ETSI TR 101 329-6 V1.1.1 (2000-07)

Technical Report

Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); End to End Quality of Service in TIPHON Systems; Part 6: Actual measurements of network and terminal characteristics and performance parameters in TIPHON networks and their influence on voice quality

Reference
DTR/TIPHON-05004

Keywords
IP, network, performance, quality, terminal, voice

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at http://www.etsi.org/tb/status/

If you find errors in the present document, send your comment to:
editor@etsi.fr

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://www.etsi.org/ipr).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Project Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON).

The present document forms one of a series of technical reports and technical specifications by Working Group 5 for TIPHON Quality of Service (QoS) classification. The structure of this work is illustrated in figure 1. The structure of the present document is very close to the structure of the TS 101 329-5 [5]. Therefore in the present document the measurement results are sorted according to main measurement methods, which are explained in TS 101 329-5 [5].

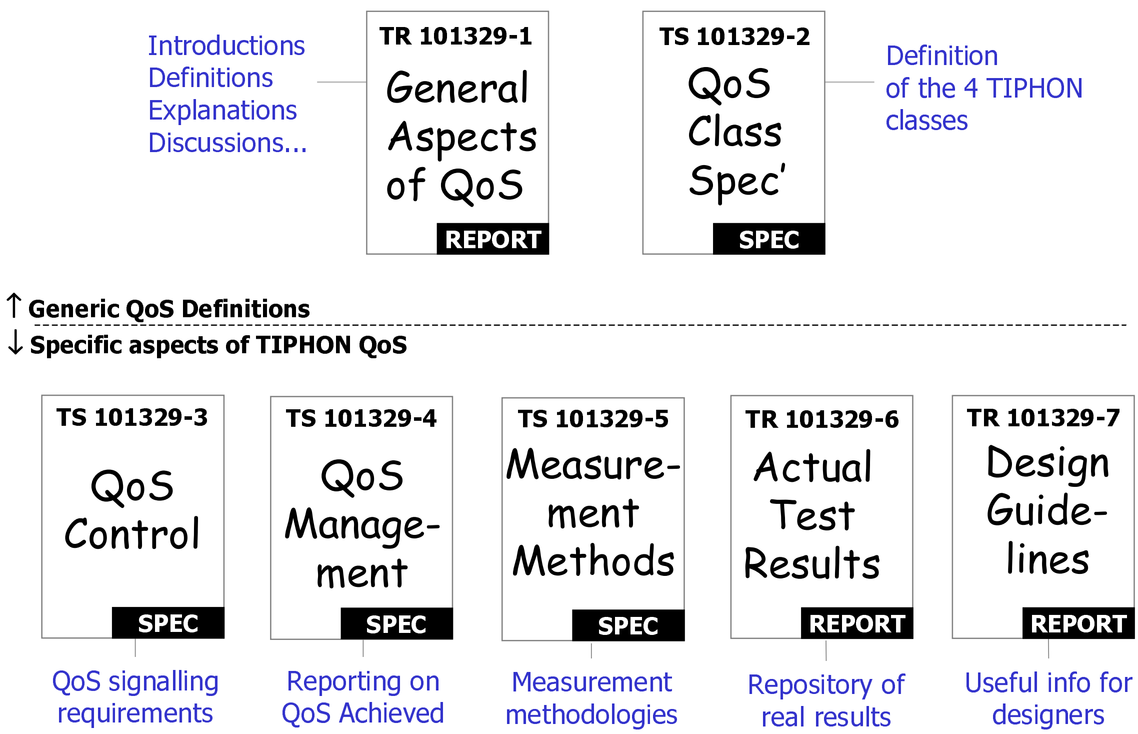


Figure 1: Structure of TIPHON QoS Documentation

1 Scope

The present document applies to IP networks that provide voice telephony in accordance with any of the TIPHON Scenarios.

The objective with the present document is to collect all results of various VoIP speech transmission quality tests and related information. This collection should be used for information and to review and discuss the values of the TIPHON QoS classes which are described in WG5 documents TR 101 329-1 [3] and TR 101 329-7 [6].

The separate measurements should give a very good opportunity to understand the goal of the measurement itself and the exact measurement set-up conditions to understand under which framework the measurements were done.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.

[1] ETSI ETR 275 (1996): "Transmission and Multiplexing (TM); Considerations on transmission delay and transmission delay values for components on connections supporting speech communication over evolving digital networks".

[2] ETSI TR 101 329 (V2.1): "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); General aspects of Quality of Service (QoS)".

[3] ETSI TR 101 329-1: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); End to End Quality of Service in TIPHON Systems; Part 1: General aspects of Quality of Service (QoS)".

[4] ETSI TR 101 329-2: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); End to End Quality of Service in TIPHON Systems; Part 2: Definition of Quality of Service (QoS) Classes".

[5] ETSI TR 101 329-5: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); End to End Quality of Service in TIPHON Systems; Part 5: Quality of Service (QoS) Measurement Methodologies in TIPHON Systems".

[6] ETSI TR 101 329-7: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); End to End Quality of Service in TIPHON Systems; Part 7: Design Guide for elements of a TIPHON connection from an end-to-end speech transmission performance point of view".

[7] ETSI ES 201 168 (V1.1): "Corporate telecommunication Networks (CN); Transmission characteristics of digital Private Branch eXchanges (PBWs)".

[8] GTS 06.10 (V3.2): "European digital cellular telecommunications system (Phase 1); GSM Full Rate Speech Transcoding (GSM 06.10)".

[9] ITU–T Recommendation G.107: "The E-model, a computational model for use in transmission planning".

[10] ITU–T Recommendation G.108: "Guidance for assessing conversational speech transmission quality effects not vocered by the E-Model".

[11] ITU–T Recommendation G.109 (09/99): "Definition of categories of speech transmission quality".

[12] ITU–T Recommendation G.113 (1996): "Transmission impairments".

[13] ITU–T Recommendation G.131 (1996): "Control of talker echo".

[14] ITU–T Recommendation G.165 (03/93): "Echo cancellers".

[15] ITU–T Recommendation G.168 (1997): "Digital network echo cancellers".

[16] ITU–T Recommendation G.711 (1988): "Pulse code modulation (PCM) of voice frequencies".

[17] ITU–T Recommendation G.721 (1988): "32 kbit/s adaptive differential pulse code modulation (ADPCM)".

[18] ITU–T Recommendation G.723.1 (03/96): "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s".

[19] ITU–T Recommendation G.726 (1990): "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)".

[20] ITU–T Recommendation G.727 (1990): "5-, 4-, 3- and 2-bits per sample embedded adaptive differential pulse code modulation (ADPCM)".

[21] ITU–T Recommendation G.728 (09/92): "Coding of speech at 16 kbit/s using low-delay code excited linear prediction".

[22] ITU–T Recommendation G.729 (1996): "C source code and test vectors for implementation verification of the G.729 8 kbit/s CS-ACELP speech coder".

[23] ITU–T Recommendation G.729A (Annex A 11/96): "C source code and test vectors for implementation verification of the G.729 reduced complexity 8 kbit/s CS-ACELP speech coder".

[24] ITU–T Recommendation G.729B (Annex B 10/96): "A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70".

[25] ITU–T Recommendation H.323 (1998): "Packet-based multimedia communications systems".

[26] ITU-T Recommendation P.57: "Artificial ears".

[27] ITU-T Recommendation P.58: "Head and torso simulator for telephonometry".

[28] ITU–T Recommendation P.64 (1997): "Determination of sensitivity/frequency characteristics of local telephone systems".

[29] ITU-T Recommendation P.501: "Test signals for use in telephonometry".

[30] ITU-T Recommendation P.502: "Objective test methods for speech communication systems, using complex test signals".

[31] ITU-T Recommendation P.581: "Use of head and torso simulator for hands-free terminal testing".

[32] ITU–T Recommendation P.800 (1996): "Methods for subjective determination of transmission quality".

[33] ITU–T Recommendation P.861 (1998): "Objective quality measurement of telephone-band (300-3 400 Hz) speech codecs".

[34] ITU–T Recommendation Q.13/12: "Rapporteur of Question 13 inside ITU-T Study Group 12; ETSI TIPHON 17TD135".

[35] SG 16, Santiago, Chile, 17-28 May 1999; D.249 (WP 3/16): A High Quality Low-Complexity Algorithm For Frame Erasure Concealment (FEC) With G.711. (Source: AT&T).

[36] T1A1.7/99-012r3; Jul-28-1999; Draft T1 Standard "American National Standard For A Packet Loss Concealment Technique For Use With ITU-T Recommendation G.711". (Source: AT&T).

[37] GSM 06.20: "European digital cellular telecommunications system (Phase 1); Half Rate Speech Transcoding GSM 06.20 - phase 2".

[38] ETSI ETS 300 726: "Digital cellular telecommunications system (Phase 2+); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60)".

3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ACR Absolute Category Rating
ASL Active Speech input Level
CAS Communication Analysis System
GSM Global System for Mobile communications
GSM FR GSM Full Rate Speech Coder
GSM EFR GSM Enhanced Full Rate Speech Coder
ISDN Integrated Services Digital Network
IP Internet Protocol
LAN Local Area Network
MOS Mean Opinion Score
PLC Packet Loss Concealment
PSTN Public Switched Telephone Network
QoS Quality of Service
SCN Switched Communications Network
VAD Voice Activity Detection
VoIP Voice over IP

4 List of Measurement Results

Nr.	Document	Source	Document Introduction	Date
1	Simulation Results of VoIP scenarios	Deutsche Telekom Berkom t.scheerbarth@berkom.de; i.kliche@berkom.de	ETSI TIPHON 11TD064	11/01/1999
2	APPENDIX I (to ITU-T Recommendation G.113 [12] - Transmission impairments)	ITU-T SG 12 Mark E. Perkins mperkins@att.com	ETSI TIPHON 11TD084	11/01/1999
3	Speech Quality Test results of IP equipment in a LAN environment	Robert Bosch GmbH Joachim.Pomy@Tenovis.com	ETSI TIPHON 14TD081	16/07/1999
4	QoS Measurements of IP-Configurations	HEAD acoustics, Robert Bosch GmbH, T-Nova (Deutsche Telekom) h.w.gierlich@head-acoustics.de	ETSI TIPHON 15 TD089	05/10/1999
5	Subjective Results on impairment effects of IP packet loss	Nortel Networks paulcov@nortelnetworks.com	ETSI TIPHON 17 TD167	14/03/2000
6	Subjective and Objective Speech Quality Evaluation on Speech Data recorded at the SuperOp 99 event in Hawaii	Rapporteur of ITU-T Recommendation Q.13/12	ETSI TIPHON 17 TD135	15/03/2000

5 General Measurement Results

5.1 Call Set-up

5.2 Subjective Testing

5.2.1 Simulation Results of VoIP Scenarios

Source: Deutsche Telekom Berkom; Simulation Results of VoIP Scenarios; ETSI TIPHON 11TD064.

5.2.1.1 Introduction

ETSI TIPHON WG 5 has defined a methodology for testing VoIP End-to-End speech quality. This methodology was used as a basis model for the T Berkom simulation processing. Figure 2 shows the methodology used for simulation.

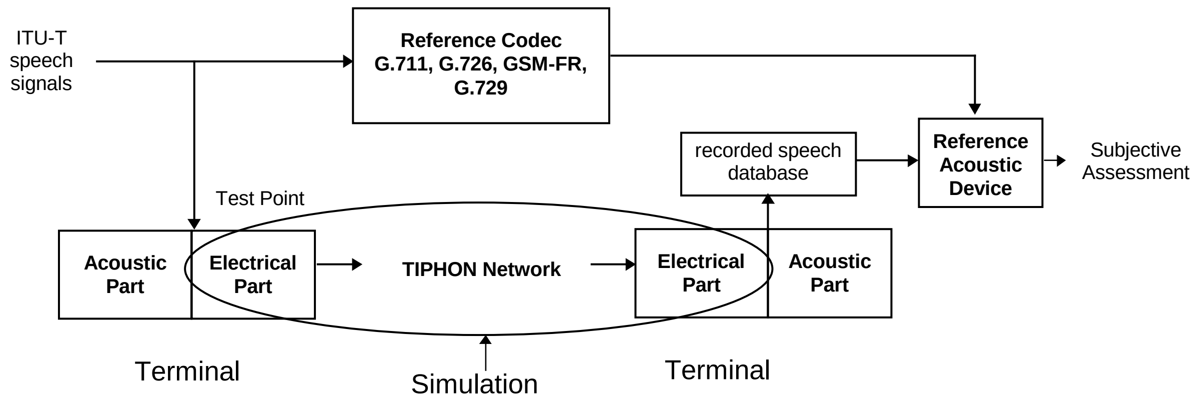


Figure 2: Simulation methodology for testing TIPHON speech quality

A set of speech signals designed according to ITU-T Recommendation P.800 [32] was used as input of the simulation path. The simulation path includes the terminal side (electrical part) and the network itself. The influence of the terminal side was focussed to the speech conversion and IP packet size issue. The influence of the network side was simulated by different packet loss rates.

After the simulation the speech samples were recorded and stored in a database.

The subjective assessment was carried out according to the ITU-T Recommendation P.800 [32] method.

5.2.1.2 Measurement Set-up

5.2.1.2.1 Basics

All source speech samples consisted of German sentences spoken by four talkers (2 male, 2 female). The input level taken for all scenarios was ASL= -26dB Ovl. In the pre- and post-processing phase the speech samples were filtered with the modified IRS transmit and receive filter.

For every test condition the speech file was encoded and then assembled in IP packets. These IP packets were assembled with different lengths, according to the concerned speech frame number per packet.

For simulation of network influences in the case of packet loss, a common channel model was designed, realized by channel files which describe the network condition with the same time resolution as the source speech sample rate. So the network has a certain condition (good or bad) for every speech sample (every 125 µs), two adjacent network states were considered as statistically independent because the network speed was assumed to be much higher than the sample rate (8000 samples per second). So for each packet loss rate one channel file was created using a random generator. The length of this channel file was exactly the same as the length of the speech file.

In a further step the speech file, assembled in IP packets, was matched to the channel file. According to the length of the IP packet (10ms, 20ms,...) the channel file was checked every time when a packet was ready to send. That means if the packet size was 10 ms the channel file was checked also every 10 ms if the condition is good or bad. In a bad case the IP packet was lost, otherwise it was further processed.

This information (IP packet lost or not) was stored in a description file which was the input of the re-assembler and speech decoder.

5.2.1.2.2 Test Cases

The test cases consisted a group of single codec scenarios (references), phone to phone scenarios in fixed network environments and a group of tandeming conditions. The tandeming conditions based on real scenarios where a mobile customer is connected to an ordinary telephone via IP. For this cases the GSM Full Rate Codec (GSM-FR) and the GSM Enhanced Full Rate Codec (EFR) were used. In such scenarios mainly the influence of the IP network was taken into account. Only one condition was chosen to simulate a voice transmission from a mobile phone to an ordinary telephone via an impaired radio channel and via an IP network with packet loss.

G.711 [16] + Codec + G.711 [16] (Phone to Phone Scenario in fixed network environments).

Codec	Packet Loss	Speech Frame Size	Nr. of Frames/Packet	Substitution
Single Codec				
G.711 [16]	5%, 10%, 15%, 20%	0,125 ms	80, 320, 480, 800	Silence
G.729B [24]	5%	10 ms	1, 4, 6, 10	G.729 [22] internal
G.723.1 [18] (5.3)	0%, 5%, 10%, 15%, 20%	30 ms	1, 2, 3	G.723.1 [18] internal
G.728 [21]	0%, 5%, 10%, 15%, 20%	0,625 ms	16, 64, 96, 160	proprietary

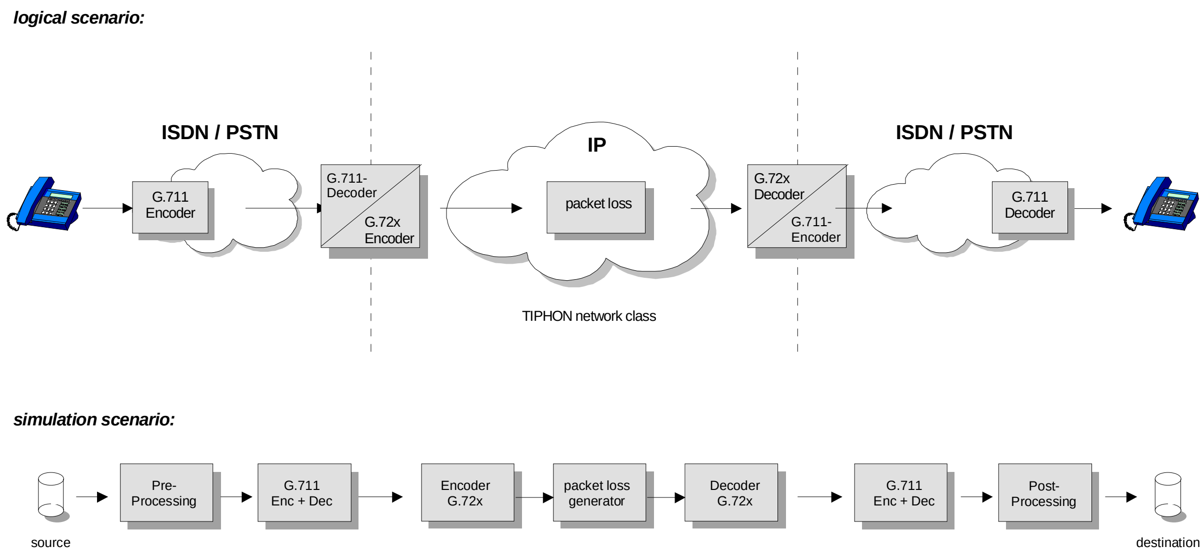
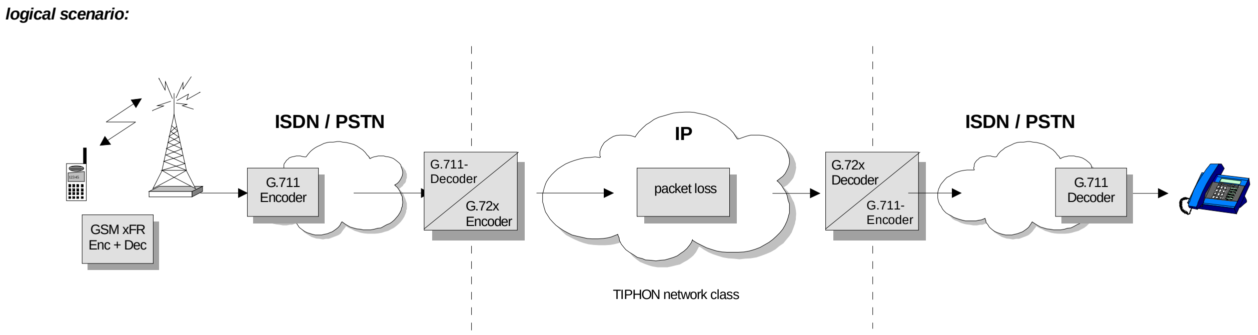


Figure 3: Processing scenario

Tandem Configuration with GSMxFR (Phone to Phone Scenario in including mobile networks)

Tandem with GSM-FR	Packet Loss	Speech Frame Size	Nr. of Frames/ Packet	Substitution
GSM FR + G.723.1 [18] (5.3)	0%, 5%, 10%, 15%	G.723.1 [18]: 30 ms	2	G.723.1 [18] internal
GSM FR + G.729B [24]	0%, 5%, 10%, 15%	G.729 [22]: 10 ms	6	G.729 [22] internal
GSM EFR + G.723.1(5.3 [18])	0%, 5%, 10%	G.723.1 [18]: 30 ms	2	G.723.1 [18] -internal
GSM EFR + G.729B [24]	0%, 5%, 10%	G.729 [22]: 10 ms	6	G.729 [22] -internal
GSM EFR EP2 + G.723.1 (5.3 [18])	0%, 5%, 10%	G.723.1 [18]: 30 ms	2	G.723.1 [18] -internal



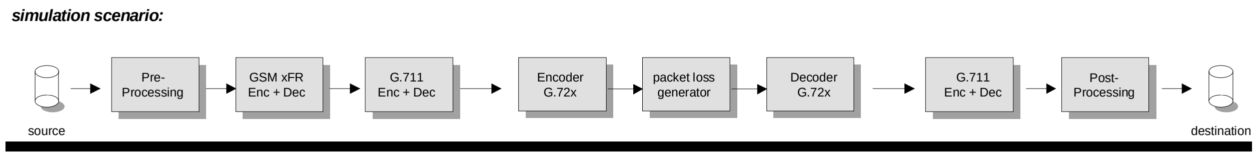


Figure 4: Simulated tandem processing scenario

5.2.1.3 Results

The following figures illustrate the subjective assessment results. On the y axis the MOS score from 1 (bad) to 5 (excellent) is shown. The x-axis shows the various end to end scenarios.

5.2.1.3.1 G.711 + Codec + G.711

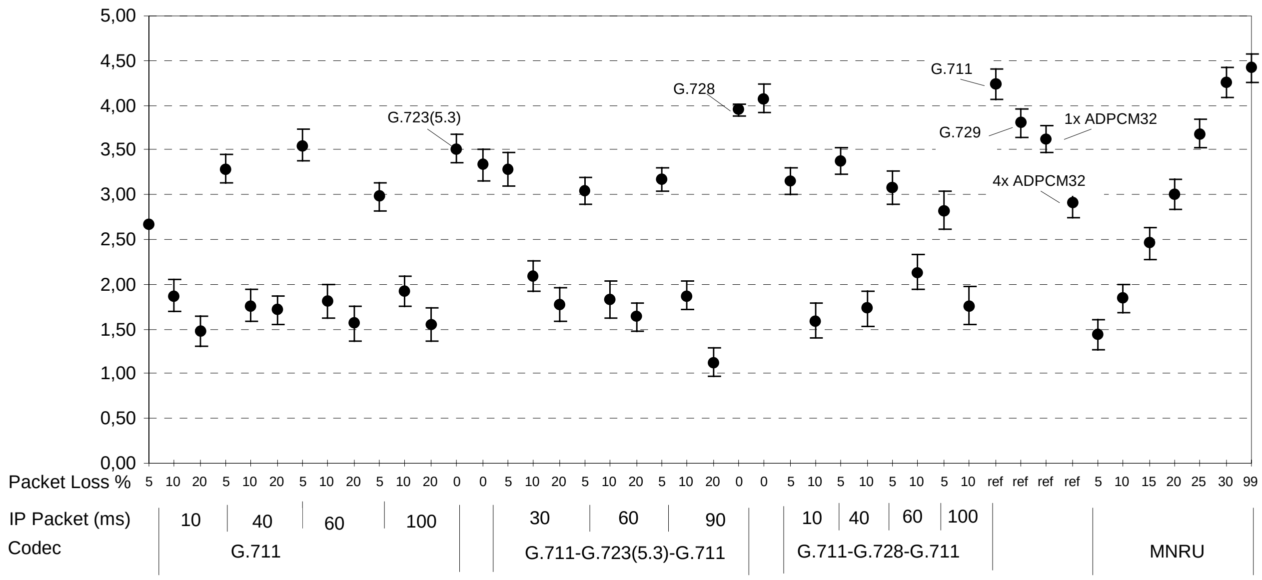


Figure 5: Subjective assessment result

The assessment of single codecs under the influence of packet loss leads to assumptions as follows:

- the packet loss rate of 5% seems to be almost the quality threshold of MOS 3,0;
- in all test cases the evaluation of voice signals with packet loss of > = 10% the MOS scores are widely below the quality threshold of MOS 3,0.

5.2.1.3.2 Tandem Conditions with GSM xFR

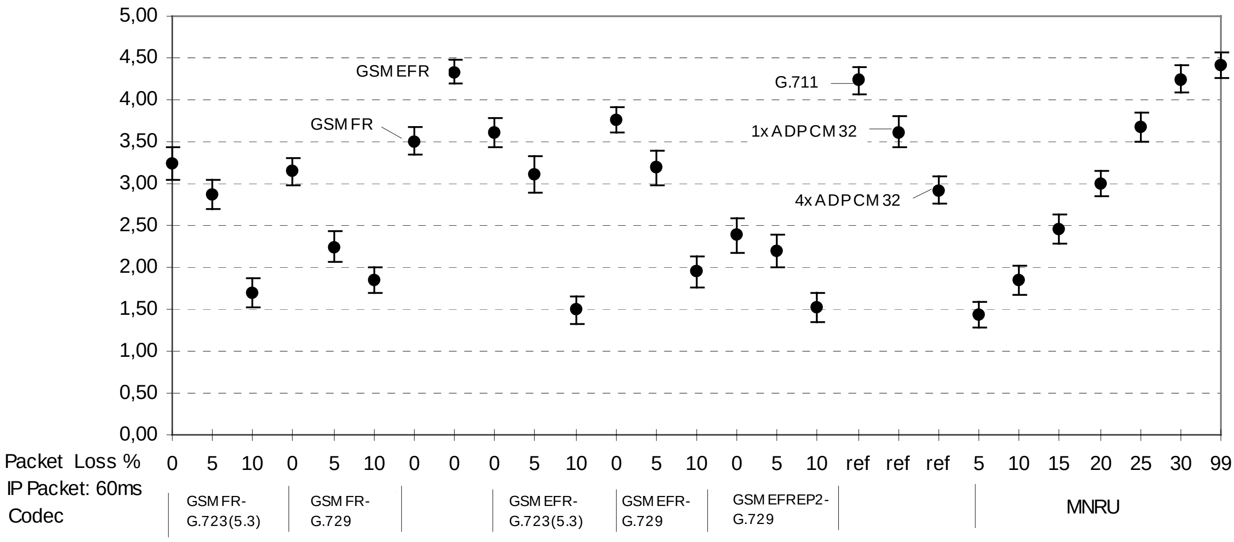


Figure 6: Tandem conditions with GSM xFR

The assessment of tandem codecs under the influence of packet loss for the G.7xx codecs, leads to assumptions as follows:

- the quality threshold for tandem connection of GSM FR and G.723.1 [18] can be seen with less than or max. 5% packet loss, under the precondition of an error free radio channel;
- the quality threshold for tandem connection of GSM FR and G.729 [22] can be seen in the range of << 5% packet loss, under the precondition of an error free radio channel;
- the quality threshold for tandem connection of GSM EFR and G.723.1 [18] can be seen in the range of 5% packet loss, under the precondition of an error free radio channel.

No acceptable speech transmission quality for tandem connection of GSM EFR and G.723.1 [18] for 0% packet loss can be provided if the radio channel induces errors.

5.2.2 Subjective Results on impairment effects of IP packet loss

Source: Nortel Networks; ETSI TIPHON 17TD167, Subjective Results on impairment effects of IP packet loss.

5.2.2.1 Introduction

With the growing interest in voice transport over Internet (IP) networks (voice over IP or VoIP), it is important to understand the effects of various impairments on voice quality. One of the important impairments is packet loss, which can be produced when voice packets are lost or delivered too late to be useful. A subjective experiment was conducted to investigate the effects of packet loss on voice quality. This experiment included a variety of codecs that are used in VoIP applications and packet loss rates ranging from 0% to 5%. Loss rates greater than 5% were not included in this study because based on previous experience such large impairments in voice quality tend to skew subjective data.

5.2.2.2 Measurement Set-up

5.2.2.2.1 Subjects

Sixty-one listeners aged 16 to 68 years (mean age 37 years) participated in the experiment. All listeners were telephony users with self-reported normal hearing drawn from Nortel Network's Subjective Assessment Lab subject pool.

5.2.2.2.2 Speech Processing

The source speech consisted of high-quality anechoic chamber recordings of North American English sentences spoken by six talkers (3 male, 3 female). Each speech sample consisted of four sentences uttered by the same talker. All the speech samples were transmit filtered before encoding and receive-filtered prior to being heard by listeners. The input level of the speech signals to the codecs was -20 dBm0. For non-G.711 [16] codec conditions, the codecs received G.711 [16] encoded/decoded speech as the input speech. The codecs tested in this experiment were G.711 [16], G.729 [22], G.729A [23], G.723.1 [18], and GSM EFR (06.60). For G.711 [16] and G.729 [22], speech frame sizes of 10 and 20 msec were tested, while G.723.1 [18] and GSM EFR were tested with their standard frame sizes (30 msec and 20 msec respectively). For all the codecs with a Voice Activity Detection (VAD) feature (G.729 [22], G.729A [23]. GSM EFR, G.723.1 [18]) the VAD was set to "OFF". In addition, two Packet Loss Concealment (PLC) schemes for G.711 [16] were tested: one scheme described by AT&T in a submission to SG 16 [1] and proposed as an ANSI standard [1] and a second proprietary scheme developed by Nortel Networks.

For the packet loss conditions, frames were removed from the speech samples randomly with a frequency determined by the test condition (e.g., 1% of the frames). A voice activity detector was used to ensure that losses always occurred during an active speech period. It is important to note that this technique for simulating packet loss may be different from other studies, so cross-experiment comparisons should be done with caution. A different random mask file was applied to the speech samples for each group of 3 listeners in order to randomize where the losses were occurring during the 4-sentence sample.

5.2.2.2.3 Procedure

Samples were played back over one channel (one side) of high-fidelity headphones to simulate handset listening. The samples were played back at 79 dB SPL, measured at the ear reference plane. Listeners heard one sample during each trial, and entered their ratings by pressing a button on a response box. The order of presentation of the samples was randomized and the ratings were stored in a file for later statistical analysis. Eight additional samples, which were not counted in the data, were presented at the beginning of the session as a warm-up. All subjects heard and rated all the conditions presented by six talkers, resulting in a total of 366 observations per condition.

Listeners rated the processed speech samples in an Absolute Category Rating (ACR) test. Samples were rated according to the telephony 5-point scale (excellent, good, fair, poor, bad). Mean-opinion-scores (MOSs) were computed from the ratings assigned to each test case.

5.2.2.2.4 MOS Data and MNRU Conditions

The MOS and MNRU equivalence data for each of the conditions are summarized in the subclause 5.2.2.3.3. The results of various conditions of interest are also summarized and discussed below.

Figure 7 shows the MOS values as a function of the Q parameter used in the MNRU conditions. This function is typical of what is seen in many subjective tests of voice quality, and a curve fit using the standard MNRU equation and the parameters A=1,6, B=1,6, C=9,5, Qm=17,3 achieved an R^2 value of 0,999.

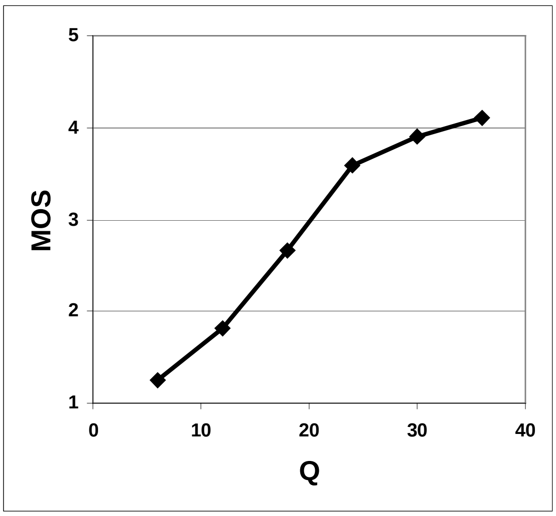


Figure 7: MOS as a function of Q in the MNRU conditions

5.2.2.3 Results

5.2.2.3.1 Reference Codec Conditions

Figure 8 summarizes the MOS ratings for the codec conditions without packet loss impairments. These results are also typical of VoIP quality experiments, with G.711 [16] and GSM EFR receiving the highest ratings and G.723.1 [18] receiving the lowest rating. The only unexpected result was that the G.729A [23] score is slightly higher than the G.729 [22] (3.78 vs. 3.65), which is inconsistent with previous findings that show that G.729A [23] is rated as slightly below or equal to G.729 [22]. The difference in the current experiment is small and only occurred for the 0% packet loss conditions.

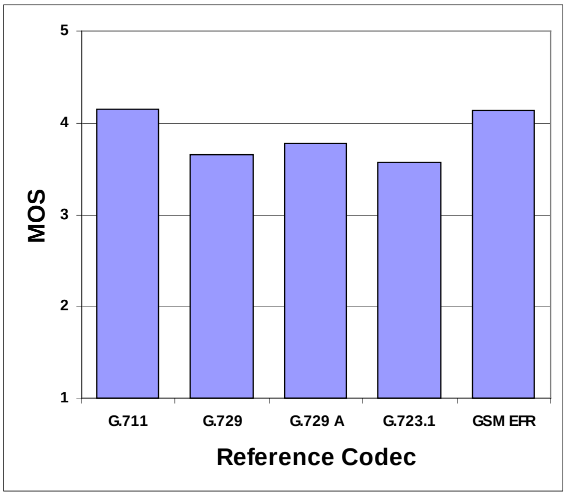


Figure 8: MOS values for each codec without packet loss impairments

5.2.2.3.2 Packet Loss Conditions

Figure 9 shows the effects of packet loss on MOS ratings for various codecs when the packet size was set to 10 msecs. It can be seen that the quality of G.711 [16] (without PLC) deteriorates rapidly with increasing loss, while G.729 [22] and G.729A [23] are more robust to packet loss. This is due to PLC (frame erasure concealment) schemes built into the G.729 [22] codecs. The two PLC schemes for G.711 [16] tested in this experiment (algorithms produced by AT&T and Nortel Networks) had large and similar beneficial effects in preserving voice quality with packet loss.

The results for 20 msec (and larger) packets are shown in figure 10. Again, G.711 [16] quality drops dramatically with increasing packet loss and this can be preserved using either of the PLC schemes. The new codecs in this figure are GSM EFR, which is somewhat resistant to packet loss, and G.723.1 [18], which deteriorates rapidly with increasing loss similar to the G.729 [22] codecs.

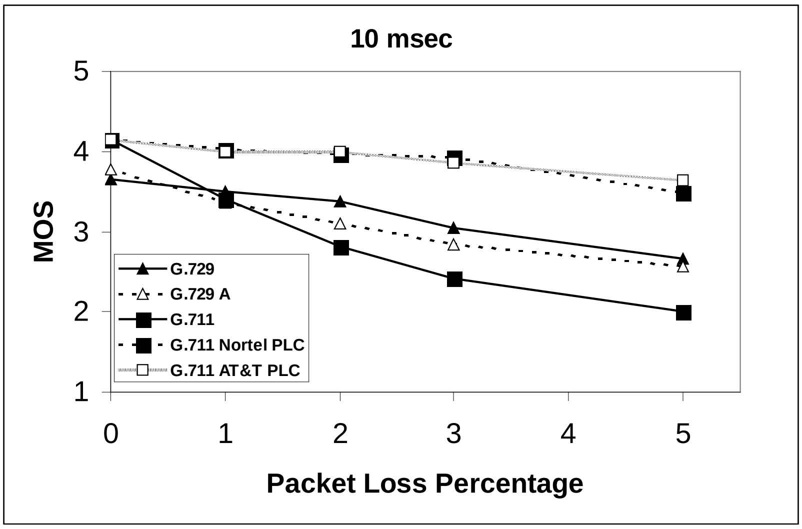


Figure 9: Effects of packet loss on voice quality with 10 msec packets

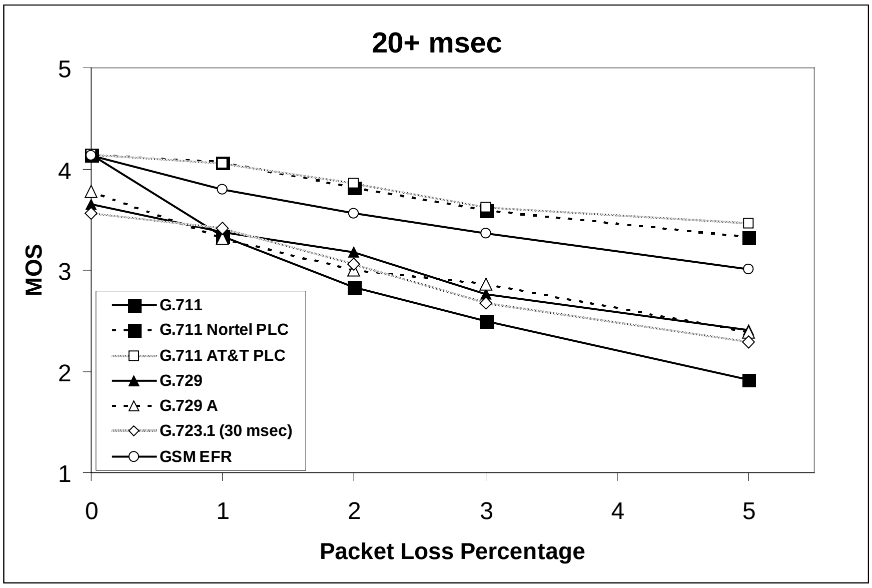


Figure 10: Effects of packet loss on voice quality with 20 msec or larger packets. All packet sizes were 20 msec except G.723.1 [18], which was 30 msec

5.2.2.3.3 Tables of Results

MOS and MNRU equivalence scores for each condition tested.
The parameters for the MNRU curve fit were: A=1,6, B=1,6, C=9,5, Qm=17,3

Packet (ms)	Codec	bps	PLC Algorithm	Packet Loss	MOS	MNRU Q
10	G.729 [22]	(8 kb/s)	included	1%	3.50	23,37
10	G.729 [22]	(8 kb/s)	included	2%	3.38	22,34
10	G.729 [22]	(8 kb/s)	included	3%	3.05	20,08
10	G.729 [22]	(8 kb/s)	included	5%	2.67	17,70
10	G.729 [22] A	(8 kb/s)	included	1%	3.37	22,27
10	G.729 [22] A	(8 kb/s)	included	2%	3.10	20,40
10	G.729 [22] A	(8 kb/s)	included	3%	2.85	18,78
10	G.729 [22] A	(8 kb/s)	included	5%	2.57	17,13
10	G.711 [16] u-law	(64 kb/s)	None	1%	3.41	22,57
10	G.711 [16] u-law	(64 kb/s)	None	2%	2.82	18,60
10	G.711 [16] u-law	(64 kb/s)	None	3%	2.41	16,18
10	G.711 [16] u-law	(64 kb/s)	None	5%	2.00	13,55
10	G.711 [16] u-law	(64 kb/s)	Nortel PLC	1%	4.02	30,83
10	G.711 [16] u-law	(64 kb/s)	Nortel PLC	2%	3.97	29,45
10	G.711 [16] u-law	(64 kb/s)	Nortel PLC	3%	3.93	28,61
10	G.711 [16] u-law	(64 kb/s)	Nortel PLC	5%	3.49	23,23
10	G.711 [16] u-law	(64 kb/s)	AT&T PLC	1%	4.00	30,16
10	G.711 [16] u-law	(64 kb/s)	AT&T PLC	2%	3.99	30,03
10	G.711 [16] u-law	(64 kb/s)	AT&T PLC	3%	3.86	27,43
10	G.711 [16] u-law	(64 kb/s)	AT&T PLC	5%	3.63	24,60
20	G.729 [22]	(8 kb/s)	included	1%	3.38	22,34
20	G.729 [22]	(8 kb/s)	included	2%	3.18	20,89
20	G.729 [22]	(8 kb/s)	included	3%	2.77	18,28
20	G.729 [22]	(8 kb/s)	included	5%	2.40	16,13
20	G.729 [22] A	(8 kb/s)	included	1%	3.32	21,90
20	G.729 [22] A	(8 kb/s)	included	2%	3.00	19,74
20	G.729 [22] A	(8 kb/s)	included	3%	2.87	18,89
20	G.729 [22] A	(8 kb/s)	included	5%	2.39	16,07
20	G.711 [16] u-law	(64 kb/s)	None	1%	3.33	21,98
20	G.711 [16] u-law	(64 kb/s)	None	2%	2.83	18,66
20	G.711 [16] u-law	(64 kb/s)	None	3%	2.50	16,69

20	G.711 [16] u-law	(64 kb/s)	None	5%	1.92	12,99
20	G.711 [16] u-law	(64 kb/s)	Nortel PLC	1%	4.06	31,96
20	G.711 [16] u-law	(64 kb/s)	Nortel PLC	2%	3.82	26,82
20	G.711 [16] u-law	(64 kb/s)	Nortel PLC	3%	3.59	24,14
20	G.711 [16] u-law	(64 kb/s)	Nortel PLC	5%	3.33	21,96
20	G.711 [16] u-law	(64 kb/s)	AT&T PLC	1%	4.05	31,76
20	G.711 [16] u-law	(64 kb/s)	AT&T PLC	2%	3.86	27,34
20	G.711 [16] u-law	(64 kb/s)	AT&T PLC	3%	3.62	24,44
20	G.711 [16] u-law	(64 kb/s)	AT&T PLC	5%	3.47	23,07
20	GSM EFR	(12.2 kb/s)	included	1%	3.80	26,48
20	GSM EFR	(12.2 kb/s)	included	2%	3.57	23,94
20	GSM EFR	(12.2 kb/s)	included	3%	3.36	22,23
20	GSM EFR	(12.2 kb/s)	included	5%	3.01	19,80
30	G.723.1 [18]	(6.3 kb/s)	included	1%	3.42	22,64
30	G.723.1 [18]	(6.3 kb/s)	included	2%	3.05	20,08
30	G.723.1 [18]	(6.3 kb/s)	included	3%	2.67	17,74
30	G.723.1 [18]	(6.3 kb/s)	included	5%	2.30	15,47
	MNRU	n/a	6 dB	n/a	1.26	
	MNRU	n/a	12 dB	n/a	1.81	
	MNRU	n/a	18 dB	n/a	2.66	
	MNRU	n/a	24 dB	n/a	3.59	
	MNRU	n/a	30 dB	n/a	3.90	
	MNRU	n/a	36 dB	n/a	4.11	
	G.729 [22]		included	None	3.65	24,74
	G.729 [22] A		included	None	3.78	26,26
	G.711 [16] u-law		None	None	4.15	36,75
	GSM EFR		included	None	4.13	35,43
	G.723.1 [18]		included	None	3.57	23,94

5.3 Objective testing

5.3.1 Objective Speech Quality Evaluation on Speech Data recorded at the SuperOp 99 event in Hawaii

Source: Rapporteur of ITU-T Q.13/12; ETSI TIPHON 17TD135.

5.3.1.1 Introduction

This subclause presents results from the ITU-T Q.13/12 evaluation for a new Recommendation for objective speech quality assessment algorithm, carried out on speech material that was recorded in a joint QoS measurement initiative between ETSI EP TIPHON and ITU-T Q.13/12 at the SuperOp Interoperability event on Hawaii in July 1999.

Due to former agreements and restrictions for publishing individual data before finalization of the new ITU-T Recommendation, subjective scores as well as anonymized objective scores are shown.

5.3.1.2 Measurement Set-up

For speech input into the simulator as well as for recording of degraded speech, the CAS (Communication Analysis System) from HEAD-acoustics was used. It was kindly provided by Robert BOSCH GmbH. The CAS provides playback and recording of stereo signals. To allow delay measurement, the input speech signal (reference) was directly recorded by the left input channel. The transmitted signal (degraded speech signal) was recorded on the right channel.

Due to technical problems and the fact that the participants of the SuperOp had scheduled very tough timeslots for all their tests, we have been able to only test one physical scenario with one set of equipment.

To collect the speech files the TIPHON scenario 2 was set-up (figure 11).

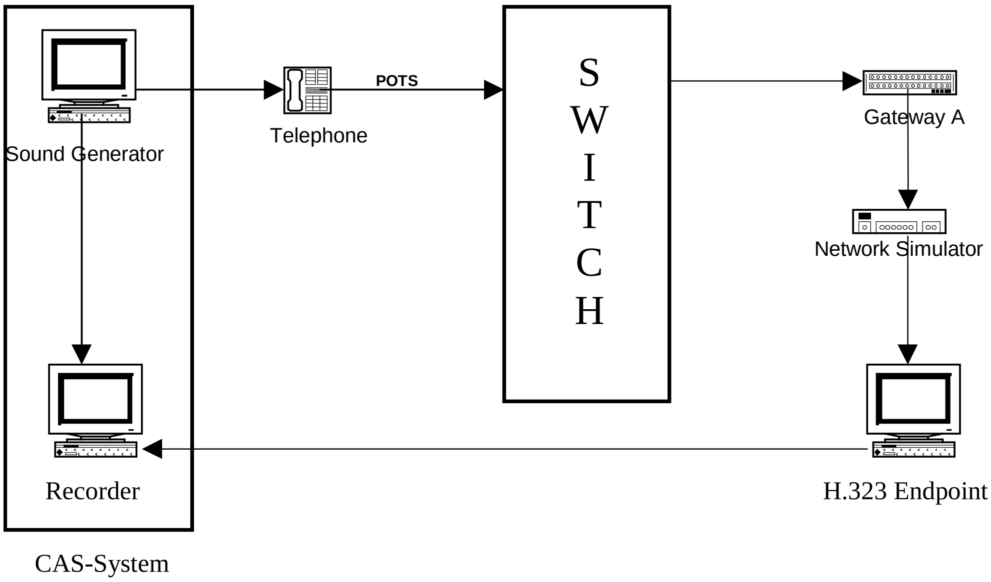


Figure 11: Recording of speech samples with TIPHON scenario 2

After recording of the speech samples, the 32s files were separated into samples of approximately 8s in order to allow per-talker-evaluations.

In addition to the processed conditions, 7 MNRU reference conditions were used in order to ensure a proper instruction and training of the test subjects. Furthermore, MNRU references are used for further comparison and quality management.

5.3.1.3 Quality Assessment

Within the ITU-T Q.13/12 model evaluation, subjective tests on the Hawaii database were performed at the subjective quality test laboratories at T-Nova Deutsche Telekom Innovationsgesellschaft mbH in Berlin. The tests were designed according to ITU-T Recommendation P.800 [32]. Each of the 208 speech samples were assessed by 24 subjects. The listener's group consists of 12 female and 12 male persons of normal hearing with no specific knowledge in speech processing and speech quality assessment. Normal hearing capability was proven by pure-tone audiometry within 125 Hz to 8 000 Hz.

The objective assessment procedure was done with 5 candidates for the new ITU-T Recommendation. The results of every objective method was compared against the subjective results and against the existing ITU-T Recommendation P.861 [33]. The correlation between the subjective and objective assessment methods for all candidates can be seen in Table 1.

5.3.1.4 Results

Within the competition of objective speech quality measurement algorithms of ITU-T Recommendation Q.13/12 [34], a large speech database with more than 2 100 speech samples was used. The database that was produced during the Hawaii QoS measurement initiative contributed one of 22 experiments to this evaluation. The objective estimates were compared against the subjective measures from the Hawaii test, and correlation coefficients were derived for each model. Table 1 shows the correlation coefficients of the objective speech quality assessment models.

In fact, only one VoIP test bed was used at the Hawaii QoS measurement initiative. Because of the good homogeneity of the recording scenario, the correlation coefficients of the best models exceed 95% - a very good measurement result for unknown data.

Table 1: Correlation coefficient for the ITU-T Q.13/12 objective speech quality assessment algorithms

Rank order	Correlation coefficient
1	0.9879
2	0.9859
3	0.9489
4	0.9275
5	0.8377
P.861 [33]	0.8706

5.4 Delay

5.4.1 Delay between two analogue PBX subscribers

Source: Robert Bosch GmbH: ETSI TIPHON 14TD081.

5.4.1.1 Introduction

The end-to-end speech transmission performance of two different IP gateways, which are commercially available on the marketplace, has been evaluated. In addition, two versions of one PC client software have been compared with respect to their impact on speech transmission quality.

The tests have been conducted in an environment which is typical for business applications, i.e. the IP network consisted of a LAN (without traffic load) and the SCN consisted of PBXs (Private Branch Exchanges). The coding scheme was in all cases selected as G.723.1 [18] (6.3 kbit/s); the Voice Activity Detection (VAD) as well as the integrated echo cancelling devices were activated in all cases.

In the following the IP Gateway from manufacturer A is denoted "Gateway type A" while the IP Gateway from manufacturer B is denoted "Gateway type B"; the two versions of the PC client software are denoted "SW version x" and "SW version y", respectively.

5.4.1.2 Measurement Set-up

Measurements have been conducted with Head Acoustics' Communication Analysis System (CAS); E-Model calculations have been performed with Alcatels' software.

The IP gateways are connected to the PBXs via 2 Mbit/s links with QSIG signaling; the measuring interfaces at the subscriber side of the PBX were analogue 4-wire interfaces in accordance with ES 201 168 [7]. Measurements at the PC client have been performed at the electric handset interface.

The telephone sets used during the tests have been standard analogue sets which comply with the default values of the E-Model (as far as applicable). In order to keep the testing effort within reasonable limits, measurement of delay, jitter and functionality of the echo cancellers has not been performed between acoustical interfaces, because measurement of such parameters between electrical interfaces gives approximately the same results.

Principal test set-up:

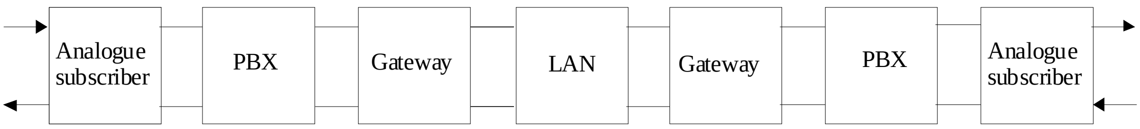


Figure 12: Set-up between two analogue PBX subscribers

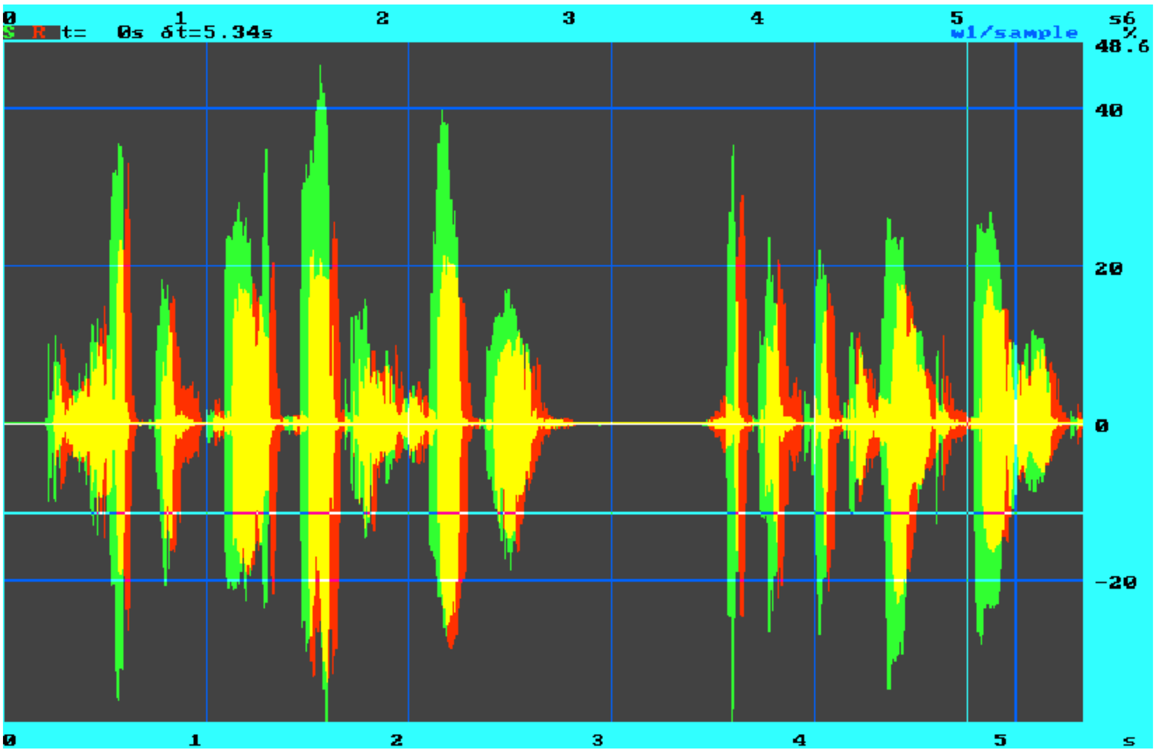
5.4.1.3 Results

Gateway Type A:

The mean one-way delay for both subscribers is variable between 125 ms and 250 ms, with an arithmetical mean value at 160 ms.

Gateway type B:

The mean one-way delay is constant with a value between 90 ms and 110 ms (see figure 13)**:**



Red = Original send signal; Green = receive signal (transmitted via the test set-up).

NOTE: Due to the selected measurement mode the original send signal (red) is displayed 150 ms later and reduced by 6 dB in level than in reality.

Figure 13

5.4.1.4 Conclusion

For transmission planning purposes the mean value of the one-way delay of both transmission directions of a connection is calculated and expressed as "Mean one-way delay". In this respect Gateway type B with 100 ms is much better than Gateway type A with approx. 160 ms (which was variable, additionally).

5.4.2 Delay between two PABX systems with analogue subscribers

Source: Head Acoustics: QoS Measurements of IP-Configurations 15TD089.

5.4.2.1 Measurement Set-up:

On both networks no additional traffic was generated such that there was no packet loss. The System Configurations (PABX [Bosch], IP-Gateway [3rd party], PC-Terminals with commercial VOIP software) were provided by BOSCH Telecom.

The acoustical access to the terminals was made using HATS (artificial head according to ITU-T Recommendation P.58 [27]) equipped with the artificial ear type 3.4 according to ITU-T Recommendation P.57 [26] and a handset positioning device according to new Recommendation P.64 [28]. This set-up guarantees a most realistic set-up for the terminals since all transmission properties are reproduced in a very realistic way. The measurements were conducted in "noisy" environments, this means the background noise in the laboratory rooms was present all the time.

The electrical access to the configurations was made using the 600 Ohm access pint of the PABX for configuration 1, for configuration 2 the soundcard was used for access. The level adjustment in this configuration was made in advance to the test in order to achieve realistic loudness ratings for sending and receiving. Although it should be noted that there was always AGC in sending and receiving present which could influence the settings during the measurements. Such settings were always checked during the measurements in order to not be misled by different settings of the devices.

5.4.2.2 Result

- One way transmission delay: 70 ms.
- Coding: ITU-T Recommendation G.729 [22].

5.4.3 Delay between two PC SW clients

Source: Head Acoustics: QoS Measurements of IP-Configurations, 15TD089.

5.4.3.1 Measurement Set-up

On both networks no additional traffic was generated such there was no packet loss. The System Configurations (PABX [Bosch], IP-Gateway [3rd party], PC-Terminals with commercial VOIP software) were provided by BOSCH Telecom.

The acoustical access to the terminals was made using HATS (artificial head according to ITU-T Recommendation P.58 [27]) equipped with the artificial ear type 3.4 according to ITU-T Recommendation P.57 [26] and a handset positioning device according to new Recommendation P.64 [28]. This set-up guarantees a most realistic set-up for the terminals since all transmission properties are reproduced in a very realistic way. The measurements were conducted in "noisy" environments, this means the background noise in the laboratory rooms was present all the time.

The electrical access to the configurations was made using the 600 Ohm access pint of the PABX for configuration 1, for configuration 2 the soundcard was used for access. The level adjustment in this configuration was made in advance to the test in order to achieve realistic loudness ratings for sending and receiving. Although it should be noted that there was always AGC in sending and receiving present which could influence the settings during the measurements. Such the settings were always checked during the measurements in order to not be misled by different settings of the devices.

5.4.3.2 Results

- One Way Transmission Time: 400-530 ms.
- Coding: ITU-T Recommendation G.723.1 [18].

5.4.4 Delay between an analogue PBX subscriber and a PC client

Source: Robert Bosch GmbH; Speech Quality Test results of IP equipment in a LAN environment; ETSI TIPHON 14TD081.

5.4.4.1 Measurement Set-up

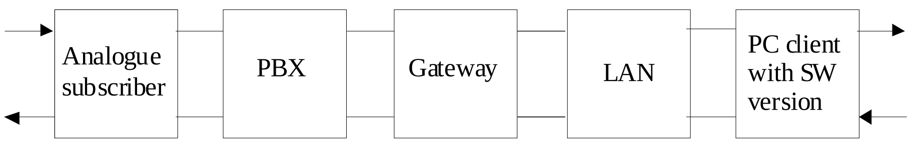
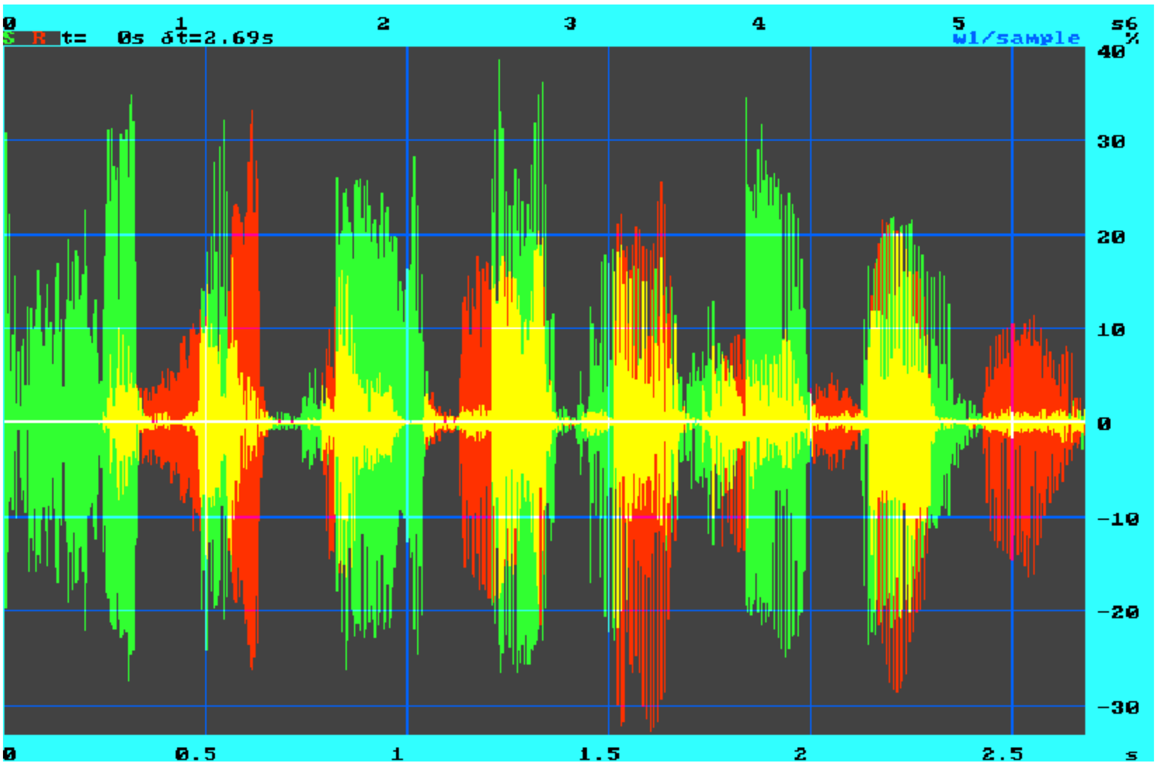


Figure 14: Set-up between an analogue PBX subscriber and a PC client

Results:

Figure 15 shows the delay behaviour of the test set-up for the transmission direction from PC client to analogue subscriber.



Red = Original send signal; Green = receive signal (transmitted via the test set-up)

NOTE: Due to the selected measurement mode the original send signal (red) is displayed 500 ms later and reduced by 6 dB in level than in reality.

Figure 15

The mean one-way delay is constant but depending on the direction of transmission:

- from analogue subscriber to PC client = 465 ms;
- from PC client to analogue subscriber = 200 ms.

Figure 15 shows that no significant delay jitter could be observed.

5.4.4.2 Conclusion

The values depend on the direction of transmission and are in the range from 200 ms to 700 ms, where the highest values occur when the PC client is at receive side of a delay measurement. The PC client functions as a combination of gateway and telephony terminal and its contribution to the total delay - in principle - should not deviate significantly from the delay contribution of a gateway plus its interconnected telephone set. The very high delay contribution of the PC client found in this test may be caused by another approach of its delay variation buffer in receive direction.

5.5 Loudness Ratings

5.6 Echo

5.6.1 Echo between two analogue PBX subscribers

Source: Robert Bosch GmbH, Speech Quality Test Results of IP equipment in a LAN environment, ETSI TIPHON 14TD081.

5.6.1.1 Measurement Set-up

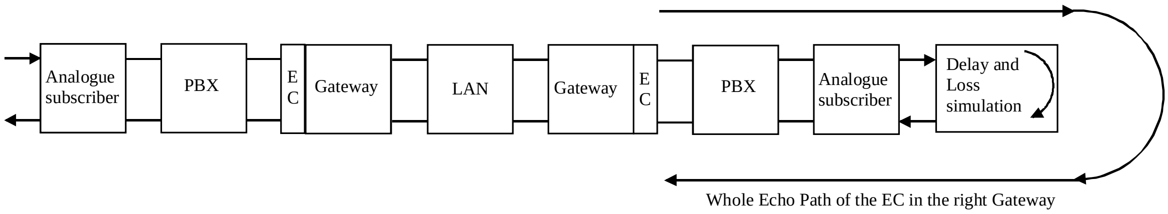


Figure 16: Echo between two analogue PBX subscribers

5.6.1.2 Results

NOTE: In case of simulated loss in the whole echo path of 7dB:
The delay and loss simulation device (see principal test set-up) has been adjusted such, that the loss of the whole echo path is 7 dB, because ITU-T Recommendation G.165 [14] specifies functionality of an EC for echo path loss > 6 dB.

In case of simulated loss in the whole echo path of 4dB:
The delay and loss simulation device (see principal test set-up) has been adjusted such, that the loss of the whole echo path is 7 dB, because ITU-T Recommendation G.168 [15] specifies functionality of an EC for echo path loss > 0 dB (which is different from G.165 [14], see above).

Table of results:

Gateway	Simulated loss in the whole echo path	
	7 dB	4 dB
A	For simulated tail delay values up to 28 ms the residual echo level is below -65 dBm0 (test signal white noise, -10 dBm0). For increased values of the tail delay the residual echo level is more than -25 dBm0. Note, that the tail delay for the EC is identical with the delay along the whole echo path.	As it can be seen from figure 17, the adaptation of the EC is very slow (upper window); sometimes the value of the remaining level of the echo signal goes below the threshold of the Non-Linear Processor (NLP) which instantly provides very good values of echo loss (lower window), because in that case the NLP clips off the remaining amplitude of the echo signal.
B	For simulated tail delay values up to 11 ms the residual echo level is below -85 dBm0 (test signal: white noise, -10 dBm0). For increased values of the tail delay the residual echo level is more than -25 dBm0.	

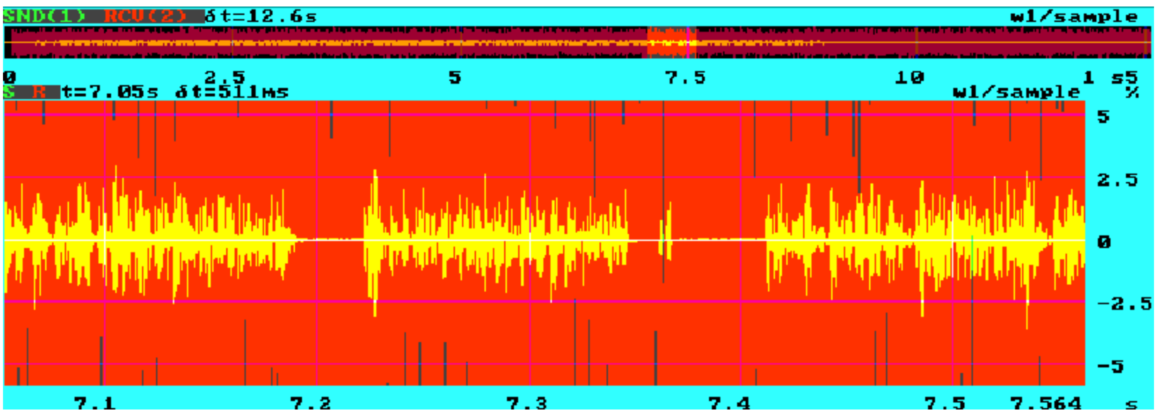


Figure 17: Red = Original send signal; Yellow = receive signal (transmitted via the test set-up)

5.6.1.3 Conclusion

The echo cancellers in both gateways do not comply with the applicable ITU-T Recommendation G.168 [15] "Digital Network Echo Cancellers".

For practical application the maximum tail delay of the echo canceller (as found in the tests) must be reduced by 6 to 8 ms; the remaining value must be divided by two (tail delay is a round-trip value) in order to get the correct value of the maximum mean one-way delay which an EC can handle.

For Gateway type A this results to (28 ms - 6 ms)/2 = 11 ms, whereas for Gateway type B (11 ms - 6 ms)/2 = 2,5 ms are available, only.

Small and Medium sized Private Networks (e.g., Corporate Networks) which in a PSTN environment are operated without the deployment of separate echo cancellers, typically add 10 ... 20 ms to the mean one-way delay of a connection; it should be recognized that the use of both types of gateways will require additional echo cancellers in such networks.

5.7 Impairment Factors

5.7.1 Transmission Impairments according to ITU-T G.113

Source: ITU-T Recommendation G.113 [12] - Transmission impairments.

5.7.1.1 Introduction

Table I.1 of Ie values refers to non-error conditions. For propagation errors and frame-erasures or packet loss, no definite values are available which would be valid for more than one codec or codec family. In order to help the transmission planner, examples of Ie values under conditions of packet loss are given in Tables I.2 and I.3, and for propagation error patterns EP1 and EP2 in Table I.4. These values are provisional only as they were determined in single or a few experiments. In Table I.5, a brief description of the codecs is provided for information.

Table I.1: Provisional Planning Values for the Equipment Impairment Factor Ie

Codec Type	Reference	Operating Rate kBit/s	Ie Value
ADPCM	G.726 [19], G.727 [20]	40	2
	G.721(1988 [17]), G.726 [19], G.727 [20]	32	7
	G.726 [19], G.727 [20]	24	25
	G.726 [19], G.727 [20]	16	50
LD-CELP	G.728 [21]	16	7
		12,8	20
CS-ACELP	G.729 [22]	8	10
	G.729A [23] + VAD	8	11
VSELP	IS-54	8	20
ACELP	IS-641	7,4	6
QCELP	IS-96a	8	19
RCELP	IS-127	8	6
VSELP	Japanese PDC	6,7	24
RPE-LTP	GSM 06.10 [8] , Full-rate	13	20
VSELP	GSM 06.20 [37], Half-rate	5,6	23
ACELP	GSM 06.60 [38], Enhanced Full Rate	12,2	5
ACELP	G.723.1 [18]	5,3	19
MP-MLQ	G.723.1 [18]	6,3	15

Table I.2: Provisional planning values for the equipment impairment factor Ie under conditions of random packet loss, codecs G.729 [22] A + VAD, G.723.1-A [18] + VAD and GSM EFR

% Packet Loss	G.729A [23] + VAD	G.723.1.A [18] + VAD 6.3 kbit/s	GSM EFR
0	11	15	5
0.5	13	17	—
1	15	19	16
1.5	17	22	—
2	19	24	21
3	23	27	26
4	26	32	—
5	—	—	33
8	36	41	—
16	49	55	—
NOTE: Number of frames per packet: • G.729A [23] + VAD: 2; • G.723.1-A [18] + VAD: 1; • GSM EFR: 1.			

Table I.3: Provisional planning values for the equipment impairment factor Ie under conditions of packet loss, codecs G.711 [16] without and with Packet Loss Concealment (PLC)

		G.711 [16] w/ PLC	
Packet Loss %	G.711 [16] w/o PLC	Random Packet Loss	Bursty Packet Loss
0	0	0	0
1	25	5	5
2	35	7	7
3	45	10	10
5	55	15	30
7	—	20	35
10	—	25	40
15	—	35	45
20	—	45	50
NOTE: Speech packet length: 10 ms.			

Table I.4: Provisional planning values for the equipment impairment factor Ie under propagation error conditions, GSM codecs

Codec type	Error pattern	Ie Range
GSM-HR	EP1	25...32
	EP2	31...42
GSM-FR	EP1	32...39
	EP2	40...45
GSM-EFR	EP1	15...22
	EP2	26...35
NOTE 1: The range given results from the difficulties in deriving exact impairment factor values for these conditions. NOTE 2: EP1 is equivalent to 10 dB C/I, EP2 is equivalent to 7 dB C/I. C/I is the carrier to interference ratio.		

5.8 R-Values

5.8.1 Analogue PBX- and SW Client scenarios

Source: Robert Bosch GmbH, Speech Quality Test Results of IP equipment in a LAN environment, ETSI TIPHON 14TD081.

5.8.1.1 Introduction

The results given in subclauses 5.4.1, 5.4.4, 5.6.1 are taken as an input to the E-Model in order to predict end-to-end speech transmission performance in terms of the E-Model Rating R as perceived by the average user and to predict user satisfaction.

The following three scenarios have been chosen to demonstrate the impact on end-to-end speech transmission performance, which are representative for the application of the gateways under test in Private Networks (e.g., Corporate Networks).

In order to include the massive impact of the low syllable recognition rate into the E-Model calculation, a preliminary Ie value for this effect had to be estimated. The syllable cut-off due to a VAD can be compared with the effect of hands free telephony. For hands free telephony it is well-known that the impact perceived by the user lies in a range Ie = 10 .... 20. Hence, for this evaluation an Ie value of Ie = 10 was chosen to consider the impact of the reduced syllable recognition rate.

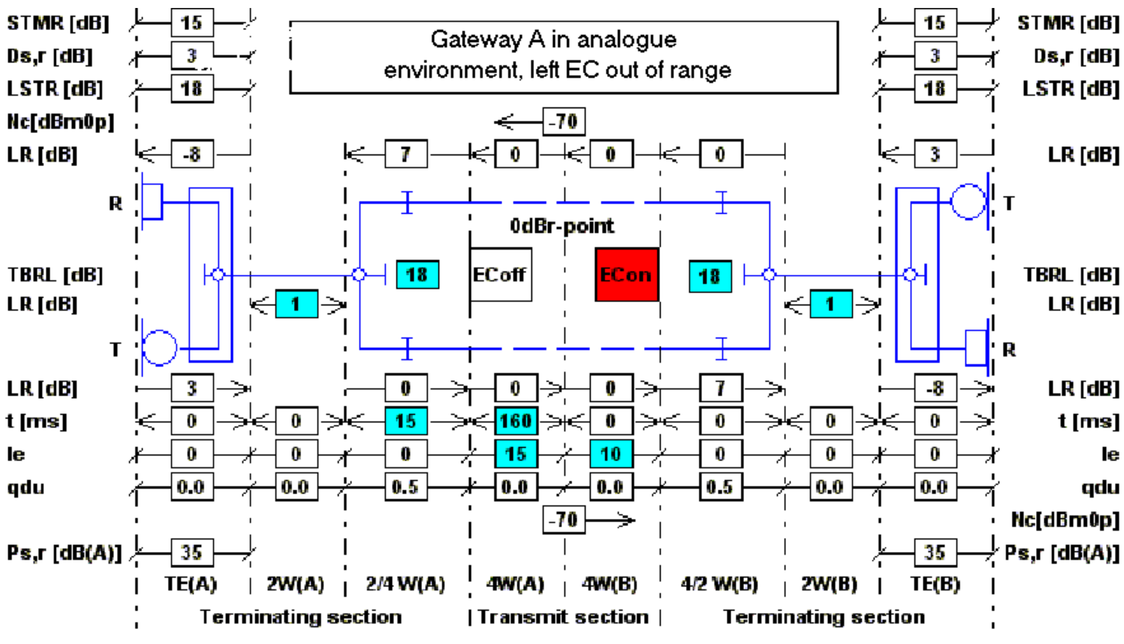
End-to-end speech transmission quality has been calculated according to ITU-T Recommendations G.107 [9] using the guidance given in ITU-T Recommendation G.108 [10] and compared with the categories of speech transmission quality defined ITU-T Recommendation G.109 [11] (and adopted in TR 101 329-2 [4], subclause 5.2, table 3). An Advantage Factor A which is sometimes discussed for Internet-Telephony does not apply for business applications.

5.8.1.2 Results

Scenario #1:

On both sides Gateway type A with analogue telephone sets as a termination of the PBXs.

The delay in the PBX network on the left side exceeds the capability of the EC (which is integrated in the gateway).



Speech transmission quality as perceived by the user at the left terminal where the EC is out of his operational range:

R-Value	Speech Transmission Quality Category	User satisfaction
67	Low	Many users dissatisfied

Speech transmission quality as perceived by the user at the right terminal where the EC is within his operational range:

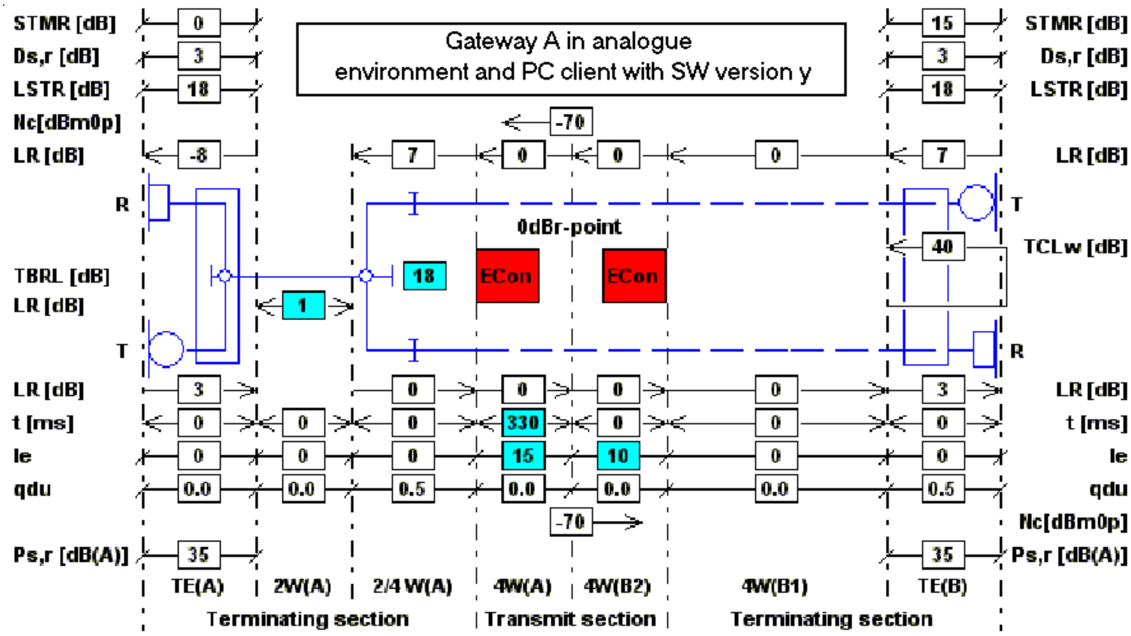
R-Value	Speech Transmission Quality Category	User satisfaction
0	values < 50 are not recommended	All users dissatisfied

NOTE: Because the left EC is out of his operational range, the speech quality perceived by the user at the right terminal is decreased.

Scenario #2:

Gateway type A with analogue subscriber behind PBX on the left side and PC client on the right side.

Both EC (which are integrated in the gateway and in the PC client) are within their operational range.



Speech transmission quality as perceived by the user at the left terminal (analogue telephone set):

R-Value	Speech Transmission Quality Category	User satisfaction
38	values < 50 are not recommended	All users dissatisfied

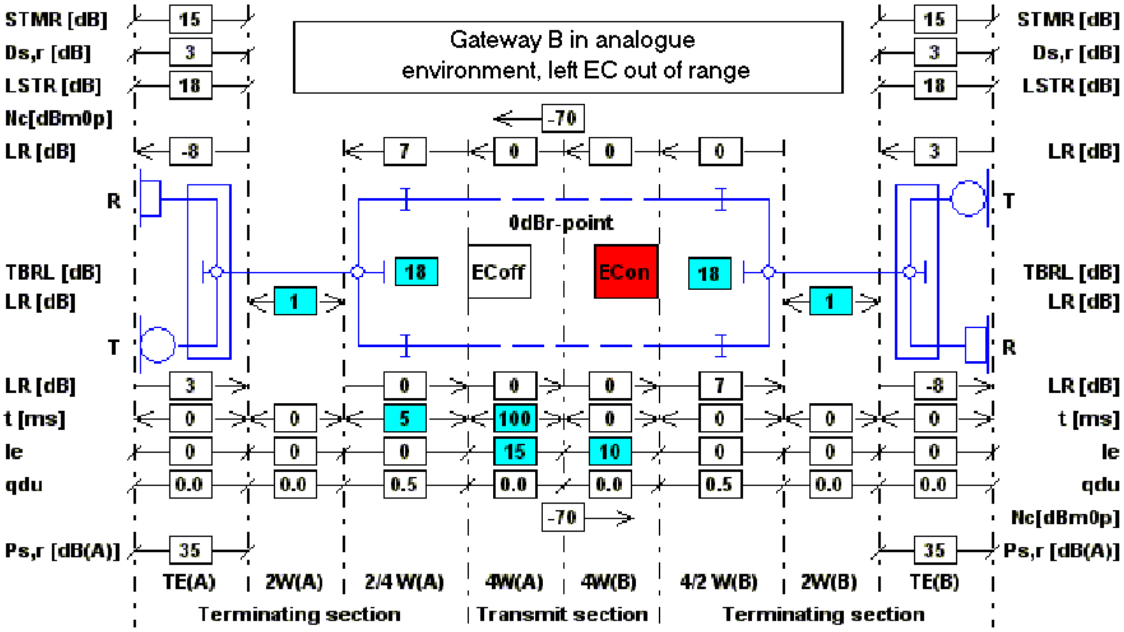
Speech transmission quality as perceived by the user at the right terminal (PC client):

R-Value	Speech Transmission Quality Category	User satisfaction
53	Poor	Nearly all users dissatisfied

Scenario #3:

On both sides Gateway type B with analogue telephone sets as a termination of the PBXs.

The delay in the PBX network on the left side exceeds the capability of the EC (which is integrated in the gateway



Speech transmission quality as perceived by the user at the left terminal where the EC is out of his operational range:

R-Value	Speech Transmission Quality Category	User satisfaction
68	Low	Many users dissatisfied

Speech transmission quality as perceived by the user at the right terminal where the EC is within his operational range:

R-Value	Speech Transmission Quality Category	User satisfaction
8	values < 50 are not recommended	All users dissatisfied

NOTE: Because the left EC is out of his operational range, the speech quality perceived by the user at the right terminal is decreased.

5.8.1.3 Conclusion

The results presented herein do not significantly differ from those we gained recently with other gateways and with other client software. Most manufacturers claim that they delivered state-of-the-art gateways and client software as an input for our evaluations.

The speech transmission end-to-end performance which was found in this evaluation is not acceptable for long-term business applications.

Hence, one may recognize a substantial gap between the TIPHON classes defined for speech quality and available solutions in practice.

5.9 Advanced Measurement Techniques

5.9.1 QoS Measurements of IP-Configurations

Source: Head Acoustics; QoS Measurements of IP-Configurations; ETSI TIPHON 15TD089.

5.9.1.1 Configuration and Measurement set-up

For the test 2 IP configurations were available which can be found in figures 18 and 19. On both networks no additional traffic was generated such there was no packet loss. The System Configurations (PABX [Bosch], IP-Gateway [3rd party], PC-Terminals with commercial VOIP software) were provided by BOSCH Telecom.

The acoustical access to the terminals was made using HATS (artificial head according to ITU-T Recommendation P.58 [27]) equipped with the artificial ear type 3.4 according to ITU-T Recommendation P.57 [26] and a handset positioning device according to new Recommendation P.64 [28]. This set-up guarantees a most realistic set-up for the terminals since all transmission properties are reproduced in a very realistic way. The measurements were conducted in "noisy" environments, this means the background noise in the laboratory rooms was present all the time.

The electrical access to the configurations was made using the 600 Ohm access pint of the PABX for configuration 1, for configuration 2 the soundcard was used for access. The level adjustment in this configuration was made in advance to the test in order to achieve realistic loudness ratings for sending and receiving. Although it should be noted that there was always AGC in sending and receiving present which could influence the settings during the measurements. Such the settings were always checked during the measurements in order to not be misled by different settings of the devices.

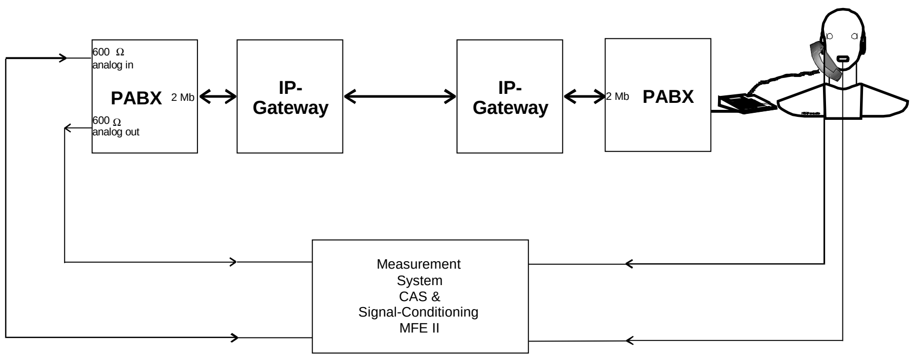


Figure 18: Configuration 1 - One way transmission delay = 70 ms, Coding ITU-T Recommendation G.729, "Standard" - handset telephone

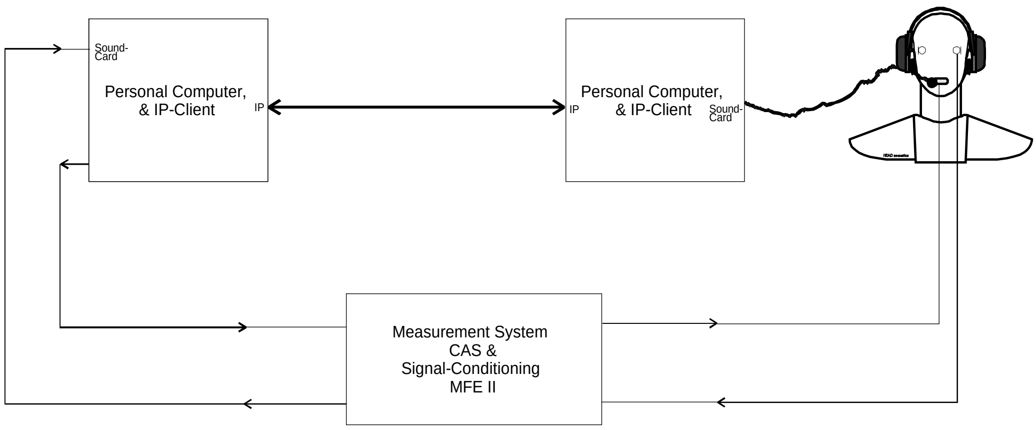


Figure 19: Configuration 2 - One way transmission time = 400 - 530 ms, variable, Coding ITU-T Recommendation G.723, PC with soundcard and headset

Figure 20: Test set-up with handset

Figure 21: Test set-up with headset

5.9.1.2 Results

5.9.1.2.1 Parameters in Single Talk Conditions

"Traditional" Parameters

Figures 22 and 23 show the measured frequency responses in sending and receiving for both, Configuration 1 and Configuration 2.

In sending the frequency responses do not indicate any problem except that the headset does not show a high pass filtering needed to exclude low frequency room noise. This can be seen for the handset at about 300 Hz.

The same statement can be made for the receiving direction. In addition the handset provides a smooth frequency response when adapted with a pressure force of 13 N to the artificial head. The headset provides a more high frequency emphasized frequency response.

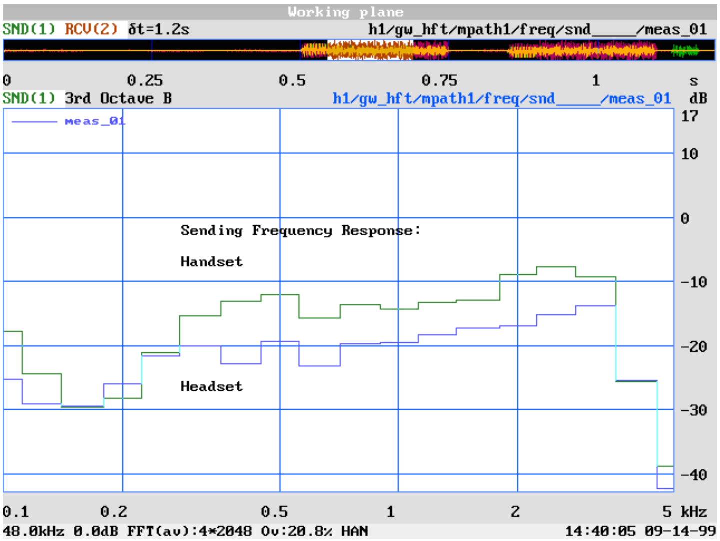


Figure 22: Sending frequency responses for configuration 1 (handset) and 2 (headset)

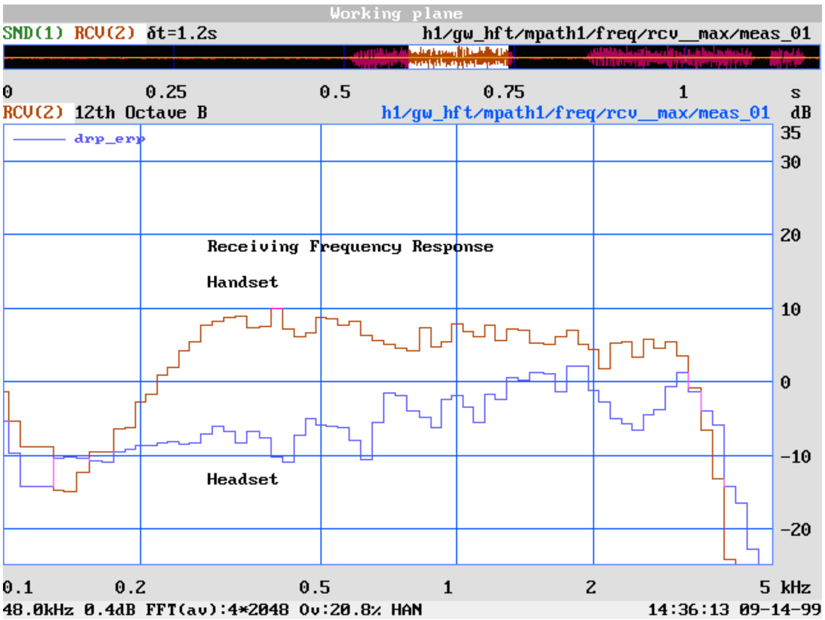


Figure 23: Receiving frequency responses for configuration 1 (handset) and 2 (headset)

More detailed results in single talk conditions switching and AGC characteristics and echo loss.

When evaluating the switching characteristics more in detail the following statements can be made:

5.9.1.2.2 Level dependant input-output characteristics

In sending and receiving direction a VAD is detectable which for both configurations indicate no or only minor distortions caused by syllable clipping. The measurement was made using an increasing voiced sound of artificial voice (ITU-T P.50), monitoring the output signal and comparing it to the input (measurement) signal.

In receiving direction the configuration 1 does not show any non linear behaviour whereas configuration 2 introduces a quite strong companding for speech as it can be seen in figure 25. In addition –depending on the input signal level- strong level variations can be seen (beginning and end of the test sequence in figure 25).

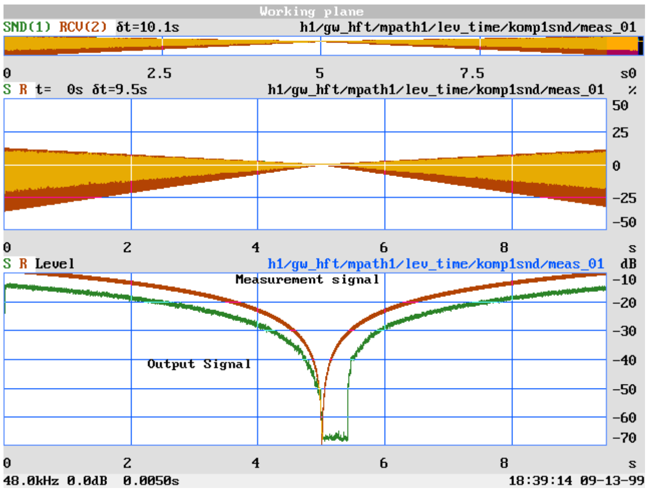


Figure 24: Level dependant input-output characteristics in sending, configuration 1 upper: time signal; lower: level vs. time

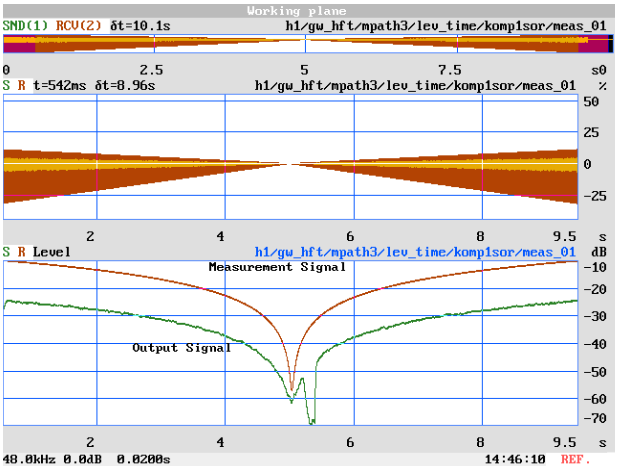


Figure 25: Level dependant input-output characteristics in sending, configuration 2 upper: time signal; lower: level vs. time

5.9.1.2.3 Echo loss and convergence

The echo loss was measured always from the electrical input to the electrical output. Certainly in single talk the characteristics of the echo canceller(s) are quite important for the subjective impression of a connection. Figures 26 and 27 show measurements using the composite source signal CSS as defined in ITU-T Recommendation P.501 [29] and ITU-T Recommendation G.168 [15] at configuration 1. Figures 26 and 27 show the convergence as a function of time, displayed in the spectral domain. Figure 26 shows the convergence. As typical for echo cancellers the echo signal is decreased, first in the low frequency domain. After about 2 s the echo signal is sufficiently low and the non linear processor (NLP) is activated. **It however should be noted that the echo loss in this condition is only 32 dB, which is sufficient for 70 ms delay but not for higher delays which may occur in the connection.** Figure 27 shows accidental switching of the NLP after the echo canceller was fully converged. This will result in "bursty" echoes during single talk and certainly degrades the speech quality. The echo loss in this conditionis only 25 dB which is not sufficient according to ITU-T Recommendation G.131 [13]. The spectral components of the echo components are the same as they can be found in figure 26 for times where the echo canceller was not fully converged.

Additional tests were carried out by introducing an additional delay between PABX and terminal and such "scratching" the limits of the processing window of the echo canceller. This leads (as expected) to poor echo loss of 12 dB (no echo loss enhancement provided by the echo canceller) and accidental switching of the NLP as seen in figure 28.

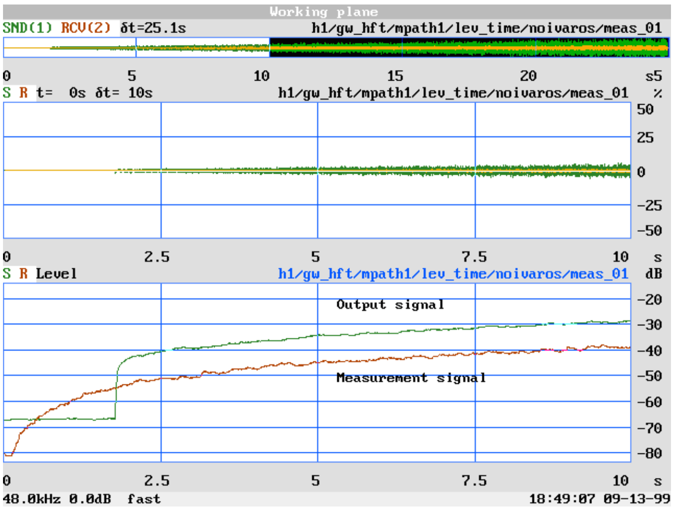


Figure 26: Convergence vs. time, configuration 1 (red: measurement signal, green: echo signal) upper: time signal; lower: spectral echo loss vs.time

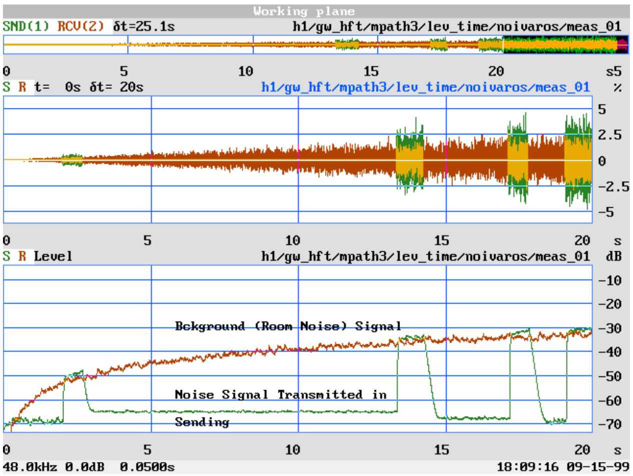


Figure 27: Echo loss vs. time, EC converged, after 11 s, configuration 1
(red: measurement signal, green: echo signal)
upper: time signal; lower: spectral echo loss vs.time

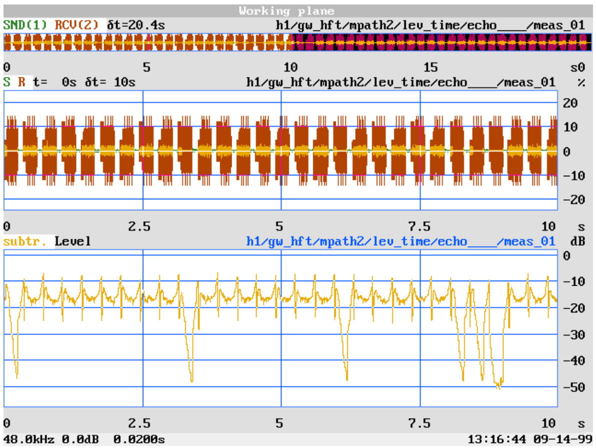


Figure 28: Echo loss vs. time when exceeding the processing window
of the echo canceller, configuration 1
(red: measurement signal, green: echo signal)
upper: time signal; lower: echo loss vs. time

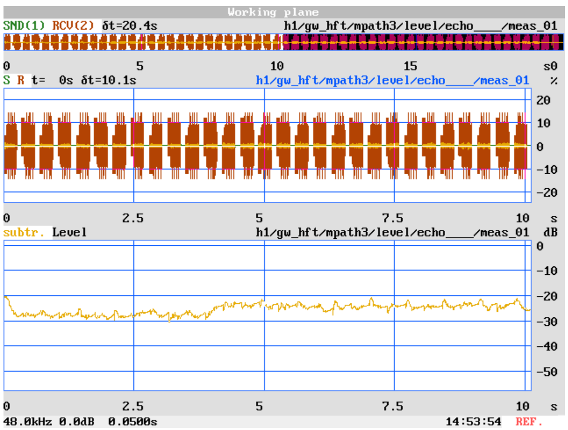


Figure 29: Echo loss vs. time, configuration 2
(red: measurement signal, green: echo signal)
upper: time signal; lower: echo loss vs. time

5.9.1.2.4 Performance in sending direction in the presence of background noise

A very importance performance parameters is the system behaviour in the presence of background noise. Some measurement examples are shown in figures 30 and 31. For configuration 1 figure 30 shows, that the background noise (indicated in red) is transmitted only when a certain level is reached. After reaching this threshold the background signal is transmitted without interruptions.

In contrast to that the background noise transmission in sending for configuration 2 is very poor. As it can be seen from figure 31, the background noise is constantly interrupted for various levels.

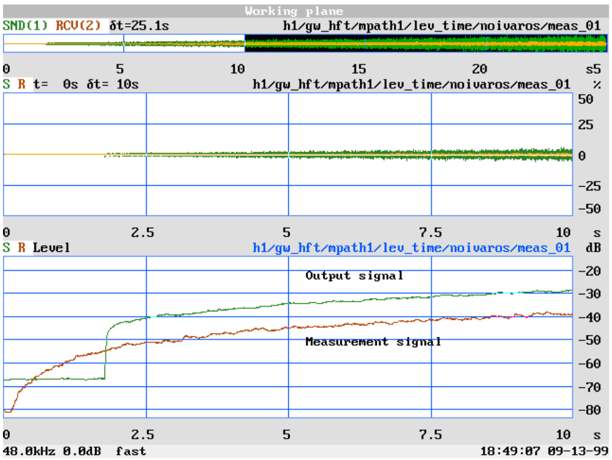


Figure 30: Background noise transmission in sending, lever versus time, configuration 1 (red: measurement signal, green: output signal at the line interface) upper: time signal; lower: level vs. time

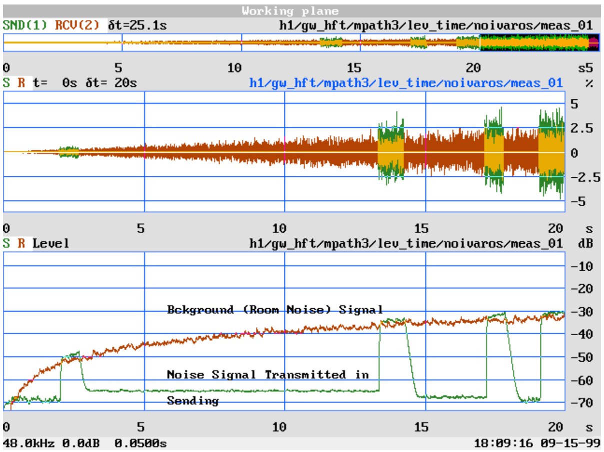


Figure 31: Background noise transmission in sending, lever versus time, configuration 2 (red: measurement signal, green: output signal at the line interface) upper: time signal; lower: level vs. time

5.9.1.2.5 Background noise performance in receiving direction

Configuration 1 behaves in receiving direction similar to the sending direction. Such no speech degradation should be expected. Configuration 2 however shows quite a strong AGC characteristics as it can be seen in figure 16.

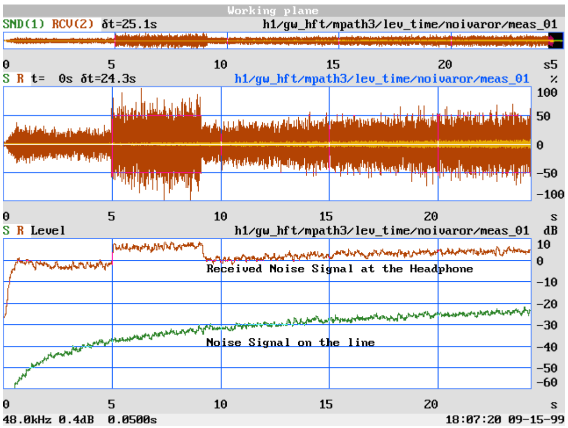


Figure 32: Background noise transmission in receiving, level versus time, configuration 2 (green: measurement signal, red: output signal at the headphone) upper: time signal; lower: level vs. time

From this evaluation it can be expected that even low level background noise signals are transmitted to the listeners ear with nearly the same level than speech! Furthermore accidental amplification of the background noise signal occurs as it can be seen e.g. in the time interval from 5 to 6 s in figure 32. This will result in a very bad background noise performance, especially if this system is connected with an terminal providing the intermittent transmission behaviour as shown in figure 31.

5.9.1.2.6 Evaluation of Double Talk Conditions

The double talk evaluations were made using test sequences as described in new ITU-T Recommendation P.502 [30] and TR 101 329 [2]. The double talk test signal is shown in figures 33 and 34 and is inserted simultaneously in sending and receiving direction.

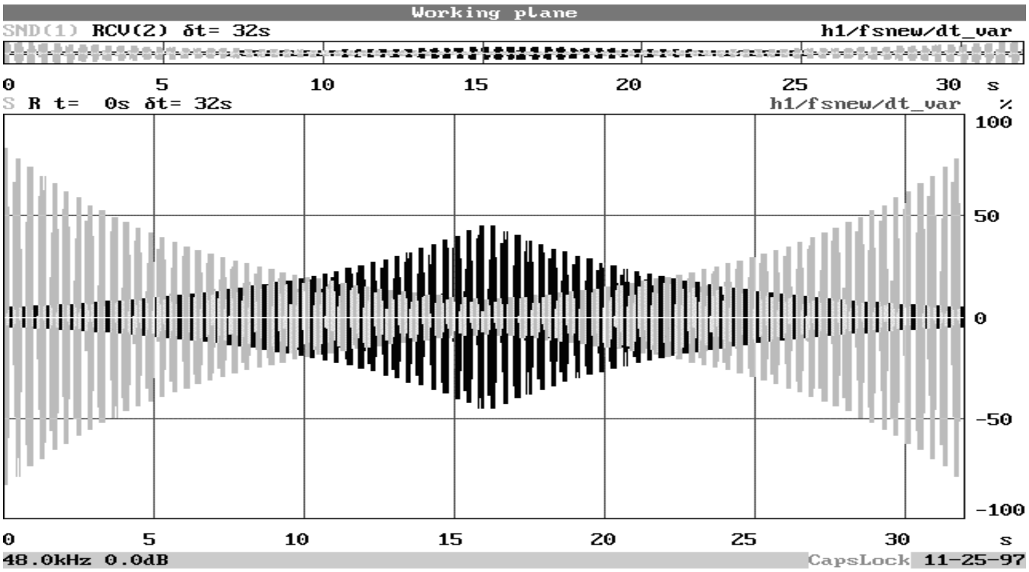
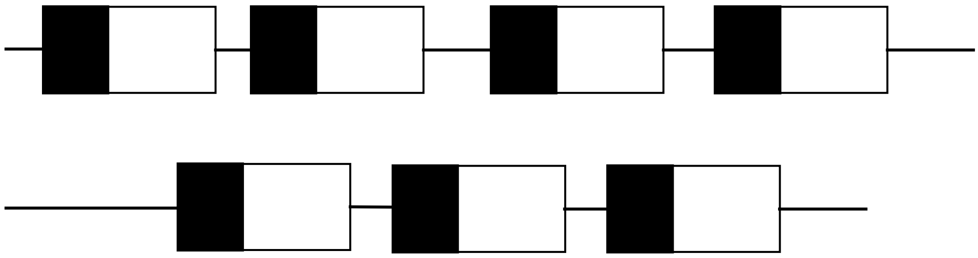


Figure 33: Overview of double talk test signal



s(t) - test signal sending direction
$s_{dt}(t)$ - double talk test signal in receiving direction

Figure 34: Cut out of the complete measurement sequence with detailed view on the overlap of sending and receiving direction signal, principle arrangement

By using this test signal the speech-like CSS Sequences are fed into sending and receiving simultaneously with varying level. The level is increasing (or decreasing) by 0,5 dB from CSS package to CSS package. Thus the level variation during double talk is 20 dB in both directions.

Some measurement results for configuration 1 are given below.

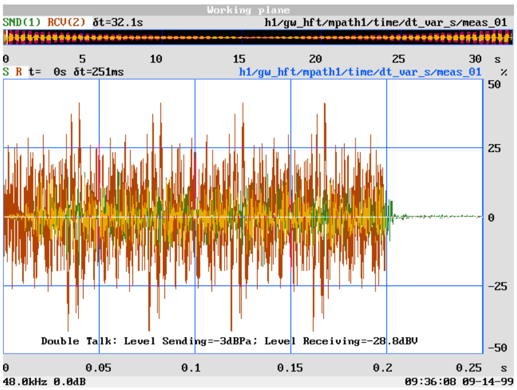


Figure 35: Transmission during double talk in sending for high sending level, configuration 1 (red: measurement signal, green: output signal at the line interface), time signal

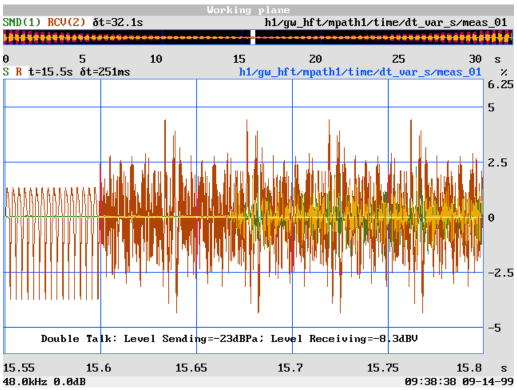


Figure 36: Transmission during double talk in sending for low sending level, configuration 1 (red: measurement signal, green: output signal at the line interface), time signal

For high signal levels in sending direction and low levels in receiving no signal clipping or echo problems are obvious. For low signal levels in sending the sending signal is subject to clipping, the switching time is about 120 ms. A similar behaviour can be found in receiving direction. Due to the low signal levels this is not critical, speech will be transmitted in double talk with no or only minor audible degradation.

The echo evaluations during double talk led to similar results than the ones found in single talk conditions and such are not discussed here in detail.

Due to the very poor echo loss of configuration 2 already in single talk conditions the more detailed investigation of the double talk characteristics is discussed not more in detail. The performance of the system in double talk is as poor as it is in single talk condition.

5.9.1.2.7 Relationship to subjective tests

The configurations investigated here were not subject to extensive subjective tests. Although some experts tests had been made which led to similar results than they were achieved in the objective measurement. All the effects which were found by the instrumental methods were the ones which were found subjectively as well. All the delay, echo and background noise problems as well as the switching found in configuration 2 were reported. In a similar way the configuration 1 was judged good which can be confirmed by the measurements conducted. The switching and echo problems for configuration 1 in the case when the processing window of the echo canceller was exceeded can be confirmed by the subjects in the same way.

5.9.1.3 Conclusion

Two IP configurations were investigated by the test methods defined in TR 101 329 [2] and the new ITU-T Recommendations P.502 [30] and P.581 [31]. The test results show that the methods seem to work good for the diagnostic evaluation of end to end scenarios using IP-connections. A detailed investigation of various impairments especially in configurations including terminals is possible by using the new methodologies and signals. The background noise performance, the switching characteristics in various conditions, the double talk performance and especially the echo characteristics can be evaluated in great detail and such a very good estimation of the various impairments introduced is possible on an objective basis.

Further work will be conducted in order to evaluate other end to end scenarios and to include the condition of packet loss and other impairments.

6 Further work

Collection of further relevant VoIP related measurements.

Bibliography

The following material, though not specifically referenced in the body of the present document (or not publicly available), gives supporting information.

- ETSI TR 101 329-6: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); End to End Quality of Service in TIPHON Systems; Part 6: Actual measurements of network and terminal characteristics and performance parameters in TIPHON networks and their influence on voice".
- ETSI ETR 250 (1996): "Transmission and Multiplexing (TM); Speech communication quality from mouth to ear for 3,1 kHz handset telephony across networks".
- ETSI EG 201 050 (V1.1): "Corporate telecommunication Networks (CN); Overall transmission planning for telephony on a Corporate Network".
- ETSI EN 201 377–1 (V1.1): "Speech Processing, Transmission and Quality Aspects (STQ); Specification and measurement of speech transmission quality; Part 1: Introduction to objective comparison measurement methods for one-way speech quality across networks".
- ETSI EG 202 306 (V1.2): "Transmission and Multiplexing (TM); Access networks for residential customers".
- ETSI I–ETS 300 245: "Integrated Services Digital Network (ISDN); Technical characteristics of telephony terminals".(all parts).
- ITU–T Recommendation E.164 (1997): "The international public telecommunication numbering plan".
- ITU–T Recommendation E.600 (1993): "Terms and definitions of traffic engineering".
- ITU–T Recommendation G.100 (1993): "Definitions used in Recommendations on general characteristics of international telephone connections and circuits".
- ITU–T Recommendation G.111 (1993): "Loudness ratings (LRs) in an international connection".
- ITU–T Recommendation G.114 (05/00): "One-way transmission time".
- ITU–T Recommendation G.121 (1993): "Loudness ratings (LRs) of national systems".
- ITU–T Recommendation G.122 (1993): "Influence of national systems on stability and talker echo in international connections".
- ITU–T Recommendation P.76 (1988): "Determination of loudness ratings; fundamental principles".
- ITU–T Recommendation P.79 (1993): "Calculation of loudness ratings for telephone sets".
- ITU–T Recommendation P.310 (1996): "Transmission characteristics for telephone band (300-3 400 Hz) digital telephones".
- ITU-T Recommendation P.340: "Transmission characteristics of hands-free telephones".
- ITU–T Recommendation P.561 (1996): "In-service, non-intrusive measurement device - Voice service measurements".
- ITU–T Recommendation P.830 (1996): "Subjective performance assessment of telephone–band and wideband digital codecs".

History

Document history		
V1.1.1	July 2000	Publication