of the measurement output signal is represented versus time (time constant = 5 ms). The attenuation range is obtained from the difference between the maximum level at full activation and the minimum level obtained immediately after switch-over.

7.2 Attenuation distribution at idle mode

The attenuation range in idle is derived from the sum of the maximum attenuation in the send direction $a_{H,S}$ and the maximum attenuation in the receive direction $a_{H,R}$. In the case of a system with no compressor, $a_H = a_{H,S} + a_{H,R}$ = constant applies. (In the case of a system including a compressor, the attenuation at maximum and minimum level due to the compressor must be obtained, see 7.4.)

For measuring attenuation distribution a signal can be used which consists of a train of activating signals (e.g. voiced sound of CSS) and pauses in both speech directions, whereby extremely large variation in transition combinations are possible. The attenuation value can thus be obtained at the appropriate transition points from idle to activation for each of the speech paths. Again, as in 7.1, measurement signal level is represented versus time.

Attenuation in the send direction $a_{H,S}$ can be read off from idle to send at the transition point, and is derived from the difference between the maximum level at full activation in the send direction and the minimum level obtained immediately after the send activation signal is present. The attenuation $a_{H,R}$ can be correspondingly read off at the transition idle/receive activation signal.

If the switch-on time of the equipment is very fast, thus making reading off difficult, the measurement technique described in 7.3 (Hang-over time) can be used. The attenuation $a_{H,S}$ or $a_{H,R}$ of each speech path can be derived by calculating the difference in level at maximum and minimum attenuation in the second part of the signal (noise signal).

7.3 Hang-over time

The transition from activation to idle can be represented by feeding in an activation signal (e.g. voiced sound of CSS) in one direction, followed by a second signal in the same direction but of lower level, which does not activate the hands-free telephone (noise signal). The second part of the signal measured thus indicates the attenuation, from which the Hang-over time (switch-off time) can be determined.

The duration of the voiced sound is 0.5 s in order to reach a final stable system condition. The level corresponds to standard level, as defined in 5.6. The second part of the signal (noise signal) has a duration of 1 s. The level must be selected low enough so as not to activate the equipment. The suggested levels to be applied are: −58.7 dBPa at HFRP for send, and −60 dBm0 for receive.

7.4 Dynamic compression

A companding or AGC response (e.g. the range of level adjustments as a function of input signal level) can be measured using an activation signal with steeply monotone rising or falling level response. The compressor is active during the period where the output signal has a constant level. The level limits within which the compressor operates can be obtained from the original signal. The compression range is obtained as the difference between the level limits. The attack time should be rather short and the release time should be sufficiently long.

Appendix I gives some more information.

Using a test signal with level steps, it is possible to determine time response of the companding or AGC (e.g. time duration for level adjustments).

7.5 Reverberance

The impulse response in the send direction is measured using maximum-length sequences, as described in [IV.5] of Appendix IV. The test signal should be composed of an activation signal (e.g. voiced sound of the CSS) and a maximal-length sequence of several periods (one segment is shown in Figure 6). The period length must be longer than the length at the impulse response of the system under test. The impulse response is calculated from the average of several periods, beginning with the second period. From this the Early Energy Balance (EEB) is calculated. The EEB represents fairly well the subjective impression of reverberance ([IV.6] of Appendix IV):

$$EBB = 10 \cdot \log \left\{ \frac{\int_0^{35ms} h^2(t)dt}{\int_0^{5ms} h^2(t)dt} \right\} dB$$

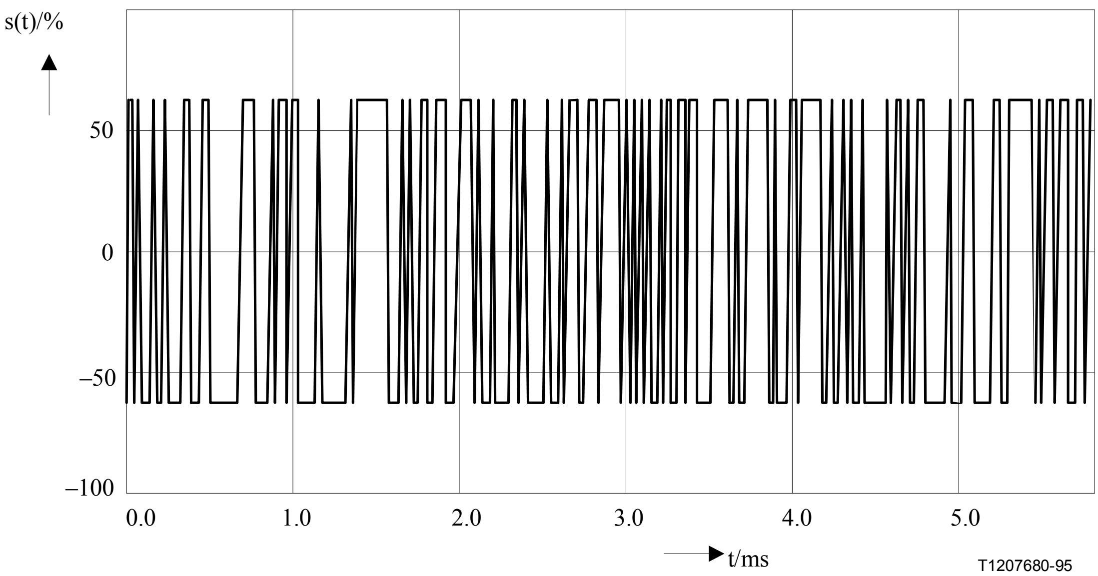


Figure 6/P.340 – Segment of maximum-length sequence

7.6 Time-dependent transfer function

In analysis, the output signals are referred to the corresponding original signals. For display of the time and frequency dependent structure of the output signal referred to the input signal, a spectrographic display may be selected where the x-axis represents time, the y-axis represents the frequency and the level is represented in different colours. In this way, any possible time or frequency variant structures can be revealed.

7.7 Duplex behaviour

Special duplex speech signals can be fed in both directions (send and receive) for investigating the duplex response. Appropriate signals are those of differently varying level, (e.g. speech sequence of fixed definition or sequences from CSS of fixed definition) fed in after each other with varying periodicity in the send and receive directions.

A double talk sequence as described in ITU-T P.501 [9] should be used. The corresponding analysis techniques are given in ITU-T P.502 [12].

The output signals are referred to the corresponding original signals and spectrographically displayed as described in 7.6. This allows to represent how often gaps or signal interrupting occur during transmission. Conclusions can thus be reached regarding the duplex capability of the equipment.

The duplex behaviour of a hands-free terminal is detailed in clause 9.

7.8 Control characteristics of echo cancellation

Subjective quality is determined by the following parameters:

– time-variable echo attenuation;
– echo attenuation over frequency;
– time response of NLP (Non-linear Processor), centre clipper;
– response in duplex operation;
– attenuation response with background noise.

Clause 10 gives more details on these characteristics.

7.9 Speech detection

Speech detection is measured in the send direction. From subjective listening and conversation tests ([IV.8] of Appendix IV), speech detection in the sending path appears as the more critical of the two transmission paths.

For terminals incorporating level detection, the level measured is the minimum at which the equipment switches through.

NOTE – Other levels and types of background noise may be used depending on the application and environment where the HFT is used.

7.10 Speech detection with ambient noise

Measurement is done in the send direction. A 10 s sequence of a complex signal should be used.

A minimum of four loudspeakers, fed with non-coherent noise signals, are simulating a room noise. They are positioned according to Figure 7. The resulting signal is measured at the point located in the middle of the test table (in absence of the table), the spectrum is a Hoth noise ([IV.7] of Appendix IV), and the level is set to −44 dBPa(A).

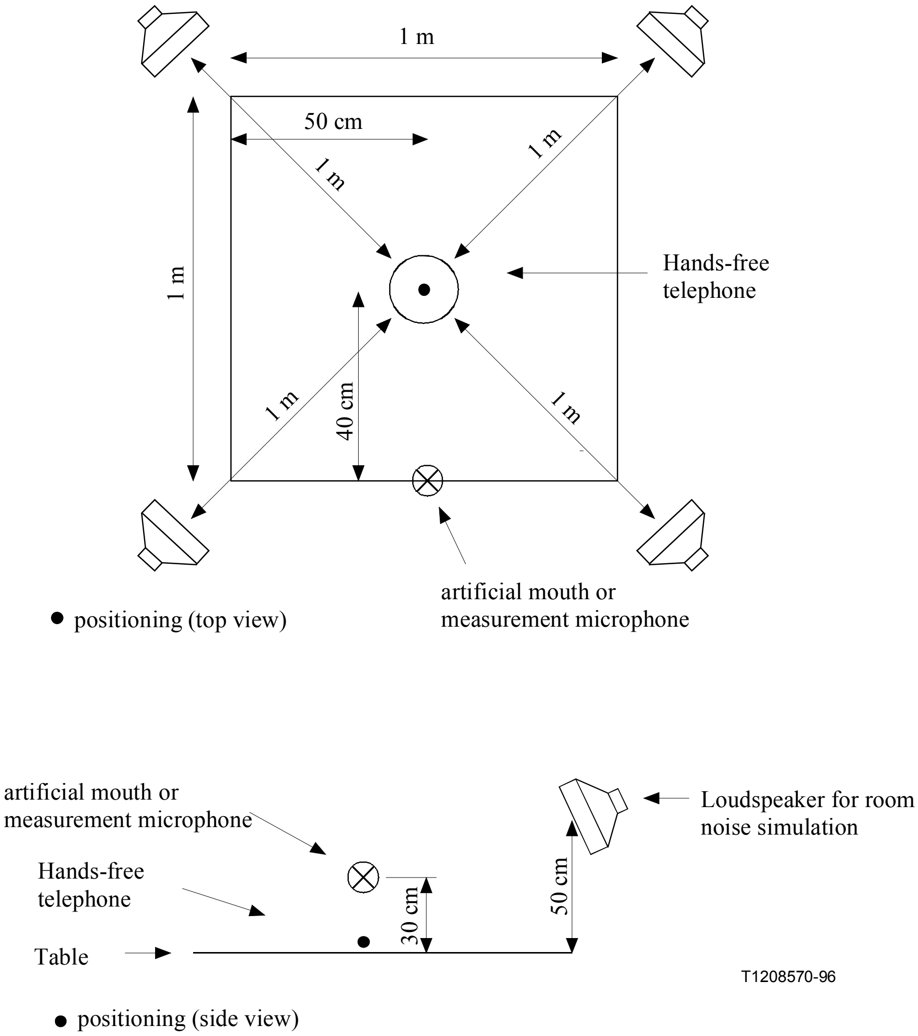


Figure 7/P.340 – Positioning of the equipment

7.11 Background noise transmission

The sources of the transmitted noise may be:

– *Acoustical sources*

Ambient noise in the test room (typically as defined in 5.4. If other situations are met, information should be given in the testing report);

Noise intentionally produced in the test room (e.g. babble noise, car noise, etc.).

– *Electrical sources*

Noise generated by the components of the terminals or/and network;

Noise intentionally generated (e.g. comfort noise, etc.).

Background noise transmission refers to the "noise intentionally produced in the test room" picked up by the microphone(s) of the hands-free terminal. See Table 3.

Table 3/P.340 – Parameters associated to background noise transmission, impacting on the quality

Parameters	Examples of recommendations
Absolute level	The transmitted background noise level should be as low as possible. The background noise transmitted signal should not be interrupted from time to time.
Level fluctuations	The level fluctuations should not be more than ±3 dB, compared to steady state conditions. NOTE – This does not apply if the residual background transmitted noise is masked (e.g. by the speech signal).
Additional parameters	Artefacts of noise reduction algorithms need to be avoided.

8 Behaviour of hands-free terminals in double talk situation

8.1 Categorization

Hands-free terminal should be categorized on the basis of "duplex capability". From [IV.9] and [IV.12] of Appendix IV, it appears that these characteristics are the most important on the perceived quality of the HFT, and that they may be characterized by the attenuation range.

So, the following categories, numbered 1, 2 (2a, 2b, 2c) and 3, associate the behaviour of the hands-free terminal and the attenuation range. See Table 4.

Table 4/P.340

Sending direction	Receiving direction
Behaviour 1: $a_{H,S,DT} \leq 3$ dB	Behaviour 1: $a_{H,R,DT} \leq 3$ dB
Behaviour 2a: 3 dB < $a_{H,S,DT} \leq 6$ dB	Behaviour 2a: 3 dB < $a_{H,R,DT} \leq 5$ dB
Behaviour 2b: 6 dB < $a_{H,S,DT} \leq 9$ dB	Behaviour 2b: 5 dB < $a_{H,R,DT} \leq 8$ dB
Behaviour 2c: 9 dB < $a_{H,S,DT} \leq 12$ dB	Behaviour 2c: 8 dB < $a_{H,R,DT} \leq 10$ dB
Behaviour 3: $a_{H,S,DT} > 12$ dB	Behaviour 3: $a_{H,R,DT} > 10$ dB

Annex A gives information on the categories and the test results used for their definition.

8.2 Terminal coupling loss in double talk situation

From the subjective tests it appears that the echo loss is also a parameter with a great influence on the quality perceived in double talk situation.

From the tests described in Annex A, the following categorization has been established for echo loss in double talk:

Behaviour 1: TELRDT ≥ 37 dB

Behaviour 2a: 37 dB > TELRDT ≥ 33 dB

Behaviour 2b: 33 dB > TELRDT ≥ 27 dB

Behaviour 2c: 27 dB > TELRDT ≥ 21 dB

Behaviour 3: TELRDT < 21 dB

9 Parameters to be evaluated for each type of HFT

The requirement of good duplex capability, the quality of room noise transmission in the send direction and other auditory quality factors result in specific requirements which should be met for each group. Different parameters are given priority in each type of equipment, according to the different quality level attaching to each type, each of which can be measured to different measurement regulations.

Table 5 defines some parameters, guidance and limits. Complementary parameters and limits are given for information in Appendix I.

Table 5/P.340

Behaviour 1 **full duplex capability**	**Behaviour 2** **partial duplex capability**	**Behaviour 3** **no duplex capability**
Reverberance The EEB value should be as small as possible.	Reverberance The EEB value should be as small as possible.	Reverberance The EEB value should be as small as possible.
Delay Extra delay for processing is defined in clause 10.	Delay Some extra delay for processing could be accepted; the maximum value is defined in clause 10.	Delay No extra delay for processing.
Time and frequency variable transfer function The transfer function is measured according to 7.6 and spectrographically displayed. The transfer function should show a constant frequency dependent response over time within ±3 dB, measured in 1/12 octaves. In the case of time and frequency variable response, auditory evaluation of the system is required.		
	Quality of speech detection in the send direction The minimum threshold level should be < –30 dBPa at HFRP.	Quality of speech detection in the send direction The minimum threshold level should be < –30 dBPa at HFRP.
	Hang-over time Requirement: T_H should be greater than 50 ms, preferably more than 100 ms.	Hang-over time Requirement: T_H should be greater than 250 ms.

Table 5/P.340 ***(concluded)***

Behaviour 1 **full duplex capability**	**Behaviour 2** **partial duplex capability**	**Behaviour 3** **no duplex capability**
	Dynamic compression Control through the use of a compressor or AGC should be undetectable. The compressor control range should be small.	Dynamic compression In the case of hands-free terminals without reliable noise detection, noise control through a compressor should not be annoying.
		Room noise detection Room noise detection must be reliable. Hoth noise ([IV.7] of Appendix IV) with a level L_N at the hands-free equipment (HFRP) is generated. No accidental switch through the hands-free equipment should be registered at a level of $L_N = -50$ dBPa(A) over a period of 10 s.

10 Acoustic echo controllers and speech enhancement devices

10.1 Functional units

The functional units of an AEC are devices or parts of devices implemented in the processing unit, which contribute to the general function of acoustic echo control. There is no restriction on how to implement them.

The functional units can be combined for better performance. They can use all the available signals in the terminal (for example, the individual signals coming from several microphones arranged in an acoustic array). Moreover, they can be mixed with other functions (for example sub-band speech coding) for efficient implementation, provided that they do not modify the proper characteristics of these functions when they are operating.

A functional block-diagram of a typical processing unit is shown in Figure 8.

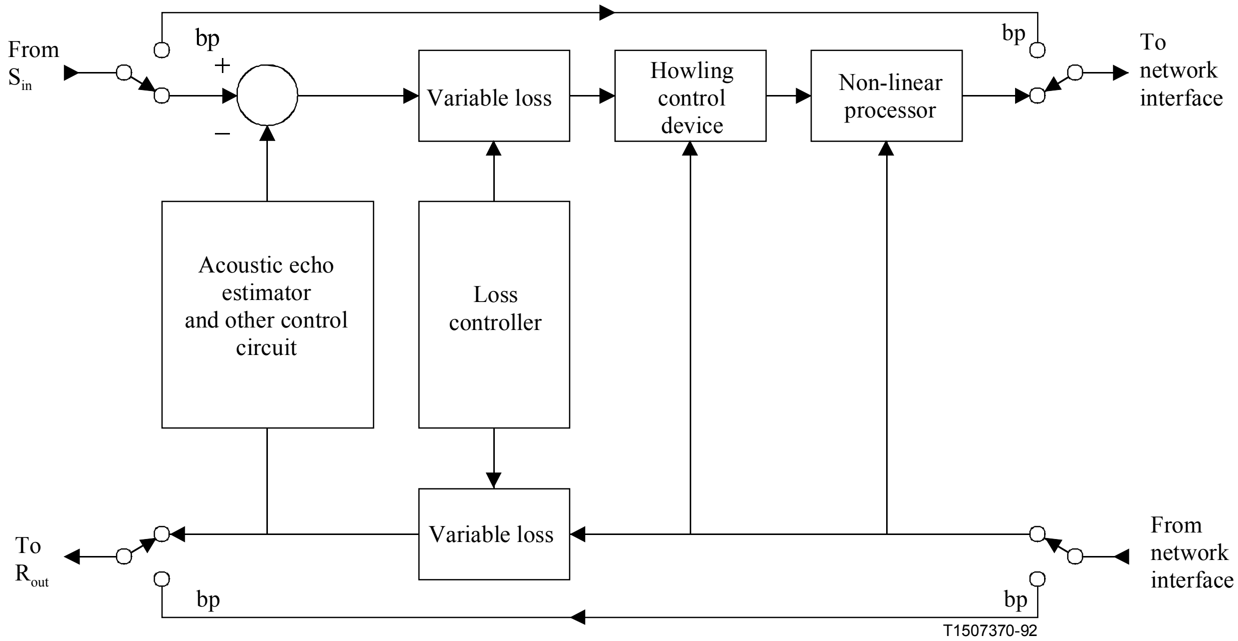


Figure 8/P.340 – Functional block-diagram of a typical processing unit (AEC part) (bp denotes bypass signal paths for testing purposes)

NOTE – Interactions between terminal AEC and network signal processing equipment are considered in I.6.6 of Appendix I/G.168 [20].

10.2 Delay

The values specified below correspond to the extra delay which can result from the AEC and noise reduction systems processing. In any case, compliance with transmission planning objectives must be achieved.

General information about transmission delays can be found in ITU-T G.114 [15]; ITU-T G.131 [16] provides rules for echo control in the network.

The sum group delays for mobile terminals, from the mouth reference point to the network interface and from the network interface to the ear reference point, of less than 20 ms delay is desirable (ITU-T G.174 [17]).

10.2.1 Processing delay

The echo cancellation and/or the noise reduction processes need some time. This time creates delay in the terminal, called below as "processing delay".

Examples of values for mobile hands-free terminal are given in Appendix II.

10.2.2 Round trip echo path delay (EPDn) – network interface

Echo audibility is dependent upon the round trip delay in echo path. This requires calculation of impulse response (echo path). To that end, stationary broadband noise should be used. The terminal can be placed in nearly any reverberant or non-reverberant environment, as the first acoustic echo will be due to direct coupling. Acoustic noise should meet the requirements defined in 5.4. AEC is first trained using a signal at R_{in}.

10.3 Acoustic echo control specifications

Performance requirements can also be found in clause 8 and test methods in ITU-T P.502 [12].

10.3.1 Acoustic echo path

The use of real rooms or enclosures with appropriate acoustic characteristics is recommended. Echo paths simulated by electronic devices like digital reverberators with non time-varying reflection patterns can be used as well if the terminal has internal accesses on the user side. In this latter case, the electronic simulator adjustments should comply with the values recommended for real rooms or enclosures; moreover, the shape of the simulated impulse response envelope should be similar to the real echo path impulse response.

– For teleconference systems, the reverberation time averaged over the transmission bandwidth shall be typically 400 ms; the reverberation time in the lowest octave shall be no more than twice this average value; the reverberation time in the highest octave shall be not less than half this value. The volume of a typical test room shall be of the order of 90 m^3.

– For hands-free terminals and videophones, the reverberation time averaged over the transmission bandwidth shall be typically 500 ms; the reverberation time in the lowest octave shall be no more than twice this average value; the reverberation time in the highest octave shall be not less than half this value. The volume of a typical test room shall be of the order of 50 m^3.

– For mobile radio terminals an enclosure simulating the interior of a car can be used; a real car can be used as well. A typical average "reverberation time" is 60 ms. The volume of the enclosure shall be of the order of 2.5 m^3.

NOTE – It is recommended to avoid extremely long rooms (Length >> Width, Height) and rooms with extremely low ceilings (Height << Length, Width), and preferably also rooms with all the side dimensions nearly identical.

Large, flat, parallel room limiting surfaces and surface areas that provide broadband sound reflection, particularly wall surfaces at an average room height (roughly 0.8 m to 1.8 m above the floor) should be avoided, since they can cause flutter echoes and flutter-echo-like disturbances (echoing, roughness), if the test setup is in an unfavourable position.

Measuring the local frequency-dependent distribution of sound pressure levels within a selected room in the steady state can help to determine the optimum position of the test setup.

As a general suggestion, the minimum distance between the test setup and room limiting surfaces should be 1 m, regardless of the acoustic properties of these surfaces. This can prevent disturbances due to initial reflections and a rise in sound pressure level that can occur locally at low frequencies. The same recommendation applies to geometrically large furniture surfaces that reflect sound.

10.3.2 Parameters and recommended limits

10.3.2.1 Weighted terminal coupling loss – single talk (TCLwst)

The weighted loss between the R_{in} and S_{out} network interfaces when the AEC is in normal operation, and when there is no signal coming from the local user[1].

Before each test the terminal is switched on.

---

[1] The weighting is made according to the rule specified in ITU-T G.122 (computation of talker echo loudness rating). Care must be taken to avoid possible masking of singing effects by the weighting (under study).

The recommended values for each type of hands-free terminal can be found in the relevant ITU-T Recommendations (e.g. P.341 for wideband hands-free terminal, P.342 for digital hands-free terminal).

10.3.2.2 Weighted terminal coupling loss – double talk (TCLwdt)

The weighted loss between the R_{in} and S_{out} network interfaces when the AEC is in normal operation, and where the local user and the far-end user are active simultaneously[1].

The recommended values for each type of hands-free terminal can be found in clause 8.

10.3.2.3 Received speech attenuation during double talk (Ardt)

The received signal attenuation (at the R_{out} point) which is inserted by the AEC during double-talk events.

The frequency response on the receive side during double talk is under study.

10.3.2.4 Sent speech attenuation during double talk (Asdt)

The sent signal attenuation (at the S_{out} point) which is inserted by the AEC during double-talk events.

The frequency response on the send side during double talk is under study.

10.3.2.5 Received speech distortion during double talk (Drdt)

The total non-linear signal distortion at the R_{out} point which can be produced by the AEC during double-talk events.

For all the applications, the supplementary distortion at R_{out} in comparison with single talk conditions should be low. The values of Drdt are under study.

10.3.2.6 Sent speech distortion during double talk (Dsdt)

The total non-linear signal distortion at the S_{out} point which can be produced by the AEC during double-talk events.

For all the applications, the supplementary distortion at S_{out} in comparison with single talk conditions should be low. The values of Dsdt are under study.

10.3.2.7 Build-up time – single talk (TRnst)

The time interval between the onset of the received signal (similarly the transmitted signal) and the instant when the attenuation on the receive path (similarly on the send path) reaches [3] dB. For this purpose, the other side is quiet.

10.3.2.7.1 Receive side (TRst-r)

For all the applications, TRst_r shall be no more than [20 ms].

10.3.2.7.2 Send side (TRst-s)

For all the applications, TRst_s shall be no more than [20 ms].

10.3.2.8 Build-up time – double talk (TRdt)

The time interval between the onset of the received signal (similarly the sent signal) and the instant when the attenuation on the receive path (similarly on the send path) reaches the value Ardt (similarly Asdt). For this purpose, the signal in the opposite direction of transmission is held at a specified level.

10.3.2.8.1 Receive side (TRdt-r)

TRdt-r should be less than [20 ms], if the attenuation is more than 6 dB.

10.3.2.8.2 Send side (Tondt-s)

TRdt-s should be less than [20 ms], if the attenuation is more than 6 dB.

10.3.2.9 Convergence time (Tc) – Initial Convergence Time (Tic)

Convergence Time is the time interval between the instant when a specified test signal is applied to the R_{in} port of the terminal (after all the functions of the AEC have been reset and then enabled), and the instant when the returned echo signal at the S_{out} port is attenuated by at least a pre-defined amount. The local user is not active.

Initial Convergence Time is the Convergence Time evaluated after switching on the HFT.

10.3.2.10 Hang-over Time after double talk (THdt)

The time elapsed between the end of a double talk event and the instant when the attenuation of the echo recovers a specified value (a signal is received continuously from the distant user).

For all the applications, the attenuation of the signal at S_{out} should be at least [20 dB] after THdt = [1] second.

10.3.2.11 Terminal coupling loss temporally weighted – single talk (TCLtst)

The echo return loss from R_{in} to S_{out} is measured according to the procedure defined for ERLtst in ITU-T Recommendation P.502 [12].

10.3.2.12 Terminal coupling loss temporally weighted echo return loss – double talk (ERLtdt)

The echo return loss from R_{in} to S_{out} is measured according to the procedure described for ERLtdt in ITU-T Recommendation P.502 [12].

ANNEX A

Test results

The test results presented in this annex are extracted from [IV.10] and [IV.11] of Appendix IV.

A.1 Categorization of hands-free terminals based on double talk performance

The results on attenuations in sending and receiving directions (Figures A.1, A.2) and on echo disturbance (Figures A.3, A.4 and A.5) have been determined from Third Party Listening Test results.

The hands-free terminal is simulated.

NOTE – It should be noted that the investigation to assess the echo disturbances during double talk are based on a one-way delay of 100 ms.

A.1.1 Results of the listening only tests with variable attenuations in sending path

The test parameter is attenuation (dB) in the sending path. Long double talk sequences are used for this test. The tests have been made for different values of the attenuation, and the curves of Figure A.1 have been interpolated from these results. A long double talk sequence is judged.

The upper curve shows the perceived quality (in MOS) on the loudness variation scale, the lower curve the overall quality.

Figure A.1 shows the perceived quality (on a MOS scale) test conditions listening only tests.

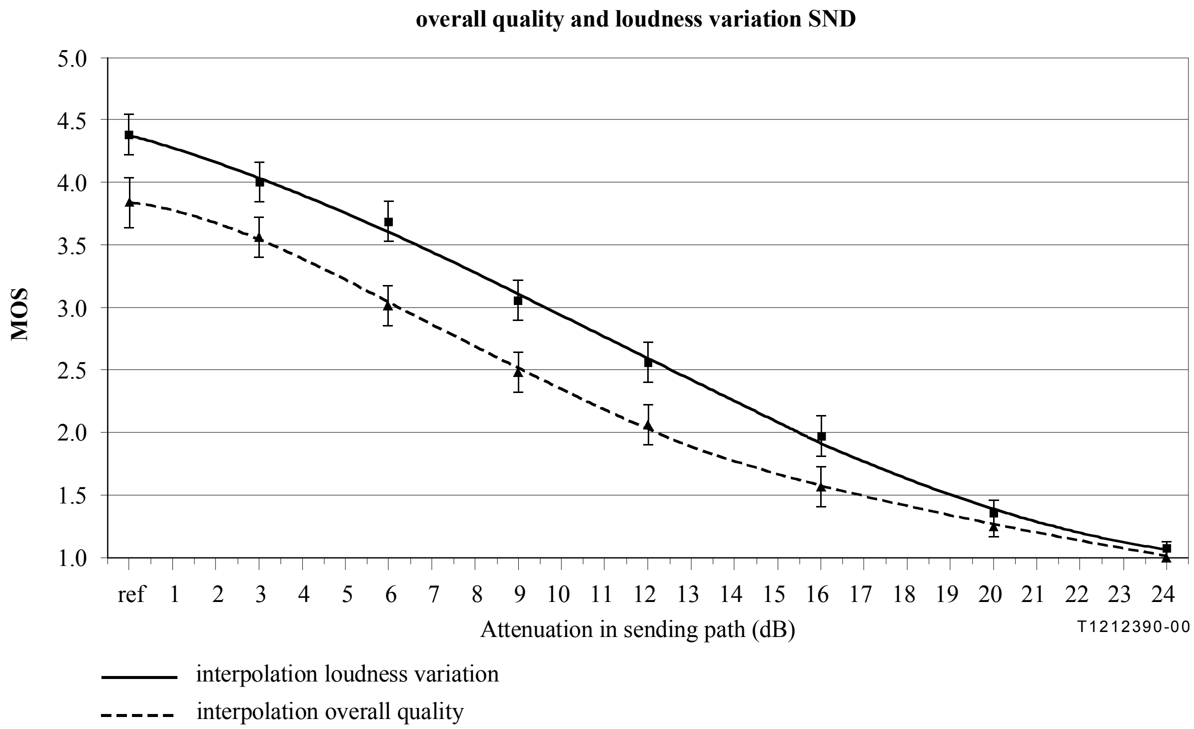


Figure A.1/P.340 – Interpolation of test results for overall quality and loudness variation in sending direction (upper line: loudness variation)

A.1.2 Results of the listening only tests with variable attenuations in receiving path

The test parameter is attenuation (dB) in the receiving path. Long double talk sequences are used for this test.

The tests have been made for different values of the attenuation, and the curve (perceived quality (in MOS) on the loudness variation scale) of Figure A.2 have been interpolated from these results.

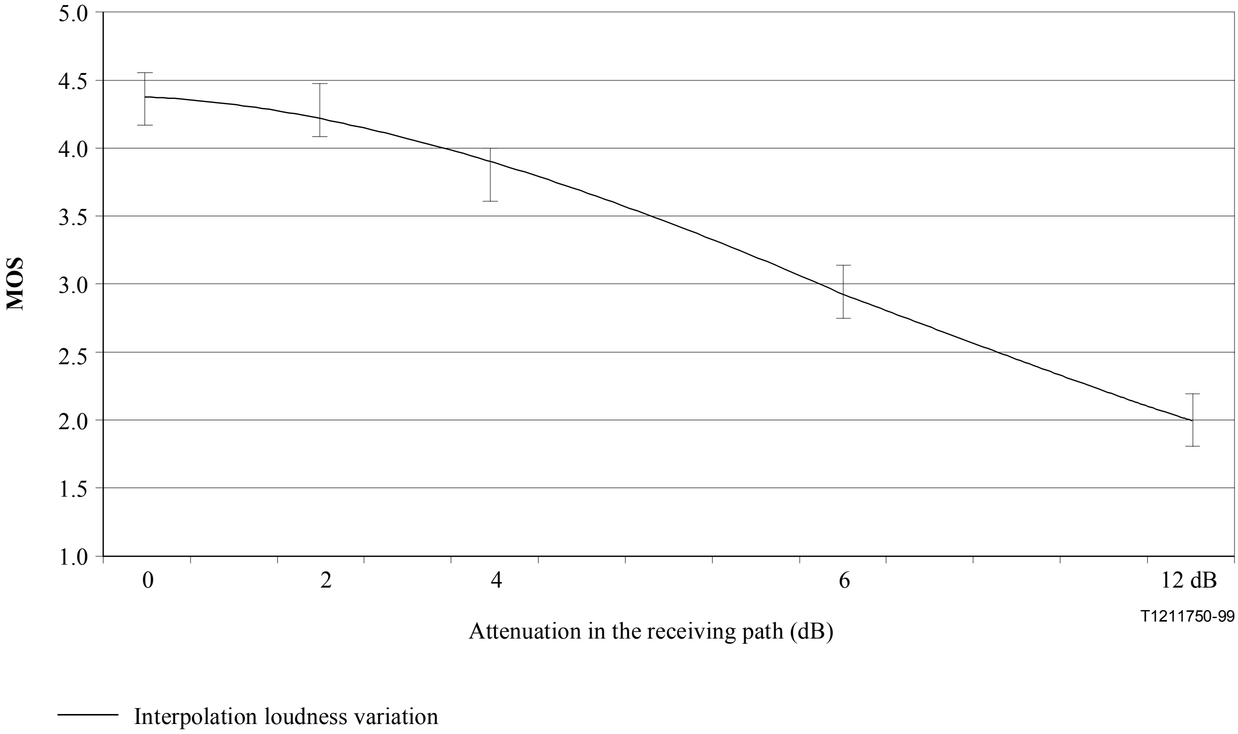


Figure A.2/P.340 – Interpolation of results for loudness variation in receiving direction

A.1.3 Results of listening only tests on echo disturbance

Test conditions (delay/TERL) for Figures A.3 and A.4 are derived from the curves 1% and 10%, as defined in ITU-T Recommendation G.131 [16].

Figure A.3 shows the results of the listening only tests in single talk.

Figure A.4 shows the results of the listening only tests in double talk.

For the results presented in Figure A.5, the delay is defined as a fix value of 100 ms. The variable parameter is TERL (dB). The tests are performed in single talk (lower line) and double talk (upper line).

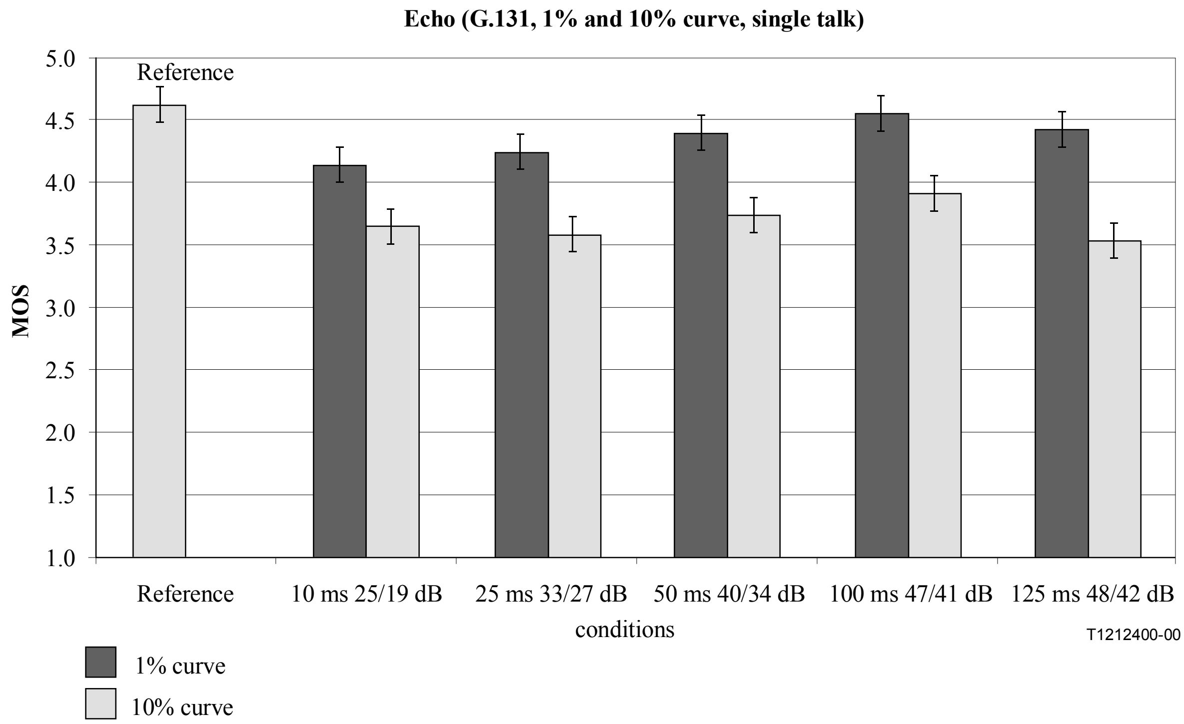


Figure A.3/P.340 – Echo disturbance for conditions according to G.131 (single talk conditions, recordings in send direction of HFT, judged at the handset)

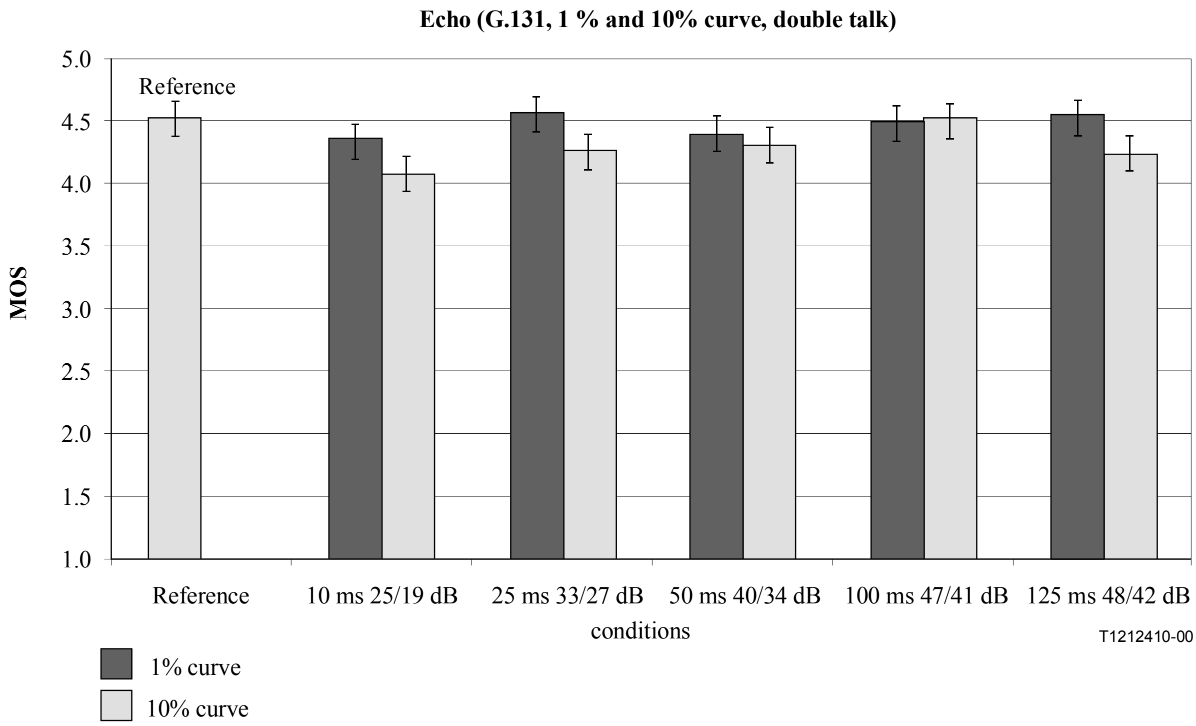


Figure A.4/P.340 – Echo disturbance for conditions according to G.131, (double talk conditions, recordings in send direction of HFT, judged at the handset)

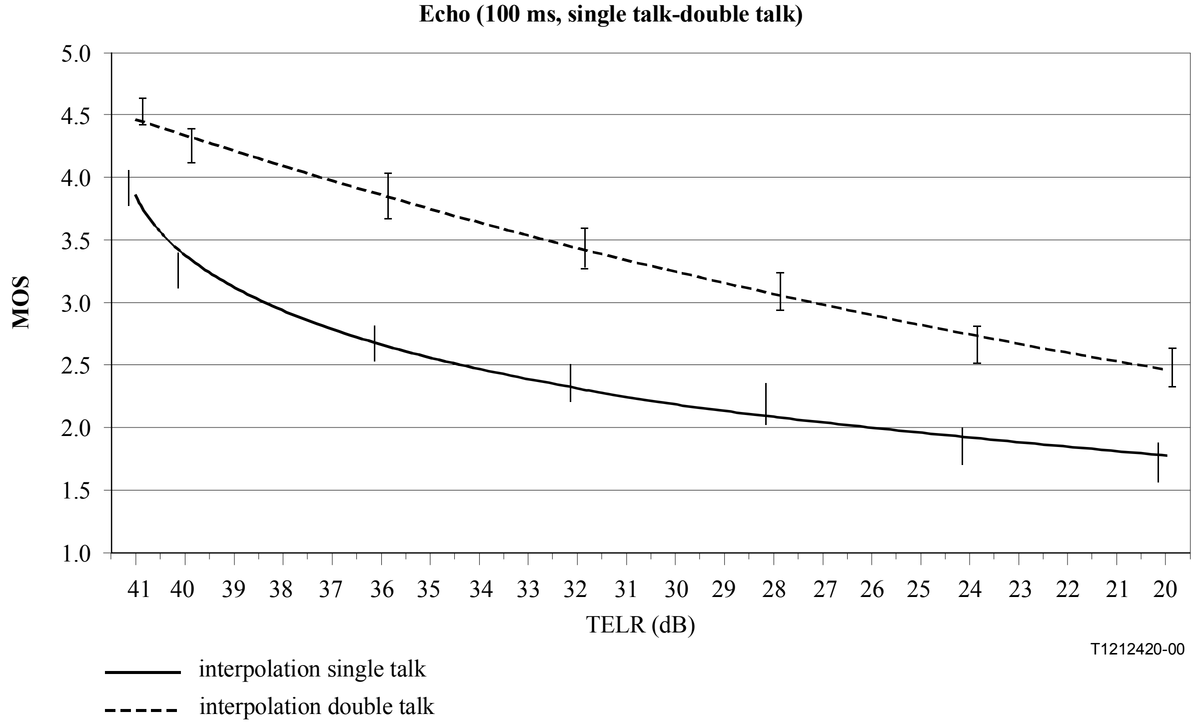


Figure A.5/P.340 – Interpolation of LOT results for echo disturbances (upper line: double talk)

Table A.1 summarizes the results and defines the correlations between the individual parameters (attenuation in sending and receiving directions, echo loss) and the subjective judgment of the quality, in double talk situation.

Table A.1/P.340 – Values for parameters determining double talk performance ($TELR_{DT}$, a_{Hrdt}, a_{Hsdt},) as a function of correlated MOS scores derived from LOT

MOS	≥4.0	4.0-3.5	3.5-3.0	3.0-2.5	2.5-2.0	≤2.0
$TELR_{DT}$ [dB]	≥37	≥33	≥27	≥21	≥13	<13
a_{Hsdt} [dB]	≤3	≤6	≤9	≤12	≤15	>15
a_{Hrdt} [dB]	≤3	≤5	≤8	≤10	≤12	>12

From Table A.1 has been derived the tables of clause 8 on the terminal categorization.

A.2 Relationship between Listening only tests, Double talk tests and Conversation tests: comparison of six different hands-free terminals

Six hands-free telephones were used in all tests. Four of them are commercially available devices, the other two were laboratory prototypes. Table A.2 gives an overview of all devices under test:

Table A.2/P.340 – Technologies implemented in each type of hands-free terminal

No.	Used algorithms
1	Level switching device, compander
2	Level switching device, echo canceller (70 ms), centre clipper
3	Echo canceller (254 ms), level switching device, centre clipper, frequency shift
4	Echo canceller
5	Compander, echo canceller, level switching device, background noise adaptation
6	Echo canceller (254 ms), speech controlled attenuation

Figures A.6 and A.7 give the test results on the quality parameter "overall quality" obtained with listening only tests and conversation tests. Globally it can be seen that the overall quality is judged more "severely" with listening only tests than with conversation tests, particularly for untrained subjects.

For trained subjects there is a rather good correlation between the results of listening only tests and conversation tests, the MOS value obtained for conversation tests being slightly higher than for listening tests.

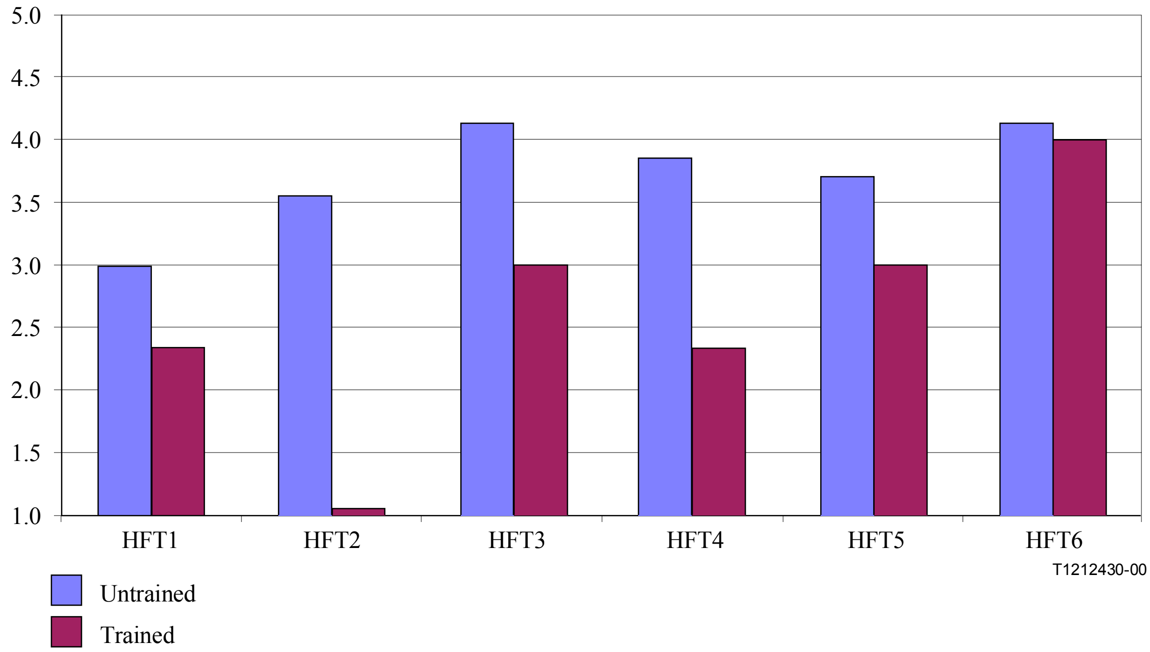


Figure A.6/P.340 – Overall quality, conversational test, room 1 – Untrained and trained subjects separated

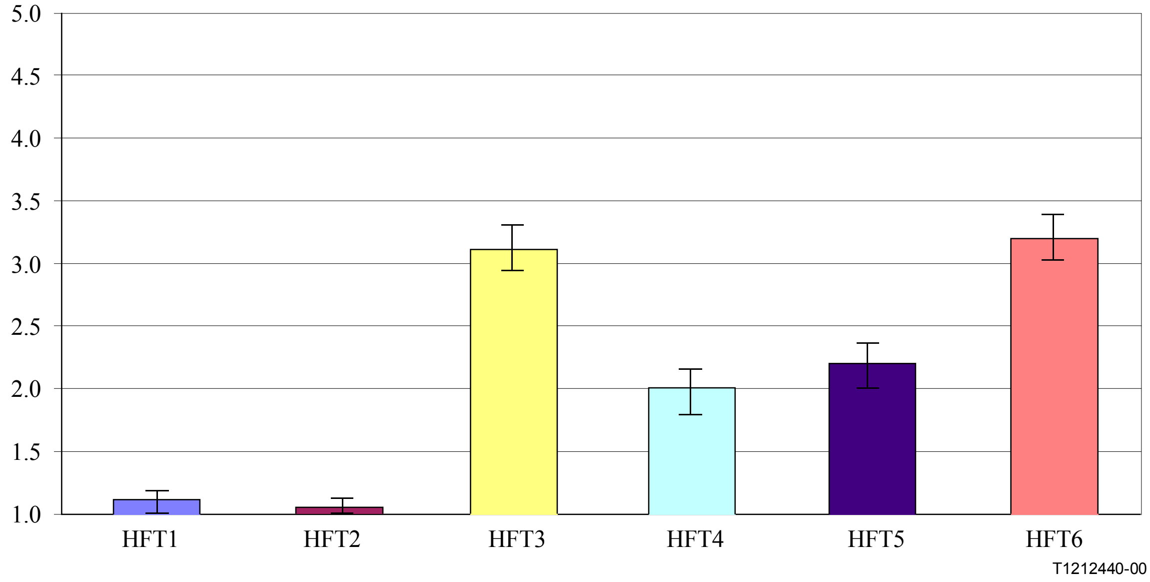


Figure A.7/P.340 – Overall quality, listening only test

Figures A.9 and A.10 present results on the parameter "double talk capability".

It can be seen in Figure A.8, and particularly in Figure A.9, that HFT1 and HFT2 have a very low double talk capability. This poor behaviour is correlated to the loudness variation in double talk (see Figure A.11). It appears from these data that the loudness variation (or attenuation range) is an important factor to qualify the double talk capability.

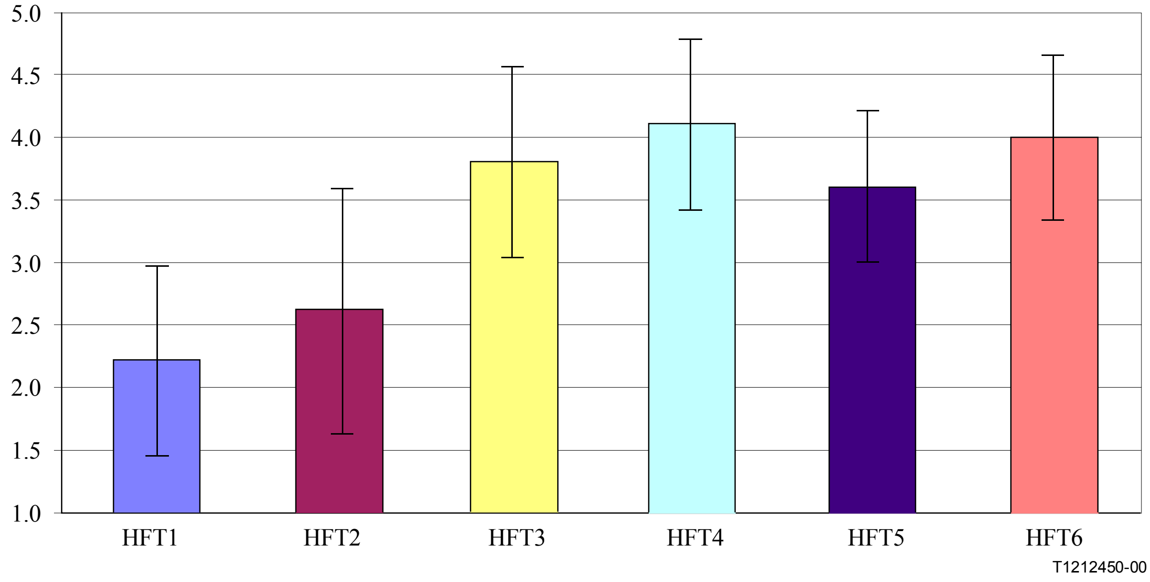


Figure A.8/P.340 – Double talk capability, conversational test, room 1

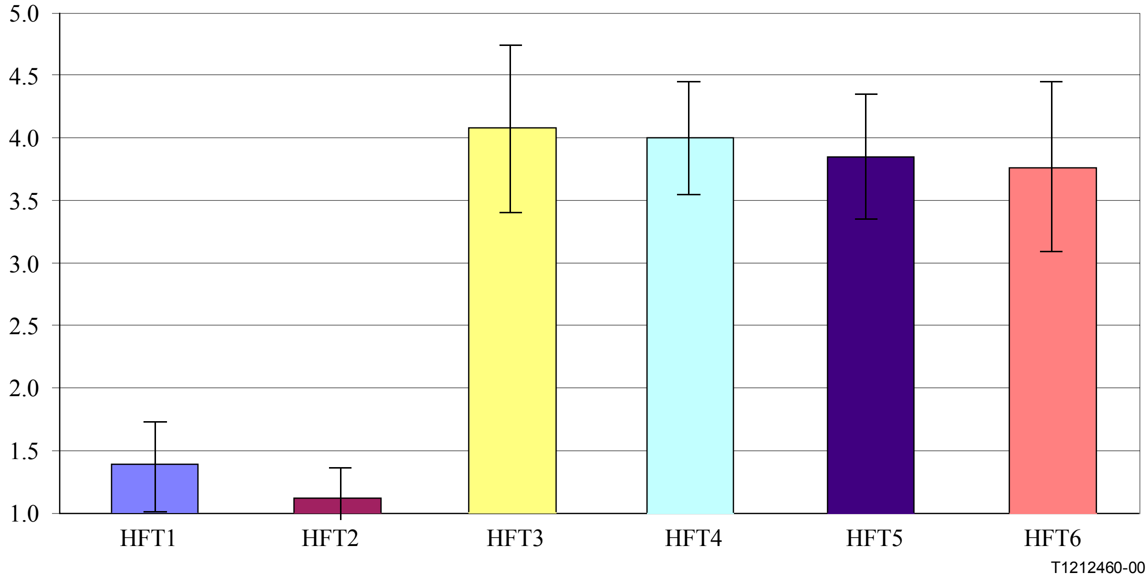


Figure A.9/P.340 – Double talk capability, double talk test, room 1

Figure A.10 presents results on the parameter "disturbance by echoes".

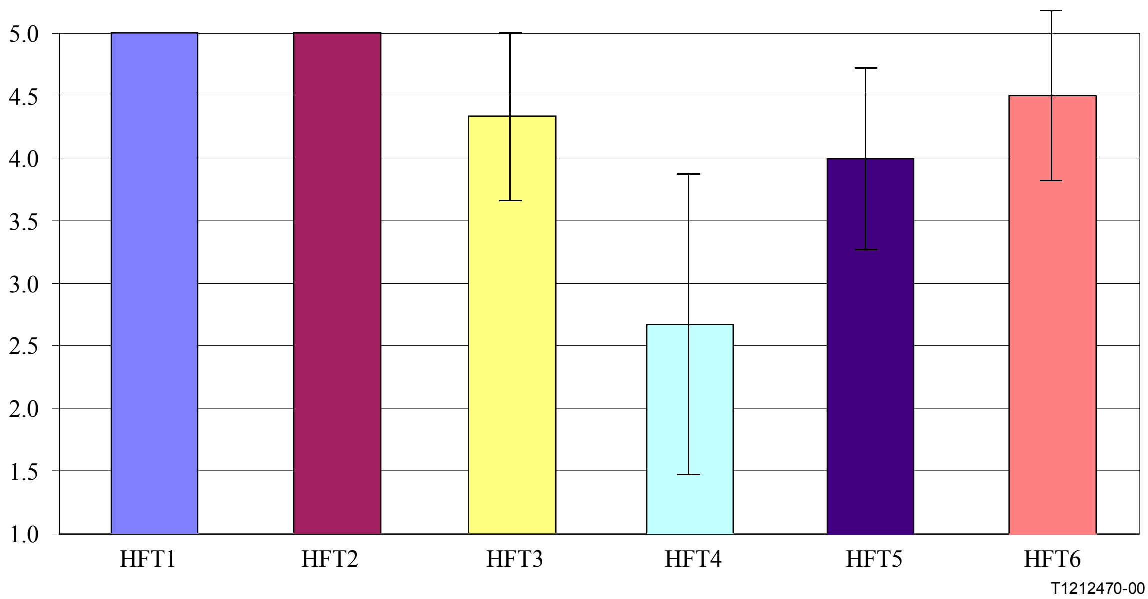


Figure A.10/P.340 – Disturbance by echoes, double talk test, room 1

Figure A.11 presents results on the parameter "loudness during double talk".

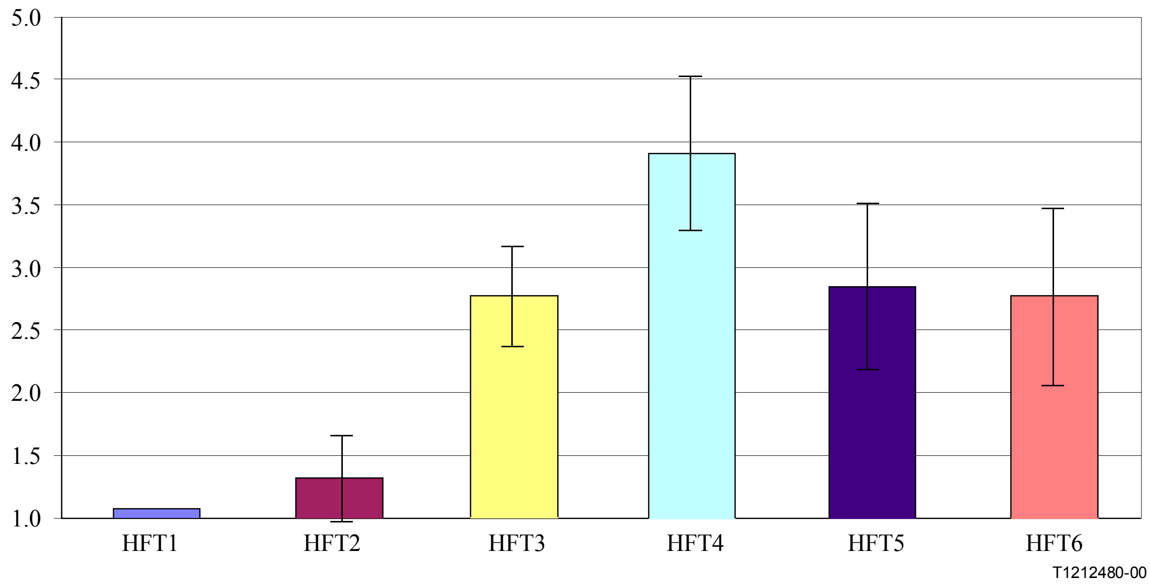


Figure A.11/P.340 – Loudness during double talk, double talk test, room 2

From Figures A.10 and A.11, if we exclude HFT1 and HFT2, it can be seen that "echo disturbance" is correlated to the double talk capability.

APPENDIX I

Preliminary values

This appendix gives some preliminary values and limits for parameters defined in clause 7. These parameters have been determined on the basis of results of subjective tests and are related to double talk performance and to the influence of companding or AGC devices on the speech quality.

I.1 Double talk performance

Parameters	From receive-to-send direction	From send-to-receive direction
Threshold (minimum activation) Level to Switch Over	(for an activation level in receive direction of –50 dBV) $L_{R\text{-}S} \leq -3$ dBPa at MRP	(for an activation level in send direction of –9.7 dBPa at MRP) $L_{S\text{-}R} \leq -2$ dBV
Switching Times	$T_{S,R\text{-}S} \leq 100$ ms	$T_{S\text{-}10dB,S\text{-}R} \leq 40$ ms (the complete attenuation except a residual value of 10 dB is removed) $T_{S,S\text{-}R} \leq 100$ ms

I.2 Build-up time characteristics

Parameters	Sending direction	Receiving direction
Build-up time characteristics	$T_{R,idle-S} \leq 15$ ms	$T_{R,idle-R} \leq 10$ ms
Threshold level and Build-up time for threshold level	$V_{Hidle-S} \leq -14$ dBPa (at MRP) $T_{R,idle-S} \leq 15$ ms	$V_{Hidle-R} \leq -48.5$ dBV $T_{R,idle-R} \leq 10$ ms

I.3 Level adjustments by companding or AGC

	Parameters in sending direction	Parameters in receiving direction
Range of level adjustments as a function of the input signal level	For decreasing send levels $L_S > -12$ dBPa (at MRP) and for increasing send level $L_S > -5$ dBPa (at MRP) The deviation of the measured output signal level referred to the input signal level should be ≤ 3 dB.	For decreasing receive levels $L_R > -44$ dBV and for increasing receive level $L_R > -42$ dBV The deviation of the measured output signal level referred to the input signal level should be ≤ 3 dB.
Time duration for level adjustments		
±5 dB (comp. to –4.7 dBPa)	For a level variation of –5 dB: No insertion of attenuation, duration for level adjustment $t_{adj,S,-5} < 10$ ms (tolerance scheme ±3 dB)	For a level variation of –5 dB: No insertion of attenuation, duration for level adjustment $t_{adj,R,-5} < 20$ ms (tolerance scheme ±3 dB)
	For a level variation of +5 dB: $t_{adj,S,+5} < 15$ ms (tolerance scheme ±3 dB)	For a level variation of +5 dB: $t_{adj,R,+5} < 15$ ms (tolerance scheme ±3 dB)
±10 dB (comp. to –4.7 dBPa)	For a level variation of –10 dB: Insertion of attenuation after hang-over time 130 ± 20 ms with (20 ± 3 dB)/s For a level variation of +10 dB: Removal of attenuation after $t_{adj,S+10} < 15$ ms (tolerance scheme ±3 dB)	For a level variation of –10 dB: No insertion of attenuation, duration for level adjustment $t_{adj,R-10} < 100$ ms for max volume $t_{adj,R,-10} < 20$ ms for min. volume (tolerance scheme ±3 dB) For a level variation of +10 dB: Duration for level adjustment $t_{adj,R,+10} < 15$ ms (tolerance scheme ±3 dB)

	Parameters in sending direction	Parameters in receiving direction
±15 dB (comp. to –4.7 dBPa)	For a level variation of –15 dB: Insertion of attenuation after hang-over time: 130 ± 20 ms with (20 ± 3 dB)/s	For a level variation of –15 dB: No insertion of attenuation, duration for level adjustment $t_{adj,R,-15}$ < 100 ms for max. volume $t_{adj,R,+15}$ < 20 ms for min. volume (tolerance scheme ±3 dB)
	For a level variation of +15 dB: Removal of attenuation after $t_{adj,S,+15}$ < 15 ms (tolerance scheme ±3 dB)	For a level variation of +15 dB: Duration for level adjustment $t_{adj,R,+15}$ < 15 ms (tolerance scheme ±3 dB)

APPENDIX II

Examples of processing delays for mobile hands-free terminal

II.1 Example 1

Tadd_proc (additional time for processing)

Tadd_proc delays providing minimal values for acceptable echo or/and noise reduction are defined as follows :

– **Tadd_proc_AEC** for Echo Cancellation when using Hands-free mobile terminal,

 Tadd_proc_AEC: 28 (to 40) **ms**.

– **Tadd_proc_NR** when using Noise Reduction (NR) and a coupling reduction processing for use of mobile handset,

 Tadd_proc_NR: 20 (to 32) **ms**.

– **Tadd_proc_HF** for Hands Free when using AEC and NR when using Hands-free mobile terminal and if it is desired to add NR for listening comfort,

 Tadd_proc_HF: 36 (to 56) **ms**.

Tadd_proc taking into account any additional and mandatory speech processing block (including noise bad effects compensation) where Tadd_proc must be decomposed as follows:

Delay for signal block size or sub-band decomposition:	16 to 24 ms
Delay for noise reduction:	12 to 16 ms
Additional delay for computation:	8 to 16 ms
Tadd_proc:	36 to 56 ms

II.2 Example 2

To fully exploit the benefits of block processing, it is important that the block length is sufficient:

– to provide enough data for statistically good estimates of the properties of noisy speech;

– to efficiently handle reverberation times in normal car cabins;

– to provide adequate resolution in the frequency domain.

It is also desirable, for reasons of efficiency in frequency domain processing, that the block length is a power of two.

A block length of 256 samples, which corresponds to a 32 ms block at 8 ksamples/s, represents a good balance in these respects.

With 32 ms blocks there will be an inherent delay of 32 ms. A reasonable figure for processing time of such blocks, using today's cost competitive DSPs, is 10 ms.

An additional delay for hands-free signal processing could be a minimum of **32 ms + 10 ms = 42 ms**.

APPENDIX III

Recommended test bed

This appendix gives some examples of recommended test beds that can be implemented for testing according to this Recommendation.

III.1 Example 1: Implemented test hybrid

The test hybrid can be constructed from a passive 2-4 wire hybrid, mated to a digital echo canceller as described in Figure III.1. The digital echo canceller is described below.

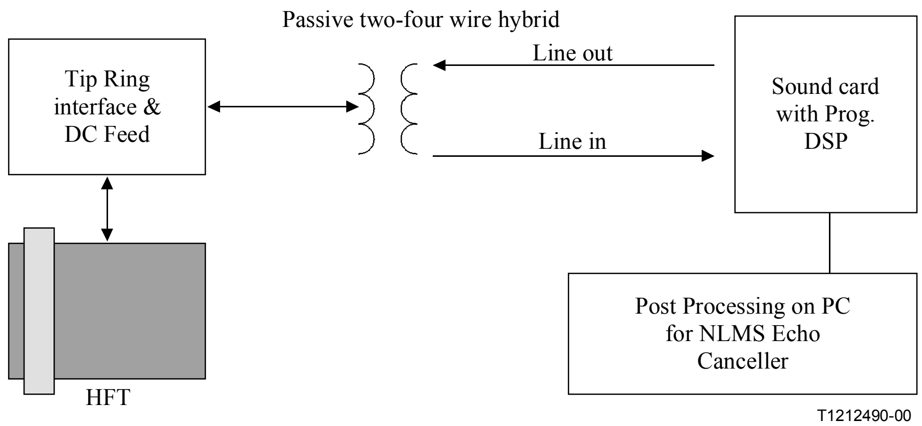


Figure III.1/P.340

The passive 2-4-wire hybrid is a transformer setup that can be balanced to provide a simulated 4-wire signal path that is then interfaced to the digital echo canceller.

The digital echo canceller can be implemented through a post processing or real time application. In either case, simultaneous (full duplex) record-playback capability is needed.

III.2 Example 2: Adaptive Echo Canceller

An NLMS (normalized least mean square) adaptive echo canceller is used to model the actual echo path formed by the 4- to 2-wire hybrid and the HFT. The block diagram is shown in Figure III.2.

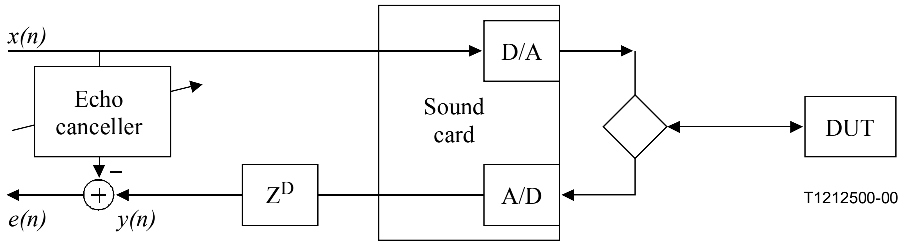


Figure III.2/P.340 – Block diagram of 2- to -4-wire conversion

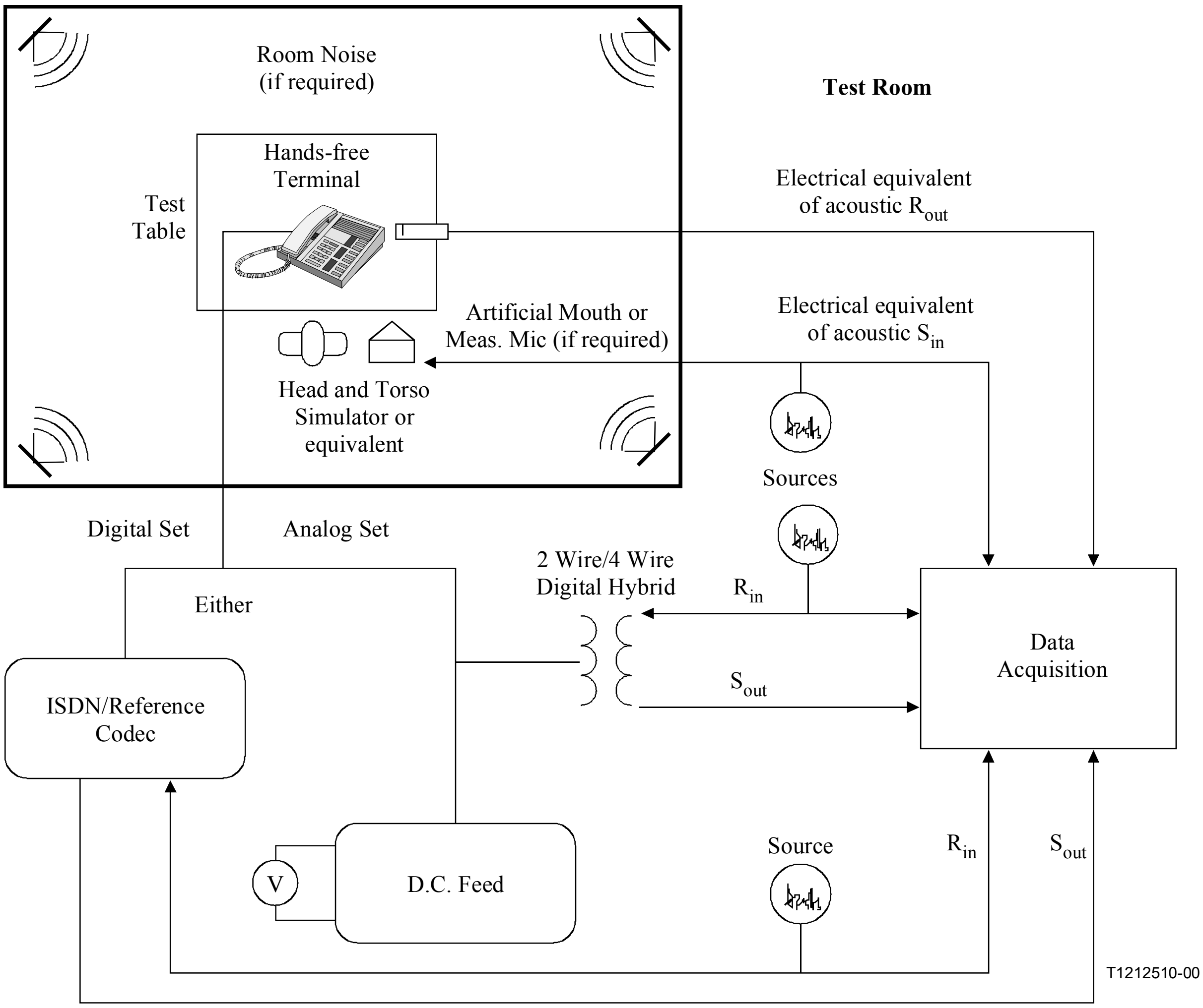


Figure III.3/P.340 – Test bed diagram

APPENDIX IV

Bibliography

[IV.1] European Committee for Standardization (CEN) Office Chair/Desk Working Position – Dimensions and Design Requirements, *CEN*: prEN91/August 1981.

[IV.2] CCITT: A Method for Measuring the Sensitivity of a Loudspeaking Telephone Set, Annex 2 to Question 17/XII, *White Book*, Vol. V, ITU, Geneva 1969.

[IV.3] ISO 1996-3:1987, *Acoustics – Description and measurement of environmental noise – Part 3: Application to noise limits*.

[IV.4] BERANEK (L.L.): Noise and Vibration Control, *McGraw Hill*, pp. 564-566, New York 1971.

[IV.5] XIANG (N.): Using M-sequences for Determining the Impulse Responses of LTI-systems, *Signal Processing*, No. 28, pp. 139-152, 1992.

[IV.6] JORDAN (V.L.): Acoustical Criteria for Concert Hall Stages, *Applied Acoustics,* No. 15, pp. 321-328, 1982.

[IV.7] HOTH (D.F.): Room Noise Spectra at Subscribers Telephone Locations. *J.A.S.A.*, Vol. 12. pp. 499-504, April 1941.

[IV.8] ITU-T COM 12-60 (1995), Subjective Evaluation of Transmission Quality for Hands-free Telephones, Federal Republic of Germany.

[IV.9] ITU-T COM 12-61 (1995), Supplementary Measurement Proposal for Certification Measurements of Hands-free Equipment, Federal Republic of Germany.

[IV.10] ITU-T COM 12-6 (1997), Subjective evaluation of hands-free telephones using conversational tests, specific double talk tests and listening only tests, Federal Republic of Germany.

[IV.11] ITU-T COM 12-102 (1999), Proposals for the definitions of different types of hands-free telephones based on double talk performance, Federal Republic of Germany.

[IV.12] ITU-T SG 12-D.83 (November 1998) Advanced Measurements for Hands-free Telephones and Non-linear/Time variant Operating devices, Deutsche Telekom Berkom/Head Acoustics).

[IV.13] ETSI 0358 601 (TR 101110) Digital Cellular Telecommunications system (Phase 2+). Characterisation, test methods and Quality assessment for hands-free mobiles.

[IV.14] IEEE 1329 Standard Method for Measuring Transmission Performance of Hands-Free Telephone Sets.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Telephone transmission quality, telephone installations, local line networks**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure and Internet protocol aspects
Series Z	Languages and general software aspects for telecommunication systems

Geneva, 2001