MARASHLIAN & DONAHUE, PLLC
THE COMMLAW GROUP

February 13, 2019

VIA Overnight Courier

Gwen R. Pinson
Executive Director
211 Sower Boulevard
P.O. Box 615
Frankfort, Kentucky 40602-0615

Re: Case No. 2018-00330
Viasat Carrier Services, Inc. – *Application for Designation as an Eligible Telecommunications Carrier*

Dear Ms. Pinson:

On behalf of Viasat Carrier Services, Inc. ("Viasat"), transmitted herewith is one (1) original and five (5) copies of Viasat's response to Commission Staff's Second Request for Information.

Please date-stamp the duplicate of this filing and return it in the self-addressed, postage-prepaid envelope. Should you have any questions regarding the contents of this letter, please do not hesitate to contact the undersigned.

Respectfully submitted,

Michael P. Donahue

Counsel for Viasat Carrier Services, Inc.

Enclosures

1420 Spring Hill Road, Suite 401
McLean, Virginia 22102
P 703.714.1300
F 703.563.6222
W www.commlawgroup.com
E mail@commlawgroup.com

COMMONWEALTH OF KENTUCKY

BEFORE THE PUBLIC SERVICE COMMISSION

In the Matter of:

PETITION OF VIASAT CARRIER SERVICES, INC. FOR DESIGNATION AS AN ELIGIBLE TELECOMMUNICATIONS CARRIER TO RECEIVE CONNECT AMERICA FUND PHASE II AUCTION SUPPORT FOR VOICE AND BROADBAND SERVICES	)))))	CASE NO. 2018-00330

RESPONSE TO STAFF'S SECOND REQUEST FOR INFORMATION
OF VIASAT CARRIER SERVICES, INC.

Jason Sophinos, Associate General Counsel of Viasat, Inc., 349 Inverness Drive South, Englewood, CO 80112, provided the information that forms the basis for each response provided.

1. Refer to the comment received by the Commission from Hughes Network Systems, LCC, on January 25, 2018.

a. State and explain whether Viasat contends that it is not required to use an independent agency or organization to perform latency testing as described in *Connect America Fund*, Order, DA 18-710, 33 FCC Rcd 6509, para. 45, https://docs.fcc.gov/public/attachments/DA-18-710A1_Rcd.pdf.

The requirement that carriers engage independent agencies or organizations to perform testing pertains to live interviews and surveys for Mean Option Score ("MOS") testing for voice services, not latency testing. In the order cited above, the FCC required that "live interviews and surveys must be conducted by an independent agency or organization (Reviewer) to determine the MOS", or Mean Opinion Score. Id. para. 45. The requirement for an independent agency or organization accordingly applies to MOS testing only. Requirements for latency testing are described throughout the order, explaining how carriers are to develop testing protocols for latency testing.

With respect to MOS testing, Hughes Network Systems, LLC's comments conveniently did not mention that Viasat filed a Petition for Reconsideration of the requirement that ETC's utilize an independent agency to conduct MOS testing.[1] In general, Viasat believes that the requirements specifically penalize satellite providers due to the nature of their service. It is Viasat's view that the FCC's decision to require satellite providers to submit to third party verification, while allowing other telecommunications providers to self verify performance was arbitrary and capricious and lacked factual support to justify such disparate treatment. Such discrimination is not in the public interest as it places unnecessary burdens on one class of providers. If the FCC's goal is truly to create technologically neutral requirements for ETCs, the Order's requirements do not support that goal. On the other hand, if the FCC determines that third party verification is needed to test performance levels, then it should apply to all ETCs.

In addition, it is Viasat's view that this requirement will impose significant costs only on Viasat. These cost concerns are related to fundamental supply and demand principles. If the FCC is to require satellite providers to submit to third party testing, Viasat will be the sole party seeking such services. With such a limited customer pool, a third party testing service would not have an incentive to provide a competitive price for its service. Viasat would be forced to pay the presumably high asking price for such testing, as a requirement to receive ETC funding. No other ETC would share this cost burden. However, if all ETCs were required to submit to third party testing, it is conceivable that a robust market would develop for testing services with competitive prices.

b. If Viasat agrees that it is required to use an independent agency or organization to perform latency testing, explain when it contends it is

---

[1] *See* Petition for Reconsideration of Viasat, Inc., WC Docket No. 10-90 (filed Sept. 19, 2018) (available here https://ecfsapi.fcc.gov/file/10920727326915/Viasat%20PFR%20(9-19-18).pdf)

required to do so and how it will satisfy that requirement.

The FCC does not require that latency testing be conducted by an independent agency or organization; rather, as noted above, the requirement applies to live interviews and surveys conducted to determine a Mean Opinion Score. In the order cited above, the FCC stated that carriers would be "required to submit the first set of testing data and accompanying certification by July 1, 2020," and that the "submission should include data for at least the third and fourth quarters of 2019." Id. para. 67.

Viasat will be able to meet the requirements that the FCC established in Order DA 18-710. As noted, Viasat believes that the requirements place an unnecessary burden upon satellite providers that have sought ETC status and sought through a Petition for Reconsideration filed with the FCC to argue that position. Viasat believes that its arguments are sound and that the FCC will reconsider the requirements as suggested by the company, but it will meet all applicable requirements established by the FCC in Order DA 18-710 regardless of the FCC's decision on Viasat's petition.

c. State whether Viasat has retained an independent agency or organization to perform the latency testing described in the Order referenced above and, if so, describe all such testing and provide the results of all such testing, including the Mean Opinion Score (MOS) of each such test.

No, Viasat has yet not retained an independent agency or organization to perform the MOS testing.

d. If Viasat contends that Viasat, Inc., was an independent agency or organization when it conducted the testing described in Viasat's response to Commission Staff's First Request for Information, provide every basis for Viasat's contention that Viasat, Inc., was an independent agency or organization as those terms are used in the Federal Communication Commission (FCC) Order described

above.

Viasat does not contend that it was an independent agency or organization when its employees conducted the MOS test described in its response.

e. Describe the voice quality test performed by Viasat, Inc.'s employees, including the number of calls made to perform the test, whether the persons called were randomly selected, the time during which the test was performed, and the MOS for the test.

Viasat conducted internal testing in July, 2018 for Viasat, Inc.'s current service, not Viasat's CAF II-supported service. However, it is Viasat's position that the results would be similar. The test consisted of Viasat employees randomly calling VoIP subscribers and conducting the MOS score test. The test was conducted using the ITU-T Recommendation P.800 conversational-opinion test and other parameters required by the FCC. The subscribers' mean score exceeded the number required by the FCC. Attachment 1(e) includes the regulations used to structure the test. These are the regulations required by the FCC in Order DA 18-170. *See* Order at paras. 44-45, Appendix A.

f. Provide any further response to the comments of Hughes Network Systems, LLC, that Viasat deems necessary.

2. Refer to the FCC's waiver of the requirement that "winning bidders seeking an FCC ETC designation... demonstrate that it will satisfy applicable consumer protection and service quality standards" as described in WCB Reminds Connect America Fund Phase II *Applicants of the Process for Obtaining Federal Designation* as *an Eligible Telecommunications Carrier,* WC Docket Nos. 09-197, 10-90, Public Notice DA 18-714 (rel. July 10, 2018) https://docs.fcc.gov/public/attachments/DA-18-714A1_Rcd.pdf. State whether Viasat contends that this waiver has any effect on the requirement in the FCC Order discussed in Item 1 above that high-latency

providers demonstrate the performance of their voice service through tests by an independent agency or organization, and explain the basis of Viasat's contention.

The FCC's waiver of the requirement that an ETC certify its compliance with consumer protection and service quality standards does not affect CAF II recipient and ETC obligations to meet applicable standards and comply with rules. The FCC waived the requirement that winning bidders demonstrate that they will satisfy applicable consumer protection and service quality standards because ETCs have an independent obligation to comply with applicable service quality standards and consumer protection rules. In its order, cited above, the FCC explained that for CAF recipients, those "requirements were no longer essential to the Commission's ability to monitor ETC use of support for its intended purposes." Id. at 4 (citing See Connect America Fund et al., ETCs Annual Reports and Certifications, WC Docket Nos. 10-90, 14-58, Report and Order, 32 FCC Rcd 5944, 5944-48, paras. 3-14 (2017)).

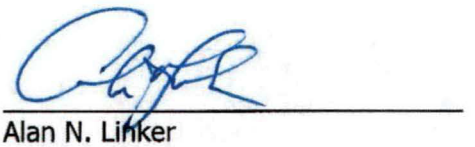
3. State whether and, if so, explain why Viasat's voice telephone service will be able to operate in an emergency as required by 47 C.F.R. § 54.202(a)(2).

VSI has been providing high speed internet service to customers on 24 hours x 365 days a year for more than thirteen years. As part of providing this commercial service, it is necessary to have in place contingency plans for credible emergency situations for each of the major network facilities that are geographically distributed across the United States. These plans contain activation, required staffing, escalation, and communication procedures to deal with such emergencies. Additionally, all the ground-based facilities are equipped with independent power generators and sufficient fuel to operate for several days so as to mitigate power outages. The design of these facilities contains multiple levels of redundancy and autonomy that also mitigate the need for dedicated human interaction. Viasat plans to apply this successful model to its CAF II Auction services and customers.

Viasat, Inc.'s ground network is extremely robust and is built with many redundancies to

design of these facilities contains multiple levels of redundancy and autonomy that also mitigate the need for dedicated human interaction. Viasat plans to apply this successful model to its CAF II Auction services and customers.

Viasat, Inc.'s ground network is extremely robust and is built with many redundancies to minimize service failures. Each ViaSat-2 spot beam is supported by multiple Satellite Access Nodes ("SANs"). Each SAN is located in a different geographic location spread across the United States. As a result, if a single SAN is out-of-service for any reason (due to a storm, cut fiber, technical failure, etc.) the remaining SANs can continue to provide service to the affected spot beam. Viasat, Inc. also has backup generators to provide power to gateways and SANs in the event of a power outage. Viasat's leased terrestrial facilities also have redundancies designed to maintain service in the event of an outage on one facility.



Alan N. Linker
Seiller Waterman LLC
Meidinger Tower, 22nd Floor
462 S. Fourth St.
Louisville, KY 40202
Office 502.584.7400
Direct Fax 502.371.9286
E-mail: alinker@derbycitylaw.com

cc: Parties of Record

*Alan N. Linker
Seiller Waterman, LLC
462 S. Fourth Street Meidinger Tower, 22nd Floor Louisville, KENTUCKY 40202

*Jason Sophinos
General Counsel
349 Inverness Drive South
ViaSat, Inc.
Englewood, COLORADO 80112

*Viasat Carrier Services, Inc.
349 Inverness Drive South
Englewood, CO 80112

*Michael P Donahue
Regulatory Counsel
Marashlian & Donahue, PLLC
The Commlaw Group
1420 Spring Hill Road
Suite 205
Tysons, VIRGINIA 22102

*Denotes Served by Email

Service List for Case 2018-00330

CERTIFICATION

STATE OF COLORADO)

COUNTY OF DOUGLAS)

I, Jason Sophinos, state that I am the Associate General Counsel of Viasat, Inc., the parent company of Viasat Carrier Services, Inc.; that I am authorized to make this Certification on behalf of Viasat Carrier Services, Inc.; that the foregoing RESPONSE TO STAFF'S FIRST REQUEST FOR INFORMATION OF VIASAT CARRIER SERVICES, INC. was prepared under my direction and supervision; and that the contents are true and correct to the best of my knowledge, information, and belief.

I declare under penalty of perjury that the foregoing is true and correct.

Executed this 13 day of February, 2019.

Jason Sophinos
Associate General Counsel
Viasat, Inc.
349 Inverness Drive South
Englewood, CO 80112

Subscribed and affirmed before me in the county of Douglas, State of Colorado, this 13th day of February, 2019.

Margaret O'Connell-Pike
(Notary's Official Signature)

March 14, 2022
(Commission Expiration)

MARGARET O'CONNELL-PIKE
Notary Public
State of Colorado
Notary ID # 20064009950
My Commission Expires 03-14-2022

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.805

(04/2007)

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

Methods for objective and subjective assessment of quality

Subjective evaluation of conversational quality

ITU-T Recommendation P.805

International Telecommunication Union

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
Transmission performance and QoS aspects of IP end-points	Series	P.1000

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

ITU-T Recommendation P.805

Subjective evaluation of conversational quality

Summary

ITU-T Recommendation P.805 describes methods and procedures for conducting conversation tests to evaluate communication quality. The methodology uses examples of scenarios, rating scales and analysis procedures to estimate the subjective quality of telecommunication services. Conversation tests allow the simulation of more realistic situations close to the actual service conditions experienced by telephone customers. In addition, conversation tests are designed to assess the effects of impairments that can cause difficulty while conversing (such as delay, packet loss, echo, interruptions, noise, clipping, etc.), and may be used to study overall system effects or specific degradations as well.

Source

ITU-T Recommendation P.805 was approved on 22 April 2007 by ITU-T Study Group 12 (2005-2008) under the ITU-T Recommendation A.8 procedure.

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 Assembly (WTSA), 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 WTSA 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.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

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, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

CONTENTS

ITU-T Recommendation P.805

Subjective evaluation of conversational quality

1 Scope

This Recommendation describes the method and procedures for generic conversational testing. The protocol described below is aimed to evaluate the effects of degradation such as delay, echo, voice clipping and dropped packets on the quality of voice communications. The methodology described in this Recommendation corresponds to the conversation-opinion tests recommended in [ITU-T P.800]. Contrary to listening tests, conversation-opinion tests are designed to assess the effects of impairments that may influence the quality of speech while conversing (such as delay).

Procedures adapted to specific equipment can be found in [ITU-T P.831] for echo cancellers, [ITU-T P.832] for hands-free terminals and ITU-T Rec. P.840 for circuit multiplication equipment.

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; 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. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

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

[ITU-T P.800.1] ITU-T Recommendation P.800.1 (2006), *Mean Opinion Scores (MOS) terminology.*

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

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

3 Definitions

This Recommendation defines the following term:

3.1 conversation test: A subjective test in which two participants have a real-time conversation, as described in Annex A to [ITU-T P.800] and in [b-Telephonometry].

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ANOVA ANalysis Of VAriance

MANOVA Multivariate ANalysis Of VAriance

MOS Mean Opinion Score

MSD Minimum Significant Difference

POTS Plain Old Telephone Service

5 Conventions

None.

6 Conversation test process

6.1 Purpose

Conversation-opinion tests allow the subjects involved to be in a more realistic situation simulating the actual service conditions experienced by telephone customers. In addition, conversation-opinion tests are designed to assess the effects of impairments that can cause difficulty while conversing (such as delay, packet loss, echo, interruptions, noise, clipping, etc.). They can be used to study overall system effects or specific degradations, such as delay.

Subjects participate in the test as paired sets of communicators. They are seated in separate sound-proof rooms and asked to hold a conversation through the transmission chain and then to give their opinion of the quality on different quality scales. Acoustic noise environments may be simulated in one or both of the rooms.

Depending on the purpose of the test, either expert, experienced or untrained (naïve) subjects may participate. Such tests can be useful to manufacturers, operators and customers, and are an important assessment tool because they provide the closest simulation of real telephony interactions between subscribers. Untrained (naïve) subjects are used when it is important to get an indication of how the general telephone-using population would rate the overall quality and difficulty in using the connection with the system under test. This can be used to give a "global" evaluation of the performance in a range of conditions. However, untrained subjects are unable to describe and identify accurately the types of degradation associated with the system under test.

The main characteristics of a conversation-opinion test are:

- To be very close to a real conversation where people are required to interact and may adapt their behaviour to accommodate the system under test.
- The use of a task to stimulate a conversation with equal participation of both parties.
- Different subjects may have variable behaviour in a conversation (due to culture, personality, etc.), which could create greater variability in subjects' responses in the assessment of speech quality.
- Since subjects have to concentrate on participating in the conversation, and are not specifically involved in assessing the quality performance during the conversation, their final measures may be less sensitive than in listening-only tests.
- Conversation tests are the most valid method for measuring the effect on acceptability of certain system impairments, such as delay.
- Devices under test and simulation tools must be available at the testing lab and must run in real time.
- This conversation test methodology can be adapted to field testing; however, it is foreseen that the control of some experimental variables (e.g., delay, packet loss, acoustic noise, etc.) would be limited.

6.2 Test facilities

A conversational test has to provide as realistic a communication environment as possible. All processes in the communication link are required to be real time.

Switching between conditions that involve different coders and/or different networks parameters must be transparent to the subjects. This may require specialized instrumentation and procedures.

Asymmetry between two subjects in a communication is typical of many actual speech communication scenarios; an asymmetric scenario may be defined by different acoustic noise environments or different transmission conditions. Special consideration may be needed to ensure accurate simulation of acoustic noise environments. For example, significant low frequency power is required for the simulation of automobile environments.

Typical test facilities are illustrated in Figure 6-1 below.

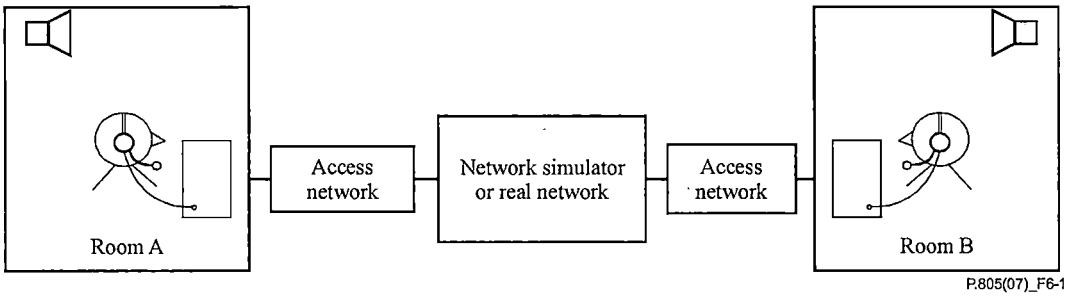


Figure 6-1 – Example of test facilities

Each subject sits in a separate sound-proof room where a variety of acoustic noise environments can be simulated. The environment in both rooms can be the same or different. Examples of different environments are quiet room, office, car, railway station, train and cafeteria. A quiet room might be simulated by the introduction of a suitable level of Hoth noise. Certain chambers also allow reverberation to be considered as an experimental variable.

A description of sound-proof rooms can be found in [ITU-T P.800].

In addition, the send and receive sensors used by the subjects may be the same or different. For example, handset, headset with microphone or microphone and loudspeaker may be used; the choice of the equipment depends on the use case.

6.3 Test design

Most of the test design issues relevant to listening-only tests are also relevant to conversation tests, for example, reference conditions and presentation order effects. A major limitation to conversational test design is the duration of each individual task, or trial, required to exercise each experimental condition. Properly exercising a communication system requires conversations lasting a minimum of 2 minutes. Typical trials require 4 to 5 minutes duration where the conversation period takes 2 to 3 minutes and the response period another 2 minutes. This would limit the total number of conditions in a subject's session to about 24 conditions which would take about 3 hours including instructions, preliminaries and breaks. Tasks designed to measure some system degradations may require conversations longer than 2 to 3 minutes. Compromises have to be made between the test duration and the choice of conditions. If more conditions are to be tested, the test must be separated into several sessions/experiments and may require different subject panels.

An example is shown below in Table 6-1.

Table 6-1 – Timetable for a 24 condition test

	Visit 1				Visit 2		
	Instruction	**Session 1**	**Break**	**Session 2**	**Session 3**	**Break**	**Session 4**
Number of conversations		7 (incl. practice)		6	6		6
Time	15 mins	35 mins	10 mins	30 mins	30 mins	10 mins	30 mins

Conditions that are identical in both directions and that use the same sensors and same acoustic noise are called symmetric conditions. Any other case is considered asymmetric. For asymmetric conditions, subject pairs should be required to swap location for each condition. This limits the total number to 12 asymmetric conditions.

It is recognized that there is a trade-off between the test resolution and the number of votes per condition. The relationship between these two parameters is given by the following general equation for minimum significant difference (*MSD*):

$$MSD = \pm C_{df,p}\sqrt{EMSq / n}$$

where:

$C_{df,p}$ is a t-like value determined by the particular statistical test, the probability level (p), and the degrees of freedom (df)

$EMSq$ is the error mean square derived from the ANOVA

n is the number of votes per condition

Some further information is given in Appendix I.

In order to achieve a sufficient resolution between conditions, it is recommended that the minimum number of subject pairs should in general be 16, but it is also recognized that this number may have to be relaxed in some circumstances in order to reduce the available time for the test but this will reduce the reliability of results.

6.4 Test conditions

Some conditions, including transmission channel and environmental noise, may vary with time. To incorporate this effect, the trial time needs to be increased to scale with the dynamics of the condition. Care should be taken by the experimenter/analyst in order not to overestimate the impact of impairments of non-linear and/or time-variant systems occurring infrequently during the conversation.

Certain types of environmental noise may require sophisticated sound reproduction systems. [b-EG 202 396-1] describes methodologies to create appropriate noise conditions. It also provides a noise database for several environmental conditions, including car simulations.

Examples of test condition variables are:

- Environmental noise (street, car, cafeteria, etc.).
- Room reverberation (none to highly reverberant).
- Transducer (hands free, headset, handset, noise canceller, microphone array, etc.).
- Frequency bandwidth (narrow-band, wideband, audio band, etc.).
- Transmission channel/network characteristics (delay, packet loss, fading, etc.).
- Terminal (mobile phone, soft phone, POTS, etc.).

- Coder (ITU-T G.700-series Recommendations, etc.).
- Etc.

All these variables may be asymmetric. A table summarizing the test conditions must be provided. An example is shown below in Table 6-2.

Table 6-2 – Example of table summarizing the test conditions

Subjects	Number	Naïve/experienced/expert subjects
Groups	Number	2 subjects/group
Rating scales	Number	
Objective of the test		Echo, delay, coder testing
Communication system	Types	Hands-free/headset (monaural/binaural)/handset/group audio terminal Transmission channel, network simulator, access interface
Communication environment		Noise

Acoustic sensors should be calibrated at the beginning of the test; subjects may or may not be allowed to modify the listening level.

When appropriate, sidetone has to be set to a level matching the tested application.

The test environment for each test room shall be defined with the following parameters:

- Room characteristics (size, reverberation time, etc.) see [ITU-T P.800].
- Background noise:
  - level of noise;
  - type of noise (car, babble, etc.);
  - frequency spectrum;
  - dynamic characteristics of the noise field.

6.5 Subjects

The choice of naïve (untrained), experienced or expert subjects depends on the questions and the required degree of precision in the results.

In general, the advice given in [ITU-T P.800] should be taken into account when selecting test subjects.

Some care should be taken when selecting subjects for conversation tests. As with any speech signal processing equipment, some potential subjects will be more experienced than others. It is recognized that the levels of experience with specific equipment or technology is a continuum ranging from those who are completely unfamiliar with technical behaviour of the equipment under test ("non-experts") to those who are thoroughly competent in the operation and maintenance of this equipment ("experts").

The age and gender of all types of subject, together with their partners, should be recorded for all types of tests, but especially for any formal conversation test as opposed to informal expert evaluations.

Unless gender, age and other socio-economic characteristics are design factors of the test, then a formal conversation test should be populated (on a best-endeavour basis) with a random mix of subjects.

6.5.1 Untrained subjects (naïve)

Untrained subjects are accustomed to daily use of a telephone. However, they are neither experienced in subjective testing methodology, nor are they experts in technical implementations of the equipment under test. Ideally, they have no specific knowledge about the device that they will be evaluating. Consistent with [ITU-T P.800], the subjects have not participated in any subjective test in the previous 6 months. In order to control experimental variability associated with subject pair familiarity, it may be appropriate to require that subject pairs are not familiar with each other. Information on familiarity should be reported (unknown, casually acquainted, friends, family members, etc.). Each subject pair is given the opportunity to become familiar with each other in a controlled period of time. Time should be allowed for instructing the subjects about the procedure of the test and the task they have to perform. Practice conditions (the result of which is not included in the result analysis) should be used at the start of the test to ensure that the subjects are comfortable with the test procedure and understand the task. The subject pool should be representative of the telecommunication user pool and the application that the experiment is designed to measure.

6.5.2 Experienced subjects

Experienced subjects are experienced in subjective testing including subjects who participate routinely in subjective testing but does not include individuals who routinely administer, design or run subjective evaluations. Experienced subjects are able to describe an auditory event in detail and are able to separate different events based on specific impairments. They are also able to describe their subjective impressions in detail. However, experienced subjects neither have a background in technical implementations of the equipment under test, nor do they have detailed knowledge of the influence of these implementations on subjective quality.

6.5.3 Experts

Experts are experienced in subjective testing. Experts are able to describe an auditory event in detail and are able to separate different events based on specific impairments. They are able to describe their subjective impressions in detail. They have a background in technical implementations of the equipment under test and do have detailed knowledge of the influence of particular implementations on subjective quality. Individuals directly involved in the design or development of the specific system under test shall be excluded from that particular test.

6.6 Tasks

In addition to the descriptions for full conversation tests in [ITU-T P.800] and [b-Telephonometry], the following consideration may be taken into account. Conversational tests have been carried out with observers (operators) present in the test room together with the subjects, but this is generally not recommended. Instead, an audio/visual link should be used to observe or communicate with the subjects. It is the task of the observers (operators) to document all comments which subjects mention during or after the test. This documentation can be useful for further analysis. In addition, audio/video recordings of the conversations can be made.

6.6.1 Requirements for tasks to be used for untrained subjects

A task should be selected that best fits the requirements of the specific objective of the experiment and the cultural factors of the subject pool. The characteristics required for selecting a task are that:

- it should allow for the generation of a sufficient number of equivalent versions. Each version should stimulate an equivalent level of conversation and interaction;
- it should stimulate semi-structured conversations (too 'open' conversations make it impossible to measure communication efficiency, but too structured communications do not leave room for the subjects to develop a balanced opinion of the channel);
- it should be easily learned;

- it should be intrinsically motivating;
- it should allow for interruptions from the subjects;
- it should be insensitive to changes in subjects' task strategy or skill in performing the task;
- it should represent a cooperative effort between the communicators rather than a competitive effort;
- it should induce the subjects to make use of a rich, varying vocabulary with sufficient two-way interaction;
- it should induce discussion that is phonetically rich and temporally widely distributed (short *and* long utterances and interruptions).

6.6.2 Examples of conversational tasks

The following conversational tasks meet the requirements given in clause 6.6.1:

- Subjects are asked to reach an agreement on an order of preference or time for a set of picture postcards as described in [b-Telephonometry].
- In the so-called "Kandinsky test" the subjects are asked to describe to their partner the position of a set of numbers on a picture. Both subjects have similar pictures, but with some of the numbers in different positions. It is recommended that the picture should be designed for the task and that both the picture and the numbers are easy to describe. This can be achieved by using pictures consisting of coloured, geometrical figures (e.g., Kandinsky or others).
- In the so-called "short conversational tests" proposed by the Ruhr University (Bochum, Germany) in [b-COM 12-35-E], scenarios developed by them are derived from typical situations of every day life: railway enquiries, rental of a car or an apartment, etc. These scenarios have been elaborated to allow a well-balanced conversation between both participants, to stimulate the discussion between persons and to facilitate the naturalness of the conversation. These conversations are approximately 2.5 to 3 minutes in duration. Examples of such scenarios are presented in Appendices IV (German), V (English) and VI (French).
- [b-Telephonometry] also gives some guidance on "simplified conversation tests", where shortcuts are suggested to reduce the time taken or to increase the number of treatments in one experiment. Subjects are asked to rate a number of individual degradations after they have given their opinions on quality and difficulty.
- In the task taken from [b-RICHARDS], random shapes are presented to the subject on a paper sheet or screen. Twenty-four shapes is a typical number on one sheet. There are no meaningful relationships between shapes and their names. The detail and concrete method of how to generate the shapes can be found in [b-RICHARDS]. The operator prepares the same set of sheets for both subjects, but with the shapes in a different order. During the conversation, each subject arbitrarily chooses one shape on the sheet and describes one of its features to his/her partner. His/her partner either guesses the name of the shape based on the information provided or requests additional information from their partner until the shape is identified. Then partners swap their role and continue with another shape. Example shapes are given in Appendix VII.
- A "game" where subjects work with their partner to complete a cooperative task or solve a problem. This approach can be used effectively to control the trial-to-trial variability. Care must be taken to ensure that the game does not limit the conversational vocabulary. The popular board game "BATTLESHIP" is an example of such a task.

In addition to such conversational tasks, specific tasks may be used which stress the interactivity of the conversation, however at the expense of being less realistic and more competitive. Such tasks may be:

- The mutual reading of random numbers or other items as fast as possible, see, e.g., [b-KITAWAKI].
- The mutual verification of numbers or other items as fast as possible, see, e.g., [b-KITAWAKI] or [b-HAMMER]. An example for such a task is given in Appendix VIII.
- More interactive versions of the short conversation test tasks, called "interactive short conversation tests", see [b-RAAKE] and [b-HAMMER]. The task consists of the fast exchange of data. Two subjects are described to be colleagues working in two different sections in one big company, exchanging, e.g., telephone numbers and email-addresses. In order to speed up the conversations, tasks are presented in terms of tabulated data which have been iteratively optimized based on a series of informal tests. These showed that the tabulated data underlying the conversations should not be too different for the two subjects, in order to avoid natural delay in the responses due to the necessity of searching for items in the tables. On the other hand, it was found that too identical list-orders lead to a training effect so that the subjects started to develop a "walkie-talkie" speaking style. As a compromise, one item in the list of each subject is chosen so that it cannot be found in the list of the other subject, with changing positions. This way, fast conversations can be achieved without a strong effect of a "walkie-talkie" style. An example for such a more interactive scenario can be found in Appendix IX.

It should be noted that the impact of, e.g., transmission delay in situations provoked by such interactive tasks may be more severe than in situations provoked by the tasks which are in accordance with clause 6.6.1. This may be due to the structure of the conversation being changed, see, e.g., [b-HAMMER] for a discussion.

6.7 Questions

[ITU-T P.800] and [b-Telephonometry] recommend both a "quality" question using a five-point scale and a "difficulty" question using a binary scale. Some organizations felt that subjects were confused by the "difficulty" question, while other organizations would still prefer to continue using it. As a result, both these scales are reproduced here but new scales are also provided. These new scales may help the subjects to formulate an overall quality judgement by initially focusing their attention on different quality dimensions.

In [ITU-T P.800] and [b-Telephonometry], the scales are as follows:

"*What is your opinion of the connection you have just been using?*"

- Excellent
- Good
- Fair
- Poor
- Bad

The experimenter allocates the following values to the categories: Excellent = 5; Good = 4; Fair = 3; Poor = 2; Bad = 1.

All further statistical processing is performed in terms of these numbers.

"*Did you or your partner have any difficulty in talking or hearing over the connection?*"

- Yes
- No

The experimenter allocates the following values to the responses: Yes = 1; No = 0.

The new scales are given below and the intention is that after each trial (corresponding to one specific condition) the subjects have to evaluate multiple aspects of the communication. The following questions are provided as examples and are representative of the multiple aspects to be considered. Several five-point category scales are provided as well as a binary response scale. The cognitive load on the subjects and therefore the number of questions asked should be minimized to reduce subject fatigue and any possible confusion.

"*How would you assess the sound quality of the other person's voice?*"

The five-point scale descriptors are:

- No distortion at all, natural
- Minimal distortion
- Moderate distortion
- Considerable distortion
- Severe distortion

"*How well did you understand what the other person was telling you?*"

The five-point scale descriptors are:

- No loss of understanding
- Minimal loss of understanding
- Moderate loss of understanding
- Considerable loss of understanding
- Severe loss of understanding

"*What level of effort did you need to understand what the other person was telling you?*"

The five-point scale descriptors are:

- No special effort required
- Minimal effort required
- Moderate effort required
- Considerable effort required
- Severe effort required

"*How would you assess your level of effort to converse back and forth during the conversation?*"

The five-point scale descriptors are:

- No special effort required
- Minimal effort required
- Moderate effort required
- Considerable effort required
- Severe effort required

"*Did you detect (insert distortion of interest here)?*"

- Yes
- No

"If yes, how annoying was it?"

The five-point scale descriptors are:

- No annoyance
- Minimal annoyance
- Moderate annoyance
- Considerable annoyance
- Severe annoyance

"*What is your opinion of the connection you have just been using?*"

The five-point scale descriptors are:

- Excellent quality
- Good quality
- Fair quality
- Poor quality
- Bad quality

The previous examples should be supplemented by the experimenter to address the needs of the specific experiment. When using multiple scales for assessing the multi-dimensional aspect of quality, care should be taken to ensure that the previous responses are not available to the subjects.

6.8 Data analysis and report

6.8.1 Analysis methods

Depending on the experimental design, t-tests, post-hoc multiple means test, ANOVA or MANOVA shall be conducted, as appropriate.

A simple correlation analysis across the various questions should support the MANOVA results. This analysis can provide a more easily understood description of the results.

6.8.2 Reported results

The report must contain the questions and the category scales used.

Summary results should include, as a minimum, mean ratings and standard deviations for all tested conditions and for all questions. If the experimenter expects a gender pairing effect, this special case should be also considered in the design of the experiment and particularly in the pairing of subjects. This effect should be analysed and reported.

Appendix I

Relationship between MSD and number of votes per condition

(This appendix does not form an integral part of this Recommendation)

Using Fisher's least significant difference (LSD) measure as one example of an MSD measure, the following calculation gives some indication of the accuracy that can be expected from a given experiment size. In the case of Fisher's test,

$$C_{df,p} = \pm\sqrt{2} \times t_{df}$$

Assuming $t_{df} \approx 2$ (for 95% confidence) for a large number of degrees of freedom, and EMSq (RMSq) is typically about 0.4, then the LSD (MSD) can be calculated for a given number of votes per condition by evaluating the expression for $C_{df,p}$ in the above equation and substituting the result into the general equation given in clause 6.3. The results are shown graphically below for the situation where the number of subjects is the same as the number of votes per condition:

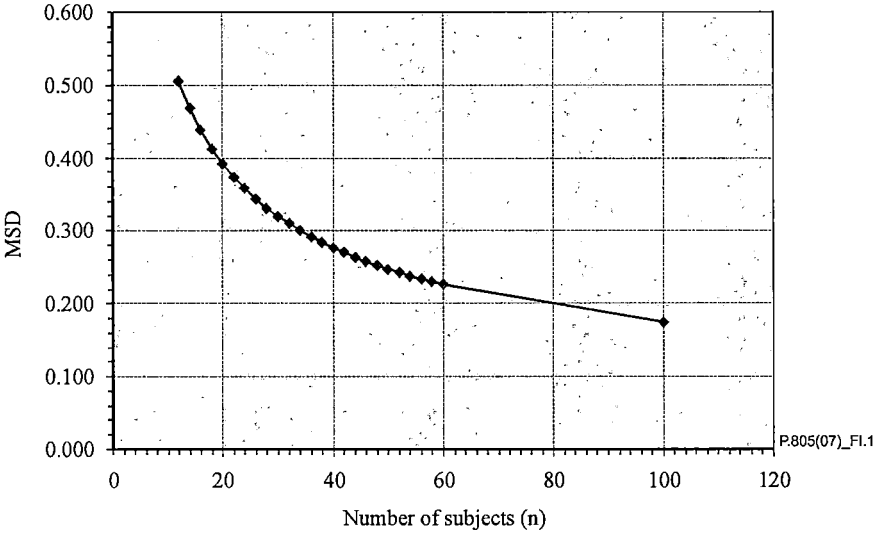


Figure I.1 – Relationship between MSD and number of votes per condition

This graph indicates that the LSD (MSD) calculated in this way for 32 subjects (16 pairs) is approximately 0.3.

Appendix II

Data analysis and presentation of results

(This appendix does not form an integral part of this Recommendation)

The test results should include a table with the following information for each condition in the experiment:

- The "conversation test mean opinion score" (MOS-CQSy) for each condition obtained from all the subjects. Where y applies to narrow-band, mixed bandwidth and wideband (refer to [ITU-T P.800.1]).
- When the conditions are symmetrical, the mean value is calculated from all the scores for the two test rooms. Asymmetric conditions have to be run in two trials in order to keep the same number of votes per condition. The results of a pair of asymmetric conditions are combined as follows:
  - The votes from the A end of condition 1 and the B end of condition 2 have to be combined and vice versa.
  - The standard deviation of the "MOS-CQSy" obtained for all the subjects, for each test condition.

II.1 Calculation and presentation of basic statistical measures

The overall mean ($\overline{Y_c}$) over N subjects for condition C can be obtained from:

$$\overline{Y}_c = \frac{\sum_{i=1}^{N} Y_{i,c}}{N}$$

The standard deviation (S_c) over N subjects for condition C can be obtained from:

$$S_c = \sqrt{\frac{\sum_{i=1}^{N} \left(Y_{i,c} - \overline{Y}_c\right)^2}{N-1}}$$

Finally, the confidence interval (CI_c) at the $(1 - \alpha)$ level can be calculated as:

$$CI_c = (t_{1-\alpha,N-1})\frac{S_c}{\sqrt{N}}$$

Some form of post-hoc multiple means tests should be used to compare test condition means.

The test results should be reported by the test laboratory and the global analysis laboratory in tabular form.

II.2 Thorough analysis

Two statistical analyses should be conducted on the data obtained with these subjective scales. The first analysis consists of a multivariate analysis of variance (MANOVA), which analyses the multiple scales, i.e., dependent variables, and determines the global effects of the experimental factors involved in the experiment. Each individual rating scale in conversation tests involving multiple questions is a separate dependent variable. We would expect the ratings on these dependent variables to show some degree of inter-correlation across test conditions. If the variables were perfectly correlated then we would conclude that each scale was measuring the same underlying variable. We could then combine them into a single measure (e.g., by averaging them) for purposes of statistical analysis and hypothesis testing. If, however, the ratings were uncorrelated, we would conclude that each scale is measuring a different underlying variable and therefore should be treated separately in subsequent analyses. In practice, the degree of inter-correlation among such dependent variables usually falls somewhere between these two extremes. MANOVA is a statistical technique designed to evaluate the results of experiments with multiple dependent variables and determine the nature and number of underlying variables. One previous standardization exercise used conversation tests involving five rating scales. However, MANOVA showed that there was only a single significant underlying variable and provided a "composite conversational quality" measure that was a weighted average of the five rating scales.

In the second statistical analysis, a specific ANOVA should be run on each dependent variable or rating scales to determine if there are significant effects of specific experimental factors for each separate dependent variable. These individual ANOVAs indicate if the differences observed between the ratings obtained for the different conditions are significant for one given dependant variable compared to the composite dependant variable(s) derived from the MANOVA. Finally, Pearson's correlation coefficients should be computed among all dependent variables, to reveal the relationships among those variables.

Appendix III

Example of instructions for the conversation test

(This appendix does not form an integral part of this Recommendation)

INSTRUCTIONS TO SUBJECTS

In this experiment we are evaluating systems that might be used for telecommunication services.

You are going to have a conversation with another user. The test situation simulates communication between two pieces of equipment under test. Most of the situations will correspond to silent environment conditions, but some will simulate more specific situations, such as in a car, in a railway station or in an office environment, when other people are talking in the background.

After the completion of each call conversation, you will have to give your opinions on the quality by answering to the following questions that will be displayed on the screen of the black box in front of you. Your judgement will be stored. You have 8 seconds to answer to each question. After "pressing" the button on the screen, another question will be displayed. You continue the procedure for the six following questions.

"How would you assess the sound quality of the other person's voice?"

No distortion at all, natural	Minimal distortion	Moderate distortion	Considerable distortion	Severe distortion

"What level of effort did you need to understand what the other person was telling you?"

No special effort required	Minimal effort required	Moderate effort required	Considerable effort required	Severe effort required

"How would you assess your level of effort to converse back and forth during the conversation?"

No special effort required	Minimal effort required	Moderate effort required	Considerable effort required	Severe effort required

"Did you detect echo?"

No	Yes

"If yes, how annoying was it?"

No annoyance	Minimal annoyance	Moderate annoyance	Considerable annoyance	Severe annoyance

"What is your opinion of the connection you have just been using?"

Excellent quality	Good quality	Fair quality	Poor quality	Bad quality

From then on you will have a break approximately every 30 minutes. The test will last a total of approximately 60 minutes.

Please do not discuss your opinions with other listeners participating in the experiment.

Appendix IV

Example scenarios in German language

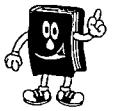
(This appendix does not form an integral part of this Recommendation)

Erläuterung der Symbole:

Als Anrufer haben Sie folgende Symbole:

Dieses Symbol bedeutet: Sie sind der Anrufer. Warten Sie bitte, bis der Versuchsleiter Sie auffordert, das Gespräch zu beginnen.

Neben diesem Symbol ist der Grund ihres Anrufes eingetragen (z.B.: Ich möchte ein Ticket kaufen!)



Neben diesem Symbol steht eine Bedingung, die beim Informationsaustausch berücksichtigt werden soll (z.B.: Ich möchte ein Ticket kaufen → ABER möglichst günstig!)

Neben diesem Symbol sollen Sie alle Informationen eintragen, die Sie von Ihrem Gesprächspartner benötigen.

Neben diesem Symbol stehen alle Informationen, die ihr Gesprächspartner benötigt und die Sie weitergeben sollen.

Wenn Sie angerufen werden, haben Sie folgende Symbole:

Dieses Symbol bedeutet: Sie werden angerufen. Warten Sie bitte bis zum Klingeln und heben Sie dann den Hörer ab.

Neben diesem Symbol stehen Informationen, aus denen Sie die Information heraussuchen sollen, die Ihr Gesprächspartner benötigt (z.B.: Ticketpreise der Bahn für Erwachsene, Schüler/Studenten, Kinder, Senioren, etc.).

Neben diesem Symbol sollen Sie alle Informationen eintragen, die Sie von Ihrem Gesprächspartner benötigen.

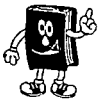
Beide Gesprächspartner haben dieses Symbol:

Neben diesem Symbol steht eine Frage, zu der weder Sie noch Ihr Gesprächspartner eine Information vorliegen hat. Sie sollen diese Frage kurz diskutieren und zu einer einvernehmlichen Lösung kommen.

Übung: Reisebüro

Ihr Name: *Waalke*

Last minute Reise für 1 Woche in den Süden



ab morgen,
nicht nach Spanien,
möglichst günstig

Preis : _____ DM

Abflugtag : _____ um _____ Uhr

Reiseziel : _____

Hotel : _____

Reservierung : Abflug Düsseldorf,
Mastercard,
Kreditkartennummer: 9685 4712 0951 2781,
Gültig bis 10/97

Wie lange vor dem Abflug sollte ich am Flughafen sein?

_____

Übung: Reisebüro

Ihr Name: *Windspielreisen Reisebüro*

Last-minute Reisen
für 1 Woche
ab Düsseldorf oder Frankfurt

Reiseziel	Abflug heute	Abflug morgen	Abflug übermorgen
Spanien (Mallorca) Hotel Playas Arenal *Halbpension*	589,- DM Abflug 19.30 Uhr	759,- DM Abflug 10.15 Uhr	929,- DM Abflug 10.15 Uhr
Italien (Sizilien) Hotel Città del Mare *Vollpension*	719,- DM Abflug 19.10 Uhr	779,- DM Abflug 8.50 Uhr	889,- DM Abflug 8.50 Uhr
Portugal (Algarve) Hotel de Lagos *Halbpension*	499,- DM Abflug 20.20 Uhr	549,- DM Abflug 9.45 Uhr	689,- DM Abflug 9.45 Uhr

(Alle Preise pro Person und incl. Mwst)

Reservierung: NAME : _____

KREDITKARTE : ☐ Mastercard ☐ Visacard

: ☐ American Express ☐ Eurocard

KREDITKARTEN-NR. : _____

GÜLTIG BIS : _____

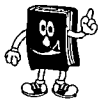
ABFLUG : ☐ Düsseldorf ☐ Frankfurt

_____

Szenario 1: Bahnauskunft

Ihr Name: *Mueller*

Gewünschte Verbindung: Bochum → München



Datum: 20.7.

Verbindung nachts, Schlafwagen

möglichst ohne Umsteigen

Abfahrt : _____ Uhr

Ankunft : _____ Uhr

Zugnummer : _____

Reservierung : 1 Bett im Schlafwagen,
1. Klasse,
Nichtraucher

Ist der Zug zuschlagpflichtig?

Szenario 1: Bahnauskunft

Ihr Name: *Bahnauskunft Deutsche Bahn AG*

Fahrplan: Bochum → München

Zug-Nummer		**D 1127**	**IC 829**	**D 1511**
Bochum	*ab*	19:18	20:50	22:09
Köln	*an*		21:49	
Zug-Nummer			**EN 225**	
Köln	*ab*		22:44	
München	*an*	4:39 Schlaf-/Liegewagen nur in der 2. Klasse	6:30 Schlaf-/Liegewagen nur in der 2. Klasse	7:16 Schlaf-/Liegewagen in 1. und 2. Klasse

Reservierung:

NAME : _____

ANZAHL BETTEN : _____

KLASSE : _____

RAUCHER : ☐

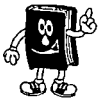
NICHTRAUCHER : ☐

_____

Szenario 2: Bahnauskunft

Ihr Name: *Schmitt*

Gewünschte Verbindung: Dortmund → Berlin



am 5.10,
Berlin Bahnhof Zoo,
Ankunft zwischen 16:00 und 16:30 Uhr

Abfahrt : _____ Uhr

Umsteigen : _____ um _____ Uhr

in/um

Ankunft : _____ Uhr

Reservierung : 2 Sitzplätze,
2. Klasse,
Nichtraucher

Reicht die Zeit zum Umsteigen?

_____

Szenario 2: Bahnauskunft

Ihr Name: *Bahnauskunft Deutsche Bahn AG*

Fahrplan: Dortmund → Berlin

Zugnummer		**IC 645**	**IC 506**	**ICE 941**
Dortmund	*ab*	9:27	11:27	14:28
Magdeburg	*an*	12:35	14:35	
Zugnummer		**IR 2341**	**IR 2343**	
Magdeburg	*ab*	12:49	14:49	
Berlin Zoo	*an*	14:01	16:01	18:51

Reservierung:

NAME : _____

ANZAHL PLÄTZE : _____

KLASSE : _____

RAUCHER : ☐

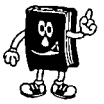
NICHTRAUCHER : ☐

_____

Szenario 3: Flugauskunft

Ihr Name: *Meyer*

Gewünschte Verbindung: Düsseldorf → Berlin



am 23. Juni,
morgens,
Direktflug bevorzugt

Abflug : _____ Uhr

Ankunft : _____ Uhr

Flugnummer : _____

Reservierung : 1 Platz,
: Economy Class

Adresse : Hohlstraße 66, 41747 Viersen
☎ 02162/2 08 33

Von welchem Flughafen kommt man besser ins Stadtzentrum: Tegel oder Schönefeld?

_____

Szenario 3: Flugauskunft

Ihr Name: *Fluginformationsdienst Flughafen Düsseldorf*

Flugplan: Düsseldorf → Berlin

		Lufthansa	British Airways	Lufthansa
Flugnummer		LH 2615	BA 381	LH 413
Düsseldorf	*ab*	6:30	6:35	8:20
Frankfurt	*an*		7:35	
Frankfurt	*ab*		7:50	
Berlin Schönefeld	*an*	7:35		
Berlin Tegel	*an*		8:55	9:25
		(*täglich*)	(*täglich*)	(*täglich*)

Reservierung:

NAME : _____

ADRESSE : _____

: _____

TELEFON : _____

ANZAHL PLÄTZE : _____

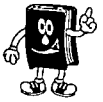
KLASSE : ☐ Business ☐ Economy

_____

Szenario 4: Flugauskunft

Ihr Name: *Janssen*

Gewünschte Verbindung: Frankfurt → New York



am 10.06,
Flughafen JFK

Abflug : _____ Uhr

Zwischenlandung in/um : _____ um _____ Uhr

Ankunft : _____ Uhr

Reservierung : 1 Platz,
MasterCard
Kreditkartennummer 2602 2704 1612 1807,
gültig bis 2/99

Wie groß ist die Zeitverschiebung zwischen Frankfurt und New York?

_____

Szenario 4: Flugauskunft

Ihr Name: *Fluginformationsdienst Flughafen Frankfurt*

Flugplan: Frankfurt → New York

		Lufthansa	United Airlines	Lufthansa
Flugnummer		LH 400	UA 129	LH 402
Frankfurt	*ab*	9:50	7:25	13:30
London Heathrow	*an*		8:00	
London Heathrow	*ab*		9:15	
New York JFK	*an*		11:40	
New York Newark	*an*	12:15 (Ortszeit)	(Ortszeit)	16:05 (Ortszeit)

Reservierung:

NAME : _____

ANZAHL PLÄTZE : _____

KREDITKARTE : _____

KARTEN-NR. : _____

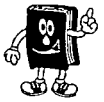
GÜLTIG BIS : _____

_____

Szenario 5: Theaterkasse

Ihr Name: *Bochhoven*

2 Karten für die "Dreigroschenoper"



nächstes Wochenende (Samstag oder Sonntag)
Studentenkarte

Tag : _____

Beginn : _____

Preis : _____ DM

Saal ☐ Großer Saal ☐ Kleiner Saal ☐ Studiobühne

Reservierung : 2 Plätze,
Parkett Mitte

Bis wann müssen die bestellten Karten im Theater abgeholt werden?

_____

Szenario 5: Theaterkasse

Ihr Name: *Schauspielhaus Bochum*

Theaterprogramm

Wochentag	Großer Saal	Kleiner Saal	Studiobühne
Montag, 20.00 Uhr	Kabale & Liebe	--	--
Dienstag, 20.00 Uhr	--	Blick zurück im Zorn	Dilletantenball III
Mittwoch, 20.00 Uhr	Romeo & Julia	--	--
Donnerstag, 20.00 Uhr	Dreigroschenoper	Blick zurück im Zorn	Der Fluch
Freitag, 20.15 Uhr	Kabale & Liebe	--	Dilletantenball III
Samstag, 19.30 Uhr	Dreigroschenoper	Sommernachtstraum	Der Fluch
Sonntag, 19.45 Uhr	Romeo & Julia	Sommernachtstraum	--

Reservierung:

NAME : _____

ANZAHL PLÄTZE : _____

PREIS : ☐ Regulär (14,- DM bis 60,- DM)

☐ Ermäßigt (12,- DM)

PLÄTZE : _____

☐ Parkett ☐ Rang

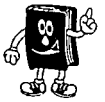
☐ Lins ☐ Mitte ☐ Rechts

_____

Szenario 6: Konzertkasse

Ihr Name: *Dr. Platvoetz*

2 Karten für Starlight Express



Samstag, den 05.07.97 oder
Sonntag, den 06.07.97,
Karten für unter 100,- DM

Datum : _____

Uhrzeit : _____

Platz-Kategorie : _____

Preis : _____ DM/Ticket

Reservierung : Düsseldorfer Straße 224,
45481 Mühlheim
☎ 0208 / 47 59 40

Wird in dem Musical auf deutsch oder auf englisch gesungen?

_____

Szenario 6: Konzertkasse

Ihr Name: *STELLA – The Musical Company*

Starlight Express

	Samstag 05.07.97 20.00 Uhr		**Sonntag 06.07.97 20.00 Uhr**	
	Preis	**Status**	**Preis**	**Status**
Parkett	140,- DM	*Ausverkauft*	110,- DM	
Tribüne	170,- DM	*Ausverkauft*	140,- DM	
Mitteltribüne	170,- DM		140,- DM	*Ausverkauft*
Seitentribüne	120,- DM		90,- DM	
Rang	90,- DM		60,- DM	*Ausverkauft*

Reservierung : ANZAHL PLÄTZE : _____

NAME : _____

ADRESSE : _____

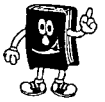
TELEFON : _____

_____

Szenario 7: Autovermietung

Ihr Name: *Pfeifer*

1 Kleintransporter für einen Umzug



für ein Wochenende (Samstag und Sonntag),
mindestens 1.80 m Ladehöhe,
möglichst günstig

Modell : _____

Ladehöhe : _____

Preis : _____ DM/Tag

Reservierung : Eurocard
Kreditkartennummer: 8823 7453 6592 0023
gültig bis 12/99
Führerschein Klasse 3

Ab wieviel Uhr kann ich den Wagen morgens abholen?

_____

Szenario 7: Autovermietung

Ihr Name: *Allround Autovermietung*

Belegungsplan Kleintransporter

Tag	Ford Transit	Mercedes Sprinter	VW Transporter
Montag	*belegt*	*belegt*	
Dienstag	*belegt*	*belegt*	
Mittwoch		*belegt*	*belegt*
Donnertag	*belegt*		*belegt*
Freitag		*belegt*	
Samstag			
Sonntag	*belegt*		
Führerschein	Klasse 3	Klasse 3	Klasse 3
Freie Kilometer	100 km	100 km	100 km
Preis pro Tag	129,- DM	169,- DM	109,- DM
Ladehöhe	2 m	2.30 m	1.90 m

Reservierung: NAME : _____

KREDITKARTE : _____

KREDITKARTE-NR. : _____

GÜLTIG BIS : _____

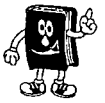
FÜHRERSCHEIN : ☐ Klasse 1 ☐ Klasse 2 ☐ Klasse 3

_____

Szenario 8: Autovermietung

Ihr Name: *Sehl*

1 PKW mieten



Donnerstag bis Samstag,
möglichst günstig

Modell : _____

Standort : _____

Preis : _____DM/Tag

Reservierung : American Express
Kreditkartennummer: 6430 5188 1926 4623
gültig bis 09/98

Ist die Versicherung im Preis enthalten?

_____

Szenario 8: Autovermietung

Ihr Name: *Intercar Autovermietung*

Belegungsplan PKW's

Tag	VW Golf	Opel Corsa	Mercedes C 240	Audi Quattro
Montag		*belegt*	*belegt*	*belegt*
Dienstag			*belegt*	*belegt*
Mittwoch	*belegt*	*belegt*	*belegt*	
Donnertag	*belegt*			
Freitag	*belegt*			
Samstag	*belegt*			
Sonntag	*belegt*		*belegt*	*belegt*
Preis/Tag	87,-DM	78,- DM	103,- DM	99,- DM
Freie Kilometer	300 km	100 km	100 km	100 km
Standort	Parkhaus Bahnhof	Parkhaus Bahnhof	Parkhaus Herner Str.	Parkhaus Herner Str.

Reservierung: NAME : _____

KREDITKARTE : ☐ MasterCard ☐ Visa ☐ American Express

KARTEN-NR. : _____

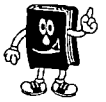
GÜLTIG BIS : _____

_____

Szenario 9: Zimmerreservierung

Ihr Name: *Waissenberk*

2 Einzelzimmer



16.-23. März,
mit Bad und/oder Balkon bevorzugt

Zimmer Nr	: _____	und	_____	
Bad	: ☐ JA	☐ NEIN	☐ JA	☐ NEIN
Balkon	: ☐ JA	☐ NEIN	☐ JA	☐ NEIN

Reservierung : 1. Zimmer für André Waissenberk
2. Zimmer für Sarah Osterloh
☎ 0 234 / 89 71 26

Liegt im März noch genug Schnee zum Skifahren?

_____

Szenario 9: Zimmerreservierung

Ihr Name: *Resort Hotel Tirol*

Zimmerbelegung

Woche	Zimmer Nr 106	Zimmer Nr 118	Zimmer Nr 235	Zimmer Nr 257
01.03.-07.03.97		*belegt*	*belegt*	*belegt*
08.03.-15.03.97	*belegt*		*belegt*	
16.03.-23.03.97				
24.03.-31.03.97	*belegt*	*belegt*		
Zimmerart	**Doppelzimmer**	**Einzelzimmer**	**Einzelzimmer**	**Doppelzimmer**
Bad	ohne	mit	mit	ohne
Balkon	mit	mit	ohne	ohne

Reservierung: NAME (n) : _____

: _____

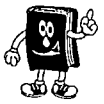
TELEFON : _____

_____

Szenario 10: Apartmentreservierung

Ihr Name: *Dr. Sieverding*

1 Ferienwohnung in Rimini/Italien



24.-31. März
für mindestens 4 Personen
möglichst in Strandnähe

Name der Wohnung : _____

Größe : _____ Personen

Preis : _____ DM/Woche

Nähe zum Strand _____ m

Reservierung : Königsallee 109
44789 Bochum
☎ 0234/75618

Ist es im März schon warm genug zum Baden?

_____

Szenario 10: Apartmentreservierung

Ihr Name: *Family Club Italien*

Apartmentbelegung

Woche	Apartment "Calderone"	Apartment "Angeli"	Apartment "Pronti"	Apartment "Riviera"
01.03.-07.03.97		*belegt*	*belegt*	*belegt*
08.03.-15.03.97	*belegt*	*belegt*	*belegt*	
16.03.-23.03.97	*belegt*			*belegt*
24.03.-31.03.97				*belegt*
Größe Personen	3	4	6	6
Nähe zum Strand	ca. 200 m	ca. 50 m	ca. 50 m	ca. 50 m
Preis pro Woche	210,- DM	379,- DM	290,- DM	569,- DM

Reservierung: NAME : _____

ADRESSE : _____

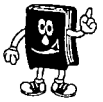
: _____

_____

Szenario 11: Pizzaservice

Ihr Name: *Clemens*

1 große Pizza



für 2 Personen

vegetarische Pizza bevorzugt

Belag : _____

_____

Preis : _____ DM

Lieferung an : Glücksburger Str. 41,
Bochum
☎: 7 34 20

Wie lange dauert es bis die Pizza geliefert wird?

_____

Szenario 11: Pizzaservice

Ihr Name: *Pizzeria Roma*

Pizza

Pizzen	1 Person	2 Personen	4 Personen
Toscana (Schinken, Champignons, Tomaten, Käse)	8,- DM	15,- DM	27,- DM
Tonno (Thunfisch, Zwiebeln, Tomaten, Käse)	10,- DM	19,- DM	35,- DM
Fabrizio (Salami, Schinken, Tomaten, Käse)	8,- DM	15,- DM	27,- DM
Vegetaria (Spinat, Champignons, Tomaten, Käse)	9,- DM	17,- DM	31,-DM

Lieferung an: NAME : _____

ADRESSE : _____

_____

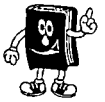
TELEFON : _____

_____

Szenario 12: Partyservice

Ihr Name: *Hessler*

Essen für eine Party



für 30 Personen
möglichst günstig

Angebot : _____

Portion : _____

Beilage : _____

Preis /Person : _____ DM

Lieferung an : Talstraße 124
44869 Bochum

Datum : Samstag, den 31.05.97 um 12.00 Uhr

Wieviel alkoholfreie Getränke sollte man für 30 Personen bestellen?

_____

Szenario 12: Partyservice

Ihr Name: *Good Time Party Service GbR*

Party Angebote

Angebot	Preis/Person	Beilage	Portion
Warmes Buffet (ab 40 Personen)	14,- DM	Salate	500 g
Kaltes Buffet (ab 30 Personen)	9,- DM	Schokoladenpudding	400 g
Gulasch (ab 25 Personen)	6,- DM	Baguette	350 g
Bockwürstchen (ab 20 Personen)	4,50 DM	Kartoffelsalat	300 g

Lieferung an: NAME : _____

ADRESSE : _____

ANZAHL PERSONEN : _____

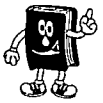
TAG, DATUM & UHRZEIT : _____

_____

Szenario 13: Bibliothek

Ihr Name: *Hüllstrung*

Buch ausleihen



Ein Krimi von Agatha Christie
falls das Buch ausgeliehen ist, bitte vormerken

Buchtitel : _____

vorgemerkt für : _____

den : _____ Tage

Ausleihdauer

Bestellung : ☎ 026 23/27 81
Kundennummer: H 1092

Wieviel kostet das Ausleihen?

_____

Szenario 13: Bibliothek

Ihr Name: *Stadtbücherei Bochum*

Lieferzeiten für bestellte Bücher

Auswahl von Autoren	Auswahl von Titeln	ausgeliehen bis nächsten	Ausleihdauer
Agatha Christie	Mord im Orientexpress	Freitag	10 Tage
	Tod am Nil	Montag	10 Tage
	Die Mausefalle	Dienstag	10 Tage
Astrid Lindgren	Pippi Langstrumpf	*erhältlich*	8 Tage
	Die Kinder von Bullerbü	Mittwoch	8 Tage
	Karlsson vom Dach	*erhältlich*	8 Tage
Michael Ende	Jim Knopf und Lukas der Lokomotivführer	Donnerstag	12 Tage
	Momo	*erhältlich*	12 Tage
	Die unendliche Geschichte	Freitag	12 Tage

Bestellung : NAME : _____

KUNDENNUMMER : _____

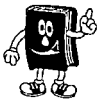
TELEFON : _____

_____

Szenario 14: Bibliothek

Ihr Name: *Jacoby*

Buch verlängern



"Pippi Langstrumpf" von Astrid Lindgren
seit 1 Woche ausgeliehen
möglichst lange verlängern

Datum heute : _____

Verlängert bis : _____

Preis : _____ DM

Reservierung : Kundennummer: J 91 38 70

Welche anderen Bücher gibt es noch von Astrid Lindgren?

_____

Szenario 14: Bibliothek

Ihr Name: *Stadtbücherei Bochum*

Verlängerungen für Bücher,
die bereits 1 Woche ausgeliehen wurden

Buchart	Verlängerung um max.	Preis
Biographien	KEINE VERLÄNGERUNG MÖGLICH	--
Fantasy	7 Tage	2,- DM
Horror	14 Tage	2,50 DM
Kinderbücher	5 Tage	1,- DM
Krimi	7 Tage	1,50 DM
Nachschlagewerke	KEINE VERLÄNGERUNG MÖGLICH	--
Sachbücher	1 Monat	3,50 DM

Reservierung: BUCHTITEL : _____

VERLÄNGERT BIS : _____

NAME : _____

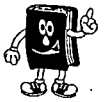
KUNDENNUMMER : _____

_____

Szenario 15: Wohnungsanzeige

Ihr Name: *Kayser*

Wohnungsanzeige aufgeben



Rubrik: Mietangebote
am Samstag,
möglichst auffällig

Textart : _____

Preis : _____ DM

Überweisung auf : _____

Text : 3 Zimmer, KDB[1], 67 m^2, 745,- DM kalt, ab 01.07.97,
☎ 0 234/71 45 30

Wie hoch ist die Auflage des Stadtspiegel am Wochenende?

_____

[1] KDB = **K**üche-**D**iele-**B**ad

Szenario 15: Wohnungsanzeige

Ihr Name: *Anzeigenannahme Stadtspiegel Bochum*

Preise pro Zeile für Wohnungsanzeigen

Tage / Textart	Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Samstag
Text fett + umrahmt	10,- DM	10,- DM	12,- DM	10,- DM	12,- DM	15,- DM
Text fett	8,- DM	8,- DM	9,50 DM	8,- DM	9,50 DM	13,- DM
Text einfach + umrahmt	7,- DM	7,- DM	8,50 DM	7,- DM	8,50 DM	11,- DM
Text einfach	5,- DM	5,- DM	6,- DM	5,- DM	6,- DM	10,- DM

Kontonummer 444 100, BLZ 430 500 01, Sparkasse Bochum

Rubrik : ☐ Mietangebote ☐ Mietgesuche ☐ Sonstige

Text : _____

: _____

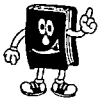
: _____

_____

Szenario 16: Wohnungssuche

Ihr Name: *Nowack*

Preise und Besichtigungstermine für Wohnungen in Bochum



3 Zimmer KDB[1],
ca. 50 bis 65 qm,
Besichtigung nur am Wochenende möglich

Größe der Wohnung : _____ m^2

Adresse : _____

Besichtigung am : _____ um _____ Uhr

Preis : _____

Telefon : ☎ 0234/80 23 63 (privat)
0234/55 00 12 (Büro)

Wie hoch ist die Kaution?

_____

[1] KDB = **K**üche-**D**iele-**B**ad

Szenario 16: Wohnungssuche

Ihr Name: *Trend Immobilien GmbH*

Freie Wohnungen in Bochum

Wohnung	2 Zimmer KDB[1]	1 Zimmer KDB	3 Zimmer KDB	3 Zimmer KDB
m^2	47 m^2	30 m^2	58 m^2	64 m^2
Adresse	Markstr. 235	Talstr. 41	Feldweg 2	Blumenallee 78
Besichtigung	Nächsten Mittwoch ab 17.30 Uhr	Nächsten Freitag ab 18.00 Uhr	Nächsten Samstag ab 10.00 Uhr	Nächsten Samstag ab 14.00 Uhr
Miete (kalt)	564,- DM	385,- DM	756,- DM	928,- DM

Termin für: NAME : _____

TELEFON : _____

TAG + UHRZEIT : _____

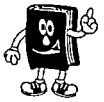
_____

[1] KDB = **K**üche-**D**iele-**B**ad

Szenario 17: Arzttermin

Ihr Name: Rheinhardt

Termin beim Augenarzt



möglichst vormittags vor 9.00 Uhr
Mittwoch oder Donnerstag

Termin : Tag : _____

Datum : _____

Uhrzeit : _____

Angaben zur Person : Krankenkasse UNIKA,
regelmäßiger Patient, letzter Besuch vor 6 Monaten,
Kontrolltermin

Wie lange muß ich bei Ihnen warten?

_____

Szenario 17: Arzttermin

Ihr Name: *Augenarztpraxis Dr. Otto*

Terminplan

1. Juniwoche

	Montag, 02.06.97	**Dienstag 03.06.97**	**Mittwoch 04.06.97**	**Donnerstag 05.06.97**	**Freitag 06.06.97**
8.30		Klos			Sting
8.45	Kohl			Kree	Gabriel
9.00	Blüm	Sammer	Chapuisat	Feiersinger	Jackson
9.15	Waigel	Kohler	Riedle		
9.30	Seehofer	Cesar		Tanko	Ami Grant
9.45	Rau	Sousa	Ricken	Tretschok	Bon Jovi
10.00	Lafontaine	Zorc	Herrlich	Bout	Clapton
10.15	Schroeder	Möller	Heinrich		Grönemeyer
10.30	Scharping	Lambert	Freund		
10.45		Reuter		De Beer	Domingo

Tag, Datum & Uhrzeit : _____

Welche Krankenkasse? : _____

Waren Sie schon einmal bei uns? : ☐ JA ☐ NEIN

Wann waren Sie das letzte Mal da? : _____

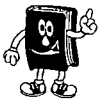
Worum geht es? : ☐ Brille ☐ Kontrolle ☐ Beschwerden

_____

Szenario 18: Arzttermin

Ihr Name: *Reineke*

Termin beim Augenarzt verschieben



möglichst nachmittags nach 17.00 Uhr
Wochentag egal
es handelt sich um einen Kontrolltermin

NeuerTermin : Datum : _____

Uhrzeit : _____

Angaben zur Person : Alter Termin: Dienstag 15.30 Uhr
Krankenkasse BUDAKA,
Kontrolltermin

Wie lange dauert der Kontrolltermin?

_____

Szenario 18: Arzttermin

Ihr Name: *Augenarztpraxis Dr. Otto*

Terminplan

1. Juliwoche

	Montag	**Dienstag**	**Mittwoch**	**Donnerstag**	**Freitag**
15.30	Klos	**Reineke**		Sousa	Sting
15.45		Kohl	*Am*	Tanko	Gabriel
16.00	Sammer	Blüm	*Mittwoch*	Ricken	Jackson
16.15	Kohler	Waigel	*nachmittag*	Chapuisat	
16.30	Cesar	Seehofer	*hat*	Riedle	
16.45	Reuter	Rau	*die*	Herrlich	Bon Jovi
17.00	Zorc	Lafontaine	*Praxis*		Clapton
17.15	Möller	Schroeder	*geschlossen*		Grönemeyer
17.30	Lambert	Scharping	*!*	Heinrich	
17.45					Domingo

Alter Termin : _____

Neuer Termin : _____

Krankenkasse : _____

Worum geht es? : ☐ Brille ☐ Kontrolle ☐ Beschwerden

_____

Appendix V

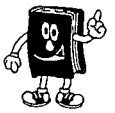
Example scenarios in English language

(This appendix does not form an integral part of this Recommendation)

If you are the caller, you have the following symbols:

This symbol means that you are the caller. Please wait until the supervisor tells you to start the conversation.

Next to this symbol, the reason for your call is given (e.g., I would like to buy a ticket).



Next to this symbol is a condition which should be applied to the exchange of information (e.g., I would like to buy a ticket→ ***but*** *as cheaply as possible).*

Next to this symbol you should write down what information you want to receive from your partner.

Next to this symbol is all the information your partner requires and which you will pass on to him/her.

If somebody rings you, you have the following symbols:

This symbol means that you are being called. Wait until it rings and pick up the receiver.

Next to this symbol is information from which you should select the details that your partner requires (e.g., train ticket prices for adults, children, special conditions for students).

Next to this symbol you should write down all the information you require from your partner.

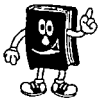
Both speakers have this symbol:

Next to this symbol is a question to which neither you nor your partner will have information. You should discuss the question briefly and find a solution that is acceptable to both of you.

Practice: Travel agent

Your name: *Walker*

One week last minute trip to the Mediterranean



Departing tomorrow

Not to Spain

As cheap as possible

Price : _____£

Day of departure : _____at_____h

Destination : _____

Hotel : _____

Reservation : Flying from Manchester

Mastercard

Credit card number: 9685 4712 0951 2781

Valid till 10/99

How long before the departure time do I have to check in?

_____

Practice: Travel agent

Your name: *Thomas Cook travel agent*

One week last minute trip

from Liverpool or Manchester

Destination	Departure today	Departure tomorrow	Departure the day after tomorrow
Spain (Mallorca) Hotel Playas Arenal *Half board*	£90 Dep. 19:30 h	£150 Dep. 10:15 h	£308 Dep. 10:15 h
Italy (Sicily) Hotel Città del Mare *Full board*	£140 Dep. 19:10 h	£160 Dep. 8:50 h	£290 Dep. 8:50 h
Portugal (Algarve) Hotel de Lagos *Half board*	£170 Dep. 20:20 h	£180 Dep. 9:45 h	£230 Dep. 9:45 h
(All prices are per person and include V.A.T.)			

Reservation: Name : _____

Credit card : ☐ Mastercard ☐ Visacard

☐ American Express ☐ Eurocard

Credit card number : _____

Valid till : _____

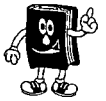
Flying from ☐ Liverpool ☐ Manchester

_____

Scenario 1: British Rail Travel Information

Your name: *Thompson*

Intended journey: London King's Cross → Inverness



Date: July 20th

Night train connection, sleeper

Through train if possible

Departure time : _____ h

Arrival time : _____ h

Train number : _____

Reservation : One berth in a sleeping-car
First class
Non-smoker

What time does the buffet car start serving breakfast?

_____

Scenario 1: British Rail travel information

Your name: *British Rail travel service*

Timetable: London → Inverness

Train number		**IC 1127**	**IC 829**	**IC 1511**
London King's Cross	*Dep.*	19:18 h	20:50 h	22:09 h
York	*Arr.*		21:49 h	
Train number			**EN 225**	
York	*Dep.*		22:44 h	
Inverness	*Arr.*	4:39 h Sleeper only 2nd class	6:30 h Sleeper only 2nd class	7:16 h Sleeper 1st and 2nd class

Reservation:

Name : _____

Number of berths : _____

Class : _____

Smoker ☐

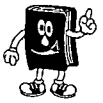
Non-smoker ☐

_____

Scenario 2: British Rail travel information

Your name: *Hill*

Intended journey: London Euston → Carlisle



Date: October 5th

Arrival between 16:00 h and 16:30 h

Departure time : _____ h

Change at : _____ at _____ h

Arrival time : _____ h

Reservation : 2 seats
Second class
Non-smoker

Is there enough time to change?

_____

Scenario 2: British Rail travel information

Your name: *British Rail travel service*

Timetable: London Euston → Carlisle

Train number		**IC 645**	**IC 506**	**IC 941**
London Euston	*Dep.*	9:27 h	11:27 h	14:28 h
Crewe	*Arr.*	11:05 h	13:35 h	
Train number		**IC 2341**	**IC 2343**	
Crewe	*Dep.*	11:20 h	13.47 h	
Carlisle	*Arr.*	14:01 h	16:01 h	18.51 h

Reservation:

Name : _____

Number of seats : _____

Class : _____

Smoker ☐

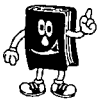
Non-smoker ☐

_____

Scenario 3: Information on flights

Your name: *Parker*

Intended journey: London Heathrow → Düsseldorf



Date: June 23rd

Morning flight

Direct flight preferred

Departure : _____ h

Arrival : _____ h

Flight number : _____

Reservation : One seat

Economy Class

Address : 47 Industry Street, Sheffield

☎ 20833

From which airport is it easier to get into Cologne city centre:

Düsseldorf or Cologne/Bonn?

_____

Scenario 3: Information on flights

Your name: *Heathrow flight information*

Flight schedule: Heathrow → Düsseldorf

		Lufthansa	British Airways	Lufthansa
Flight number		LH 2615	BA 381	LH 413
Heathrow	*Dep.*	6:30 h	6:35 h	8:20 h
Brussels	*Arr.*		7:35 h	
Brussels	*Dep.*		7:50 h	
Düsseldorf	*Arr.*	7:35 h (*daily*)	8:55 h (*daily*)	9:25 h (*daily*)

Reservation:

Name : _____

Address : _____

: _____

Telephone number : _____

Number of seats : _____

Class ☐ Business ☐ Economy

_____

Scenario 4: Information on flights

Your name: *Kelly*

Intended journey: Frankfurt → New York

Date: June 6th

To JFK airport

Departure : _____ h

Stopover at : _____ at _____ h

Arrival : _____ h

Reservation : One seat

MasterCard

Credit card number: 2602 2704 1612 1807

Valid till 2/99

What is the time difference between Frankfurt and New York? Are they ahead of us or behind?

_____

Scenario 4: Information on flights

Your name: *Flight information Frankfurt airport*

Timetable: Frankfurt → New York

		Lufthansa	United Airlines	Lufthansa
Flight number		LH 400	UA 129	LH 402
Frankfurt	*Dep.*	9:50 h	7:25 h	13:30 h
London Heathrow	*Arr.*		8:00 h	
London Heathrow	*Dep.*		9:15 h	
New York JFK	*Arr.*		11:40 h	
New York Newark	*Arr.*	12:15 h (local time)	(local time)	16:05 h (local time)

Reservation:

Name : _____

Number of seats : _____

Credit card : _____

Credit card number : _____

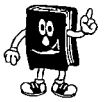
Valid till : _____

_____

Scenario 5: Theatre box office

Your name: *Richardson*

2 tickets for the "Three Penny Opera"



Next weekend (Saturday or Sunday)

Student reduction

Day : _____

Starting : _____h

Price : _____ £

Auditorium ☐ Main auditorium ☐ Small auditorium ☐ Studio

Reservation : Two seats

Middle section of the stalls

By what time do we have to collect tickets we have reserved?

_____

Scenario 5: Theatre box office

Your name: *Palladium Theatre, London*

Theatre programme

Weekday	Main auditorium	Small auditorium	Studio
Monday, 20:00 h	Hamlet	--	--
Tuesday, 20:00 h	--	Look Back in Anger	The Crucible III
Wednesday, 20:00 h	Romeo and Juliet	--	--
Thursday, 20:00 h	Three Penny Opera	Look Back in Anger	My Fair Lady
Friday, 20:15 h	Hamlet	--	The Crucible III
Saturday, 19:30 h	Three Penny Opera	A Midsummer Night's Dream	My Fair Lady
Sunday, 19:45 h	Romeo and Juliet	A Midsummer Night's Dream	--

Reservation:

Name : _____

Number of seats : _____

Price : ☐ Adults (£5 to £20) ☐ Reduction (£4)

Seats : ☐ Stalls ☐ Row

☐ Right ☐ Middle

_____

Scenario 6: Concert box office

Your name: *Dr Spencer*

2 tickets for Starlight Express

Saturday, July 5th or
Sunday, July 6th
Tickets costing less than £35

Date : _____

Time : _____h

Seat category : _____

Price : _____£ per ticket

Reservation : 12 Industry Street
London
☎ 47 59 40

Will the musical be in German or in English?

_____

Scenario 6: Concert box office

Your name: *STELLA – The Musical Company*

Starlight Express

	Saturday, July 5th 20:00 h		Sunday, July 6th 20:00 h	
	Price	**Availability**	**Price**	**Availability**
Front stalls	£50	*Sold out*	£37	
Dress circle	£60	*Sold out*	£47	
Gallery	£60		£47	*Sold out*
Side gallery	£40		£30	
Row at the rear	£30		£20	*Sold out*

Reservation : Number of seats : _____

Name : _____

Address : _____

Telephone number : _____

_____

Scenario 7: Renting a car

Your name: Phillips

One small removal van

For one weekend (Saturday and Sunday)

At least 1.80 m headroom inside

As cheap as possible

Type : _____

Headroom inside : _____m

Price : _____£ per day

Reservation : Eurocard

Credit card number: 8823 7453 6592 0023,

Valid till 12/99

HGV driving licence

How early can I collect the van?

_____

Scenario 7: Renting a car

Your name: *AVIS Car Rentals*

Booking sheet for small vehicle

Day	Ford Transit	Mercedes Sprinter	VW Transporter
Monday	*Booked*	*Booked*	
Tuesday	*Booked*	*Booked*	
Wednesday		*Booked*	*Booked*
Thursday	*Booked*		*Booked*
Friday		*Booked*	
Saturday			
Sunday	*Booked*		
Type of licence	HGV	HGV	HGV
Kilometres free of charge	100 km	100 km	100 km
Price per day	£40	£60	£35
Headroom inside	2 m	2.30 m	1.90 m

Reservation:

Name : _____

Credit card : _____

Credit card number : _____

Valid till : _____

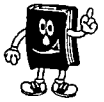
Driving licence ☐ class 1 ☐ class 2 ☐ class 3

_____

Scenario 8: Renting a car

Your name: *Pickering*

Rent a car



Thursday to Saturday

As cheaply as possible

Model : _____

Collection : _____

Point : _____

Price _____ £ per day

Reservation : American Express

Credit card number: 6430 5188 1926 4623

Valid till 09/99

Is insurance included in the price?

_____

Scenario 8: Renting a car

Your name: *Hertz Car rentals*

Booking sheet for cars

Day	VW Golf	Vauxhall Corsa	Mercedes C 240	Audi Quattro
Monday		*Booked*	*Booked*	*Booked*
Tuesday			*Booked*	*Booked*
Wednesday	*Booked*	*Booked*	*Booked*	
Thursday	*Booked*			
Friday	*Booked*			
Saturday	*Booked*			
Sunday	*Booked*		*Booked*	*Booked*
Price/day	£ 30	£ 25	£ 35	£ 33
Kilometres free of charge	300 km	100 km	100 km	100 km
Collection point	Multi-storey car park in station	Multi-storey car park in station	Car park York Street	Car park York Street

Reservation:

Name : _____

Credit card : ☐ MasterCard ☐ Visa ☐ American Express

Credit card Number : _____

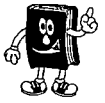
Valid till : _____

_____

Scenario 9: Room reservation

Your name: *Francis*

Two singles



Dates: From March 16th to 23rd

With a bathroom/balcony if possible

Room number : _____ and _____

Bath	: ☐ yes	☐ no	☐ yes	☐ no
Balcony	: ☐ yes	☐ no	☐ yes	☐ no

Reservation : 1. Room for Andrew Francis

2. Room for Sarah Harrison

☎ 897126

Is there still enough snow in March to go skiing?

_____

Scenario 9: Room reservation

Your name: *Resort Hotel Tirol*

Room plan

Week	Room No. 106	Room No. 118	Room No. 235	Room No. 257
From 1st to 7th March		*Booked*	*Booked*	*Booked*
From 8th to 14th March	*Booked*		*Booked*	
From 15th to 21st March				
From 22nd to 28th March	*Booked*	*Booked*		
Type of room	**Double**	**Single**	**Single**	**Double**
Bath	Without	With	With	Without
Balcony	With	With	Without	Without

Reservation: Name : _____

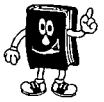
Telephone number : _____

_____

Scenario 10: Booking an apartment

Your name: *Dr Silversmith*

One holiday apartment in Rimini, Italy



From March 24th to 31st

To sleep at least 4 people

If possible, near the beach

Name of apartment : _____

Size : _____ persons

Price : _____ £ per week

Distance from Beach : _____ m

Reservation : 233 Middle Street
Oxford
☎ 75618

Is it warm enough to go swimming in March?

_____

Scenario 10: Booking an apartment

Your name: *Family Club Italy*

Apartment booking sheet

Week	Apartment "Calderone"	Apartment "Angeli"	Apartment "Pronti"	Apartment "Riviera"
From 1st to 7th of March		*Booked*	*Booked*	*Booked*
From 8th to 14th of March	*Booked*	*Booked*	*Booked*	
From 15th to 21st of March	*Booked*			*Booked*
From 22nd to 28th of March				*Booked*
Number of beds	3	4	6	6
Distance from beach	Approx. 200 m	Approx. 50 m	Approx. 50 m	Approx. 50 m
Price per week	£ 70	£ 120	£ 90	£ 180

Reservation: Name : _____

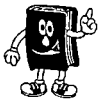
Address : _____

_____

Scenario 11: Pizza service

Your name: *Clemence*

One large pizza



For 2 people

Vegetarian pizza preferred

Topping : _____

_____

Price : _____£

Delivery to : 109 Weather Street
Brighton
☎: 73420

How long will we have to wait for the pizza to be delivered?

_____

Scenario 11: Pizza service

Your name: *Pizzeria Roma*

Pizza

Pizzas	1 person	2 persons	4 persons
Toscana (ham, mushrooms, tomatoes, cheese)	£3.20	£5.95	£10.50
Tonno (tuna, onions, tomatoes, cheese)	£3.95	£7.50	£13.95
Fabrizio (salami, ham, tomatoes, cheese)	£4.20	£7.95	£14.95
Vegetarian (spinach, mushrooms, tomatoes, cheese)	£4.50	£8.50	£15.95

Delivery to: Name : _____

Address : _____

_____

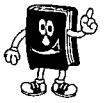
Telephone number : _____

_____

Scenario 12: Party service

Your name: *Chatwin*

Food for a party



For 30 people

As cheap as possible

Offer : _____

Portion : _____

Side dish : _____

Price per person : _____ £

Deliver to : 436 Turner Street
Dover

Date : Saturday, April 5th, 12.00 h

How much beer should be ordered for 30 people?

_____

Scenario 12: Party service

Your name: *Good Time Party Service Ltd.*

Special offers

Offer	Price per person	Side dishes/extras	Portion
Warm buffet (min. 40 people)	£5	Salads	500 g
Cold buffet (min. 30 people)	£3	Chocolate pudding	400 g
Irish stew (min. 25 people)	£2	Baguette	350 g
Hotdogs (min. 20 people)	£1	Potato salad	300 g

Deliver to:

Name : _____

Address : _____

Number of people : _____

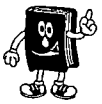
Day, date and time : _____

_____

Scenario 13: Library

Your name: *Mackintosh*

Borrow a book



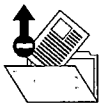
A detective novel by Agatha Christie

Please reserve it if it has already been taken out

Title of book : _____

Reserved for : _____

Borrowing period : _____ days



Order : ☎ 262271

Membership number: H 1092

How many books can I take out?

_____

Scenario 13: Library

Your name: *City Library*

Delivery times for ordered books

Choice of authors	Choice of titles	Borrowed until	Lending period
Agatha Christie	Murder on the Orient Express	Friday	10 days
	Death on the Nile	Monday	10 days
	The Mouse trap	Tuesday	10 days
Astrid Lindgren	Pippi Longstocking	*Available*	8 days
	The Six Bullebly Children	Wednesday	8 days
Michael Ende	Jim Button and Luke the Engine Driver	Thursday	12 days
	Momo	*Available*	12 days
	The Never-ending story	Friday	12 days

Order : Name : _____

Membership number : _____

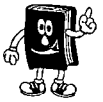
Telephone number : _____

_____

Scenario 14: Library

Your name: *Jacobs*

Extend the length of time I can borrow a book



"Pippi Longstocking" by Astrid Lindgren
Taken out a week ago
Keep out for as long as possible

Date today : _____

Extension until : _____

Price : _____ £

Reservation : Membership number: J 91 38 70

Have you got other books by Astrid Lindgren?

_____

Scenario 14: Library

Your name: *City Library*

Extended borrowing period for books that have already been out for one week

Type of book	Extension of borrowing period (max.)
Biography	No extension possible
Fantasy	7 days
Horror	14 days
Children's book	5 days
Detective novel	7 days
Reference books	No extension possible
Non-fiction	1 month

Reservation:

Book title : _____

Extension till : _____

Name : _____

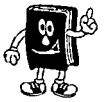
Membership number : _____

_____

Scenario 15: Flat to let

Your name: *Christie*

Put in an advertisement about a flat



Section: property to let

On Saturday

In as prominent a place as possible

Kind of text : _____

Price : _____ £

Cheque payable to : _____

Text : 3 rooms, kitchen, hall, bathroom, 67 m^2, £250 excluding bills

From June 7th

☎ 714530

How many copies of the Evening Chronicle are sold at the weekend?

_____

Scenario 15: Flat to let

Your name: *Private Advertising Department, Evening Chronicle*

Price per line to adverts concerning flats to be let

Style of text \ day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Bold print + framed	£4	£4	£5	£4	£5	£6
Bold print	£3	£3	£4	£3	£4	£5
Normal print + framed	£3	£3	£3	£3	£3	£4
Normal print	£2	£2	£2	£2	£2	£3

Cheques payable to: _____

Section ☐ To Let ☐ Accommodation wanted ☐ Miscellaneous

Print : _____

: _____

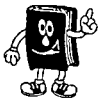
: _____

_____

Scenario 16: Flat to let

Your name: *Nelson*

Price and appointments for viewing flats in London



3 rooms, hall, bathroom

Approx. 50 m by 65 m

Viewing only possible at the weekend

Size of flat : _____ m^2

Address : _____

Viewing on : _____ at _____ h

Price : _____ £

Telephone : ☎ 802363 (home)

☎ 550012 (office)

Do I have to pay a deposit?

_____

Scenario 16: Flat to let

Your name: *Trandy Property Developers Ltd.*

Vacant flats in London

Flat	Flat with 2 rooms	Flat with 1 room	Flat with 3 rooms	Flat with 3 rooms
Size	47 m^2	30 m^2	58 m^2	64 m^2
Address	235 Parker Street	41 Morning Street	2 Pitt Street	7 Flower Street
Viewing on	Next Wednesday from 17.30 h	Next Friday from 18.00 h	Next Saturday from 10.00 h	Next Saturday from 14.00 h
Rent excluding bills	£300	£230	£340	£400

Appointment for: Name : _____

Telephone number : _____

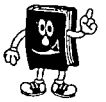
Day and time : _____

_____

Scenario 17: Eye specialist's appointment

Your name: *Robinson*

Appointment at an eye specialist



If possible, before 9 am

Wednesday or Thursday

Appointment : Day : _____

Date : _____

Time : _____

Personal details : Last visit 6 months ago

Check-up

Will I have to wait even though I have an appointment?

_____

Scenario 17: Eye specialist's appointment

Your name: *Practice Dr Wilson*

Appointments

First week in June

	Monday, 2nd	Tuesday, 3rd	Wednesday, 4th	Thursday, 5th	Friday, 6th
8.30 h		Flowers			Sting
8.45 h	Marshall			Kelly	Gabriel
9.00 h	Hill	Samson	Chapman	Flintstone	Jackson
9.15 h	Walsh	Coolman	Riddle		
9.30 h	Sims	King		Tanko	Grant
9.45 h	Reed	Simons	Robson	Tretschok	Bon Jovi
10.00 h	Owry	Zorco	Hermas	Bond	Clapton
10.15 h	Collins	Miller	Edwards		Dobson
10.30 h	Sharp	Lambert	French		
10.45 h		Reuter		Debenham	Domingo

Day, date and time : _____

Have you been here before? ☐ yes ☐ no

When was your last visit? : _____

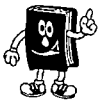
Reason for visit? ☐ New glasses ☐ Check-up ☐ Problems/pain

_____

Scenario 18: Eye specialist's appointment

Your name: *Robson*

Change an appointment



After 5 pm, if possible

Does not matter which day

It is only a check-up

New appointment time : Date : _____

Time : _____

Personal details : Old appointment: Tuesday 15.30 h

Check-up

How long does a check-up take?

_____

Scenario 18: Eye specialist's appointment

Your name: *Practice Dr Miller*

Appointments

First week in June

	Monday	Tuesday	Wednesday	Thursday	Friday
15.30	Marshall	Robson			Sting
15.45		Ewing		Kells	Gabriel
16.00	Summer	Dickson		Carrington	Jackson
16.15	Hill	Coolman		Major	
16.30	Sims	King		Riddle	
16.45	Walsh	Snow	*Closed*	Windsor	Johnson
17.00	Zorro	Lowry			Clapton
17.15	Miller	Shriver			Parker
17.30	Lambert	Clark		Banks	
17.45					Debenham

Old appointment : _____

New appointment : _____

Reason for visit? ☐ New glasses ☐ Check-up ☐ Problems/pain

_____

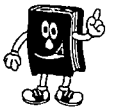
Appendix VI

Example scenarios in French language

Symboles pour la personne qui appelle:

Ce symbole signifie que vous êtes la personne qui appelle. Attendez s'il vous plaît que la personne supervisant l'opération vous invite à commencer la discussion.

La raison de votre appel est décrite à côté de ce symbole (par exemple: je voudrais acheter un billet).



À côté de ce symbole est décrite une condition qui doit être prise en considération lors de l'échange d'informations (par exemple: je voudrais acheter un billet → MAIS le moins cher possible).

Vous devez, à côté de ce symbole, écrire toutes les informations que vous voulez apprendre de votre interlocuteur.

À côté de ce symbole sont énumérées toutes les informations dont votre interlocuteur a besoin et que vous devez lui fournir.

Symboles pour la personne qui est appelée:

Ce symbole signifie que vous êtes la personne appelée. Attendez s'il vous plaît que le téléphone sonne puis décrochez.

À côté de ce symbole sont énumérées des informations parmi lesquelles vous devez chercher celles dont votre interlocuteur a besoin (par exemple: le prix des billets de train pour les adultes, pour les lycéens et étudiants, pour les enfants, pour les personnes âgées, etc.)

Vous devez, à côté de ce symbole, écrire les informations que votre interlocuteur doit vous donner.

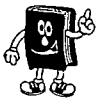
Symbole pour la personne qui appelle <u>et</u> pour la personne appelée:

À côté de ce symbole est posée une question pour laquelle ni vous ni votre interlocuteur n'avez d'information – Vous devez dans ce cas discuter brièvement la question et trouver une solution satisfaisante pour les deux parties.

Exercice: Agence de voyage

Votre nom: ***Chevallier***

Voyage pour une semaine en Méditerranée



voyage de dernière minute: à partir de demain,
pas en Espagne,
le moins cher possible

Prix : _____ FF

Jour de départ : _____ à _____ h _____

Destination : _____

Hôtel : _____

Réservation : Départ de Brest,
Mastercard,
N° de la carte de crédit: 9685 4712 0951 2781,
Valable jusqu'en 10/99

Combien de temps avant le départ dois-je être à l'aéroport?

_____

Exercice: Agence de voyage

Votre nom: *Agence de voyage Look Saint-Brieuc*

Voyages de dernière minute
pour une semaine
départ de Brest ou de Rennes

Destination	Départ aujourd'hui	Départ demain	Départ après-demain
Espagne (Majorque) Hôtel Playas Arenal *Demi-pension*	1999 FF départ 19 h 30	2499 FF départ 10 h 15	2799 FF départ 10 h 15
Italie (Sicile) Hôtel Città del Mare *Pension complète*	2199 FF départ 19 h 10	2589 FF départ 8 h 50	2749 FF départ 8 h 50
Portugal (Algarve) Hôtel de Lagos *Demi-pension*	1549 FF départ 20 h 20	1699 FF départ 9 h 45	2099 FF départ 9 h 45

(Tous les prix sont donnés par personne et TVA incluse)

Réservation:

NOM : _____

CARTE DE CRÉDIT : ☐ Mastercard ☐ Eurocard

: ☐ American Express ☐ Autre

N° CARTE DE CRÉDIT : _____

VALABLE JUSQU'À : _____

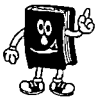
DÉPART DE : ☐ Brest ☐ Rennes

_____

Scénario 1: Renseignements ferroviaires

Votre nom: *Mueller*

Vous voulez aller de: Montauban (82) à Paris



Date: 21/10
Voyage de nuit, wagon-lit
si possible sans changement de train

Heure de départ : _____h_____

Heure d'arrivée : _____h_____

N° de train : _____

Réservation : 1 couchette,
1ère classe
non-fumeur

Doit-on payer un supplément pour ce train ?

_____

Scénario 1: Renseignements ferroviaires

Votre nom: *Renseignements pour les voyageurs de la S.N.C.F.*

Horaire des trains: Montauban → Paris

N° train		**Train Vert 4424**	**Train Vert 4424**	**Train 174**
Montauban	*départ*	23 h 16	23 h 16	23 h 51
Limoges	*arrivée*		2 h 48	
N° train			**Train 472**	
Limoges	*départ*		3 h 47	
Paris Austerlitz	*arrivée*	6 h 59	7 h 29	7 h 05
Réservation		Recommandée	Recommandée	Obligatoire
		1ère classe: couchettes	1ère classe: indisponible	1ère classe: COMPLET
		2ème classe: places assises, assises inclinables ou couchettes au choix	2ème classe: places assises ou assises inclinables	2ème classe: couchette ou voiture lit

Réservation:

NOM : _____

NOMBRE DE COUCHETTES : _____

CLASSE : _____

FUMEUR : ☐

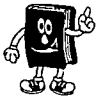
NON-FUMEUR : ☐

_____

Scénario 2: Renseignements ferroviaires

Votre nom: *Le Roy*

Vous voulez aller de: Saint-Brieuc à Paris



le 5/10,
Gare Montparnasse,
Arrivée entre 18 heures et 18 h 30

Départ : _____h_____

Changement à : _____ à _____h_____

Arrivée : _____h_____

Réservation : 2 places assises,
2ème classe,
Non-fumeur

A-t-on assez de temps pour changer de train?

_____

Scénario 2: Renseignements ferroviaires

Votre nom: *Renseignements pour les voyageurs de la S.N.C.F.*

Horaire des trains: Saint-Brieuc → Paris

N° de train		**TGV 8742**	**TGV 8736**	**Train 88526**	**TGV 8754**
Saint-Brieuc	*départ*	14 h 20	13 h 25	12 h 47	16 h 22
Rennes	*arrivée*		14 h 15	14 h 10	
N° de train			**Train vert 3734**	**TGV vert 8746**	
Rennes	*départ*		14 h 38	15 h 59	
Paris Montparnasse	*arrivée*	17 h 20	18 h 02	18 h 15	19 h 25
Particularités		réservation obligatoire	réservation obligatoire de Saint-Brieuc à Rennes		réservation obligatoire

Réservation:

NOM : _____

NOMBRE DE PLACES : _____

CLASSE : _____

FUMEUR : ☐

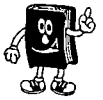
NON-FUMEUR : ☐

_____

Scénario 3: Renseignements aériens

Votre nom: *Benhamou*

Vous souhaitez un vol: Lyon → Berlin



le 23 décembre,
le matin,
vol direct de préférence

Départ : _____h_____
Arrivée : _____h_____
N° du vol : _____

Réservation : 1 place,
: en classe économique
Adresse : 17, rue Jean Savidan 22300 Lannion
☎ 02 96 37 25 94

Quel est l'aéroport le plus près du centre ville de Berlin?

_____

Scénario 3: Renseignements aériens

Votre nom: *Renseignements aériens, aéroport de Lyon Satolas*

Horaire des vols: Lyon → Berlin

		Air inter Europe	TAT	Lufthansa
N° des vols		IT 2615	IJ 381	LH 413
Lyon Satolas	*départ*	6 h 30	6 h 35	8 h 20
Paris Orly	*arrivée*		7 h 35	
Paris Orly	*arrivée*		7 h 50	
Berlin Schönefeld	*arrivée*	7 h 55		
Berlin Tegel	*arrivée*		9 h 15	9 h 45
		(*tous les jours*)	(*tous les jours*)	(*tous les jours*)

Réservation:

NOM : _____

ADRESSE : _____

_____

N° DE TÉLÉPHONE : _____

NOMBRE DE PLACES : _____

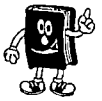
CLASSE : ☐ Affaire ☐ Economique

_____

Scénario 4: Renseignements aériens

Votre nom: *De Konynk*

Vous souhaitez un vol Paris → New York



le 10 octobre,
Aéroport John Fitzgerald Kennedy (JFK)

Départ : _____h_____

Escale à : _____ à _____h_____

Arrivée : _____h_____

Réservation : 1 place,
Mastercard
N° de carte de crédit: 2602 2704 1612 1807,
valable jusqu'en 02/99

Quel est le décalage horaire entre Paris et New York?

_____

Scénario 4: Renseignements aériens

Votre nom: *Renseignements aériens, aéroport de Roissy Charles de Gaulle*

Horaire des vols : Roissy → New York

		Lufthansa	United Airlines	Air France
N° de Vol		LH 400	UA 129	AF 242
Paris Roissy	*départ*	9 h 50	7 h 25	13 h 30
London Heathrow	*arrivée*		8 h 00	
London Heathrow	*départ*		9 h 15	
New York JFK	*arrivée*		11 h 40	
New York Newark	*départ*	12 h 15 (heure locale)	(heure locale)	16 h 05 (heure locale)

Réservation:

NOM : _____

NOMBRE DE PLACES : _____

CARTE DE CRÉDIT : _____

N° CARTE DE CRÉDIT : _____

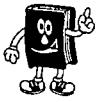
VALABLE JUSQU'À : _____

_____

Scénario 5: Places de spectacle

Votre nom: *Vernhes* (prononcer "vergne")

2 places pour "l'Opéra de quat' sous"



le week-end prochain (Samedi ou Dimanche)

Carte d'étudiant

Date (jour) : _____

Début de la représentation : _____

Prix : _____ FF

Salle ☐ Grande Salle ☐ Petite Salle ☐ Scène Studio

Réservation : 2 places,
Milieu orchestre

Jusqu'à quelle heure peut-on retirer des billets réservés?

_____

Scénario 5: Places de spectacle

Votre nom: *Salle Le Quartz à Brest*

Programme des représentations

Jour de la semaine	Grande Salle	Petite Salle	Scène Studio
Lundi, 20 h 00	Bernard Alisson	--	--
Mardi, 20 h 00	--	Match d'improvisation	Les Escrocs
Mercredi, 20 h 00	Roméo et Juliette	--	--
Jeudi, 20 h 00	Opéra de quat' sous	Danses de Hongrie	Romain Didier
Vendredi, 20 h 15	Opéra de quat' sous	--	Les Têtes Raides
Samedi, 19 h 30	Opéra de quat' sous	Songe d'une nuit d'été	Daft Punk
Dimanche, 19 h 45	Roméo et Juliette	Songe d'une nuit d'été	--

Réservation:

NOM : _____

N° DES PLACES : _____

PRIX : ☐ Normal (de 50 à 200 FF)

☐ Réduit (40 FF)

PLACES : ☐ Orchestre

☐ Balcon

☐ Gauche ☐ Milieu ☐ Droite

_____

Scénario 6: Places de concert

Votre nom: *Dr Le Quesnoy (prononcer Le Quénoi)*

2 places pour "Starmania"

Samedi 3 octobre ou
Dimanche 4 octobre 1998,
Places en dessous de 350 FF

Date : _____

Début de la représentation : _____

Catégorie de la place : _____

Prix : _____ FF/billet

Réservation : 13 Place de l'église,
22700 Perros-Guirec
☎ 02 96 23 21 76

La comédie musicale est-elle en français ou anglais?

_____

Scénario 6: Places de concert

Votre nom: *STARMANIA En Tournée*

Starmania

	Samedi 03.10.98 20 h 00		**Dimanche 04.10.98 20 h 00**	
	Prix	**Statut**	**Prix**	**Statut**
Orchestre	450 FF	*Complet*	360 FF	
Tribune	550 FF	*Complet*	450 FF	
Tribune milieu	550 FF		450 FF	*Complet*
Tribune côté	380 FF		300 FF	
Balcon	300 FF		200 FF	*Complet*

Réservation:

NOMBRE DE PLACES : _____

NOM : _____

ADRESSE : _____

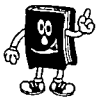
N° DE TÉLÉPHONE : _____

_____

Scénario 7: Location Auto

Votre nom: *Pélissier*

1 petit utilitaire pour un déménagement



pour un week-end (samedi et dimanche),
Hauteur de chargement: 1m 80 au moins,
le moins cher possible

Modèle : _____

Hauteur de chargement : _____

Prix : _____ FF/jour

Réservation : Eurocard,
Numéro de la carte de crédit: 8823 7453 6592 0023,
Valable jusqu'en décembre 1999
Permis de conduire B

A partir de quelle heure le matin puis-je venir chercher le véhicule?

_____

Scénario 7: Location Auto

Votre nom: *Location auto "Europcar"*

Planning utilitaires

Jour	Ford Transit	Peugeot J9	Iveco Turbo Daily
Lundi	*réservé*	*réservé*	
Mardi	*réservé*	*réservé*	
Mercredi		*réservé*	*réservé*
Jeudi	*réservé*		*réservé*
Vendredi		*réservé*	
Samedi			
Dimanche	*réservé*		
Permis	B	B	B
Kilométrage max.	100 km	100 km	100 km
Tarif/jour	409 FF	529 FF	389 FF
Hauteur de chargement	2 m	2.30 m	1.90 m

Réservation: NOM : _____

CARTE DE CRÉDIT : _____

N° CARTE DE CRÉDIT : _____

VALABLE JUSQU'À : _____

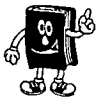
PERMIS DE CONDUIRE : ☐ A ☐ B ☐ E

_____

Scénario 8: Location Auto

Votre nom: *Pinaut*

Location d'un véhicule de tourisme



de jeudi à samedi,
le moins cher possible

Modèle : _____

Lieu : _____

Prix : _____ FF/jour

Réservation : American Express,
N° de carte de crédit: 6430 5188 1926 4623,
valable jusqu'en septembre 1999

L'assurance est-elle comprise dans le prix?

_____

Scénario 8: Location Auto

Votre nom: *location Auto "Hertz"*

Planning véhicules de tourisme

Jour	Renault Clio	Peugeot 106	Citroën Xantia	Volkswagen Golf
Lundi		*réservé*	*réservé*	*réservé*
Mardi			*réservé*	*réservé*
Mercredi	*réservé*	*réservé*	*réservé*	
Jeudi	*réservé*			
Vendredi	*réservé*			
Samedi	*réservé*			
Dimanche	*réservé*		*réservé*	*réservé*
Prix/jour	310 FF	275 FF	370 FF	340 FF
Kilométrage max.	300 km	100 km	100 km	100 km
Lieu	Gare routière	Gare routière	Aéroport	Aéroport

Réservation: NOM : _____

CARTE CRÉDIT : ☐ MasterCard ☐ Visa ☐ American Express

N° CARTE CRÉDIT : _____

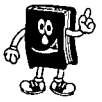
VALABLE JUSQU'EN : _____

_____

Scénario 9: Location d'une chambre d'hôtel

Votre nom: *Weill (prononcer Veille)*

2 chambres simples



du 15 au 28 mars,
avec douche et/ou balcon de préférence

N° chambre	: _____	et	_____	
Douche	: ☐ OUI	☐ NON	☐ OUI	☐ NON
Balcon	: ☐ OUI	☐ NON	☐ OUI	☐ NON

Réservation : 1ère chambre pour André Weill
2ème chambre pour Ghislain Chautard
☎ 05 65 81 54 72

Y a-t-il en mars encore assez de neige pour faire du ski?

_____

Scénario 9: Location d'une chambre d'hôtel

Votre nom: *Hôtel des Cimes à Morzine*

Planning d'occupation des chambres

Semaine	Chambre N° 106	Chambre N° 118	Chambre N° 235	Chambre N° 257
01.03.-07.03.		*réservé*	*réservé*	*réservé*
08.03.-14.03.	*réservé*		*réservé*	
15.03.-21.03.				
22.03.-28.03.	*réservé*	*réservé*		
29.03.-04.04.	*réservé*	*réservé*		
Chambre	**Chambre double**	**Chambre simple**	**Chambre simple**	**Chambre simple**
Douche	non	oui	oui	non
Balcon	oui	oui	non	non

Réservation: NOM : _____

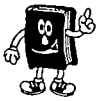
N° DE TÉLÉPHONE : _____

_____

Scénario 10: Réservation d'un appartement

Votre nom: *M. Rocques*

1 appartement (vacances) à Rimini (Italie)



du 22 au 28 mars,
pour au moins 4 personnes,
si possible à proximité de la plage

Nom de l'appartement : _____

Nombre de personnes : _____ Personnes

Prix : _____ FF/semaine

Plage à _____ m

Réservation : 104 avenue Saint-Yves
22220 Tréguier
☎ 02 96 92 16 34

Est-ce qu'il fait déjà assez chaud en mars pour se baigner?

_____

Scénario 10: Réservation d'un appartement

Votre nom: *Club Famille Italie*

Planning des appartements

Semaine	Appartement "Calderone"	Appartement "Angeli"	Appartement "Pronti"	Appartement "Riviera"
01.03.-07.03.		*réservé*	*réservé*	*réservé*
08.03.-14.03.	*réservé*	*réservé*	*réservé*	
15.03.-21.03.	*réservé*			*réservé*
22.03.-28.03.				*réservé*
29.03.-04.04.				*réservé*
Nombre de personnes	3	4	6	6
Plage à	env. 200 m	env. 50 m	env. 50 m	env. 50 m
Prix/semaine	750 FF	1150 FF	1050 FF	1700 FF

Réservation: NOM : _____

ADRESSE : _____

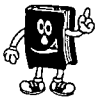
: _____

_____

Scénario 11: Pizzaservice

Votre nom: *Lapeyre*

1 grosse pizza



pour deux personnes,

pizza végétarienne de préférence

Garniture : _____

_____

Prix : _____ FF

Livraison à l'adresse : 6, rue principale (face à l'école)
Loguivy les Lannion
☎: 02 96 48 45 31

Dans combien de temps la pizza sera-t-elle livrée?

_____

Scénario 11: Pizzaservice

Votre nom: *Pizzeria Roma*

Pizza

Type de pizza	1 personne	2 personnes	4 personnes
Toscana (jambon, champignons, tomates, fromage)	25 FF	47 FF	84 FF
Tonno (thon, oignons, tomates, fromage)	34 FF	61 FF	112 FF
Fabrizio (salami, jambon, tomates, fromage)	25 FF	47 FF	84 FF
Vegetaria (épinards, champignons, tomates, fromage)	31 FF	56 FF	107 FF

Livraison à:

NOM : _____

ADRESSE : _____

_____

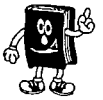
N° DE TÉLÉPHONE : _____

_____

Scénario 12: Réception

Votre nom: *Ayrault*

Repas pour une fête



pour 30 personnes,
le moins cher possible

Offre : _____

Portion : _____

Supplément : _____

Prix / personne : _____ FF

Livraison à : 24, quai d'Aiguillon
22300 Lannion

Date : Samedi 31 octobre 1998 à 12 h 00

Combien de boissons non alcoolisées doit-on commander pour 30 personnes ?

_____

Scénario 12: Réception

Votre nom: *Traiteur Le Bihan S.A.R.L.*

Offres

Offre	Prix/personne	Supplément	Portion
Buffet chaud (à partir de 40 personnes)	45 FF	Salade	500 g
Buffet froid (à partir de 30 personnes)	30 FF	Mousse au chocolat	400 g
Goulasch (à partir de 25 personnes)	20 FF	Pain	350 g
Plateau de charcuterie (à partir de 20 personnes)	15 FF	Salade de pommes de terre	300 g

Livraison à: NOM : _____

ADRESSE : _____

NOMBRE DE PERSONNES : _____

JOUR, DATE ET HEURE : _____

_____

Scénario 13: Bibliothèque

Votre nom : *Le Gall*

Emprunt d'un livre

Un roman policier d'Agatha Christie

au cas où le livre serait déjà emprunté, le réserver

Titre de l'ouvrage : _____

réservé pour le : _____

Durée de l'emprunt : _____ jours

Réservation : ☎ 02 96 37 35 36

N° de client: H 1092

Combien cela coûte-t-il d'emprunter un livre?

_____

Scénario 13: Bibliothèque

Votre nom: *Bibliothèque municipale de Lannion*

Service de prêt

Auteurs	Titres	emprunté jusqu'à ... prochain	Durée de l'emprunt
Agatha Christie	Le crime de l'Orient Express	Vendredi	10 jours
	Mort sur le Nil	Lundi	10 jours
	Dix petits nègres	Mardi	10 jours
Maurice Leblanc	l'arrestation d'Arsène Lupin	*non emprunté*	8 jours
	Arsène Lupin contre Sherlock Holmes	Mercredi	8 jours
	Arsène Lupin et l'aiguille creuse	*non emprunté*	8 jours
Arthur Conan Doyle (Sherlock Holmes)	Le chien des Baskerville	Jeudi	12 jours
	Une étude en rouge	*non emprunté*	12 jours
	Le ruban moucheté	Vendredi	12 jours

Réservation : NOM : _____

N° DE CLIENT : _____

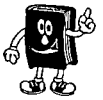
N° DE TÉLÉPHONE : _____

_____

Scénario 14: Bibliothèque

Votre nom: *Jacoby*

Prolonger un emprunt d'ouvrage



La Chartreuse de Parme, de Stendhal
emprunté depuis 1 semaine,
prolongation la plus longue possible

Date d'aujourd'hui : _____

Prolongation jusqu'à : _____

Prix : _____ FF

Réservation : N° de client: J 91 38 70

Quels autre livres du même auteur avez-vous?

_____

Scénario 14: Bibliothèque

Votre nom: *Bibliothèque municipale de Lannion*

Prolongation des ouvrages déjà empruntés depuis 1 semaine

Type d'ouvrage	Prolongation maximale	Prix
Biographies	AUCUNE PROLONGATION POSSIBLE	--
Romans	7 jours	5 FF
Horreur	14 jours	7 FF
Ouvrages pour enfants	5 jours	3 FF
Policier	7 jours	4 FF
Ouvrages de référence	AUCUNE PROLONGATION POSSIBLE	--
Essais, documents	1 mois	10 FF

Réservation: TITRE DE L'OUVRAGE : _____

PROLONGATION JUSQU'À : _____

NOM : _____

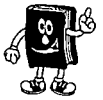
N° D'ABONNEMENT : _____

_____

Scénario 15: Petites annonces immobilières

Votre nom: *Briand*

Insertion d'une annonce immobilière



Rubrique: locations

le samedi,

le plus voyant possible

Type de texte : _____

Prix : _____ FF

Virement sur le compte : _____

Texte : 3 pièces, 67 m^2, 2200 FF (hors charges), à partir du 1ère novembre 1998,

☎ 02 96 48 45 30

Quel est le tirage du Télégramme le week-end?

_____

Scénario 15: Petites annonces immobilières

Votre nom: *Réception téléphonique de l'antenne de Lannion du Télégramme*

Prix par ligne pour les annonces immobilières

Jours / Type de texte	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
Caractères gras + encadrés	30 FF	30 FF	35 FF	30 FF	35 FF	45 FF
Caractères gras	25 FF	25 FF	28 FF	25 FF	28 FF	40 FF
Caractères normaux + encadrés	22 FF	22 FF	26 FF	22 FF	26 FF	33 FF
Caractères normaux	15 FF	15 FF	18 FF	15 FF	18 FF	30 FF

N° de compte: 444 100, Code banque: 430 500 01, Caisse d'épargne Lannion

Rubrique : ☐ Location (offres) ☐ Location (recherches) ☐ Vente ou achat

Texte : _____

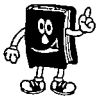
: _____

_____

Scénario 16: Recherche d'appartement

Votre nom: *Nowack*

Prix et rendez-vous pour visiter des appartements à Lannion



3 pièces,
env. 50 à 65 m^2,
Visites possibles seulement le week-end

Superficie de l'appartement : _____ m^2

Adresse : _____

Visite le : _____ à _____h_____

Prix : _____FF

N° de téléphone : ☎ 02 96 48 96 45 (privé)
02 96 05 34 31 (professionnel)

À combien se monte la caution?

_____

Scénario 16: Recherche d'appartement

Votre nom: *Agence Melscoët*

Appartements à louer à Lannion

Appartement	2 pièces	Studio	3 pièces	3 pièces
m^2	47 m^2	30 m^2	58 m^2	64 m^2
Adresse	Rés. Haute Rive	37, rue Ernest Renan	12, impasse des peupliers	Le Rusquet
Visite	mercredi prochain, à partir de 17 h 30	vendredi prochain, à partir de 18 h 00	samedi prochain, à partir de 10 h 00	samedi prochain, à partir de 14 h 00
Loyer (sans les charges)	1700 FF	1100 FF	2300 FF	2700 FF

Rendez-vous pour: NOM : _____

N° DE TÉLÉPHONE : _____

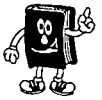
JOUR & HEURE : _____

_____

Scénario 17: Rendez-vous chez le dentiste

Votre nom: *Besnard* (prononcer Bénar)

Rendez-vous chez le dentiste



si possible le matin avant 9 h 00,

mercredi ou jeudi

Rendez-vous : Jour : _____

Date : _____

Heure : _____

Renseignements personnels : N° Sécurité Sociale: 2 57 12 22 002 125 (pour une femme) ou 1 58 05 22 157 084 (pour un homme),

Patient venant régulièrement,

dernière visite il y a 6 mois,

Visite de routine

Combien de temps dois-je attendre avant d'obtenir un rendez-vous?

_____

Scénario 17: Rendez-vous chez le dentiste

Votre nom: *Cabinet du docteur Manac'h*

Emploi du temps

1ère semaine de juin

	Lundi, 19.09.98	**Mardi 20.09.98**	**Mercredi 21.09.98**	**Jeudi 22.09.98**	**Vendredi 23.09.98**
8 h 30		Paranthoën			Patard
8 h 45	Le Guen			Christin	Guillaumin
9 h 00	Latrasse	Billiou	Chapuis	Chanteur	Arzur
9 h 15	Moreau	Auclerc	Amaury		
9 h 30	Gillet	Chevalier	Delmotte	Krawczyk	Le Braz
9 h 45	Mérignac	Champigny	Leroi	Delorme	Jolivet
10 h 00	De Wolf	Rouquié	Gonçalves	Boullet	Jacob
10 h 15	Anselmini	Maubert	Péron		Jouan
10 h 30	Bailly	Lambert	Philippon		
10 h 45		Roulet		Le Gall	Domingo

Jour, date & heure : _____

N° de Sécurité Sociale? : _____

Êtes-vous déjà venu? : ☐ OUI ☐ NON

Quand êtes-vous venu pour la dernière fois? : _____

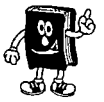
Raison de la visite? : ☐ Douleurs ☐ Routine ☐ Prothèse

_____

Scénario 18: Rendez-vous chez l'oculiste

Votre nom: *Rouger*

Repousser le rendez-vous chez l'oculiste



si possible après 17 h 00,
n'importe quel jour de la semaine,
il s'agit d'une visite de routine

Nouveau rendez-vous Date : _____

Heure : _____

Renseignements personnels : Ancien rendez-vous: Mardi 15 h 30

N° Sécurité Sociale: 2 72 10 75 174 075 (pour une femme) ou 1 69 02 29 023 002 (pour un homme),

Visite de contrôle

Combien de temps dure une visite de routine ?

_____

Scénario 18: Rendez-vous chez l'oculiste

Votre nom: *Cabinet du docteur Anthelme*

Emploi du temps

1ère semaine de juillet

	Lundi	Mardi	Mercredi	Jeudi	Vendredi
15 h 30	Duclos	**Rouger**		De Sousa	Gicquel
15 h 45		Chaussier		Pinault	Louvion
16 h 00	Samois	Blum		Le Maux	Jacq
16 h 15	Poletti	Turpin		Gros	
16 h 30	Auguste	Trichet	*fermé!*	De la Bessière	
16 h 45	Etchegoyen	Hunault		Levillain	Bon
17 h 00	Le Fur	Lafontaine			Capelle
17 h 15	Mueller	Le Meur			Fernandez
17 h 30	Lambert	Auriac		Amiel	
17 h 45					Dominique

Ancien rendez-vous : _____

Nouveau rendez-vous : _____

N° de Sécurité Sociale : _____

Raison de la visite ? : ☐ Lunettes ☐ Contrôle ☐ Troubles

_____

Appendix VII

Example of a Richard's task

(This appendix does not form an integral part of this Recommendation)

Random shapes are shown on each sheet. Twenty-four shapes is a typical number on one sheet. There are no meaningful relationships between shapes and their names. The detailed method used for generating shapes can be found in [b-RICHARDS]. The operator prepares the same set of sheets for both subjects. If the same 24 shapes are given to both subjects the order of the shapes should be different. An example sheet of random shapes is shown in Figure VII.1.

During the conversation, each subject arbitrarily chooses a shape on the sheet and talks about one of its features to his/her partner. His/her partner either guesses the name of the shape based on the information provided or requests additional information from the questioner.

The conversation may proceed as shown in the following example:

[Questioner]		[Respondent]
"Upper part is protruding like a thumb"	→	"How about left hand side?"
"Left is very sharply protruding"	→	"How about right hand side?"
...		
"Is it TEA?"	→	"You are correct"

To help switch the roles of questioner and respondent, the operator carefully instructs both subjects during the preliminary training and tells them not to take too long discussing each shape.

The roles of respondent and questioner are switched every time the respondent provides the correct answer. The operator forces them to stop their conversation after the defined conversation time is passed and prompts them to shift quality evaluation stage.

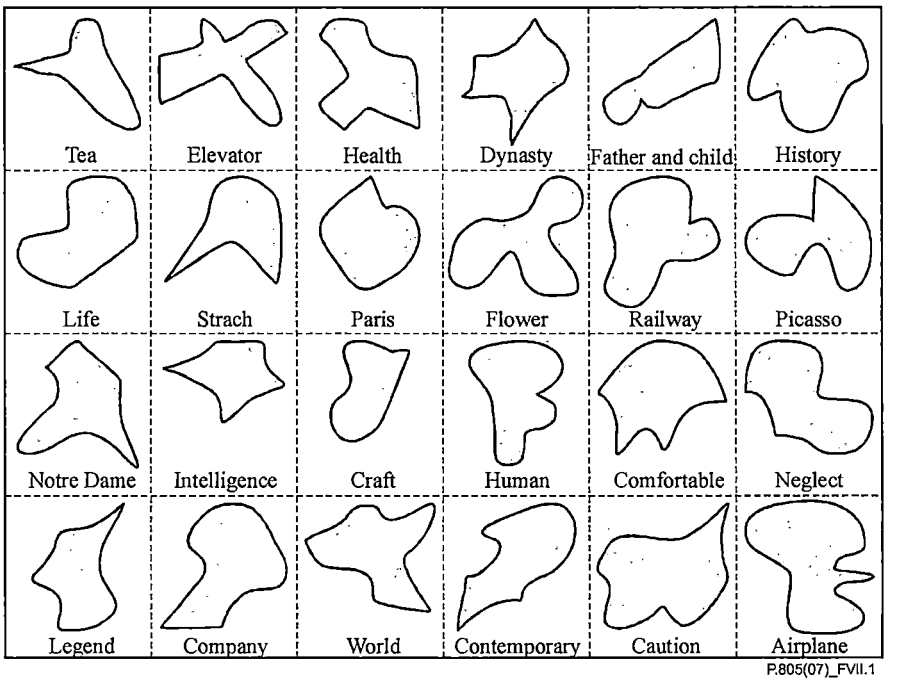


Figure VII.1 – Example sheet of random shapes for use in a Richard's task

Appendix VIII

Example scenarios for random number verification tasks

(This appendix does not form an integral part of this Recommendation)

Instructions: "Your conversation partner is also provided with such a list. Some of the numbers in your list do not correspond with those of your conversation partner. Find the wrong numbers as quickly as possible by taking turns reading them line by line. Acknowledge by saying "yes" or "no", and cross out the wrong numbers. You will read the bold red numbers and your conversation partner will read the non-bold blue ones".

You are the "caller".

18	**88**	**80**	**74**	**55**	**7**
15	29	14	37	17	82
20	**95**	**36**	**77**	**34**	**83**
46	84	30	67	25	99
28	**27**	**36**	**96**	**60**	**97**
55	10	87	53	43	98

Instructions: "Your conversation partner is also provided with such a list. Some of the numbers in your list do not correspond with those of your conversation partner. Find the wrong numbers as quickly as possible by taking turns reading them line by line. Acknowledge by saying "yes" or "no", and cross out the wrong numbers. You will read the bold blue numbers and your conversation partner will read the non-bold red ones".

Please wait to be called.

18	84	80	74	55	7
15	**29**	**14**	**67**	**17**	**82**
36	95	36	77	53	83
46	**88**	**30**	**37**	**25**	**99**
28	27	20	96	60	97
55	**10**	**87**	**34**	**43**	**98**

Appendix IX

Example scenarios for interactive short conversation tests

(This appendix does not form an integral part of this Recommendation)

Explanation of the symbols:

As calling subject:

*This symbol means: You **call**. Before you do so, please wait until the test operator tells you to start the telephone conversation.*

Here, the reason for your call is provided (e.g., "I need the up-to-date stock exchange data!").

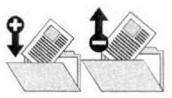
As called subject:

*This symbol means: You **are** called. Please wait until the telephone rings and then pick up the receiver.*

Here, the reason for the telephone conversation is provided (e.g., your conversation partner needs up-to-date stock exchange data, and you want to know the identification number his company uses for the respective rates!).

Both subjects:

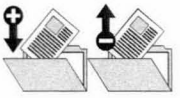
These symbols indicate the information on who you are in the current scenario. Please remember that you have your own name in all scenarios! ("Where do you work?", etc.).



Besides this symbol you will find the table of data which need to be communicated to your conversation partner. Please fill in the data you obtain from your partner!

Human Resource Department, Company X

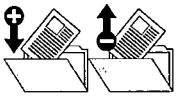
Exchange of identification numbers and email-addresses of new company members



Name	Member ID	Email
Fachmann	536-952487	ks.fachmann
Bauer	258-761926	
Dreierlein	536-879177	
Gerhards	258-327431	
Hamberg	668-215623	
Tuchmeyer	536-412142	

Network Central, Company X

Exchange of identification numbers and email-addresses of new company members



Name	Member ID	Email
Fachmann	536-952487	ks.fachmann
Bauer		fp.bauer
Dreierlein		ps.dreierlein
Gerhards		jf.gerhards
Kobalt		kh.kobalt
Tuchmeyer		ag.tuchmeyer

Bibiliography

[b-COM 12-35-E] ITU-T COM 12-35-E (1997), *Development of scenarios for short a conversation test.*

[b-COM 12-98-E] ITU-T COM 12-98-E (1999), *User expectation for telephony from computer terminals.*

[b-Telephonometry] Handbook on Telephonometry (1992), *Measurement methods: telephonometry.*

[b-EG 202 396-1] ETSI EG 202 396-1 (2006), *Speech processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database.*

[b-HAMMER] HAMMER (F.): Quality Aspects of Packet-Based Interactive Speech Communication, *PhD Thesis*, University of Technology at Graz 2006.

[b-KITAWAKI] KITAWAKI (N.) and ITOH (K.): Pure Delay Effects on Speech Quality in Telecommunications, *IEEE Journal on Selected Areas in Communications*, vol. 9 (4), pp. 586-593, 1991.

[b-MÖLLER] MÖLLER (S.): Assessment and Prediction of Speech Quality in Telecommunications, *Kluwer Academic Publishers*, pp. 75-81, Boston MA 2000.

[b-RAAKE] RAAKE (A.): Speech Quality of VoIP: Assessment and prediction, *John Wiley and Sons Ltd.*, Chichester 2006.

[b-RICHARDS] RICHARDS (D.L.): The transmission performance of telephone networks, *The Butterworth Group*, pp. 199-203, London 1973.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
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	Cable networks and 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	Telecommunication management, including TMN and network maintenance
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, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2007

INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.800

(08/96)

SERIES P: TELEPHONE TRANSMISSION QUALITY

Methods for objective and subjective assessment of quality

Methods for subjective determination of transmission quality

ITU-T Recommendation P.800

(Previously CCITT Recommendation)

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY

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 ITU-T List of Recommendations.

ITU-T RECOMMENDATION P.800

METHODS FOR SUBJECTIVE DETERMINATION OF TRANSMISSION QUALITY

Summary

This Recommendation describes methods and procedures for conducting subjective evaluations of transmission quality. The main revision encompassed by this version of this Recommendation is the addition of an annex describing the Comparison Category Rating (CCR) procedure. Other modifications have been made to align this Recommendation with recent revision of Recommendation P.830.

Source

ITU-T Recommendation P.800 was revised by ITU-T Study Group 12 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 30th of August 1996.

Keywords

Absolute Category Rating, Comparison Category Rating, conversational test, Degradation Category Rating, listening test, subjective evaluation, Subjective testing

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. The 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 their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1 (Helsinki, March 1-12, 1993).

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.

CONTENTS

Introduction

Modern telecommunication networks provide a wide array of voice services using many transmission systems. In particular, the rapid deployment of digital technologies has led to an increased need for evaluating the transmission characteristics of new transmission equipment. In many circumstances, it is necessary to determine the subjective effects of some new transmission equipment or modification to the transmission characteristics of a telephone network. This Recommendation describes methods for obtaining subjective evaluations of transmission systems and components. Recommendation G.113 contains useful information on the impairments that can occur. Recommendation P.11 discusses the effects that transmission impairments may have on the users of telecommunication networks and services. The methods described in this Recommendation may be used to estimate the equipment impairment factors (eifs) or quantization distortion units (qdus) that are described in Recommendation G.113.

Recommendation P.800[1]

METHODS FOR SUBJECTIVE DETERMINATION OF TRANSMISSION QUALITY

(Amended at Helsinki, 1993; revised in Geneva, 1996)

1 Scope

This Recommendation contains advice to Administrations on conducting subjective tests of transmission quality in their own laboratories. It does not however deal with types of tests described in detail in other ITU–T Recommendations and documentation, namely:

a) determination of Reference and Relative Equivalents – see *Handbook on Telephonometry*, Geneva, 1993;

b) determination of Loudness Ratings – see Recommendation P.78;

c) determination of Articulation Ratings (A.E.N. values) – see *Handbook on Telephonometry*, Geneva, 1993.

Neither does it deal with the various kinds of specialized tests used in the course of developing items of telephone equipment, for the purpose of diagnosing faults and shortcomings, such as Diagnostic Rhyme Tests [1] and other tests dedicated to the study of specific aspects of speech output.

This Recommendation gives the approved methods which are considered to be suitable for determining how satisfactorily given telephone connections may be expected to perform.

The methods indicated here are intended to be generally applicable whatever the form of degradation factors present. Examples of degrading factors include: loss (often frequency dependent); circuit noise; transmission errors (random bit errors as well as erased frames that occur in systems such as mobile communications); environmental noise; sidetone; talker echo; non-linear distortion of various kinds including low bit-rate encoding; propagation time; harmful effects of voice-operated devices; distortions of the time scale arising from packet switching; and time-varying degradations of the communication channel, including those arising in loudspeaking sets. Combinations of two or more of such factors also have to be catered for. Further guidance for specific applications is available in Recommendations P.830 (digital speech codecs), P.84 (DCME/PCME), and P.85 (speech output devices).

2 References

The following Recommendations and other references contain provisions that, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated are 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 listed below. A list of the currently valid ITU–T Recommendations is regularly published.

– IEC Publication 1260: 1995, *Electroacoustics – Octave-band and fractional – Octave-band filters*.

– IEC Publication 581-5: 1981, *High fidelity audio equipment and systems; Minimum performance requirements – Part 5: Microphones*.

---

[1] Formerly Recommendation P.80.

- IEC Publication 651: 1979, *Sound level meters. (Amendment 1-1993) (Corrigendum March 1994).*
- ISO 266: 1975, *Acoustics – Preferred frequencies for measurements.*
- ISO 1996-1: 1982, *Acoustics – Description and measurement of environmental noise – Part 1: Basic quantities and procedures.*
- ISO 1996-2: *Acoustics – Description and measurement of environmental noise – Part 2: Acquisition of data pertinent to land use.*
- ISO 1996-3: 1987, *Acoustics – Description and measurement of environmental noise – Part 3: Application to noise limits.*
- ITU-T Recommendation G.113 (1996), *Transmission impairments.*
- CCITT Recommendation G.722 (1988), *7 kHz audio-coding within 64 kbit/s.*
- CCITT Recommendation G.726 (1990), *40, 32, 24 and 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM).*
- CCITT Recommendation G.728 (1992), *Coding of speech at 16 kbit/s using low-delay code excited linear prediction.*
- ITU–T Recommendation G.729 (1996), *Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP).*
- ITU–T Recommendation P.10 (1993), *Vocabulary of terms on telephone transmission quality and telephone sets.*
- ITU–T Recommendation P.11 (1993), *Effect of transmission impairments.*
- CCITT Recommendation P.48 (1988), *Specification for an intermediate reference system.*
- ITU–T Recommendation P.56 (1993), *Objective measurement of active speech level.*
- ITU–T Recommendation P.78 (1993), *Subjective testing method for determination of loudness ratings in accordance with Recommendation P.76.*
- ITU–T Recommendation P.810 (1996), *Modulated Noise Reference Unit (MNRU).*
- CCITT Recommendation P.82 (1984), *Method for evaluation of service from the standpoint of speech transmission quality.*
- ITU–T Recommendation P.830 (1996), *Subjective performance assessment of telephone-band and wideband digital codecs.*
- ITU–T Recommendation P.84 (1993), *Subjective listening test method for evaluating digital circuit multiplication and packetized voice systems.*
- ITU–T Recommendation P.85 (1994), *A method for subjective performance assessment of the quality of speech voice output devices.*

3 Definitions

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

3.1 **dBov**: dB relative to the overload of a digital system.

3.2 ***Q***: The ratio, in dB, of speech power to modulated noise power in the Modulated Noise Reference Unit, as described in Recommendation P.810.

4 Abbreviations

For the purposes of this Recommendation, the following abbreviations are used:

ACR	Absolute Category Rating
ADPCM	Adaptive Differential Pulse Code Modulation
BER	Bit Error Rate
CCR	Comparison Category Rating
CMOS	Comparison Mean Opinion Score
DCR	Degradation Category Rating
DMOS	Degradation Mean Opinion Score
FER	Frame Erasure Rate
IRS	Intermediate Reference System (Recommendation P.48)
MNRU	Modulated Noise Reference Unit (Recommendation P.810)
MOS	Mean Opinion Score
PCM	Pulse Code Modulation
SNR	Signal-to-Noise Ratio

5 Conventions

Subjective evaluation of telecommunications equipment and systems may, in principle, be conducted using listening-only or conversational methods of subjective testing. As a practical matter, listening-only tests may be the only feasible method of subjective testing during the development of new transmission equipment or telecommunication services. This Recommendation describes recommended procedures for conversational and listening-only methods of subjective evaluation.

6 Recommended methods

6.1 Conversation-opinion tests

Laboratory conversation tests are intended – as far as possible – to reproduce, in the laboratory situation, the actual service conditions experienced by telephone customers. To this end it is necessary to choose the circuit conditions and subjects suitably, and to administer the tests in an appropriate manner.

It is important that the conditions simulated in the test are correctly specified and set up, and measured accurately before and after each experiment; that auxiliary facilities such as dialling and ringing are provided; and that faithful records of the output of each test are kept. Detailed description of the method, considerations and precautions are found in Annex A.

6.2 Listening-opinion tests

Listening-opinion tests are not expected to reach the same standard of realism as conversation tests, and the restrictions are therefore less severe in some respects; but the artificiality that has to be accepted brings with it a necessity for strict control of many things which in conversation tests are allowed to find their own equilibrium.

The recommended test method for listening-only tests is the "Absolute Category Rating" (ACR) method described in Annex B, which is in conformance with the Category Judgement method recommended for conversation tests (see Annex A), and adopted partly for the same reasons.

Category ratings are applied to short groups of unrelated sentences, each of which has been passed through a number of standard processes as well as the processes under test. This method is well-established, and has been applied to analogue and digital telephone connections and to telecommunications devices, such as digital codecs. In the work leading to Recommendations G.726 32 kbit/s ADPCM, G.728, G.729, and G.722, for example, laboratories in different countries performed subjective tests by the same method on the same physical conditions and on identical transmission systems, and the results showed a high degree of consistency.

Other methods commonly used are the Quantal-Response Detectability Method, Degradation Category Rating (DCR), Comparison Category Rating (CCR) and the Threshold Method.

Annex C describes Quantal-Response Detectability Tests, which are suitable for evaluating threshold values of certain quantities and their associated probabilities. For example, the level above which single-frequency interference has a given probability of being objectionable or detectable, or the probability that crosstalk in a given range of levels is intelligible, can best be determined by this method.

An alternative to the Absolute Category Rating method is the Degradation Category Rating (DCR) method which is described in detail in Annex D. The DCR method compares the system under test with a high quality fixed reference and the degradation (from "Inaudible" to "Very annoying") is rated on a five-point scale. This method is suitable when the impairment (especially digital impairments) is small. It may therefore be particularly useful for evaluating similar digital speech processing algorithms. Thus, the DCR method may serve as a means for system optimization once it has been shown by the methods of Annexes A and B that the worst-case connection incorporating the degradation in question is within acceptable limits.

Annex E describes a variation of the DCR procedure called the Comparison Category Rating (CCR) method. As in the DCR, the CCR method compares the system under test with a high quality fixed reference (in the CCR case on a scale from "Much Better" to "Much Worse"). This procedure may be particularly suitable for systems that improve the quality of the input speech (e.g. noise cancellation systems).

The Threshold method, also suitable for system optimization, is described in Annex F. By direct comparison of the system under test with a reference system, such as the Modulated Noise Reference Unit (MNRU, as described in Recommendation P.810), it is possible to equate the value of the reference condition (Q for digital processes) which equals the performance of the system under test.

Information on other types of subjective test methods, which include scaling methods, can be found in 2.6 of the *Handbook on Telephonometry.*

Listening tests have direct applications in the assessment of physical transmission systems which are essentially unidirectional. Examples include broadcast circuits, public address systems and recorded announcement systems in which listening degradations such as loss, noise and distortion may be present.

Results of listening-only tests can be applied, but only with certain reservations, to the prediction of the assessment for conversation conducted over a two-way system, such as a connection in a public switched telephone network. The provisos are that the effects of the following additional factors are duly taken into account:

- talking degradations (e.g. sidetone and echo);
- conversation degradations (e.g. propagation time and mutilation of speech by the action of voice-operated devices).

The annexes to this Recommendation provide information on preparation of speech material, processing of speech material, experiment philosophy (including choice of circuit conditions), listening test procedure and treatment of results.

6.3 Interview and survey tests

If the rather large amount of effort needed is available and the importance of the study warrants it, transmission quality can be determined by "service observations". Recommended ways of performing these, including the questions to be asked when interviewing customers, are given in Recommendation P.82. To maintain a high degree of precision a total of at least 100 interviews per condition is required.

A disadvantage of the service-observation method for many purposes is that little control is possible over the detailed characteristics of the telephone connections being tested. However, this method does afford a global appreciation of how the "equipment" performs in the real environment.

Further information can be found in 2.5.8.3 of the *Handbook on Telephonometry.*

6.4 Other tests

Reference [2] gives information of a method that largely overcomes the disadvantages of the interview technique of 6.3, yet retains many of the advantages. This method, termed SIBYL, allows a small proportion of a user's ordinary calls to be passed through special arrangements which modify the normal quality of transmission according to a test programme. If a particular call has been so treated, the volunteer is asked to vote by dialling one of a set of digits to indicate his opinion. In this way, all results are recorded by the controlling computer and complete privacy is maintained.

Annex A

Conversation-opinion tests

A.1 Test facilities

A.1.1 Physical conditions

A.1.1.1 Test cabinets

The two subjects are seated in separate sound-proof cabinets near the point from which the experiment is controlled. The volume of the room is not less than 20 m^3, with a reverberation time less than 500 ms (normally in the range of 200-300 ms), for handheld systems such as telephone handsets, or for headset systems; and not less than 30 m^3 for handsfree systems (extra care is exercised if reverberation time is an experimental variable).

The internal dimensions of the cabinet are such that standing-wave pattern effects are kept to a minimum. A typical ratio is 5:4:3.

The physical construction of the rooms should be such that sufficient sound attenuation of the outside noise environment is achieved so that the requirements of A.1.1.2.1 are met.

The cabinets are favourably decorated to recreate a natural environment.

A.1.1.2 Noise

A.1.1.2.1 Noise floor

The ambient noise level (when no environmental noise is deliberately introduced) is kept as low as possible. For practical reasons, such as regular changes of fresh air in the cabinet, the target is an upper limit of NC25 [3] or NR25 (see ISO 1996). These values approximate the noise level in homes (sleeping areas), hospitals and libraries.

A.1.1.2.2 Environmental noise

Environmental noise is fed in with the required spectrum (e.g. Hoth spectrum to represent typical room noise – see A.1.1.2.2.1) at the required level (e.g. 50 dBA) measured with a Precision Sound Level Meter conforming to IEC Publication 651, used with the "A weighting" and the "fast" meter characteristic. If different conversations in the same experiment require different room noise levels, then care is taken to prevent the transitions from being too obvious to the subjects. Ideally, room noise should be changed only when subjects are out of the sound-proof rooms. If this is not possible, then changes of level are carried out gradually (at a rate not exceeding 4 dB per second), at a time when no experimental conversation is in progress and when the subjects' attention is otherwise occupied – by communicating with the operator, for example.

Spectra with appropriate long-term characteristics are given in A.1.1.2.2.1 and A.1.1.2.2.2.

For some purposes it is necessary to use noise that fluctuates in level or spectrum, such as tape recordings of actual office noise or traffic noise. In such cases it should be ensured that the statistical characteristics are stable when averaged over a reasonably short period of time such as one minute.

It is recommended that the noise level and spectrum are measured at least twice; at the beginning and end of the experiment. Any significant variation in the two measurements, when compared with each other, must be assessed by the experimenter as it may cast doubt on the validity of the experiment.

It is essential to ensure that the loudspeakers and amplifiers are capable of faithfully reproducing the required noise.

A.1.1.2.2.1 Room noise

The room noise shall have a power density spectrum corresponding to that published by Hoth [4]. Table A.1 gives the spectrum density adjusted in level to produce a reading of 50 dBA on a sound level meter conforming to IEC Publication 651. This is produced in Figure A.1. This spectrum is independent of level, i.e. for 40 dBA the level in each band shall be 10 dB less than that shown in Table A.1. Additional information on the power in each one-third octave band is also given in Table A.1.

TABLE A.1/P.800

Room noise spectrum

Frequency (Hz)	Spectrum density (dB SPL/Hz)	Bandwidth $10 \log_{10} \Delta f$ (dB)	Total power in each 1/3rd octave band (dB SPL)	Tolerance (dB)
100	32.4	13.5	45.9	
125	30.9	14.7	45.4	
160	29.1	15.7	44.9	
200	27.6	16.5	44.1	
250	26.0	17.6	43.6	
315	24.4	18.7	43.1	
400	22.7	19.7	42.3	
500	21.1	20.6	41.7	
630	19.5	21.7	41.2	±3
800	17.8	22.7	40.4	
1000	16.2	23.5	39.7	
1250	14.6	24.7	39.3	
1600	12.9	25.7	38.7	
2000	11.3	26.5	37.8	
2500	9.6	27.6	37.2	
3150	7.8	28.7	36.5	
4000	5.4	29.7	34.8	
5000	2.6	30.6	33.2	
6300	–1.3	31.7	30.4	
8000	–6.6	32.7	26.0	

NOTES

1 The electrical input signal, e.g. white noise, shall be band-limited to the 1/3rd octave bands centred on the ISO preferred frequencies (ISO 266) between 100 Hz and 8000 Hz with the band edges conforming to the filters described in IEC 1260.

2 The acoustical room noise is difficult to control at low frequencies, especially in the unspecified region below 100 Hz because of the dimensions of typical test cabinets, poor attenuation of such cabinets and the influence of extraneous noises, e.g. air-conditioning plant. It is therefore desirable to select a test cabinet that keeps these unwanted low frequency sound pressure levels to a minimum.

A.1.1.2.2.2 Internal vehicle noise

Two spectra representing internal vehicle noise [5], [6] are recommended. They are adequately represented by simplified curves [7]: one spectrum for moving vehicles and the other for stationary vehicles. Table A.2 gives the spectrum densities together with additional information on the power in each one-third octave band. The spectrum density for moving vehicles is shown in Figure A.2 a) and for stationary vehicles in Figure A.2 b). These spectra are independent of level.

NOTE – The noise spectra in Table A.2 should be considered provisional. More detailed specifications are under study.

Table A.3 gives the computed values of the unweighted sound pressure levels for various speeds calculated over the ISO one-third octave frequency bands centred on 63 Hz to 8000 Hz.

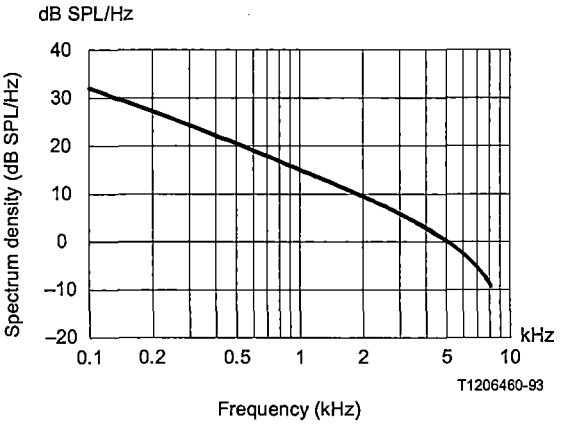


FIGURE A.1/P.800

Room noise spectral density

TABLE A.2/P.800

Internal vehicle noise spectra

Frequency (Hz)	Spectrum density (dB SPL/Hz)		Bandwidth $10 \log_{10} \Delta f$	Total power in each 1/3rd octave band (dB SPL)		Tolerance (dB)
	Moving	Stationary	(dB)	Moving	Stationary	
63	72.3	58.3	11.7	84.0	70.0	
80	69.3	55.0	12.7	82.0	66.7	
100	66.5	49.8	13.5	80.0	63.3	
125	63.3	45.1	14.7	78.0	60.0	
160	60.3	42.0	15.7	76.0	56.7	
200	57.5	36.8	16.5	74.0	53.3	
250	54.4	34.7	17.6	72.0	52.3	
315	51.3	32.6	18.7	70.0	51.3	
400	48.3	30.6	19.7	68.0	50.3	
500	45.4	28.7	20.6	66.0	49.3	
630	42.3	26.6	21.7	64.0	48.3	±3
800	39.3	24.6	22.7	62.0	47.3	
1000	36.5	22.8	23.5	60.0	46.3	
1250	33.3	20.6	24.7	58.0	45.3	
1600	30.3	18.6	25.7	56.0	44.3	
2000	27.5	16.8	26.5	54.0	43.3	
2500	24.4	14.7	27.6	52.0	42.3	
3150	21.3	12.6	28.7	50.0	41.3	
4000	18.3	10.6	29.7	48.0	40.3	
5000	15.4	8.7	30.6	46.0	39.3	
6300	12.3	6.6	31.7	44.0	38.3	
8000	9.3	4.6	32.7	42.0	37.3	

TABLE A.3/P.800

Computed sound pressure levels of spectra

Spectra		Sound pressure level, unweighted (dB SPL)
Moving	30 km/h	80
	80 km/h	85
	110 km/h	90
Stationary		75

NOTES to Tables A.2 and A.3:

1 These values apply for typical vehicles. Discretion may be used to adjust the levels downwards for luxury vehicles and upwards for noisier vehicles.

2 Because of the practical difficulty of generating such high sound pressure levels at low frequencies, and because normal speech contains no apparent energy below about 63 Hz in which range of frequencies the ear is also comparatively insensitive it is probably advisable to restrict the recommended noise spectrum to frequencies above 63 Hz. However, it should be borne in mind that low and medium frequency vibrations have important physiological and psychological effects which should be studied in their own right.

3 The electrical input signal, e.g. white noise, shall be band-limited to the 1/3rd octave bands centred on the ISO preferred frequencies (ISO 266) between 63 Hz and 8000 Hz with the band edges conforming to the filters described in IEC 1260.

4 The acoustical room noise is difficult to control at low frequencies especially in the unspecified region below 63 Hz because of the dimensions of typical test cabinets, poor attenuation of such cabinets and the influence of extraneous noises, e.g. air-conditioning plant. It is therefore desirable to select a test cabinet that keeps these unwanted low frequency sound pressure levels to a minimum.

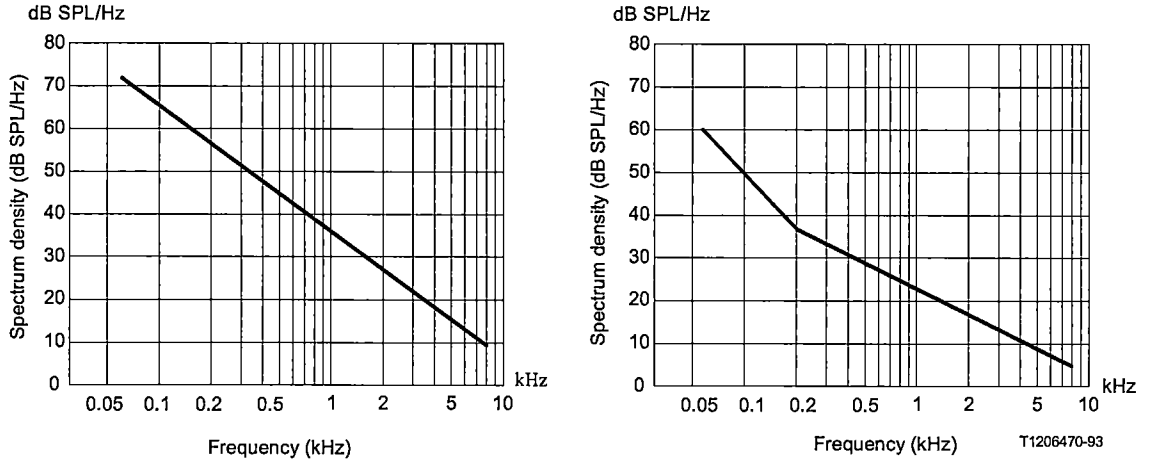


FIGURE A.2/P.800

Vehicle noise spectral density

A.1.1.3 Noise measurement position

It is recommended that the measurement of Sound Pressure Level (SPL) in the test cabinets (see A.1.1.1) shall be made as follows:

- furniture in position;
- no subject or test personnel present;
- the SPL shall be measured at a vertical distance of 740 mm above the centre of the seat of the subject's chair with a meter conforming to Recommendation P.54 using "A" weighting;
- the spectrum of the environmental noise shall be measured in one-third octaves, centred at the preferred frequencies as defined in ISO 266, and must stay within the specified tolerances, e.g. ±3 dB for Hoth noise (see A.1.1.2.2.1);
- in rooms where more than one subject is to be tested the difference in dBA, for all subject positions, shall not vary by more than ±2 dB.

NOTE – It is suggested that the minimum distance between each loudspeaker and the measurement position should be 1.5 m.

A.1.2 Establishing the connection

When establishing the laboratory connection, the following must be taken into consideration:

- telephone sets;
- initial call set-up;
- laboratory representation of telephone connections.

It is recommended that the sensitivity/frequency characteristic of the connection is measured at least twice; at the beginning and end of the experiment. Any significant variation in the two measurements, when compared with each other, must be assessed by the experimenter as it may cast doubt on the validity of the experiment.

Detailed information on this aspect can be found in 2.5.8.2 of the *Handbook on Telephonometry.*

A.1.3 Monitoring

Monitoring can take many forms but the three most commonly used are:

- *Intercommunication system* – Essential to allow the subject and experimenter to communicate with each other.
- *Visual monitoring* – This has two purposes; the first being safety and the second to observe the peculiarities of the subject, e.g. how they hold the telephone handset.
- *Tape recordings and recording system* – This facilitates other useful information to be gleaned, e.g. duration of the call, speech voltage, speech activity.

Detailed information on this aspect can be found in 2.5.8.2 of the *Handbook on Telephonometry.*

A.2 Experiment design

There are a number of methods suitable for use in designing experiments, e.g. Latin Squares, Youden Squares, Balanced Incomplete Blocks and Randomization with Replication to name but a few. The experimenter must decide for himself the method to be used taking into account the number of test conditions, accuracy of results and the ability to make sound judgements from the findings.

Suitable designs include those of the n × n graeco-latin square type. A detailed description of these designs can be found in 2.5.8.2 of the *Handbook on Telephonometry.*

A.3 Conversation task

Every effort is to be made to ensure that conversations are purposeful, and that subjects have full opportunity to exploit the transmission capabilities of the test circuit.

The general rule is that every conversation should have a natural beginning and a natural ending. Unless it is absolutely necessary, the conversation must never be terminated in the middle of the task (with the exception of Simplified Conversation Tests which are described in 2.5.8.2 of the *Handbook on Telephonometry*).

An example of a conversational task can be found in 2.5.8.2d of the *Handbook on Telephonometry*.

A.4 Test procedure

A.4.1 Eligibility of subjects

Subjects taking part in the conversation tests are chosen at random from the normal telephone using population, with the provisos that:

a) they have not been directly involved in work connected with assessment of the performance of telephone circuits, or related work such as speech coding; and

b) they have not participated in any subjective test whatever for at least the previous six months, and not in a conversation test for at least one year.

If the available population is unduly restricted, then allowance must be made for this fact in drawing conclusions from the results.

No steps are taken to balance the numbers of male and female subjects unless the design of the experiment requires it. Subjects are arbitrarily paired in the experimental design prior to the test and remain thus paired for its duration.

A.4.2 Opinion scale

The following opinion scales are those recommended by the ITU-T.

A.4.2.1 Conversation opinion scale

Various five-point category-judgement scales may be used for different purposes. The layout and wording of opinion scales, as seen by subjects in experiments, is very important, and should follow the standard arrived at through years of experience. The following opinion scale is the most frequently used for ITU-T applications and equivalent wording should be used depending on language which might result in small variations to the original English text.

This is a category rating obtained from each subject at the end of each conversation.

Opinion of the connection you have just been using

Excellent

Good

Fair

Poor

Bad

The experimenter allocates the following values to the scores:

Excellent = 5; Good = 4; Fair = 3; Poor = 2; Bad = 1

and all further statistical processing is performed in terms of these numbers. The arithmetic mean of any collection of these opinion scores is called the mean conversation-opinion score, and is represented by the symbol MOS_C (or, where suffix notation is not available, the symbol MOSc).

NOTE – In the past, the equivalences Excellent = 4, Good = 3, Fair = 2, Poor = 1, Bad = 0 have often been used. Anyone using results from earlier experiments must be aware that the mean scores must all be increased by one to be comparable with those now obtained; otherwise there is no difference in the numerical processing that may be applied.

A.4.2.2 Difficulty scale

This is a binary response obtained from each subject at the end of each conversation.

Did you or your partner have any difficulty in talking or hearing over the connection?

Yes

No

The experimenter allocates the following values to the scores:

Yes = 1 No = 0

The quantity evaluated (percentage of "yes" responses) is called percentage Difficulty or per cent "Difficult", and is denoted by the symbol %D. The corresponding simple proportion is denoted by the symbol d; in other words, %D = 100d.

NOTE – It is often the case that the nature of the difficulty is required and then it is usual for the experimenter to ask the subject to describe in his/her own words their perception of the difficulty.

The layout and wording of the opinion scale, as seen by subjects in experiments, are very important, and should follow the standard arrived at through years of experience: see A.4.3.

A.4.2.3 Other opinion scales

Other opinion scales that may be suitable are variants of the methods of "magnitude estimation" and "cross-modality matching" [8]. The responses on these scales may be one of the following:

a) one of a numerical series of categories labelled 1, 2, 3, 4, 5 (and denoted as such to the subject), but with descriptions attached only to the first and the last, to identify the subjective dimension;

b) a numerical mark on a scale from one to a number much greater than five – say 10 or 100; or

c) a length proportional to some property (e.g. quality), marked manually along a given straight line.

A survey of experimental methods can be found in 2.6.2 of the *Handbook on Telephonometry.*

A.4.3 Instructions to subjects

Instructions are given to subjects on arrival for their first visit. It is normal for the subject to receive a letter prior to arrival which contains non-technical information on the experiment and what will be expected of them. An example of such a letter can be found in 2.5.8.2, Table 3/2.5 of the *Handbook on Telephonometry.*

They are asked whether they have read and understood the letter. Any obscurities are clarified, and opportunity is given for asking questions. The sound-proof rooms and their facilities are demonstrated. Subjects are informed how many calls will be comprised in this visit. On subsequent visits the subjects are merely informed that the procedure will be the same as before, with possibly a different number of calls. An example of some operational details required by an experiment can be found in 2.5.8.2, Table 4/2.5 of the *Handbook on Telephonometry.*

A.4.4 Data collection

Speech levels, and related data such as durations and activity factors, may be derived from the tape recordings, but are now normally measured on-line, by computer-controlled meters, and stored directly into computer files for subsequent analysis.

Two subjective responses are collected per conversation per subject by the experimenter. The essential data consists of the conversation-opinion score and the Difficulty decision. These responses may be collected using any suitable means, including pencil and paper, electronic buttons, keyboards, keypads, or computer touch-screen terminals. A sample response form can be found in 2.5.8.2, Table 5/2.5 of the *Handbook on Telephonometry*.

A.4.5 Treatment of results

This is a very extensive subject, and only a brief outline can be given here.

Each conversation gives rise to two conversation opinions on the scale: Excellent – Good – Fair – Poor – Bad (scored respectively 5, 4, 3, 2, 1), two votes on the Difficulty scale (scored 1 = Yes, 0 = No), two measured active speech levels and one value of duration. In particular cases information may be collected about other variables; e.g. video recordings may be made in order to observe how subjects hold their handsets, or other data may be derived from the opinion forms or from the audio recordings.

The average of the opinion scores should be calculated for each test condition. Confidence limits should be evaluated and significance tests performed by conventional analysis-of-variance techniques.

The usual assumptions underlying the analysis of variance are sufficiently nearly satisfied in the case of opinion score, active speech level and most other variables of interest; but they are not satisfied – particularly the assumption of constant residual variance – in the case of a binary variant like Difficulty score. In spite of this, experience confirms the observation made in other fields [9] that the analysis of variance technique is robust enough to give reasonable results even with such extreme departures from the statistically ideal conditions. The results from the first stage of the analysis of variance of the Difficulty scores should be regarded with some reserve; but once it is established that there are no unexplained abnormalities in the results and no unexplained conflicts with the outcome of the corresponding analysis of the MOS_C results, then the second stage (detailed analysis of the averages from the End-Condition combinations) can be confidently undertaken with the aid of a mathematical transformation.

Detailed descriptions of the analysis can be found in 2.5.9 of the *Handbook on Telephonometry*.

As a further aid to the review of the data, graphs if appropriate should be plotted showing the mean opinion score as a function of the parameter under test, e.g. MOS_C versus circuit attenuation. On the graph the vertical axis should always be MOS_C.

Annex B

Listening tests – Absolute Category Rating (ACR)

B.1 Source recordings

In order to eliminate unwanted variability in the speech source, samples of speech having the desired standardized properties should first be prepared in recorded or stored form, as follows.

B.1.1 Recording environment

The talker should be seated in a quiet room with volume between 30 and 120 m^3 and a reverberation time less than 500 ms (preferably in the range 200-300 ms). The room noise level must be below 30 dBA with no dominant peaks in the spectrum.

The room noise characteristic should be reported in as complete a form as possible, e.g. dBA, long-term spectrum, and amplitude-time distribution. It is desirable to record a 30-second sample of the room noise for detailed investigation if this proves to be necessary.

B.1.2 Sending system

Whatever sending system is chosen, e.g. local telephone or Intermediate Reference System (IRS) as specified in Recommendation P.48, the system should be calibrated according to the relevant Recommendation (e.g. Recommendation P.64), and the sending sensitivity-frequency characteristic should be reported in full. Annex D/P.830 describes the "modified IRS" that has been deemed appropriate for evaluation of all-digital connections using speech codecs.

It is recommended that the sending sensitivity characteristic of the connection is measured at least twice; at the beginning and end of the experiment. Any significant variation in the two measurements, when compared with each other, must be assessed by the experimenter as it may cast doubt on the validity of the experiment.

B.1.3 Recording system

The recording system must be of high (studio) quality and can be any of the following:

a) A conventional two-track tape recorder. The type of equalization must be stated, but IEC is recommended. High grade tape (low print-through, low noise) should be used at all times.

b) A two-channel digital audio processor with a high quality video cassette recorder or Digital Audio Tape (DAT) machine.

c) A computer-controlled digital storage system.

The third system is the best and most versatile, but practical reasons often dictate the choice of one of the other systems. In these, one of the two tracks should be used for recording the speech and the other for inserting control signals at a level and a frequency chosen to avoid crosstalk problems.

B.1.4 Speech material

The speech material should consist of simple, meaningful, short sentences, chosen at random as being easy to understand (from current non-technical literature or newspapers, for example). These sentences should be made up into lists in random order in such a way that there is no obvious connection of meaning between one sentence and the next. Very short and very long sentences should be avoided, the aim being that each sentence when spoken should fit into a time-slot of 2–3 seconds. Examples of sentences are shown in Table B.1.

The experimenter must decide how many sentences are required in each group to constitute a speech sample. A minimum of two and a maximum of five are recommended. The time interval between sentences, during which circuit noise may be heard and adaptive processes settle into new states, is

also important. It is advisable to record the longest groups that may be needed, as it is always possible to obtain shorter groups by copying or replaying parts of longer ones.

Groups are combined into lists consisting of five or ten groups each, so that a complete list can be used as a series of samples subjected to the same treatment but with listening level or some other parameter varied when the list is reproduced.

TABLE B.1/P.800

Examples of speech material

You will have to be very quiet.
There was nothing to be seen.
They worshipped wooden idols.
I want a minute with the inspector.
Did he need any money?

B.1.5 Recording procedure

The following recording scheme has been extensively used and is recommended.

Speech is recorded from a linear microphone and low-noise amplifier with a flat frequency response as specified in IEC Publication 581-5. The microphone is positioned between 140 mm and 200 mm from the talker's lips. In some applications, it may be necessary to use a windscreen, it is used if breath puffs from the talker are noticed.

The same speech may be recorded simultaneously from the sending output of an Intermediate Reference System (IRS, see Recommendation P.48), with the handset held in the normal manner. If the investigation in view specifically requires it, another telephone instrument may be used in place of the IRS.

Two separate recording systems are used simultaneously: one for recording the wideband speech in one channel, and the other for recording the telephone speech in the corresponding channel. The other channel of each recording system is used for recording control signals as explained in 2.5 of the *Handbook on Telephonometry.*

This dual recording system ensures that the same speech is recorded in two forms (telephone speech and wideband speech). Normally only one of these is required in any one experiment, but there are occasions when it is necessary to use both, and it is an advantage in any case to be able to make comparative measurements on the two versions.

The active speech level, as defined in Recommendation P.56, is observed during recording. Care is taken during the recording process that the active speech level in both recording systems is between 20 and 30 dB below the overload point of the recording system for each sentence measured separately. Any group of sentences for which this does not hold is re-recorded.

It is recommended that the ratio of the active speech level to psophometrically weighted noise level (for definition see 8.2.3/P.830), SNR(p), on the recording media should be > 40 dB with an objective of 50 dB.

All speech samples used in one experiment may be different: this is essential for Listening-Effort tests and desirable for other types.

B.1.6 Talkers

There must be as many talkers as required in the design of the experiment (see B.3).

Talkers should pronounce the sentences fluently but not dramatically, and have no speech deficiencies such as stutter; they should adopt a speaking level that is comfortable to them as individuals and which they can maintain fairly constantly.

B.1.7 Speech levels

The recordings when completed are played back, and the active speech level of each sentence is measured with a meter conforming to Recommendation P.56. The lists (announcements, sentences and control tones) are then re-recorded on to a second system with the necessary gain adjustments, so as to bring each group of sentences to the standardized active speech level specified below, and still preserve the proper time relationships between the sentences and the tone signals in the other channel.

For the narrow-band speech, the standardized level is derived by measuring and adjusting the narrow-band recorded signal directly; the recommended target is –26 dB (+0.5 dB) relative to the peak overload level of the recording system. The calibration tone has its r.m.s. level equal to the mean active level of the re-recorded speech.

For the wideband speech, account must be taken of the intended use of the recordings. It is sometimes appropriate to adopt the same levels as for telephone speech, but if the recording is intended for playback through a loudspeaker or artificial mouth, then the individual target speech levels should be such that equality is maintained at the output of the whole electric-acoustic replay chain.

B.1.8 Calibration signal

At the beginning of each resulting recording, 20 seconds or more of tone are inserted at the re-recording stage (for calibration purposes) at a level that is in a known relationship to the mean active speech level (most conveniently, equal to it). This calibration tone is normally at 1000 Hz, but may be at some other frequency if the recordings are intended for playing through systems (such as certain types of sub-band coder) that respond to 1000 Hz in a special manner.

This tone can then be used later to adjust the mean input speech levels (see B.4.3).

B.2 Selection of circuit conditions

B.2.1 Speech input and listening levels

In the selection of circuit conditions particular attention should be applied to:

- range of input levels;
- range of listening levels:
  - there is no universal optimum listening level;
  - a variety of listening levels will occur in practice;
  - comparability considerations;
  - interactions may occur.

A detailed explanation of these aspects can be found in 2.5.8.1 of the *Handbook on Telephonometry.*

B.2.2 Talkers

Since sophisticated processes often affect male and female voices differently, the experimental design should provide for two types of voice as a balanced factor; scores for male and female speech

should be evaluated separately, only to be averaged if they yield main effects and interactions that are not statistically different.

Moreover, to reduce the danger that the results may depend heavily on peculiarities of the voices chosen, it is essential for more than one male and more than one female voice to be used in a balanced design.

B.2.3 Reference conditions

Every experiment should include reference conditions so that experiments made in different laboratories or at a different time in the same laboratory can be sensibly compared. Such reference conditions will depend on what is being assessed. For a digital system the reference conditions may include the Modulated Noise Reference Unit (MNRU) conforming to Recommendation P.810; other controlled degradations are appropriate in other cases (e.g. signal-to-noise ratio, see 8.2.3/P.830).

B.2.4 Other conditions

Besides the requirements of B.2.1 to B.2.3 inclusive, other conditions will be included depending on the purpose of the test. For instance, room noise might be a variable as well as bit-error rate for a digital system or Rayleigh fading for a radio system.

B.3 Design of experiment

The design of the experiment uses the same principles as given in A.2.

In addition, the experiment design must cater for the following:

a) requirements of B.2;

b) order-of-presentation effect.

For a given sample of subjects the test is limited in size by the maximum length of session possible without fatigue. If the experiment is too large to be catered for in one session then it is prudent to sub-divide into two or more sessions. Ideally no session should last for more than 20 minutes and in no case should a session exceed 45 minutes.

B.4 Listening test procedure

B.4.1 Listening environment

The listening room should meet the same conditions as the recording room (see B.1.1) with the exception that the environmental noise (see A.1.1.2.2) should be set to the appropriate level. See A.1.1.2.2.1 and A.1.1.2.2.2 for examples of noise spectra.

It is recommended that the noise level and spectrum are measured at least twice; at the beginning and end of the experiment. Any significant variation in the two measurements, when compared with each other, must be assessed by the experimenter as it may cast doubt on the validity of the experiment.

B.4.2 Listening system

Whatever listening system is chosen (e.g. local telephone system, Intermediate Reference System as specified in Recommendation P.48, a loudspeaker system), the system should be calibrated according to the relevant Recommendation (e.g. Recommendation P.64) and the receiving sensitivity/frequency characteristic should be reported in full. Annex D/P.830 describes the "modified IRS" that has been deemed appropriate for evaluation of all-digital connections using speech codecs.

It is recommended that the receiving sensitivity/frequency characteristic of the connection is measured at least twice; at the beginning and end of the experiment. Any significant variation in the

two measurements, when compared with each other, must be assessed by the experimenter as it may cast doubt on the validity of the experiment.

B.4.3 Listening level

The gain of the system should be set in such a way that the calibration tone (see B.1.8) played back from the processed tapes produces the required listening level.

Variations in listening level, as required by the experiment design, can either be accommodated by use of:

a) attenuators/amplifiers in the listening system; or

b) included in the processing or re-recording stage.

The second method is not recommended because it is difficult to maintain a high enough signal-to-noise ratio at low levels, and because flexibility and variety in the randomization are greatly reduced.

The listening level should always be recorded. Information on this subject can be found in 2.5 of the *Handbook on Telephonometry*.

B.4.4 Listeners

Subjects taking part in listening tests are chosen at random from the normal telephone using population, with the provisos that:

a) they have not been directly involved in work connected with assessment of the performance of telephone circuits, or related work such as speech coding;

b) they have not participated in any subjective test whatever for at least the previous six months, and not in any listening-opinion test for at least one year; and

c) they have never heard the same sentence lists before.

If the available population is unduly restricted, then allowance must be made for this fact in drawing conclusions from the results.

In some cases screening of subjects may be necessary and a method based on Annex B/P.78 may be applicable.

B.4.5 Opinion scales recommended by the ITU-T

Various five-point category-judgement scales may be used for different purposes. The layout and wording of opinion scales, as seen by subjects in experiments, is very important, and should follow the standard arrived at through years of experience. The following opinion scales are those most frequently used for ITU-T applications and equivalent wording should be used depending on language which might result in small variations to the original English text:

a) **Listening-quality scale**

Quality of the speech	*Score*
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

The quantity evaluated from the scores (mean listening-quality opinion score, or simply mean opinion score) is represented by the symbol MOS.

b) **Listening-effort scale**

The heading of the listening-effort opinion scale is particularly important. Without it, the other descriptions are liable to be seriously misunderstood.

Effort required to understand the meanings of sentences	*Score*
Complete relaxation possible; no effort required	5
Attention necessary; no appreciable effort required	4
Moderate effort required	3
Considerable effort required	2
No meaning understood with any feasible effort	1

The quantity evaluated from the scores (mean listening-effort opinion score) is represented by the symbol MOS_{LE} but where suffix notation is not available, the symbol MOSle is used.

c) **Loudness-preference scale**

Loudness preference	*Score*
Much louder than preferred	5
Louder than preferred	4
Preferred	3
Quieter than preferred	2
Much quieter than preferred	1

The quantity evaluated from the scores (mean loudness-preference opinion score) is represented by the symbol MOS_{LP} but where suffix notation is not available, the symbol MOSlp is used.

NOTE – Examples of alternative subjective scales, which should only be used if the above three opinion scales do not meet the needs of the experimenter, are given in 2.6 of the *Handbook on Telephonometry* and CCIR Report 751, Volume VIII.3, 1986.

B.4.6 Instructions to subjects

An example of typical instructions is given in Table B.2. The instructions must be given (verbally as well, if necessary) prior to commencement of the experiment. When the subject has understood the instructions, he/she should listen to the preliminary list and give his opinions. No suggestion should be made to the subjects that the preliminary samples include the best or worst in the range to be covered, or exhaust the range of conditions they can expect to hear. After the preliminary list, there should be sufficient time allowed for answering possible questions by the subjects. Questions about procedure or about the meaning of the instructions should be answered, but any technical questions must be met with the response, "We cannot tell you anything about that until the experiment is finished".

TABLE B.2/P.800

Example of instructions to subjects

LISTENING EXPERIMENT No. ...

In this experiment you will be listening to short groups of sentences via the telephone handset, and giving your opinion of the speech you hear.

On the table in front of you is a box with five illuminated press buttons. When all the lamps go on, you will hear ... sentences. Listen to these, and when the lamps go out, press the appropriate button to indicate your opinion on the following scale.

EFFORT REQUIRED TO UNDERSTAND THE MEANINGS OF SENTENCES

5	Complete relaxation possible; no effort required.
4	Attention necessary; no appreciable effort required.
3	Moderate effort required.
2	Considerable effort required.
1	No meaning understood with any feasible effort.

The button you have pressed will light up for a short time. Then the lamp will go out, and there will be a brief pause before all the lamps go on again for the next group of ... sentences.

There will be a longer pause after every ... groups (each calling for an opinion). There will be a total of ... groups in this visit, and a similar number in your subsequent visit(s).

Thank you for your help in this experiment.

B.4.7 Statistical analysis and reporting of results

The numerical mean (over subjects) should be calculated for each condition at each listening level, and these means listed for initial inspection (so that effects such as those due to male and female speech can be seen).

Calculation of separate standard deviations for each condition is not recommended. Confidence limits should be evaluated and significance tests performed by conventional analysis-of-variance techniques.

NOTE – In the past, the equivalences, for example, Excellent = 4, Good = 3, Fair = 2, Poor = 1, Bad = 0 have often been used. Anyone using results from earlier experiments must be aware that the mean scores must all be increased by one to be comparable with those now obtained; otherwise there is no difference in the numerical processing that may be applied.

The method of analysis of the opinion scales of B.4.5 follows the principles stated in A.4.5.

As a further aid to the review of the data, graphs if appropriate should be plotted showing the mean opinion score as a function of the parameter under test, e.g. MOS versus circuit attenuation. On the graph the vertical axis should always be MOS.

Averaging the scores of the male and female talkers should be made with care and does not imply that this step would be warranted for a detailed study and interpretation of results unless the significance tests justify it.

Annex C

Quantal-Response Detectability Tests

The best method for obtaining information on the detectability or some analogous property of a sound (such as echo) as a function of some objective quantity (such as listening level), is a quantal-response method similar in principle to that described in 2.2 of the *Handbook on Telephonometry.*

The main difference is that the subject's response is not a decision in the form "Reference" or "Test" (the designation of the louder of the two circuits), but a vote on a scale such as [10]:

Detectability opinion scale

A Objectionable

B Detectable

C Not detectable

where B is understood to mean "Detectable but not objectionable".

Scales of this sort, usually with three points, may be used in a variety of quantal-response tests; for example the scale as shown above may be used where the stimulus is echo, reverberation, sidetone, voice-switching mutilation, or interfering tones, while crosstalk and perhaps echo in some circumstances may be judged on the scale Intelligible – Detectable – Not detectable.

It is sometimes permissible to regard these votes as opinion scores, with values 2, 1, 0 respectively, and treat them in the same sort of way as listening or conversation opinion scores. But this is often unsatisfactory because the decisions on such a scale as detectability (see above) are not really equivalents of responses on a continuous scale – as votes on such scales as "Loudness preference" (see B.4.5), may be legitimately taken to be – but effectively embody two distinct dichotomies (for example detectable/not detectable and objectionable/not objectionable), which though not independent may nevertheless call different psychological processes into action: in other words, Objectionability or Intelligibility differs in kind, not merely in degree, from Detectability. For this reason a more profitable method of analysis is to express the probability of response according to each dichotomy separately, as a function of some objective variable, by fitting probit or logit equations, and then using the quantiles or other parameters as a basis of comparison between circuit conditions, in a manner analogous to that used in applying articulation scores.

The actual conduct of experiments of this type resembles that of listening-effort tests (see Annex B), but there are some differences. In particular, it is advisable that the first presentation of the signal in each run should be at a high listening level, so that the listener is left in no doubt what kind of signal is a candidate for his decisions. Where sidetone or echo is involved, the subject will be required to talk as well as listen.

Simple audiometric measurements, as described in Recommendation P.78, are usually performed on subjects who participate in these experiments, so that results can be expressed relative to their threshold of hearing.

For examples of the application of these techniques, see [11].

Noise, fading and other disturbances are sometimes investigated by means of responses on a scale with many more points; for example [12]:

A *Inaudible* – Noise completely undetectable.

B *Just audible* – Noise can just be detected by listening carefully.

C *Slight* – Noise detectable, but not disturbing.

D *Moderate* – Noise slightly disturbing.

E *Rather loud* – Noise causes appreciable disturbance.

F *Loud* – Noise very disturbing, but call would be continued.

G *Intolerable* – Noise so loud that the call would be abandoned, or operator asked to change the line.

These scales are more nearly of the quantized-continuum type, like the Loudness preference scale, and can be treated similarly.

Annex D

Degradation Category Rating (DCR) method

D.1 Introduction

The Absolute Category Rating (ACR) method described in Annex B tends to lead to low sensitivity in distinguishing among good quality circuits. A modified version of the ACR procedure, called the Degradation Category Rating (DCR) [13] procedure, affords higher sensitivity. This procedure is adapted from the CCIR Recommendation [14] for evaluation of good quality circuits. The DCR procedure, which in particular uses an annoyance scale and a quality reference before each configuration to be evaluated, seems to be suitable for evaluating good quality speech.

D.2 Degradation Category Rating (DCR) procedure

D.2.1 Speech samples

Each configuration is evaluated by means of judgements on speech samples from at least four talkers. Each sample should be composed of two sentences separated by approximately 0.5 s of silence. These two samples (S1, S2), hence four different sentences, should be selected from a wider corpus composed of phonetically balanced sentences so that the mean score obtained in evaluating reference (e.g. MNRU for digital processes) circuits for these sentences is about the same as that obtained for the wider corpus. Therefore the corpus consists of eight samples defined as follows:

- talker T1 reading samples S1, S2;
- talker T2 reading samples S1, S2;
- talker T3 reading samples S1, S2;
- talker T4 reading samples S1, S2;
- etc.

This results in a repetition of the two samples during the test. It is felt that this is not a critical factor for the procedure where a degradation is evaluated with regard to the reference. This is especially true for good telephone quality, where the intelligibility of speech is nearly perfect. The use of different samples for each configuration, as is often done in ACR experiments (where the speaker and the sentence effects are confounded), could be one of the reasons for lack of sensitivity in the ACR method.

Some variations of this basic scheme are allowed: increase the number of talkers, mix sentence and talker effects. However, it is important that all configurations are evaluated on the same corpus.

D.2.2 Reference conditions

Reference conditions shall be included, e.g. for digital processes multiplicative noise with Q values within the range 10 to 30 dB with a minimum of four steps is desirable.

A quality reference should be chosen to be inserted before each judgement. Usually source conditions are used, i.e. samples with no more degradation than those introduced by sending systems and limitations of frequency bandwidth. Thus, the choice of the quality reference depends on the application, i.e. for standard telephony, the source signal is 3.4 kHz bandwidth limited, for wideband telephony it is 7 kHz band limited and for high quality sound, the signal is 15 or 20 kHz band limited.

D.2.3 Stimulus presentation

The stimuli are presented to listeners by pairs (A-B) or repeated pairs (A-B-A-B) where A is the quality reference sample and B the same sample processed by the system under evaluation. The

purpose of the reference sample is to anchor each judgement of the listeners. Some "null pairs" (A-A), at least one for each talker, are included to check the quality of anchoring. Using a reference and subjective judgements with respect to that reference is quite a common procedure in psychoacoustics. It tends to result in a good sensitivity for the overall evaluation by listeners. Samples A and B should be separated by 0.5–1 s. In a repeated pair procedure (A-B-A-B), the separation between the two pairs should be 1–1.5 s.

The order effect observed in a one-sample listening tests (e.g. ACR) is not observed with the DCR procedure. Thus, only one random order of presentation can be used. Therefore the basic test and reference conditions will be eight times (four talkers × two samples) the number of nominal conditions.

D.2.4 Test instructions

The subjects should be instructed to rate the conditions according to the five point degradation category scale as follows:

5 Degradation is inaudible.

4 Degradation is audible but not annoying.

3 Degradation is slightly annoying.

2 Degradation is annoying.

1 Degradation is very annoying.

The quantity evaluated from the scores (degradation mean opinion score) is represented by the symbol DMOS.

D.3 Statistical analysis

Sensitivities can be quantified by means of a statistical multiple comparison test. When an *a posteriori* comparison of circuits is needed a Tukey [15] Honestly Significant Difference (HSD) test can be applied effectively. The HSD test is designed to make all pair-wise comparisons among the means and to determine the significance of the differences in the mean values.

Annex E

Comparison Category Rating (CCR) method

E.1 Introduction

The Comparison Category Rating (CCR) method is similar to the Degradation Category Rating (DCR) method described in Annex D. Listeners are presented with a pair of speech samples on each trial. In the DCR procedure, a reference (unprocessed) sample is presented first, followed by the same speech sample, which has been processed by some technique. In the DCR method, listeners always rate the amount by which the processed (second) sample is *degraded* relative to the unprocessed (first) sample. In the CCR procedure, the order of the processed and unprocessed samples is chosen at random for each trial. On half of the trials, the unprocessed sample is followed by the processed sample. On the remaining trials, the order is reversed. Listeners use the following scale to judge the quality of the second sample relative to that of the first:

The Quality of the Second Compared to the Quality of the First is:

3 Much Better

2 Better

1 Slightly Better

0 About the Same

–1 Slightly Worse

–2 Worse

–3 Much Worse

In effect, listeners provide two judgements with one response: "Which sample has better quality?" and "By how much?" The DCR and the CCR methods are particularly useful for assessing the performance of telecommunications systems when the input has been corrupted by background noise. However, an advantage of the CCR method over the DCR procedure is the possibility to assess speech processing that either degrades or improves the quality of the speech.

The quantity evaluated from the scores (comparison mean opinion score) is represented by the symbol CMOS.

NOTE – Caution should be exercised when using the CCR method. Some laboratories have found the method to be useful in evaluating noise reduction systems. However, when this method was used in the recent subjective evaluations of the G.729 (8 kbit/s) codec, the method was found to be too sensitive when evaluating the performance of the codec for speech embedded in background noise.

E.2 Quality reference

The reference (unprocessed) sample (Quality reference or Direct connection) is presented either before or after the processed or degraded signal. The reference sample is generated using the same talker and speech material as used for the processed sample. This reference sample will be corrupted by the same noise (if any) and processed through the same preliminary processes, such as transmitter characteristic, logarithmic companding, etc. Thus, there will be a different quality reference for each of the test conditions.

E.3 MNRU references

MNRU reference conditions should be included to calibrate the judgement scale. These multiplicative noise references are used without being further mixed with environmental noises.

E.4 Presentation to listeners

Each of the speech samples is presented to the listener through the quality reference condition and through a test codec or reference condition (e.g. Recommendation G.726, MNRU). In addition, a "Null pair" should be included for each of the quality references. On these trials, the quality reference is presented twice.

Listeners should judge the quality of the second sample relative to the quality of the first sample. This judgement is made on the 7-point scale shown in E.1. Sample instructions for the listeners are shown in Table E.1.

E.5 Data analysis

Some care must be exercised when analysing the data from a CCR experiment. As half of the trials for any test condition are presented in the order (unprocessed, processed), and the other half are

presented in the opposite order, simple averaging of the numerical scores should yield a CMOS of approximately 0 for all conditions. It is necessary to recode the raw data. If the order of presentation is (processed, unprocessed), then the sign of the numerical score must be reversed (i.e. $-1 \rightarrow 1$, $-2 \rightarrow 2$, ..., $2 \rightarrow -2$, $1 \rightarrow -1$). The recoded scores may be used to compute CMOS, standard deviations, etc. Thus, results are presented in terms of the (unprocessed, processed) order. Appropriate Analysis of Variance, or other statistical tests, may also be performed on the recoded scores. However, comparison opinion scores may not be presumed to represent a linear interval scale. Therefore, statistics for ordinal scales may need to be applied instead.

TABLE E.1/P.800

Example of instructions to subjects

INSTRUCTIONS TO LISTENERS

Comparison category rating test

"Evaluation of the influence of various environmental noises on the quality of different telephone systems"

In this experiment you will hear pairs of speech samples that have been recorded through various experimental telephone equipment. You will listen to these samples through the telephone handset in front of you.

What you will hear is one pair of sentences, a short period of silence, and another pair of sentences. You will evaluate the quality of the second pair of sentences compared to the quality of the first pair of sentences.

You should listen carefully to each pair of samples. Then, when the green light is on, please record your opinion about the quality of the second sample relative to the quality of the first sample using the following scale:

The Quality of the Second Compared to the Quality of the First is:

3:	Much Better
2:	Better
1:	Slightly Better
0:	About the Same
–1:	Slightly Worse
–2:	Worse
–3:	Much Worse

You will have five seconds to record your answer by pushing the button corresponding to your choice. There will be a short pause before the presentation of next pair of sentences.

We will begin with a short practice session to familiarize you with the test procedure. The actual tests will take place during sessions of 10 to 15 minutes.

Annex F

The threshold method for comparison of transmission systems with a reference system

F.1 Introduction

By direct comparison of a transmission system with a reference system, it is possible to assess the performance of the system under test in terms of a degradation characteristic of the reference system which can be varied and set to defined values. An example of such a characteristic is signal-to-noise ratio (for definition see 8.2.3/P.830), SNR(p). The method described here leads to a threshold of equality defined as 50% preference level between the MNRU and the digital system.

F.2 Testing procedure

A listening-only test procedure is used. A signal pair consisting of a reference signal and a test signal is presented to listeners, who are then asked to indicate which of the signals in the pair they judge to have the highest quality (preference rating). Subjective equivalence is defined as the reference value corresponding to the intersection point of the regression curve of the preference scores at the 50% preference level. An example of equivalent SNR obtained with hypothetical preference scores is shown in Figure F.1.

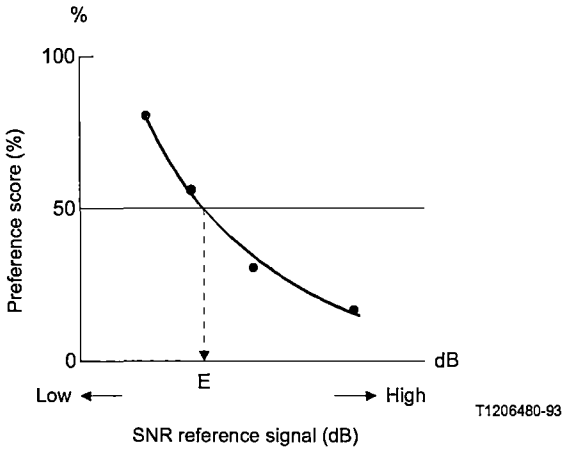


FIGURE F.1/P.800

Example of an equivalence threshold E with hypothetical preference scores

F.3 Presentation of signals

Reference signal A and test signal B are arranged in an equal number of A-B pairs and B-A pairs, and presented in random order. Several degradation levels spaced for example, at 2 dB intervals, are introduced in the reference path so that the range of preference scores extends from 20% to 80%, where the 50% preference lies in the middle of the degradation range. A timing diagram of the presentation is shown in Figure F.2.

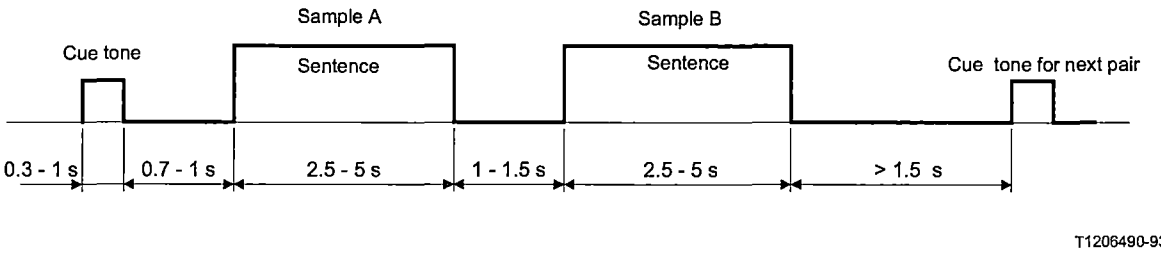


FIGURE F.2/P.800

Timing diagram of the presentation

The subject is required to make a judgement and respond by saying "A is better" or "B is better" (forced choice). The response "A equals B", or "No difference" is forbidden. The duration of the presentation should be limited to about six minutes in order not to tire the listeners. More listening samples may be presented after a suitable rest period. At least two, preferably four or five replications (repetitions of identical presentations) are recommended.

NOTE – If the reference system is available in hardware and the degradation characteristic can easily be changed between presentations, a simplified procedure can be used. In this case the balancing to equally perceived quality is done by the subject. The adjustment is made during the pause between the pairs. The reference signal is always presented first. Presentation continues until the subject reports that the equality threshold has been reached.

F.4 Speech sources

It is necessary to use short sentences spoken by at least two males and two females, preferably four or six of each; different sentences are required for each speaker. The duration should be 2.5-5 seconds for speech and less than 10-15 seconds for music signals. Clicks at the beginning and end of the samples must be avoided. A linear microphone of sufficient bandwidth should be used to record the source signals in a sound-absorbent room having an ambient noise of less than 20 dBA and a reverberation time of less than 0.3 seconds in the band 125 - 8000 Hz. If digital recording equipment is used, the quantizing noise level should be less than the noise level in 14-bit linear PCM.

F.5 Listening environment

A high-fidelity sound reproduction system should be used for the listening test. When listening is carried out with loudspeakers, the reproduction equipment should be studio-quality and the listening room should conform to CCIR Report 797 or IEC 268-13. If headphones are used, diotic (binaural) listening is preferable. The bandwidth shall be at least as wide as that of the system under test.

F.6 Listeners

Although it is preferred that listeners should be selected according to the description in the ACR method (see Annex B), this is not a strict condition in the pair comparison test. If the purpose of the listening test is to obtain the opinions of untrained listeners, untrained subjects are necessary. However, if this is not the purpose of the test, then trained listeners can be used and the reliability of the listening test can be extended by increasing the number of replications for each listener. The minimum number of listeners is six, but should preferably be twelve or more. Several subjects may listen simultaneously but it must be ensured that their responses are obtained independently.

F.7 Reliability

Since variations in preference score in subjective tests are assumed to conform to a t-distribution, the score variation width r which yields 95% reliability at score u ($0 \leq u \leq 1$) over the number (n) of trials (i.e. the number of repetitions for each presentation pair multiplied by the number of subjects and number of source signals) is presented in equation (E-1).

$$r = \pm t(n-1, 0.05) \cdot \sqrt{u(1-u)/(n-1)} \qquad \text{(E-1)}$$

NOTE – The threshold method is expected to give stable and reliable results even for high quality systems with little degradation.

Degradation can be introduced in the reference system, e.g. by addition of white noise. For digital systems, multiplicative noise as defined in Recommendation P.810 (MNRU) is recommended. For

wideband digital speech coders, the use of a wideband MNRU, as described in Recommendation P.810, is recommended. For some purposes shaped noise instead of white may be appropriate.

Bibliography

[1] VOIERS (W.D.): Evaluating processed speech using the Diagnostic Rhyme Test, *Speech Technology*, Volume 1, No. 4, pp. 30-39, January-February 1983.

[2] CCITT Supplement No. 5 to Recommendation P.74, *The SIBYL method of subjective testing, Red Book*, Volume V.

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

[4] HOTH (D.F.): Room noise spectra at subscribers' telephone locations, *J.A.S.A.*, Volume 12, pp. 99-504, April 1941.

[5] CCITT Question 24/XII, Contribution COM XII-120, *Noise inside light motor vehicles*, study period 1981-1984.

[6] CCITT Question 24/XII, Contribution COM XII-134, *Internal vehicle noise spectra*, study period 1981-1984.

[7] CCITT Contribution COM XII-208, *Comparison of the results of vehicle noise submitted by France and BT*, study period 1981-1984.

[8] STEVENS (S.S.): Psychophysics – Introduction to its perceptual, neural and social prospects, *John Wiley and Sons*, 1975.

[9] CLARINGBOLD (P.J.): The within-animal bioassay with quantal responses, *Journal of the Royal Statistical Society,* Series B, Volume 18, No. 1, pp. 133-137, 1956.

[10] RICHARDS (D.L.): Telecommunication by speech, subclause 3.5.2, *Butterworths*, London, 1973.

[11] *Ibid*, subclauses 3.5.3 and 4.5.1.

[12] *Ibid*, subclause 4.2.1.6.

[13] COMBESCURE (P.) *et al*: Quality evaluation of speech coded at 32 kbit/s by means of degradation category ratings, *Proc. ICASSP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, May 1982.

[14] CCIR Document 11/17, *Subjective assessment of the quality of television pictures (EBU)*, study period 1978-1982.

[15] TUKEY (J.W.): The problem of multiple comparisons, *Ditton*, Princeton University, Ed. 1953.

[16] GABRIELSSON (A.): Statistical treatment of data from listening tests on sound-reproducing systems, Report TA No. 92, *KTH Karolinska Institutet*, Department of Technical Audiology, S-10044 Stockholm, Sweden, November 1979.

[17] IEC Publication 268-13, Annex 3, subclause 3.3 (a condensed version of [16]).

ITU-T RECOMMENDATIONS SERIES

Series A	Organization of the work of the ITU-T
Series B	Means of expression
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Telephone network and ISDN
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media
Series H	Transmission of non-telephone signals
Series I	Integrated services digital network
Series J	Transmission of sound-programme and television signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	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**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminal equipment and protocols for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communication
Series Z	Programming languages

International Telecommunication Union

ITU-T P.800.1

(07/2016)

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Methods for objective and subjective assessment of speech and video quality

Mean opinion score (MOS) terminology

Recommendation ITU-T P.800.1

International Telecommunication Union

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30
		P.300
Reference systems	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 speech quality	Series	P.80
Methods for objective and subjective assessment of speech and video quality	**Series**	**P.800**
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100
Models and tools for quality assessment of streamed media	Series	P.1200
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400
Methods for objective and subjective assessment of quality of services other than speech and video	Series	P.1500

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

Recommendation ITU-T P.800.1

Mean opinion score (MOS) terminology

Summary

Recommendation ITU-T P.800.1 provides a terminology which shall be used in conjunction with audio, video and audiovisual quality expressions in terms of mean opinion score (MOS). This terminology is motivated by the intention to avoid misunderstanding as to whether specific values of MOS for audio are related to listening quality, talking quality or conversational quality, whether they originate from subjective tests, from objective models or from network planning models, as well as extending the concept to video and audiovisual quality. Furthermore, this Recommendation provides identifiers regarding the audio bandwidth, the type of interface (electrical or acoustical) and the video resolution. Further guidance on the interpretation of MOS is given in Recommendation ITU-T P.800.2.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.800.1	2003-03-16	12	11.1002/1000/6257
2.0	ITU-T P.800.1	2006-07-14	12	11.1002/1000/8860
3.0	ITU-T P.800.1	2016-02-29	12	11.1002/1000/12749
4.0	ITU-T P.800.1	2016-07-29	12	11.1002/1000/12972

Keywords

Audio, audio-visual, MOS, notation, video.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). 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 Assembly (WTSA), 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 WTSA 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.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

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, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T P.800.1

Mean opinion score (MOS) terminology

1 Scope

This Recommendation provides a terminology which shall be used in conjunction with audio, video and audio-visual quality expressions in terms of mean opinion score (MOS).

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; 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. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T G.107] Recommendation ITU-T G.107 (2015), *The E-model, a computational model for use in transmission planning.*

[ITU-T G.113 Ap.I] Recommendation ITU-T G.113 (2007), Appendix I, *Provisional planning values for the equipment impairment factor, Ie, and packet-loss robustness factor, Bpl.*

[ITU-T G.1070] Recommendation ITU-T G.1070 (2012), *Opinion model for video-telephony applications*

[ITU-T G.1071] Recommendation ITU-T G.1071 (2015), *Opinion model for network planning of video and audio streaming applications.*

[ITU-T J.140] Recommendation ITU-T J.140 (1998), *Subjective picture quality assessment for digital cable television systems.*

[ITU-T J.144] Recommendation ITU-T J.144 (2004), *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.*

[ITU-T J.247] Recommendation ITU-T J.247 (2008), *Objective perceptual multimedia video quality measurement in the presence of a full reference.*

[ITU-T J.341] Recommendation ITU-T J.341 (2016), *Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference.*

[ITU-T J.343] Recommendation ITU-T J.343 (2014), *Hybrid perceptual bitstream models for objective video quality measurements.*

[ITU-T P.10] Recommendation ITU-T P.10/G.100 (2006), *Vocabulary for performance and quality of service.*

[ITU-T P.562] Recommendation ITU-T P.562 (2004), *Analysis and interpretation of INMD voice-service measurements.*

[ITU-T P.563] Recommendation ITU-T P.563 ((2004), *Single-ended method for objective speech quality assessment in narrow-band telephony applications.*

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

[ITU-T P.800.2] Recommendation ITU-T P.800.2 (2016), *Mean opinion score interpretation and reporting.*

[ITU-T P.830] Recommendation ITU-T P.830 (1996), *Subjective performance assessment of telephone-band and wideband digital codecs.*

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

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

[ITU-T P.833] Recommendation ITU-T P.833 (2001), *Methodology for the derivation of equipment impairment factors from subjective listening-only tests.*

[ITU-T P.834] Recommendation ITU-T P.834 (2015), *Methodology for the derivation of equipment impairment factors from instrumental models.*

[ITU-T P.835] Recommendation ITU-T P.835 (2003), *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm.*

[ITU-T P.840] Recommendation ITU-T P.840 (2003), *Subjective listening test method for evaluating circuit multiplication equipment.*

[ITU-T P.862] Recommendation ITU-T P.862 (2001), *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.*

[ITU-T P.862.1] Recommendation ITU-T P.862.1 (2003), *Mapping function for transforming P.862 raw result scores to MOS-LQO.*

[ITU-T P.862.2] Recommendation ITU-T P.862.2 (2007), *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs.*

[ITU-T P.863] Recommendation ITU-T P.863 (2014), *Perceptual objective listening quality assessment.*

[ITU-T P.910] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications.*

[ITU-T P.911] Recommendation ITU-T P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications.*

[ITU-T P.912] Recommendation ITU-T P.912 (2016), *Subjective video quality assessment methods for recognition tasks.*

[ITU-T P.913] Recommendation ITU-T P.913 (2016), *Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment.*

[ITU-T P.920] Recommendation ITU-T P.920 (2000), *Interactive test methods for audiovisual communications.*

ITU-T P.1201] Recommendation ITU-T P.1201 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality.*

[ITU-T P.1202] Recommendation ITU-T P.1202 (2012), *Parametric non-intrusive bitstream assessment of video media streaming quality.*

[ITU-T P.1301] Recommendation ITU-T P.1301 (2012), *Subjective quality evaluation of audio and audiovisual multiparty telemeetings.*

[ITU-R BT.500] Recommendation ITU-R BT.500 (2012), *Methodology for the subjective assessment of the quality of television pictures.*

[ITU-R BT.709-6] Recommendation ITU-R BT.709-6 (2015), *Parameter values for the HDTV standards for production and international programme exchange.*

[ITU-R BT.2020-2] Recommendation ITU-R BT.2020-2 (2015), *Parameter values for ultra-high definition television systems for production and international programme exchange.*

3 Definitions

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACR Absolute Category Rating

HD High Definition

IRS Intermediate Reference System

MOS Mean Opinion Score

SD Standard Definition

UHD Ultra High Definition

5 Conventions

In this Recommendation the following conventions are used:

MOS-AVQE is used to refer to MOS estimated audiovisual quality;

MOS-AVQO is used to refer to MOS objective audiovisual quality;

MOS-AVQS is used to refer to MOS subjective audiovisual quality;

MOS-CQE is used to refer to MOS estimated conversational quality;

MOS-CQO is used to refer to MOS objective conversational quality;

MOS-CQS is used to refer to MOS subjective conversational quality;

MOS-LQE is used to refer to MOS estimated listening quality;

MOS-LQO is used to refer to MOS objective listening quality;

MOS-LQS is used to refer to MOS subjective listening quality;

MOS-TQE is used to refer to MOS estimated talking quality;

MOS-TQO is used to refer to MOS objective talking quality;

MOS-TQS is used to refer to MOS subjective talking quality;

MOS-VQE is used to refer to MOS estimated video quality;

MOS-VQO is used to refer to MOS objective video quality;

MOS-VQS is used to refer to MOS subjective video quality.

6 Limitations and important notes

This Recommendation applies generic abbreviations to simplify the reporting of MOS scores. While these abbreviations are useful for giving a quick overview of the context under which a reported MOS was generated, these abbreviations are insufficient for completely describing the context and can by no means replace a detailed description as recommended in [ITU-T P.800.2].

It is well known that users rate quality differently depending on the rating context and their expectations. This leads to different scoring of, e.g., audio/video quality for entertainment services (e.g., OTT video streaming), or audio/video quality during a teleconference. Comparing MOS-VQO results from one application scenario with MOS-VQO obtained for another scenario is therefore considered invalid, even though both values may still be reported as MOS-VQO. Again, full guidance on how to report such values is given in [ITU-T P.800.2].

Similarly, this Recommendation describes MOS suffixes to specify viewing devices. These suffixes only set the context to a more mobile or more TV specific environment since this has the strongest impact on the subjects' behavior. However, there are other factors such as screen size, screen resolution and viewing distance, which should be reported together with MOS scores. Trying to describe the complete MOS context with suffixes is impractical and hence is not recommended here. Again, [ITU-T P.800.2] should be followed instead.

7 Recommended MOS terminology for audio

Mean opinion score (MOS) is defined in [ITU-T P.10] as "The mean of opinion scores" where [ITU-T P.10] further defines opinion score as "The value on a predefined scale that a subject assigns to his opinion of the performance of the telephone transmission system used either for conversation or for listening to spoken material."

Apart from subjective opinion, the abbreviation MOS is also used for scores that originate from objective models or network planning models. The following identifiers, listed in Table 1, are recommended to be used together with the abbreviation MOS in order to distinguish the area of application, where N refers to 'narrow-band', W refers to 'wideband', LQ refers to 'listening quality', CQ refers to 'conversational quality', TQ refers to talking quality, S refers to 'subjective', O refers to 'objective', and E refers to 'estimated'.

For the purposes of this Recommendation, the term 'Audio' for this clause primarily relates to speech quality. Care should be taken that general audio signals, such as music or mixed speech and music, are not used with objective models with a scope limited to speech only, for example ITU-T P.863. The type of audio used, e.g., speech, music or mixed speech and music, should also be indicated when reporting a MOS value.

Please note that apart from the audio bandwidth, other factors like level, application or listening device and environment also have an impact on the absolute MOS value and should be reported according to [ITU-T P.800.2].

Table 1 – Audio identifiers

	Listening-only	Conversational	Talking
Subjective	MOS-LQSy	MOS-CQSy	MOS-TQSy
Objective	MOS-LQOy	MOS-CQOy	MOS-TQOy
Estimated	MOS-LQEy	MOS-CQEy	MOS-TQEy

NOTE – The letter "y" at the end of above acronyms is a placeholder for the descriptor of the respective audio bandwidth, see the following provisional instructions.

– N for MOS scores obtained for audio signals up to narrow-band (300-3400 Hz) relative to a narrow band high quality reference. This is applicable for instance to narrow-band only subjective tests or to [ITU-T P.862.1] or [ITU-T P.863] (narrow-band operational mode) scores.
– W for MOS scores obtained for audio signals up to wideband (50-7000 Hz) relative to a wideband high quality reference. This is applicable for instance to wideband only subjective tests or to ITU-T P.862.2 scores.
– S for MOS scores obtained for audio signals up to super-wideband (20-14000 Hz) relative to a super-wideband high quality reference. This is applicable for instance to super-wideband only subjective tests or to ITU-T P.863 scores.
– F for MOS scores obtained for audio signals up to fullband (10-20000 Hz) relative to a fullband high quality reference.

In cases where the bandwidth denominators N, W, S or F do not properly reflect the actual situation, it is suggested that provisionally the placeholder "y" be replaced by a proper notation.

7.1 MOS related to listening-only situations

These MOS scores are applicable to a listening-only situation. Three different cases have to be distinguished.

7.1.1 MOS-LQS

The score has been collected in a laboratory test by calculating the arithmetic mean value of subjective judgments on a 5-point absolute category rating (ACR) quality scale, as it is defined in [ITU-T P.800]. Subjective tests carried out according to [ITU-T P.830], [ITU-T P.835] and [ITU-T P.840] give results in terms of MOS-LQS.

7.1.2 MOS-LQO

The score is calculated by means of an objective model which aims at predicting the quality for a listening-only test situation. Objective measurements made using the models given in [ITU-T P.862.1], [ITU-T P.862.2] and [ITU-T P.863] give results in terms of MOS-LQO.

In this context, it should be noted that there is a necessary distinction between intrusive end-to-end measurements of listening quality aiming at:

- **MOS-LQO (electrical)**

  This kind of measurement is performed at electrical interfaces only. In order to predict the listening quality as perceived by the user, assumptions for the terminals are made in terms of intermediate reference system (IRS) or corrected IRS frequency response; this implicitly includes the assumption of a sealed condition between the handset receiver and the user's ear. [ITU-T P.862] falls into this category.

- **MOS-LQO (acoustical)**

  This kind of measurement is performed at acoustical interfaces. In order to predict the listening quality as perceived by the user, this measurement includes the actual telephone set products provided by the manufacturer or vendor. In combination with the choice of the acoustical receiver in the laboratory test ("artificial ear"), there will be a more or less leaky condition between the handset's receiver and the artificial ear. Consequently, for more realistic test scenarios, there may be a degradation of the measured MOS value, while for more artificial test scenarios there may be a negligible difference.

7.1.3 MOS-LQE

The score is calculated by a network planning model which aims at predicting the quality in a listening-only application situation.

7.2 MOS related to conversational situations

These MOS scores are applicable to a conversational situation. Three different cases have to be distinguished.

7.2.1 MOS-CQS

The score has been collected in a laboratory test by calculating the arithmetic mean value of subjective judgments on a 5-point ACR quality scale, as it is defined in [ITU-T P.800]. Subjective conversation tests carried out according to [ITU-T P.800], [ITU-T P.831] and [ITU-T P.832] give results in terms of MOS-CQS.

7.2.2 MOS-CQO

The score is calculated by means of an objective model which aims at predicting the quality for a conversational test situation. Objective measurements made using the model given in [ITU-T P.562] and [ITU-T P.563] give results in terms of MOS-CQO.

7.2.3 MOS-CQE

The score is calculated by a network planning model which aims at predicting the quality in a conversational application situation. Estimates of conversational quality carried out according to [ITU-T G.107], when transformed to mean opinion score, give results in terms of MOS-CQE.

7.3 MOS related to talking situations

Talking quality describes the quality of a telephone call as it is perceived by the talking party only. Talking quality will be mainly affected by the annoyance of the echo signal and effects like background noise switching and double talk.

7.3.1 MOS-TQS

The score has been collected in a laboratory test by calculating the arithmetic mean value of subjective judgments on a 5-point ACR quality scale, as it is defined in [ITU-T P.800].

7.3.2 MOS-TQO

The score is calculated by means of an objective model which aims at predicting the quality for a talking-only test situation. Methods generating a MOS-TQO are currently under development and not yet standardized.

7.3.3 MOS-TQE

The score is calculated by a network planning model which aims at predicting the quality in a talking-only application situation. No methods generating a MOS-TQE are currently standardized.

7.4 Relationship between some audio MOS qualifiers

Figure 1 gives an overview of the relation between MOS-LQS, MOS-LQO and MOS-LQE.

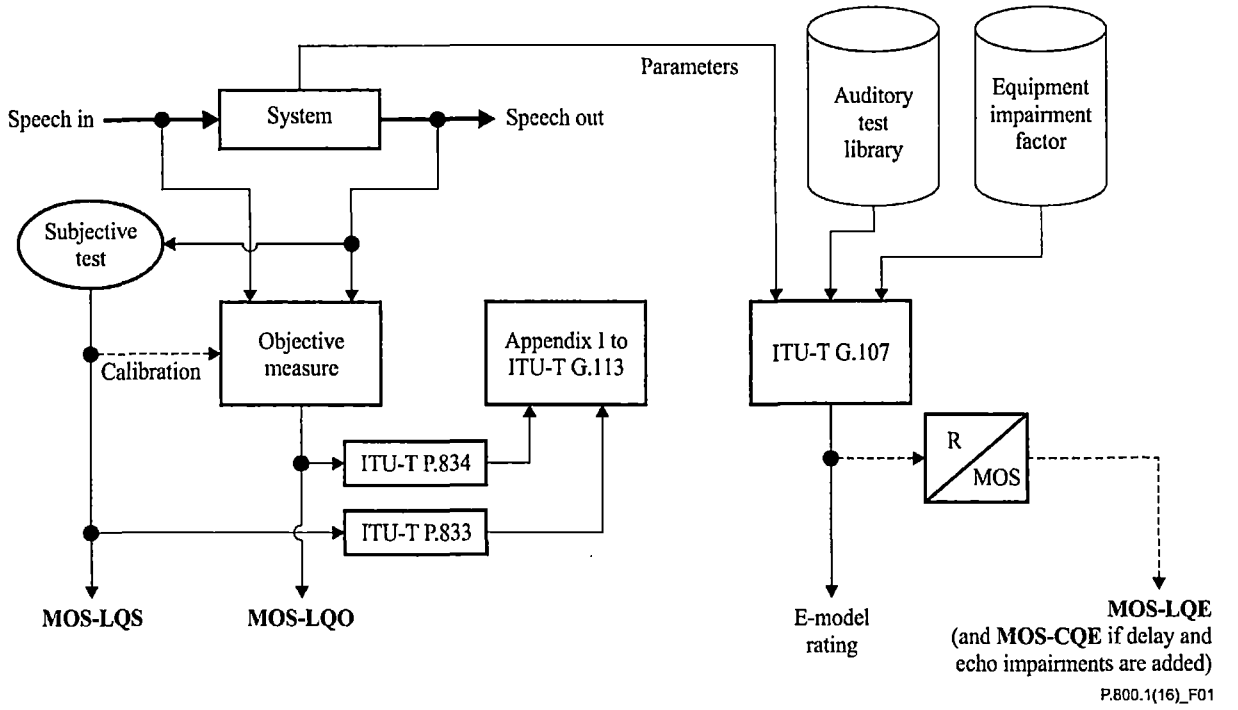


Figure 1 – Relationship between some audio MOS qualifiers

8 Recommended MOS terminology for video

The following identifiers, listed in Table 2, are recommended to be used together with the abbreviation MOS in order to distinguish the area of application of MOS for video, where VQ refers to 'video quality', S refers to 'subjective', O refers to 'objective' and E refers to 'estimated'. As noted in [ITU-T P.800.2], display resolution, display size, viewing device as well as the application also have an influence on the absolute value of video MOS and should be reported as well.

Table 2 – Video identifiers

	Video
Subjective	MOS-VQSz
Objective	MOS-VQOz
Estimated	MOS-VQEz

NOTE – The letter "z" at the end of above acronyms is a placeholder for the descriptor of the respective viewing context, see the following provisional instructions.

To account for differentiation in perceived quality for mobile and fixed devices and to allow for proper handling of different use-cases, the suffix z can be used, where z can be

- M for MOS scores obtained for mobile screen such as a smartphone or tablet (approximately 25 cm or less)
- T for MOS scores obtained for PC/TV monitors

In cases where the resolution denominators M or T do not properly reflect the actual situation, it is suggested that provisionally the placeholder "z" be replaced by a proper notation.

8.1 MOS-VQS

The score has been collected in a laboratory test by calculating the arithmetic mean value of subjective judgments, typically on a 5-point quality scale. Subjective tests carried out according to [ITU-T P.910], [ITU-T P.912], [ITU-T P.913], [ITU-T P.920], [ITU-T J.140] and [ITU-R BT.500] give results in terms of MOS-VQS.

8.2 MOS-VQO

MOS-VQO refers to objective video quality, which represents the MOS score obtained from an algorithmic quality evaluation model. The evaluation model uses real-time objective metrics that can be obtained from information carried in the video streams and corresponding networks. [ITU-T J.144], [ITU-T J.247], [ITU-T J.341], [ITU-T J.343], [ITU-T P.1201], [ITU-T P.1202] are examples of such models.

8.3 MOS-VQE

The score is calculated by a network planning model which aims at predicting the quality in a video application situation, based on non real-time parameters. [ITU-T G.1070] and [ITU-T G.1071] are examples of such models.

8.4 Relationship between some video MOS qualifiers

Figure 2 gives an overview of the relation between MOS-VQS, MOS-VQO and MOS-VQE.

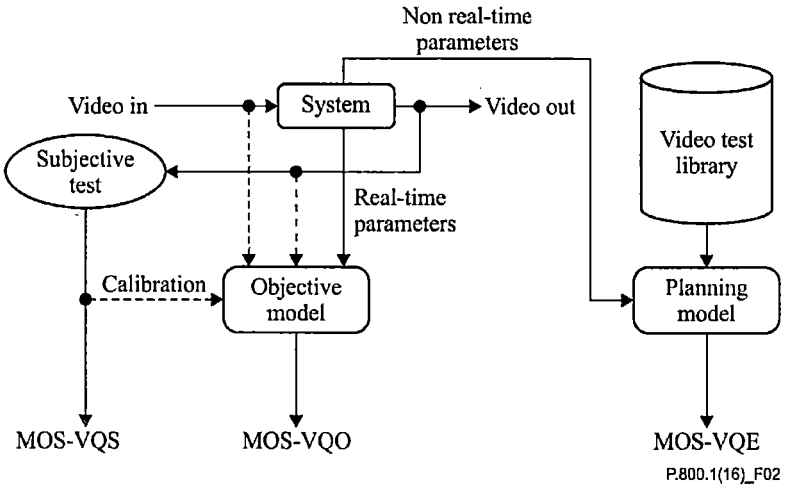


Figure 2 – Relationship between some video MOS qualifiers

9 Recommended MOS terminology for audiovisual services

The following identifiers, listed in Table 3, are recommended to be used together with the abbreviation MOS in order to distinguish the area of application of MOS for audiovisual services, where AVQ refers to 'audiovisual quality', i.e., a single score taking into account the combination of audio and video quality. S refers to 'subjective', O refers to 'objective' and E refers to 'estimated'.

Table 3 – Audiovisual identifiers

	Video
Subjective	MOS-AVQSyz
Objective	MOS-AVQOyz
Estimated	MOS-AVQEyz

NOTE 1 – The letter "y" at the end of above acronyms is a placeholder for the descriptor of the respective audio bandwidth in the audiovisual combination, see the following provisional instructions.

- N for MOS scores obtained for audio signals up to narrow-band (300-3400 Hz) relative to a narrow band high quality reference. This is applicable for instance to narrow-band only subjective tests or to ITU-T P.862.1 or ITU-T P.863 (narrow-band operational mode) scores.
- W for MOS scores obtained for audio signals up to wideband (50-7000 Hz) relative to a wideband high quality reference. This is applicable for instance to wideband only subjective tests or to ITU-T P.862.2 scores.
- S for MOS scores obtained for audio signals up to super-wideband (20-14000 Hz) relative to a super-wideband high quality reference. This is applicable for instance to super-wideband only subjective tests or to ITU-T P.863 scores.
- F for MOS scores obtained for audio signals up to fullband (10-20000 Hz) relative to a fullband high quality reference.

In cases where the bandwidth denominators N, W, S or F do not properly reflect the actual situation, it is suggested that provisionally the placeholder "y" be replaced by a proper notation.

NOTE 2 – The letter "z" at the end of above acronyms is a placeholder for the descriptor of the respective viewing context, see the following provisional instructions.

To account for differentiation in perceived quality for mobile and fixed devices and to allow for proper handling of different use-cases, the suffix z can be used, where z can be

- M for MOS scores obtained for mobile screen such as a smartphone or tablet (approximately 25 cm or less)
- T for MOS scores obtained for PC/TV monitors

In cases where the resolution denominators M or T do not properly reflect the actual situation, it is suggested that provisionally the placeholder "z" be replaced by a proper notation.

9.1 MOS-AVQS

The score has been collected in a laboratory test by calculating the arithmetic mean value of subjective judgments, typically on a 5-point quality scale. Subjective tests carried out according to [ITU-T P.911], [ITU-T P.913], [ITU-T P.920] and [ITU-T P.1301] give results in terms of MOS-AVQS.

9.2 MOS-AVQO

MOS-AVQO refers to objective audiovisual quality, which represents the MOS score obtained from an algorithmic quality evaluation model. The evaluation model uses real-time objective metrics that can be obtained from information carried in the audiovisual streams and corresponding networks.. [ITU-T P.1201] is an example of such a model.

9.3 MOS-AVQE

The score is calculated by a network planning model which aims at predicting the quality in an audiovisual application situation, based on non real-time parameters. [ITU-T G.1070] is an example of a model providing a two-way MOS-AVQE score, while [ITU-T G.1071] is an example of a model giving a one-way MOS-AVQE score.

9.4 Relationship between some audiovisual MOS qualifiers

Figure 3 gives an overview of the relation between MOS-AVQS, MOS-AVQO and MOS-AVQE.

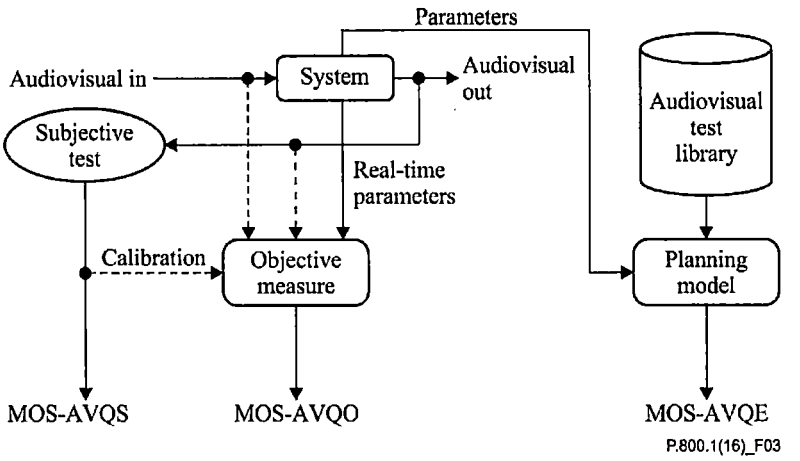


Figure 3 – Overview of the relation between MOS-AVQS, MOS-AVQO and MOS-AVQE

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
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	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Terminals and subjective and objective assessment methods**
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, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2016