INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.108.1

(05/2000)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

International telephone connections and circuits – General definitions

---

Guidance for assessing conversational speech transmission quality effects not covered by the E-model

ITU-T Recommendation G.108.1

(Formerly CCITT Recommendation)

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
General definitions	**G.100–G.109**
General Recommendations on the transmission quality for an entire international telephone connection	G.110–G.119
General characteristics of national systems forming part of international connections	G.120–G.129
General characteristics of the 4-wire chain formed by the international circuits and national extension circuits	G.130–G.139
General characteristics of the 4-wire chain of international circuits; international transit	G.140–G.149
General characteristics of international telephone circuits and national extension circuits	G.150–G.159
Apparatus associated with long-distance telephone circuits	G.160–G.169
Transmission plan aspects of special circuits and connections using the international telephone connection network	G.170–G.179
Protection and restoration of transmission systems	G.180–G.189
Software tools for transmission systems	G.190–G.199
INTERNATIONAL ANALOGUE CARRIER SYSTEM	
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
DIGITAL TRANSMISSION SYSTEMS	
TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999

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

ITU-T Recommendation G.108.1

Guidance for assessing conversational speech transmission quality effects not covered by the E-model

Summary

This ITU-T Recommendation provides guidance for transmission planners on how to evaluate those effects impacting end-to-end speech transmission performance which are not covered by the E-model (ITU-T Recommendation G.107 [2] – The E-model, a computational model for use in transmission planning) and its associated Planning Guide (ITU-T Recommendation G.108 [3] – Application of the E-model – A planning guide). Procedures for informal subjective and objective evaluations that can be used to complement the E-model are provided here.

Source

ITU-T Recommendation G.108.1 was prepared by ITU-T Study Group 12 (1997-2000) and approved under the WTSC Resolution 1 procedure on 18 May 2000.

Keywords

Conversational impacts, E-model, speech transmission quality, voice quality.

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
1	Scope	1
2	References	1
3	Abbreviations and definitions	2
3.1	Abbreviations	2
3.2	Definitions	2
4	General considerations	2
4.1	Parameters describing speech transmission quality (including terminals)	6
4.2	Test setup for terminals	7
4.3	Test setup for echo cancelling devices	8
4.4	Test signals for conversational evaluation	8
5	Evaluation of the conversational situation	8
5.1	Double talk performance	9
5.2	Subjective evaluations	9
	5.2.1 Conversational evaluation	9
	5.2.2 Specific double talk tests and listening only tests	11
5.3	Objective evaluations	12
6	Guidance on the improvement of conversational speech quality	12
6.1	Delay and echo	13
6.2	Background noise transmission	13
6.3	Double talk performance	14
6.4	Quality of speech sound and loudness	14

Introduction

This ITU-T Recommendation provides guidance for transmission planners on supplementary conversational parameters impacting end-to-end speech transmission performance which are not covered by ITU-T Recommendation G.108, since:

- these supplementary factors are not covered by the E-model as it stands;
- the guidelines and principles in ITU-T Recommendation G.108 are based on the use of the E-model, which is applicable to 3.1 kHz handset telephony, only;
- the current E-Model cannot completely predict the conversational effects of electric or acoustic echo cancelling devices, which may affect quality during only some time segments of the conversation;
- conversational impacts on end-to-end speech transmission performance will occur in conjunction with conversation over hands-free terminals.

ITU-T Recommendation G.108.1

Guidance for assessing conversational speech transmission quality effects not covered by the E-model

1 Scope

This ITU-T Recommendation is intended to provide guidance on the conversational impairments which are not covered by the E-model and which are thus not included in ITU-T Recommendation G.108 [3]. Those impairments have been investigated thoroughly by ITU-T during the study period 1997-2000.

2 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 Recommendation G.100 (1993), *Definitions used in Recommendations on general characteristics of international telephone connections and circuits.*

[2] ITU-T Recommendation G.107 (2000), *The E-model, a computational model for use in transmission planning.*

[3] ITU-T Recommendation G.108 (1999), *Application of the E-model: a planning guide.*

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

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

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

[7] ITU-T Recommendation P.57 (1996), *Artificial ears.*

[8] ITU-T Recommendation P.58 (1996), *Head and torso simulator for telephonometry.*

[9] ITU-T Recommendation P.59 (1993), *Artificial conversational speech.*

[10] ITU-T Recommendation P.64 (1999), *Determination of sensitivity/frequency characteristics of local telephone systems.*

[11] ITU-T Recommendation P.340 (2000), *Transmission characteristics and speech quality parameters of hands-free terminals.*

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

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

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

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

[16] ITU-T Recommendation P.831 (1998), *Subjective performance evaluation of network echo cancellers.*

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

[18] ITU-T Recommendation P.861 (1998), *Objective quality measurement of telephone-band (300-3400 Hz) speech codecs.*

[19] ETSI EG 201 377-1 (1999), *Speech Processing, Transmission and Quality Aspects (STQ); Specification and measurement of speech transmission quality; Part 1: Introduction to objective comparison measurement methods for one-way speech quality across networks.*

3 Abbreviations and definitions

3.1 Abbreviations

This ITU-T Recommendation uses the following abbreviations:

DCME	Digital Circuit Multiplication Equipment
ECD	Echo Cancelling Device
HATS	Head and Torso Simulator
HEC	Half Echo Canceller
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union – Telecommunication Standardization Sector (former CCITT)
NLP	Non-Linear Processor
RCV	Receive
SCT	Short Conversational Test
SND	Send
TCLw	Terminal Coupling Loss weighted
VAD	Voice Activity Detection

3.2 Definitions

This ITU-T Recommendation defines the following term:

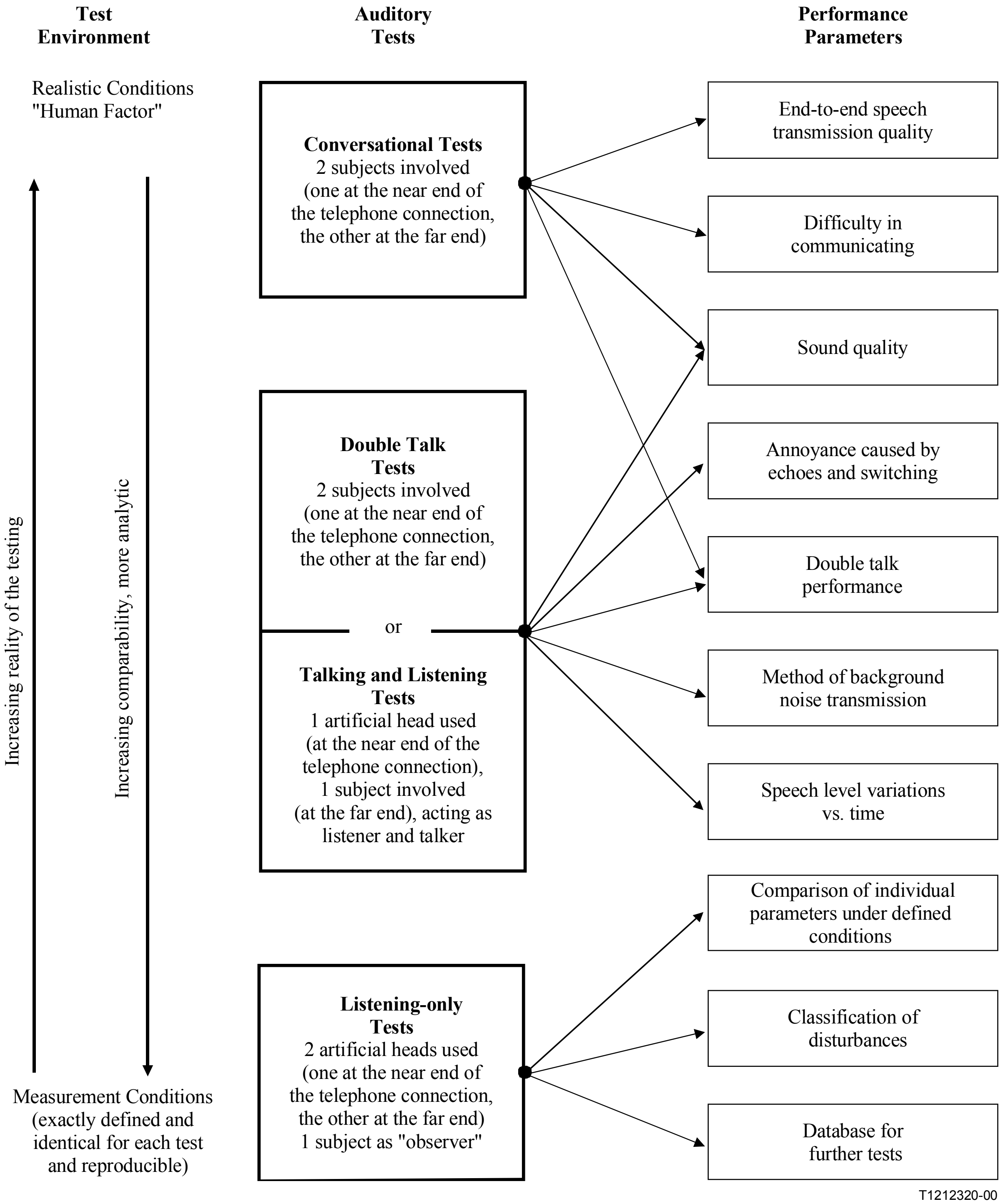
The term "**Lab Conditions**" is used in this ITU-T Recommendation in order to describe a testing environment which is well controllable and which allows for speech quality testing under defined and reproducible conditions.

4 General considerations

When evaluating the end-to-end speech transmission quality, it may be found that networks and terminals impact the speech quality of a telephone connection quite significantly: coding and decoding processes, introduction of additional delay, packetization and signal processing techniques as implemented, e.g. in echo cancelling devices or DCME are mainly deployed in the network domain but can increasingly be found in terminal devices as well. The frequency response and loudness ratings of a connection are mainly determined by the terminals, the background noise and the background noise transmission are highly influenced by the terminal and the acoustical environment the terminal is exposed to. The conversational properties which are the most important ones in a conversation are determined by the terminal in combination with the network: double talk capability, switching characteristics and delay are dominant impairments which are often introduced.

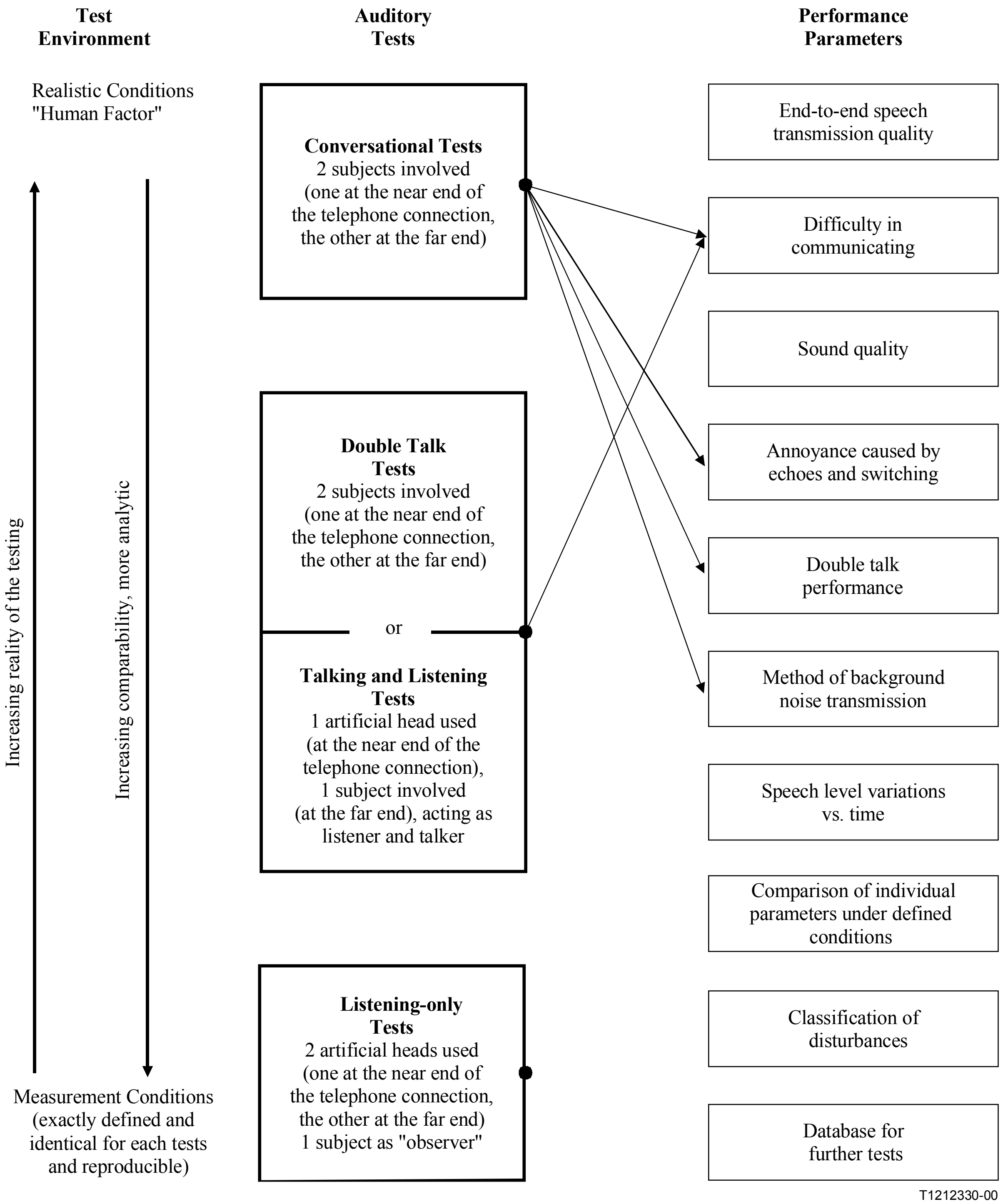
In order to evaluate the factors which determine end-to-end speech transmission quality a set of subjective test procedures has been developed. These procedures allow to extract the dominant quality aspects: conversational tests, talking and listening tests, double talk tests and listening only tests as described in ITU-T Recommendations P.800 [15], P.831 [16] and P.832 [17] are the basis for the parameter extraction procedure.

An overview of the methodologies is given in Figures 1 through 3.



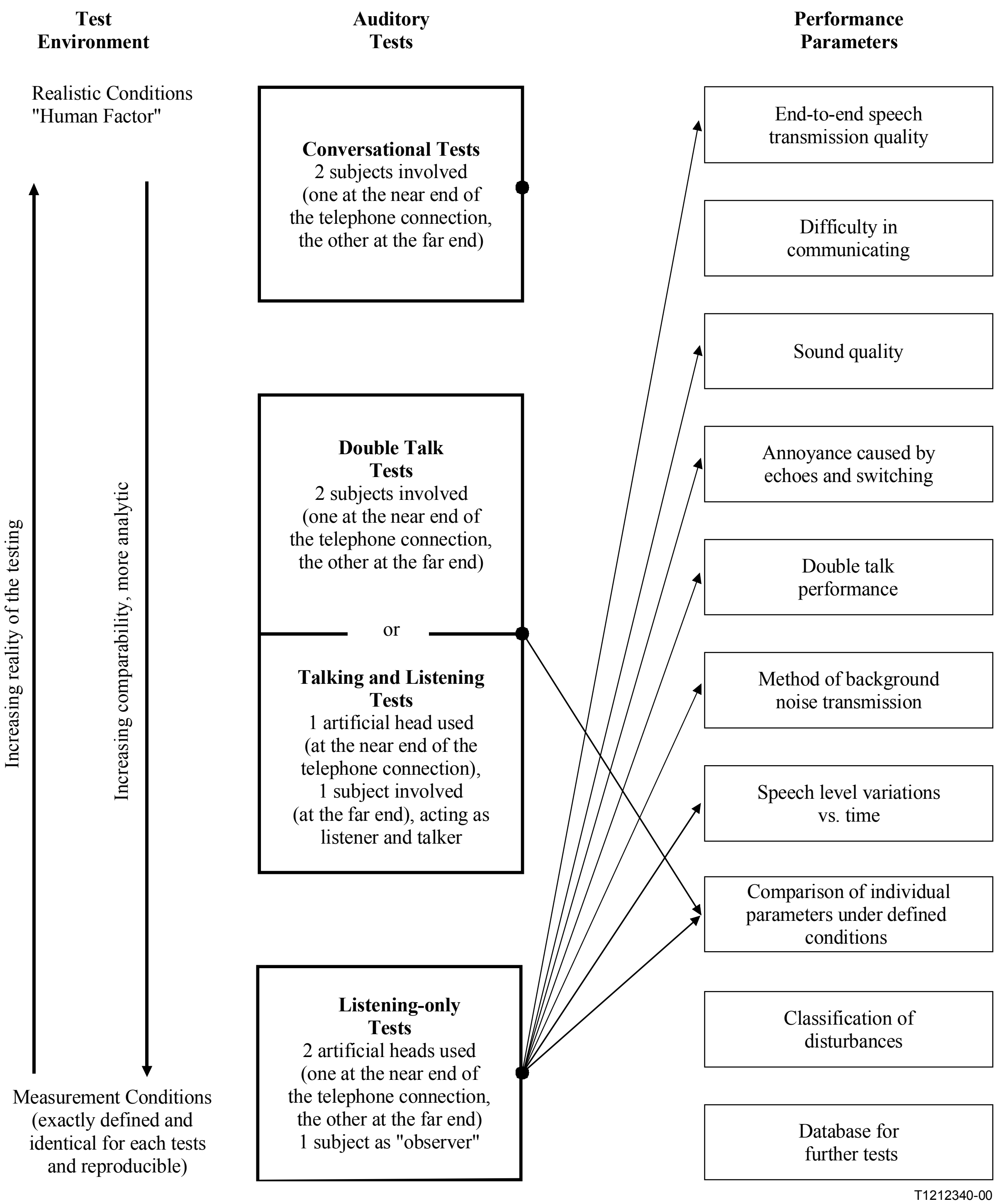
NOTE – The assignment of "near end" and "far end" is chosen according to the E-model (ITU-T Recommendation G.107).

Figure 1/G.108.1 – Overview of test methods used for subjective evaluation – direct parameter access



NOTE – The assignment of "near end" and "far end" is chosen according to the E-model (ITU-T Recommendation G.107).

Figure 2/G.108.1 – Overview of test methods used for subjective evaluation – Parameter access via interviews



NOTE – The assignment of "near end" and "far end" is chosen according to the E-model (ITU-T Recommendation G.107).

Figure 3/G.108.1 – Overview of test methods used for subjective evaluation – Parameter access by using reference conditions

The subjectively relevant parameters determining the "speech transmission quality" are as follows:

- sound quality;
- method of background noise transmission at idle, under single talk and under double talk conditions;
- speech level variations during single talk and double talk conditions;
- disturbances caused by switching during single talk and double talk conditions (completeness of speech transmission);
- disturbances caused by echoes during single talk and double talk conditions;
- disturbances caused by delay and jitter;
- time-variant impairments due to transmission errors.

Consequently, for objective evaluations it should be distinguished between evaluations under single talk conditions and between double talk evaluation. Furthermore, a third type of evaluations is required during periods of silence where only background noise is present.

In order to serve as an approach to end-to-end scenarios a test setup should consider all components involved in the mouth-to-ear transmission; hence it should include the terminals "attached". Thereby a realistic replica of a user and his typical environment is possible.

Some parameters mentioned in the list above are already dealt with by the E-model.

4.1 Parameters describing speech transmission quality (including terminals)

The parameters affecting the subjectively perceived speech quality and their correlating objective parameter are in general very important. An overview of the most important parameters – found in extensive auditory tests – and indication whether the parameters are covered by the E-model is provided in Table 1.

Table 1/G.108.1 – Correlation between subjective and objective parameters and indication whether or not parameters are covered by the E-model

Subjectively relevant parameter	More detailed description	Correlating objective parameter	Covered by the E-model
Echo	Talker echo and listener echo	• echo loss • TCLw • delay • switching characteristics	• Yes • Yes • Yes • No
Absolute delay	One way transmission time caused by signal processing, packetizing and transmission	• delay	• Yes
Method of background noise transmission	Typically the transmission in SND direction • at idle mode • with far end speech • with near end speech	• attenuation range • attenuation in SND direction • switching characteristics • minimum activation level in SND direction • frequency response • design of NLP or center clippers in conjunction with ECDs • design of noise reduction systems • sensitivity of background noise detection (activation level, absolute level, level fluctuations)	• No

Table 1/G.108.1 – Correlation between subjective and objective parameters and indication whether or not parameters are covered by the E-model *(concluded)*

Subjectively relevant parameter	More detailed description	Correlating objective parameter	Covered by the E-model
Double talk performance	Typically in SND and RCV direction • loudness variation between single and double talk periods • loudness variation during double talk • echo disturbances • occurrence of speech gaps	• attenuation range • attenuation in SND/RCV direction during double talk • switching characteristics • minimum activation level to switch over from RCV to SND direction and from SND to RCV direction • echo attenuation • spectral and time dependent echo characteristics • design of NLP or center clippers in conjunction with ECDs	• No
Echo disturbances under single talk conditions	Measured between RCV and SND direction	• echo level • echo level fluctuation vs. time • spectral echo attenuation	• Yes • No • No
Speech sound quality	In SND and RCV direction	• frequency responses • distortions	• No • Yes
Loudness	In SND and RCV direction	• loudness ratings in SND and RCV	• Yes
Noise	In SND and RCV direction	• noise level • level fluctuations • spectral characteristics	• Yes • No • No

These parameter correlations have been identified in tests with hands-free terminals and network echo cancelling devices. It should be noted, that these parameters are of general nature, the type of device which introduces such a degradation into a telephone connection is of minor importance for the subject exposed to this degradation

4.2 Test setup for terminals

The general test setup for the assessment of terminals can be found in ITU-T Recommendations P.64 [10], P.340 [11] and P.581 [14]. In addition, ITU-T Recommendations P.57 [7] and P.58 [8] describe HATS and the artificial ears which are recommended to be used for this kind of conversational application.

These Recommendations – dealing with standardized methods for tests under lab conditions – provide nevertheless useful information on how test setups in such an environment where the terminals are used, can be designed in practice. Basically, the individual conditions, such as background noise conditions, typical type of use of the equipment and other individual factors need to be adequately taken into account and should be reflected in the design of the test setup. Such test setups may differ from those described in the Recommendations mentioned afore but are closer to the real use condition of the connection and/or equipment under test.

4.3 Test setup for echo cancelling devices

Echo cancellers – as a general rule – are realized as so-called "Half Echo Canceller" devices (HEC), i.e. operative into one of either directions of transmission only. HEC devices may be deployed at various places within an end-to-end telephone connection:

- in the centre part of the network facing either terminal;
- at the edge of the network part facing the near-end terminal;
- at the edge of the network part facing the far-end terminal;
- in either of both terminals involved in the connection.

In combination with devices which are introducing significant delay Half Echo Cancelling devices are often deployed in the network. If contained in the terminal they are intended for cancelling the acoustic echo or the (electrical) hybrid echo. Appropriate test setups for the individual devices can be found in ITU-T Recommendations G.167 [4], G.168 [5] and P.340 [11].

For end-to-end scenarios those Recommendations give basic information on how the setup should be made. In practice all equipment typically involved in a connection needs to be included in the test setup. For end-to-end scenarios it is obvious that the terminals should be included. If individual sections of a connection are subject to a test, it is recommended to make – in principle – use of the same end-to-end test setups. Examples for such tests are the following:

- performance test of a network echo canceller in conjunction with a variety of typically used equipment in the end echo path;
- performance test of an acoustic canceller contained in a terminal in the specific room where the terminal is used – in conjunction with the equipment involved in the connection; or
- the performance test of echo cancellers in tandem.

4.4 Test signals for conversational evaluation

The basis for test signals can be found in ITU-T Recommendations P.50 [6], P.59 [9], P.501 [12] and P.502 [13]; according test signals are recommended to be used for objective evaluation of the conversational situation. The signals and test methods described here can be used for the objective assessment.

For subjective evaluation speech signals and test methods are required as described below.

5 Evaluation of the conversational situation

In an end-to-end telephone connection the conversational situation is the most important one to be considered and the most difficult one to evaluate. This is especially true if non-linear and/or time variant systems and devices are involved in a connection. The non-linear or time variant process may be integrated in either of the terminals involved in the connection or in one of the devices forming the network.

A very extreme situation where all types of signal processing may be present is the situation where hands-free terminals are involved in the connection. In general, the use of hands-free telephones leads to acoustical stability problems. Therefore, various types of signal processing devices such as speech detectors, level switching, echo cancellation, non-linear processes, dynamic level control and others are included in most hands-free terminals.

As discussed above not only the echo problem causes a degradation of end-to-end speech transmission quality. The use of one of the mentioned signal processing devices – or even a multitude of them – impacts the quality in various ways. Besides echo disturbances the remarkable parameters are:

- conversational capability;
- double talk performance;
- sound quality in send and receive direction;
- audible level variations;
- quality of background noise transmission; or
- reverberation in the listening situation.

Some of these parameters come immediately to the attention of the subscriber and impact the quality during the whole course of a conversation. Other parameters are important only in the listening situation. In addition, there is a third group of disturbances which occurs only if either one or both subscribers involved in a connection are talking (such as echo, modulated background noise, double talk performance).

5.1 Double talk performance

A very high influence on subjectively perceived speech quality of a telephone connection results from its double talk capability; this may be due to absolute delay or by signal processing, switching, etc. Although double talk situations are not predominant in an average conversation, double talk performance is one of the main factors determining the end-to-end speech transmission quality as perceived by the user.

If double talk capability in a telephone conversation is not given or is very restricted only, then both subscribers involved in that particular call will realize this impact immediately. In these situations a higher concentration is required from either subscriber during the entire telephone call and the naturalness of the conversation decreases. The double talk performance of the evaluated telephone connection has an influence on the quality ratings received from conversational tests.

This (subjective) quality parameter "Double Talk Performance" is determined by several (objective) technical parameters and by the combination of various types of signal processing; accordingly, the complexity of required quality evaluation rises dramatically. These influences on double talk performance have to be distinguished according to different aspects. Subjective procedures are necessary in order to determine the significance of each parameter which should be sensitive and efficient. While evaluating the quality subjectively, specific double talk tests and third party listening only tests are applicable – besides the complete conversational tests.

5.2 Subjective evaluations

As far as the following subclauses do refer to subjective evaluations or auditory testing procedures it should be clearly understood that the procedures recommended herein, do neither replace nor supersede the appropriate ITU-T Recommendations in force. The only intention of those subclauses is to provide practical procedures in addition to the E-model for quick access to conversational impacts outside of subjective test labs.

NOTE – Care should be taken not to express the results in terms of MOS since typically neither the conditions can be controlled in a proper way nor those conducting the test do typically have the necessary background knowledge for conducting these tests formally correct. Nevertheless such tests may be quite useful to evaluate problems seen in the field more in detail.

5.2.1 Conversational evaluation

Conversational evaluations in complex scenarios are a critical issue since – in general – no lab conditions can be achieved. Therefore, while a controlled evaluation is not possible, nevertheless a good estimate of the speech quality can be the outcome when applying the test methods as described in ITU-T Recommendations P.800 [15], P.831 [16] and P.832 [17].

A very simple and easy to use method for such tests are the so called short conversational tests, SCT.

These scenarios are structured to a high degree and are natural, while resulting in conversations of approx. 2.5 min. each. For purposes of conversational tests, they seem to be more easy to use than other scenarios. A comparison with the so-called "Kandinsky" scenarios leads to similar useful test results, whereas the required conversation time is being reduced.

A general structure for such a test is given in Figure 4. The subject for the conversations are enquiry and booking situations, where the telephone serves as a means of information exchange, i.e. a very typical purpose. The dialogue structure is provided with the test instructions, which include parts with longer monologues as well as others with various turn-takings, and some parts which are intended to evoke double talk.

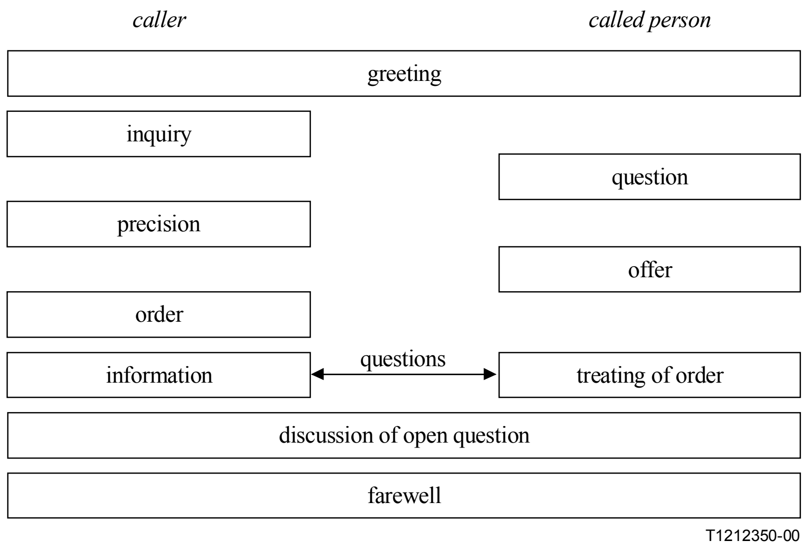


Figure 4/G.108.1 – Possible theoretical structure of a SCT dialogue

Due to their fixed structure, the dialogues during SCTs may be significantly shorter than during more traditional types of conversational tests; while offering this advantage SCTs are still maintaining the various dialogue sections required for concise conversational tests. Various possible situations (e.g. inquiries with railway information, travel agency, pizza service, theatre reservation, medical appointment, flight schedule, car rental, etc.) increase the probability that the conversations carried out remain interesting for the test subjects, and that different type of vocabulary is used. Pictograms, tables, etc., can be used as formal and quasi-standardized means of test instruction. As none of the subjects involved in a conversational test has *a priori* knowledge which information the other subject is requesting, they cannot shorten the conversation after some dialogues (as it happens when the same scenario is being repeated several times).

These types of tests have been used by several Administrations with success, e.g. for the subjective evaluations of hands-free terminals and echo cancellers and are for further study.

One Administration found a very useful multi-criteria approach which had been employed to collect the opinion of the subjects involved in the test; this approach follows the suggestions in ITU-T Recommendation P.831 [16]. The following two types of questionnaire had been prepared for conducting this test:

- a first type of questionnaire for each of the two subjects involved in the conversation under test;
- a second type of questionnaire for the observer who has the task to listen to the very same conversation under test as a third party.

The first type of questionnaire which either of the subjects involved in the conversation had been requested to fill in after having talked with their respective conversation partner consists of six questions; whereas the second type of questionnaire which the third party observer had to fill in consists only of a subset of four questions (see Table 2). The specific question (i.e. Question No. 4) related to the sound level is not intended for speech performance evaluation; but as it is supposed to get a constant mean score (requires appropriate test setup) it may be used for normalization purposes.

Table 2/G.108.1 – Questionnaires for multi-criteria approach

Questionnaire for either of the two subjects involved in the conversation		**Questionnaire for the third party observer (not actively involved in the conversation)**	
Q.1	Fidelity of your partner's voice	**Q.1**	Fidelity of the voices
Q.2	Annoyance due to the perception of your own voice delayed		
Q.3	Annoyance due to various perceived degradations	**Q.2**	Annoyance due to various perceived degradations
Q.4	Effort to interrupt your partner		
Q.5	Sound level	**Q.3**	Sound level
Q.6	Overall quality	**Q.4**	Overall quality

NOTE – The third party observer in the test described here should be attached to the conversation under test by using a passive circuitry – thus not influencing the conversation. The use of conference or multi-line circuits – which may actively influence the two-party conversation – should not be considered for this type of test.

5.2.2 Specific double talk tests and listening only tests

Although a continuous process of improvement of objective measurement techniques for the assessment of speech transmission quality is under way, subjective tests are still required. This subclause gives an overview of different subjective evaluation procedures for hands-free telephones and network echo cancellers which have been developed by ITU-T during the study period 1997-2000. The objective of these subjective tests is the evaluation of the performance under double talk conditions. Based on a brief overview of possible influences of hands-free telephones on the perceived quality, different kinds of subjective test procedures are described. Each of these procedures has a certain purpose and the combination of all tests provides a very comprehensive evaluation tool.

The double talk test procedure as described in ITU-T Recommendation P.832 [17] allows a very specific evaluation of the double talk situation and can be used for complete end-to-end scenarios; some administrations have applied this method successfully, e.g. for the hands-free situation. Typical parameters found in such tests are:

- double talk capability;
- completeness of speech transmission;
- loudness during double talk;
- echo level;
- echo characteristics;
- sound quality;
- transmission of background noise.

Listening only tests as described in ITU-T Recommendations P.800 [15], P.831 [16] and P.832 [17] can be employed to investigate speech impacting parameters with a very refined granularity. More easily than other types of tests, these tests will allow to identify impacts in detail and assign them to specific technical implementations.

As a matter of fact such listening tests are less realistic (see Figures 1 through 3) than other, e.g. conversational tests; but especially in third party listening tests it is possible to investigate even conversational situations. Further guidance on these issues is provided by ITU-T Recommendations P.831 [16] and P.832 [17].

5.3 Objective evaluations

It is important that the design of objective measurements for terminals, networks and end-to-end scenarios be chosen appropriately in order to evaluate each of the important parameters. These parameters have been identified by and derived from subjective tests which were conducted for hands-free terminals and network echo cancellers.

The application of objective tests can be directed towards the assessment of values for these parameters as well as towards compliance checks with respect to requirement limits for each of these parameters; whereby the definition of the requirement limits itself is also based on subjective tests results. This combination of subjective testing for parameter identification and value selection on the one hand and the more efficient objective laboratory tests on the other hand provide a good correlation of objective measures with subjectively perceived speech quality.

The objective test methods derived accordingly are described in ITU-T Recommendation P.502 [13]. Such test methods can be applied to all scenarios where the impact on speech quality which may be described by parameters listed in Table 1. The objective test methods take into account background noise scenarios, single talk conditions and the conversational situation. For all scenarios a detailed, diagnostic evaluation can be made.

Besides these parameter oriented methods, a more general investigation with respect to end-to-end speech transmission performance may be conducted by applying methods which are based on modelling the human perception of speech; they are described in ITU-T Recommendation P.861 [18] and other methods which were under study by ITU-T during the study period 1997-2000. Further guidance on perceptual speech quality measurement methods is provided in ETSI EG 201 377-1 [19].

6 Guidance on the improvement of conversational speech quality

Based on the investigations made for hands-free terminals and network echo cancelling devices, the important parameters impacting the speech transmission quality are well known. It should be noted, that these parameters are of a general nature and do not specifically refer to a certain technical implementation. From a subjective point of view there is no perception with regard to the type, number or location of any network or terminal elements that cause any impact to the telephone connection, as e.g. echo or switching; only the parameters measured with regard to this impact are important for the speech quality as perceived by the subscriber.

The perceived speech quality parameters are listed below in detail.

The test signals can be found in ITU-T Recommendations P.50 [6], P.59 [9] and P.501 [12], the test methods can be found in ITU-T Recommendation P.502 [13].

Requirement limits for the individual parameters can be found in ITU-T Recommendation P.340 [11]. Although not all requirement limits for the individual parameters have been evaluated yet, many important parameters are already defined. ITU-T Recommendation P.340 [11] provides information on echo loss requirements during double talk and on switching parameters. Although these parameters have been evaluated in subjective tests for combinations of hands-free terminals and handset terminals, the values provided herein can be applied to other scenarios with a sufficient degree of confidence.

In addition, requirement limits for echo cancellers, especially on their convergence behaviour, their performance in the presence of background noise and in double talk situations can be found in ITU-T Recommendation G.168 [5]. These requirement limits are also based on subjective evaluations of the various parameters.

6.1 Delay and echo

Impacts to end-to-end speech transmission performance due to high values of absolute delay or due to talker or listener echo are the most obvious to the subscribers of a telephone connection. For simple situations where a constant delay and a non-time variant echo loss can be assumed, the E-model gives good estimations of the speech degradation introduced by delay and echo.

If, however, the echo loss is time variant and associated with switching and or time variant loudness variation (caused, e.g. by VAD, non-linear processors and other types of signal processing) a more detailed evaluation is necessary.

In this case it is recommended to employ the methods described in clause 6.

6.2 Background noise transmission

Impacts due to background noise transmission may turn out to be the most disturbing ones. In general, the background noise should be transmitted with a low level and with as less variations – in the time and in the frequency domain – as possible. Any injection of so-called "comfort-noise" should be rendered inoperative in cases where the time or frequency characteristic cannot be reproduced correctly; it should be noted that comfort-noise injection typically occurs in conjunction with switching effects.

In the presence of background noise any speech detection device should work reliable in order not to cut off syllables or parts thereof.

Furthermore, background noise performance of low-bitrate codecs might in some cases not be correctly covered by the corresponding Ie (equipment impairment factor) value and the Ro value (taking the room noise at send side, Ps, into account) in the E-model. This holds true, if the additivity of both impairments is not strictly satisfied. In such cases, an integral Ie value may be determined for the codec working under background noise conditions. For the derivation of such an Ie value, the Lombard effect (the effect of a person adapting its speaking behaviour in a noisy environment) should be included in the subjective test setup, namely the recordings should be made in the noisy environment. This is important in order to cover the conversational impacts of the background noise in a realistic way. An Ie value derived by this procedure can then be used as an input to the E-model, setting the Ps parameter to its default value (Ps = 35 dB(A)).

Background noise in many situations is being considered as a plain signal in cases where there is no other signal present beside the background noise.

The requirement limits for the background noise transmission have not yet been evaluated in detail and thus are for further study. However, the test procedures have been defined and can be found in ITU-T Recommendation P.502 [13].

While evaluating configurations with respect to the presence of background noise at the far-end subscriber side no additional (circuit or near-end) noise should be inserted.

The transmission method of background noise (from the near end in sending direction) can be evaluated:

- at idle mode;
- with far end speech;
- with near end speech.

In all these cases important parameters are:

- the sensitivity of background noise detection in terms of activation level;
- the absolute level of the transmitted signal;
- level fluctuations;
- variation of the spectral content of the background noise.

6.3 Double talk performance

The double talk situation is the most critical one for configurations where various types of signal processing (e.g. echo cancellation, level dependant switching and attenuation) are involved. Conversational tests in combination with double talk tests and third party listening tests proved the importance of the double talk situation in a complete conversation (see annexes to Recommendation P.340 [11]). The most annoying effects during double talk are:

- sentences, words, syllables or parts thereof being interrupted or being not completely transmitted during or shortly after/before double talk;
- transmission of speech and/or background noise with time variable level causing annoying "level variations during double talk";
- echo during double talk.

The most critical situations during double talk are the time intervals shortly before and shortly after double talk.

In this case it is recommended to employ the methods described in clause 6.

6.4 Quality of speech sound and loudness

Speech sound quality are mainly determined by the terminals. For handsets it is important to measure these parameters in a most realistic condition. This requires the measurement setup using HATS in combination with type 3.3 or 3.4 ears and positioning according to ITU-T Recommendation P.64 [10]. By employing this setup the acoustical coupling of the signal as well as of the noise is achieved in a very realistic manner. Speech sound quality can – in a first step – be determined by measuring the standard parameters frequency response and loudness rating.

Care should be taken in order to use the appropriate test signals to receive the correct estimations for frequency responses and loudness ratings.

Speech or speech-like signals should be used in test scenarios where one or more of the involved network or terminal elements are unknown (e.g. proprietary low bit-rate coding).

In this case it is recommended to employ the methods described in clause 6.

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