ETSI TR 101 329-6 V2.1.1 (2002-02)

Technical Report

Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; 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

RTR/TIPHON-05013

Keywords

IP, network, performance, QoS, quality, speech, 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://portal.etsi.org/tb/status/status.asp

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://webapp.etsi.org/IPR/home.asp).

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 is part 6 of a multi-part deliverable covering End-to-end Quality of Service in TIPHON systems, as identified below:

TR 101 329-1: "General aspects of Quality of Service (QoS)";

TS 101 329-2: "Definition of speech Quality of Service (QoS) Classes";

TS 101 329-3: "Signalling and control of end-to-end Quality of Service (QoS)";

TS 101 329-5: "Quality of Service (QoS) measurement methodologies";

TR 101 329-6: "Actual measurements of network and terminal characteristics and performance parameters in TIPHON networks and their influence on voice quality";

TR 101 329-7: "Design guide for elements of a TIPHON connection from an end-to-end speech transmission performance point of view".

Quality of Service aspects of TIPHON Release 4 and 5 Systems will be covered in TS 102 024 and TS 102 025 respectively, and more comprehensive versions of the Release 3 documents listed above will be published as part of Release 4 and 5 as work progresses.

Introduction

The present document forms one of a series of technical specifications and technical reports produced by TIPHON Working Group 5 addressing Quality of Service (QoS) in TIPHON Systems. The structure of this work is illustrated in Figure 1.

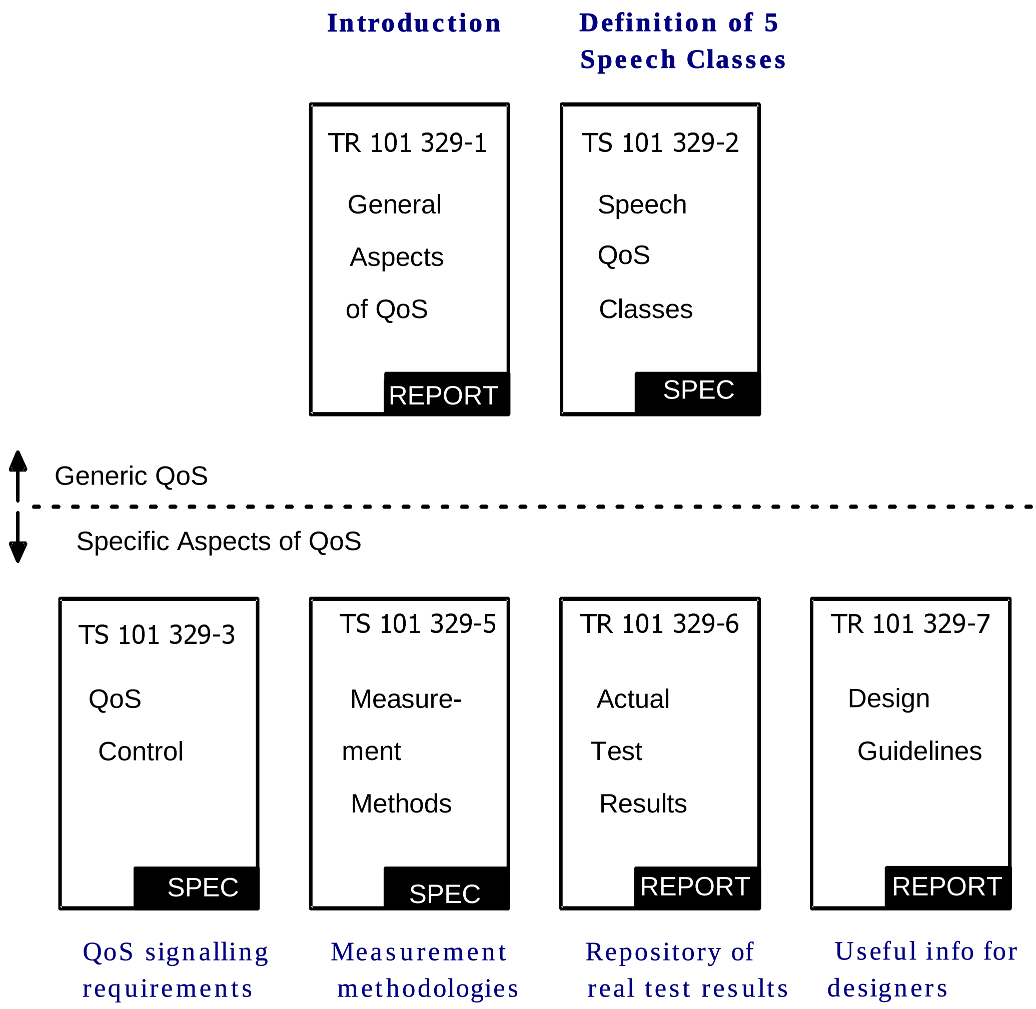


Figure 1: Structure of TIPHON QoS Documentation for Release 3

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.

The present document covers measurement results provided to TIPHON during the years 1999 to 2001, which have contributed to the measurement methodologies in 101 329-5 [5] as well as providing design parameters in 101 329-7 [6].

2 References

For the purposes of this Technical Report (TR) the following references apply:

[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 (V2.1.1): "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) Release 3; 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) Release 3; End-to-end Quality of Service in TIPHON systems; Part 2: Definition of speech Quality of Service (QoS) classes".

[5] ETSI TR 101 329-5: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 5: Quality of Service (QoS) measurement methodologies".

[6] ETSI TR 101 329-7: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; 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 (V1.1.1): "Corporate Telecommunication Networks (CN); Transmission characteristics of digital Private Branch eXchanges (PBXs)".

[8] ETSI GTS 06.10 (V3.2.0): "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: "Application of the E-model: A planning guide".

[11] ITU–T Recommendation G.109 (1999): "Definition of categories of speech transmission quality".

[12] ITU–T Recommendation G.113 (2001): "Transmission impairments due to speech processing".

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

[14] ITU–T Recommendation G.165 (1993): "Echo cancellers".

[15] ITU–T Recommendation G.168 (2000): "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 (1996): "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-bit/ sample embedded adaptive differential pulse code modulation (ADPCM)".

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

[22] ITU–T Recommendation G.729 (1996): "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)".

[23] ITU–T Recommendation G.729A (Annex A - 1996): "Reduced complexity 8 kbit/s CS-ACELP speech codec".

[24] ITU–T Recommendation G.729B (Annex B - 1996): "A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70".

[25] ITU–T Recommendation H.323 (2000): "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 (1999): "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 (HATS) 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 Proposed American National Standard - A Packet Loss Concealment Technique for Use with ITU-T Recommendation G.711" (Source: AT&T).

[37] ETSI ETS 300 581-2: "Digital cellular telecommunications system (Phase 2) (GSM); Half rate speech; Part 2: Half rate speech transcoding (GSM 06.20 version 4.3.1)".

[38] ETSI EN 300 726: "Digital cellular telecommunications system (Phase 2+) (GSM); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60 version 8.0.1 Release 1999)".

[39] ITU-T Recommendation P.862 (2001): "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs".

[40] ITU-T Recommendation G.723: "Extensions of Recommendation G.721 adaptive differential pulse code modulation to 24 and 40 kbit/s for digital circuit multiplication equipment application".

[41] ITU-T Recommendation G.723.1-A: "Speech coders : Silence compression scheme".

[42] ITU-T Recommendation P.50: "Artificial voices".

[43] ITU-T Recommendation G.114: "One-way transmission time".

3 Abbreviations

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

ACR	Absolute Category Rating
ASL	Active Speech Level
CAS	Communication Analysis System (HEAD acoustics test system)
CSS	Composite Source Signal
EC	Echo Canceller
EP	Error Pattern
GSM	Global System for Mobile communications
GSM EFR	GSM Enhanced Full Rate Speech Coder
GSM FR	GSM Full Rate Speech Coder
HATS	Head And Torso Simulator
IP	Internet Protocol
IRS	Intermediate Reference System
ISDN	Integrated Services Digital Network
JLR	Junction Loudness Rating
LAN	Local Area Network
MNRU	Modulated Noise Reference Unit
MOS	Mean Opinion Score
NLP	Non-Linear Processor
OLR	Overall Loudness Rating
OVL	Over-Load Point
PESQ	Perceptual Evaluation of Speech Quality (see ITU-T Recommendation P.862)
PLC	Packet Loss Concealment
PSTN	Public Switched Telephone Network
PVS	PC Voice Switch
QoS	Quality of Service
RLR	Receive Loudness Rating
SCN	Switched Communications Network
SLR	Send Loudness Rating
TMOS	TOSQA Mean Opinion Score (output of TOSQA)
TOSQA	Telecommunication Objective Speech Quality Assessment
VAD	Voice Activity Detection

4 List of Measurement Results

Table 1: 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]	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 15TD089	05/10/1999
5	Subjective Results on impairment effects of IP packet loss	Nortel Networks paulcov@nortelnetworks.com	ETSI TIPHON 17TD167	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 [34]	ETSI TIPHON 17TD135	15/03/2000
7	Anonymous Test report of ETSI Speech Quality Test Event 2000	Deutsche Telekom, T-Nova; HEAD acoustics	ESTI TIPHON 22TD38	26/03/2001
8	Problems with the behaviour of jitter buffers and their influence on the end-to-end speech quality	Pieter Veenstra p.k.veenstra@kpn.com	ESTI TIPHON 22TD47	26/03/2001

5 General Measurement Results

5.1 Subjective Testing

5.1.1 Simulation Results of VoIP Scenarios

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

5.1.1.1 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.

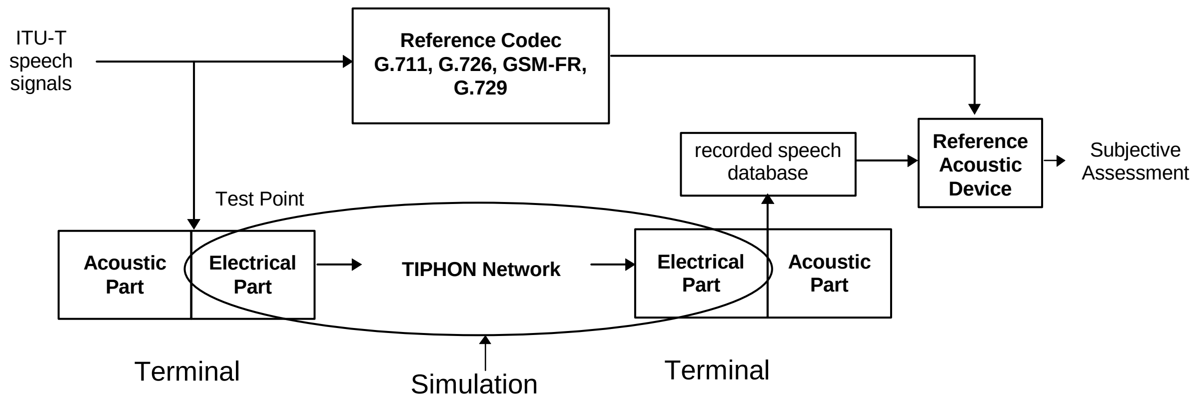


Figure 2: Simulation methodology for testing TIPHON speech quality

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.

5.1.1.2 Measurement Set Up

5.1.1.2.1 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 = -26 dB 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 (8 000 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 (10 ms, 20 ms,...) 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.

5.1.1.2.2 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.

Table 2: G.711 [16] + Codec + G.7xx (Phone to Phone Scenario in fixed network environments)

Codec	Packet Loss	Speech Frame Size	Nr. of Frames/Packet	Substitution
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

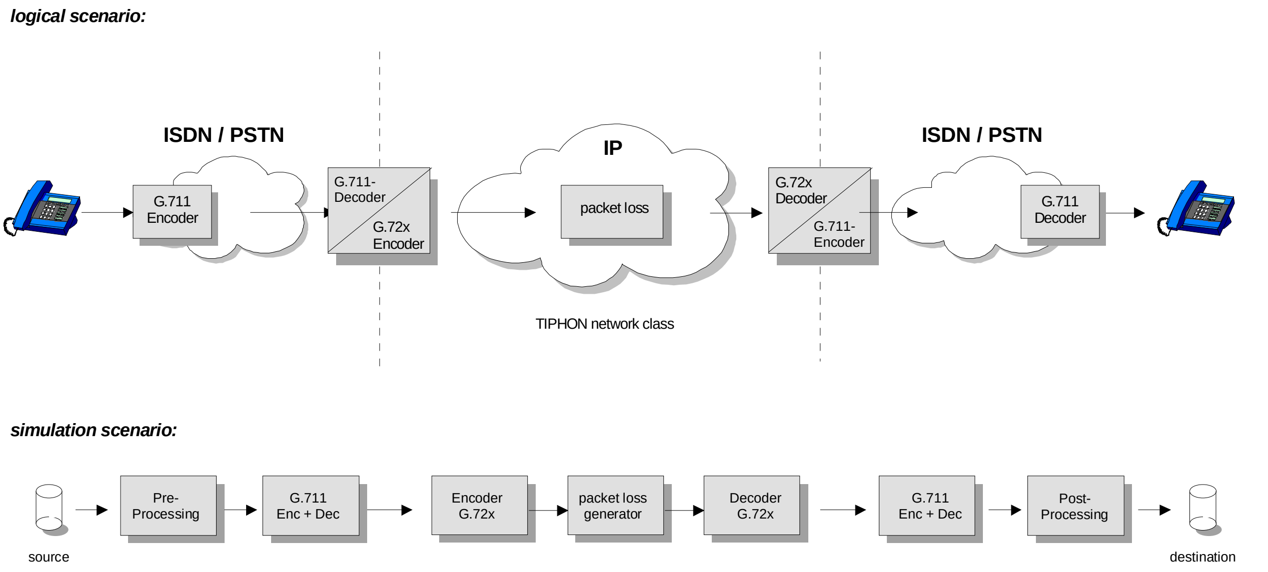


Figure 3: Processing scenario for single codec conditions

Table 3: Tandem Configuration with GSM x FR (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 [18] (5.3)	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 [18] (5.3)	0 %, 5 %, 10 %	G.723.1 [18]: 30 ms	2	G.723.1 [18] internal

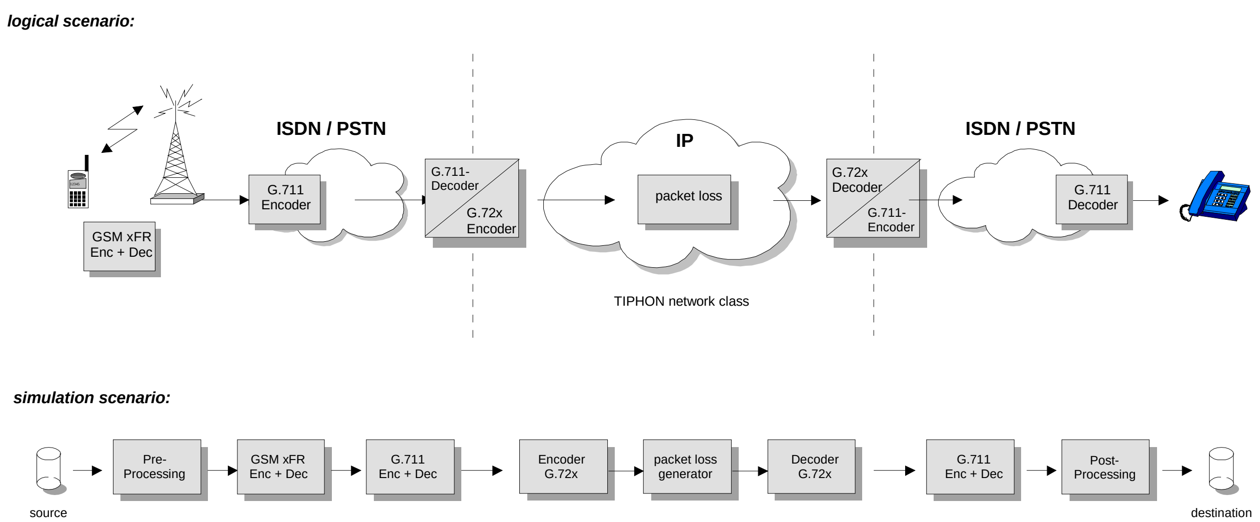


Figure 4: Processing scenario for tandem codec conditions

5.1.1.3 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.

5.1.1.3.1 G.711 + Codec + G.711

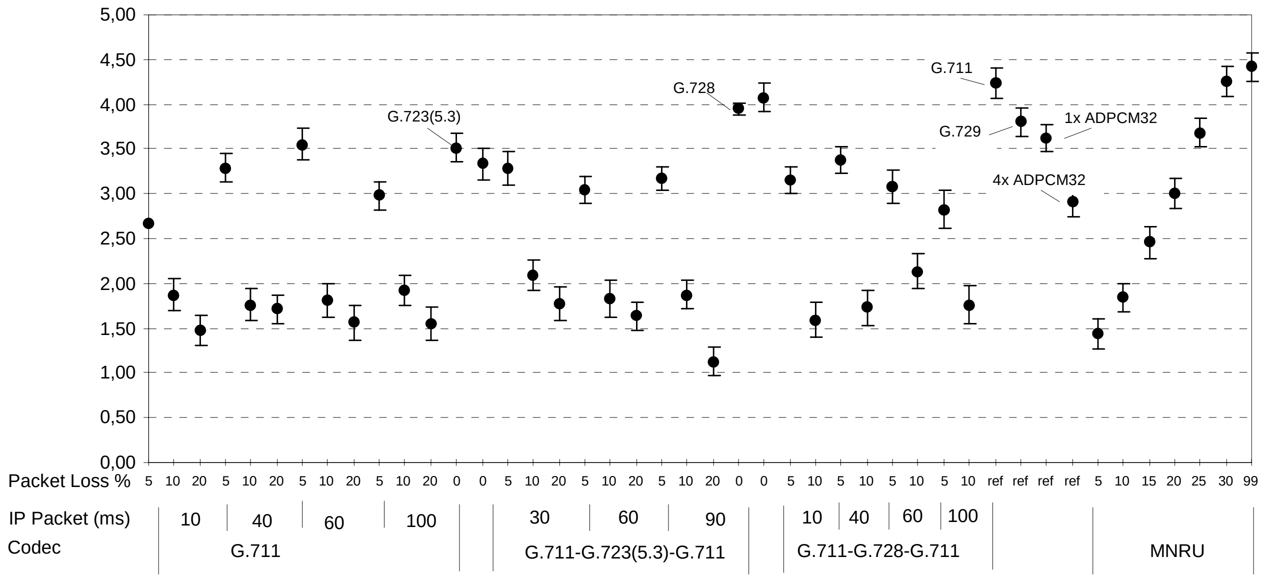


Figure 5: Subjective assessment result

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.

5.1.1.3.2 Tandem Conditions with GSM x FR

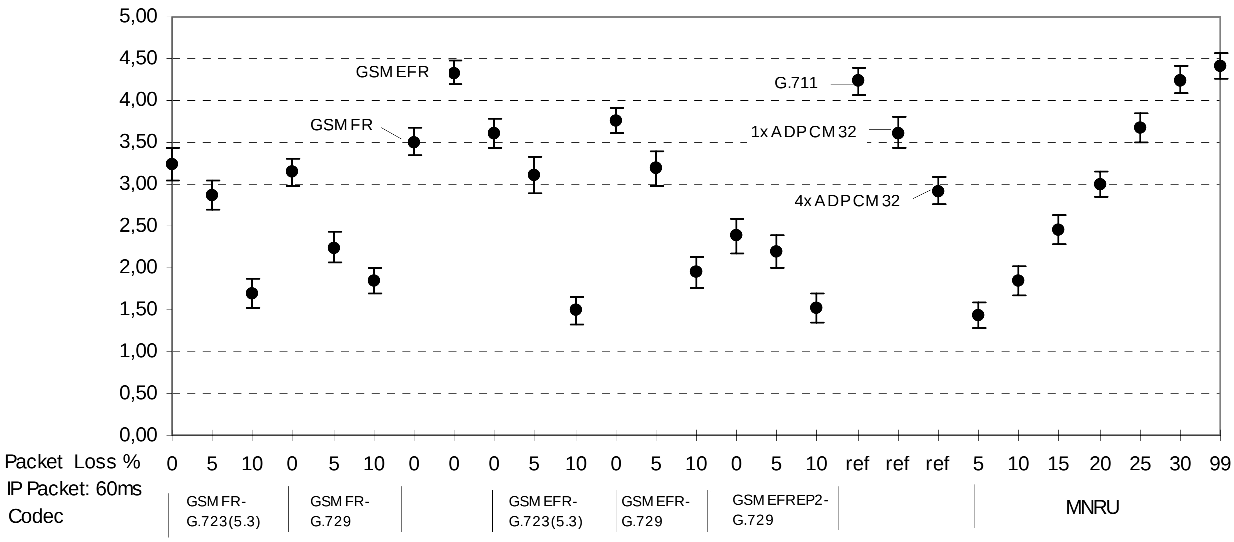


Figure 6: Tandem conditions with GSM x FR

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.

5.1.2 Subjective Results on impairment effects of IP packet loss

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

5.1.2.1 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.

5.1.2.2 Measurement Set Up

5.1.2.2.1 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.

5.1.2.2.2 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) [38]. For G.711 [16] and G.729 [22], speech frame sizes of 10 ms and 20 ms were tested, while G.723.1 [18] and GSM EFR were tested with their standard frame sizes (30 ms and 20 ms 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 [35] and proposed as an ANSI standard 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.

5.1.2.2.3 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 (MOS) were computed from the ratings assigned to each test case.

5.1.2.2.4 MOS Data and MNRU Conditions

The MOS and MNRU equivalence data for each of the conditions are summarized in clause 5.1.2.3. 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.

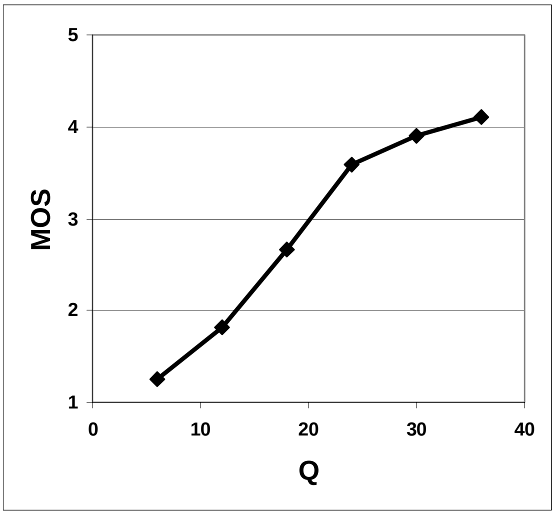


Figure 7: MOS as a function of Q in the MNRU conditions

5.1.2.3 Results

5.1.2.3.1 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.

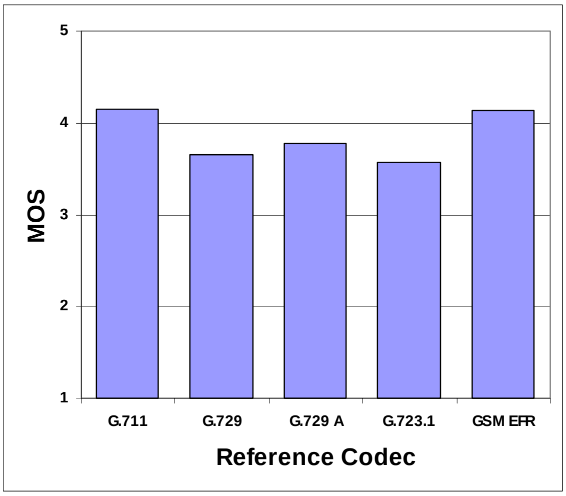


Figure 8: MOS values for each codec without packet loss impairments

5.1.2.3.2 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 ms. 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 ms (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.

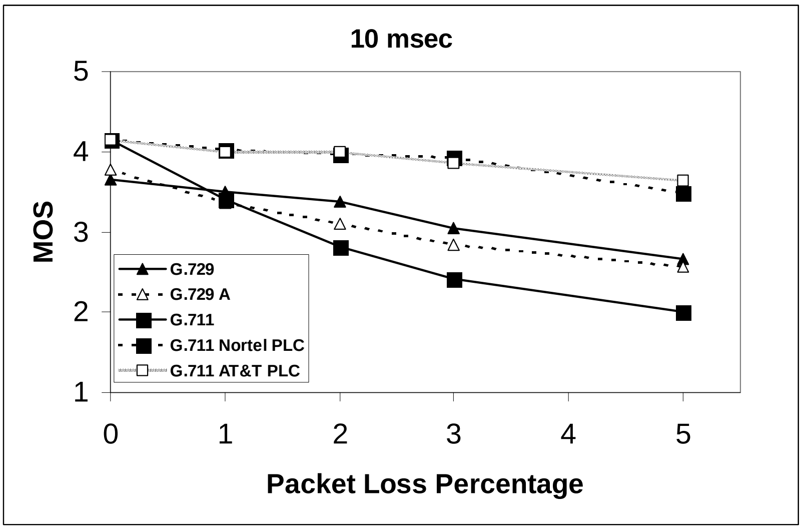
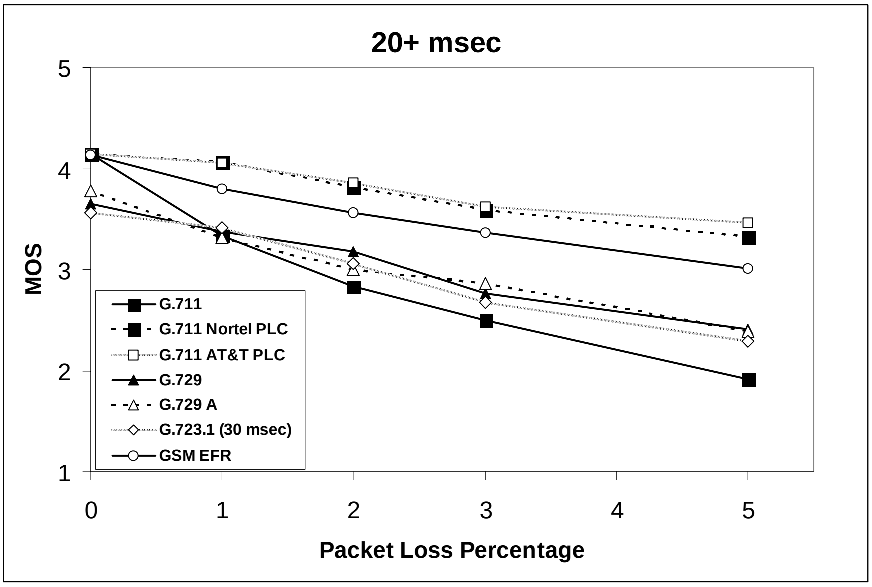


Figure 9: Effects of packet loss on voice quality with 10 ms packets



NOTE: All packet sizes were 20 ms except G.723.1 [18], which was 30 ms.

Figure 10: Effects of packet loss on voice quality with 20 ms or larger packets

5.1.2.3.3 Tables of Results

The parameters for the MNRU curve fit were: A = 1,6; B = 1,6; C = 9,5; Qm = 17,3.

Table 4: MOS and MNRU equivalence scores

Packet (ms)	Codec	bps	PLC Algorithm	Packet Loss	MOS	MNRU Q
10	G.729 [22]	(8 kbit/s)	included	1 %	3,50	23,37
10	G.729 [22]	(8 kbit/s)	included	2 %	3,38	22,34
10	G.729 [22]	(8 kbit/s)	included	3 %	3,05	20,08
10	G.729 [22]	(8 kbit/s)	included	5 %	2,67	17,70
10	G.729A [23]	(8 kbit/s)	included	1 %	3,37	22,27
10	G.729A [23]	(8 kbit/s)	included	2 %	3,10	20,40
10	G.729A [23]	(8 kbit/s)	included	3 %	2,85	18,78
10	G.729A [23]	(8 kbit/s)	included	5 %	2,57	17,13
10	G.711 [16] u-law	(64 kbit/s)	none	1 %	3,41	22,57
10	G.711 [16] u-law	(64 kbit/s)	none	2 %	2,82	18,60
10	G.711 [16] u-law	(64 kbit/s)	none	3 %	2,41	16,18
10	G.711 [16] u-law	(64 kbit/s)	none	5 %	2,00	13,55
10	G.711 [16] u-law	(64 kbit/s)	Nortel PLC	1 %	4,02	30,83
10	G.711 [16] u-law	(64 kbit/s)	Nortel PLC	2 %	3,97	29,45
10	G.711 [16] u-law	(64 kbit/s)	Nortel PLC	3 %	3,93	28,61
10	G.711 [16] u-law	(64 kbit/s)	Nortel PLC	5 %	3,49	23,23
10	G.711 [16] u-law	(64 kbit/s)	AT&T PLC	1 %	4,00	30,16
10	G.711 [16] u-law	(64 kbit/s)	AT&T PLC	2 %	3,99	30,03
10	G.711 [16] u-law	(64 kbit/s)	AT&T PLC	3 %	3,86	27,43
10	G.711 [16] u-law	(64 kbit/s)	AT&T PLC	5 %	3,63	24,60
20	G.729 [22]	(8 kbit/s)	included	1 %	3,38	22,34
20	G.729 [22]	(8 kbit/s)	included	2 %	3,18	20,89
20	G.729 [22]	(8 kbit/s)	included	3 %	2,77	18,28
20	G.729 [22]	(8 kbit/s)	included	5 %	2,40	16,13
20	G.729A [23]	(8 kbit/s)	included	1 %	3,32	21,90
20	G.729A [23]	(8 kbit/s)	included	2 %	3,00	19,74
20	G.729A [23]	(8 kbit/s)	included	3 %	2,87	18,89
20	G.729A [23]	(8 kbit/s)	included	5 %	2,39	16,07
20	G.711 [16] u-law	(64 kbit/s)	None	1 %	3,33	21,98
20	G.711 [16] u-law	(64 kbit/s)	None	2 %	2,83	18,66
20	G.711 [16] u-law	(64 kbit/s)	None	3 %	2,50	16,69
20	G.711 [16] u-law	(64 kbit/s)	None	5 %	1,92	12,99
20	G.711 [16] u-law	(64 kbit/s)	Nortel PLC	1 %	4,06	31,96
20	G.711 [16] u-law	(64 kbit/s)	Nortel PLC	2 %	3,82	26,82
20	G.711 [16] u-law	(64 kbit/s)	Nortel PLC	3 %	3,59	24,14
20	G.711 [16] u-law	(64 kbit/s)	Nortel PLC	5 %	3,33	21,96
20	G.711 [16] u-law	(64 kbit/s)	AT&T PLC	1 %	4,05	31,76
20	G.711 [16] u-law	(64 kbit/s)	AT&T PLC	2 %	3,86	27,34
20	G.711 [16] u-law	(64 kbit/s)	AT&T PLC	3 %	3,62	24,44
20	G.711 [16] u-law	(64 kbit/s)	AT&T PLC	5 %	3,47	23,07
20	GSM EFR	(12,2 kbit/s)	included	1 %	3,80	26,48
20	GSM EFR	(12,2 kbit/s)	included	2 %	3,57	23,94
20	GSM EFR	(12,2 kbit/s)	included	3 %	3,36	22,23
20	GSM EFR	(12,2 kbit/s)	included	5 %	3,01	19,80
30	G.723.1 [**18**]	(6,3 kbit/s)	included	1 %	3,42	22,64
30	G.723.1 [**18**]	(6,3 kbit/s)	included	2 %	3,05	20,08
30	G.723.1 [**18**]	(6,3 kbit/s)	included	3 %	2,67	17,74
30	G.723.1 [**18**]	(6,3 kbit/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.729A [23]		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

5.2 Objective Testing

5.2.1 Objective Results of SuperOp event 1999 in Hawaii

Source: Rapporteur of ITU-T Recommendation Q.13/12 [34]; ETSI TIPHON 17TD135.

5.2.1.1 Introduction

This clause presents results from the ITU-T Recommendation Q.13/12 [34] 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 Recommendation Q.13/12 [34] 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.

5.2.1.2 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.

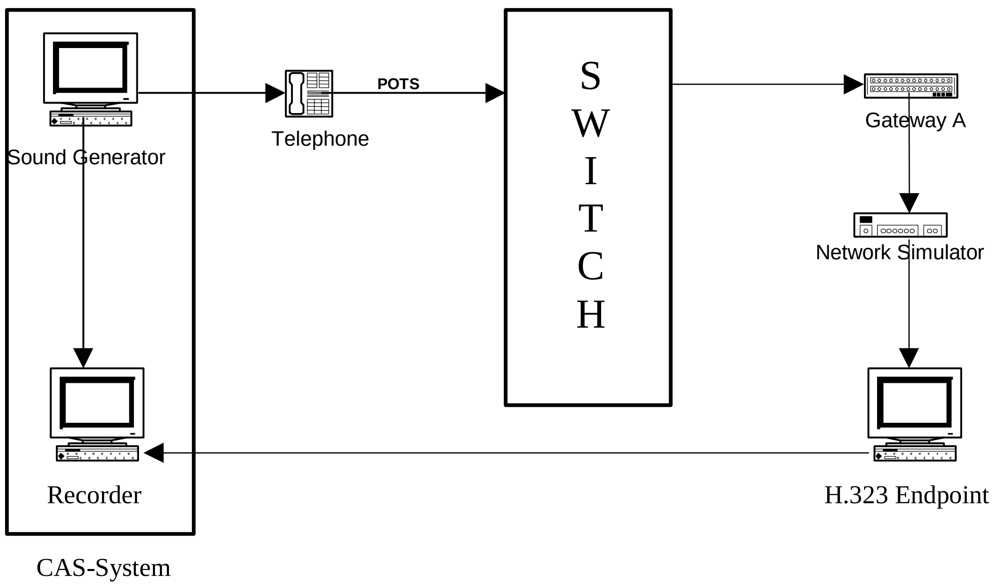


Figure 11: Recording of speech samples with TIPHON scenario 2

After recording of the speech samples, the 32 s files were separated into samples of approximately 8 s 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.

5.2.1.3 Quality Assessment

Within the ITU-T Recommendation Q.13/12 [34] 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 was 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 result 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 5.

5.2.1.4 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 5: Correlation coefficient for the ITU-T Recommendation Q.13/12 [34] objective speech quality assessment algorithm

Rank order	Correlation coefficient
1	0,9879
2	0,9859
3	0,9489
4	0,9275
5	0,8377
P.861 [33]	0,8706

Table 5 shows clearly that the P.861 [33] algorithm is less accurate then other instrumental assessment methods.

P.861 [33] is withdrawn by ITU-T in 2001. It has been replaced by P.862 [39] in 2001.

5.2.2 Objective Results of ETSI Speech Quality Test Event 2000

Source: Anonymous Test Report of the 1st ETSI Speech Quality Test Event; T-Nova and HEAD acoustics, 22TD38.

5.2.2.1 Introduction

This clause describes the test methodologies and the objectice results of the measurements which were carried out during the 1st ETSI Speech Quality Test Event. The tests were conducted by T-Nova Deutsche Telekom Innovations-gesellschaft mbH Berkom (T-Nova) in collaboration with HEAD acoustics GmbH (HEAD acoustics).

The aim of the test event was to determine the speech quality of various voices over IP equipment under certain IP network conditions. During the test event, speech material as well as measurement data were collected by transferring voice samples and artificial signals across the Voice over IP Set Up. This material was analysed and the results are reported in this chapter.

The one-way speech transmission quality was evaluated by processing real speech samples and analysing it using the TOSQA algorithm. TOSQA leads to MOS-comparable results. To validate the TOSQA results an auditory reference test was carried out.

5.2.2.2 Measurement Set Up

Measurements for one way speech transmission quality were mainly conducted by "electrical-electrical" measurements. For the "electrical-electrical" measurements two kinds of input signals were used: a) speech samples designed according to ITU-T Recommendation P.800 [32]; and b) test signals according to P.501 [29].

The input signals were transmitted and recorded simultaneously, i.e. the sending and receiving process were started at the same time. Therefore exact delay measurements were possible.

For all kind of measurements a packet loss generator (NIST Net) and a packet loss monitor were used. Both entities were controlled by the PVS measurement equipment. The measurement set up is shown in figure 12.

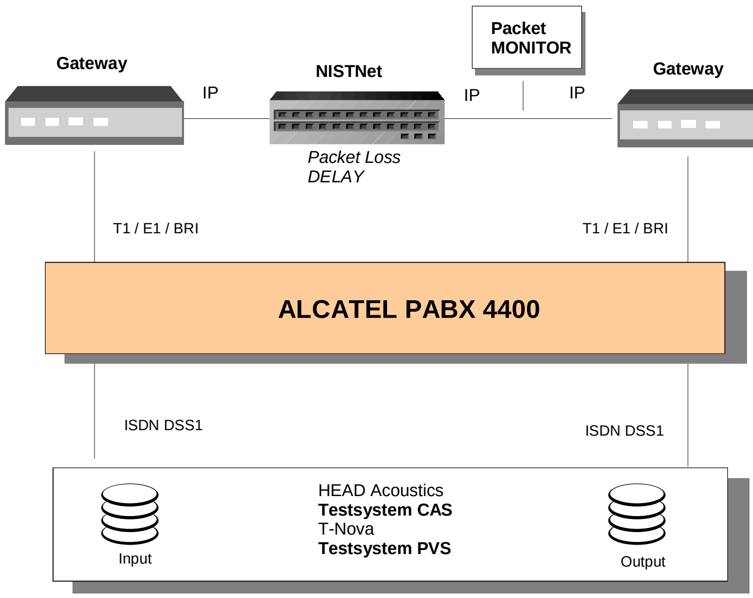


Figure 12: Electrical – Electrical Measurement Set Up

For all kinds of "electrical-electrical" measurements the following IP network conditions were used.

Table 6: Network Conditions for Electrical-Electrical Measurements

Condition	Packet Loss (Equal)	Additional Delay (see note)	Delay Variation
1	0	0	No
2	1 %	0	No
3	2 %	0	No
4	3 %	0	No
5	5 %	0	No
6	1 %	50 ms	20 ms
NOTE: Additional IP network delay was introduced by NIST Net.			

The additional delay in condition 6 was intended to have a fix delay value in order to ensure proper delay variation (jitter) handling.

In such jitter condition the test network can cause situations where packets are reordered, if the packet size is very small. This effect can be avoided with high probability by using a certain packet length, which should be at least three times higher than the delay variation itself.

5.2.2.3 Quality Assessment

The speech quality assessment of the speech samples was performed using the method for instrumental speech quality estimation of Deutsche Telekom, TOSQA. For verifying the instrumental results auditory reference assessments were performed using speech material in German.

The auditory test was performed in the Berlin speech quality labs at T-Nova Deutsche Telekom Innovationsgesellschaft mbH and was carried out according to ITU-T Recommendation P.800 [32]. Each circuit condition was tested with four different speech samples (short, concatenated sentences, read by 2 male and 2 female speakers each). Speech samples were played back to the subjects via a conventionally shaped standard telephone handsets in a low-noise test cabinet (room noise floor < 30 dB(A)), at a listening level of about 79 dB SPL. Test subjects were required to judge the overall listening quality on a 5-point ACR quality scale, as recommended in ITU-T Recommendation P.800 [32], with German scale labels. Before starting the first test session, 8 example speech samples were played back to the test subjects, which were expected to range in quality from good to bad.

In addition to the tested VoIP scenarios also 11 reference conditions were integrated in the test. These are single speech codecs (G.723.1 [18], G.729 [22]) as well as different MNRU conditions.

To avoid effects caused by presentation order, groups of test subjects listen to the samples in different randomized orders. All in all 25 test subjects took part on the auditory test. Four votes per each condition and test subject yield the Mean Opinion Sore (MOS) for the condition under test. So the MOS per condition is an average of 100 single votes.

In the similar manner as subjects assess the speech quality in the listening test, TOSQA rates the four speech samples for each condition. Here exactly the same speech material from the auditory test was used. TOSQA compares this transmitted and possibly distorted speech material with the clean input speech material as reference.

The four TOSQA results (TMOS) of the four speech samples were averaged and are the basis for comparison with the MOS values gained by the auditory test.

The graphic in figure 13 plot shows mean opinion scores gained by auditory experiments on the X axis. The Y axis shows the speech quality value TMOS calculated by TOSQA. The TMOS values were transformed in a common way by a third order monotonous mapping.

In an ideal prediction of quality, the instrumental results would be equal to the MOS-values. In this case all symbols would be on the 45° line in the diagram. All in all TOSQA shows a good prediction of speech quality, the correlation coefficient after third order mapping is 0,916.

This result is based on the four German samples which are used in the auditory reference test. In addition to the four German samples, also four English samples were transmitted via each network condition, followed by another 8 German samples; all of them were assessed by TOSQA. The resulting mean TMOS averaged over all 16 samples was also compared with MOS values gained by the auditory test. In this case the correlation coefficient between MOS and "mean TMOS" increases slightly to 0,921.

This value should be also compared with the requirement which was defined in the ITU-T Q13/12 for such kind of conditions (cf. ITU-T document COM12-117E, March 2000). During the competition for the new standard for objective speech quality measurement a minimum correlation coefficient of 0,80 was required for real VoIP recordings if the test corpus was not used for training the objective model. This is exactly the same situation we had in our VoIP test here.

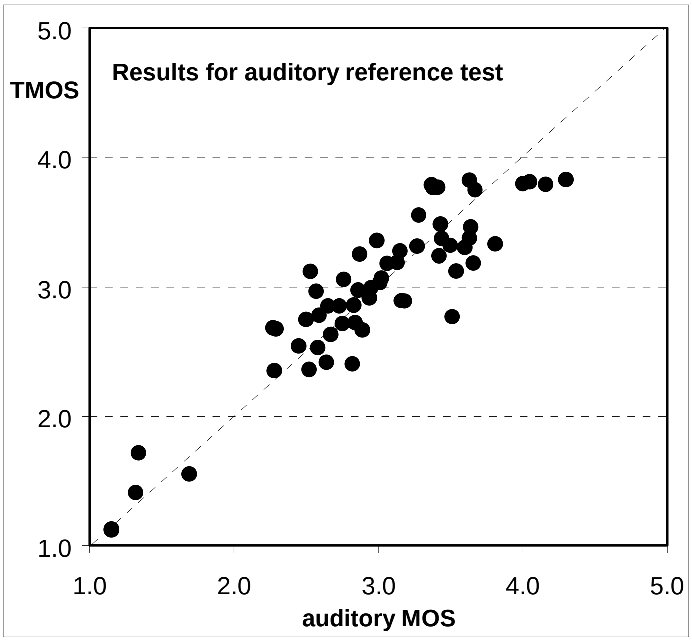


Figure 13: MOS versus TMOS for VoIP conditions and references

Figure 14 shows a more detailed scatter plot with the G.711 [16] conditions as unfilled circles, all other VoIP conditions as filled circles and the reference conditions as grey diamonds.

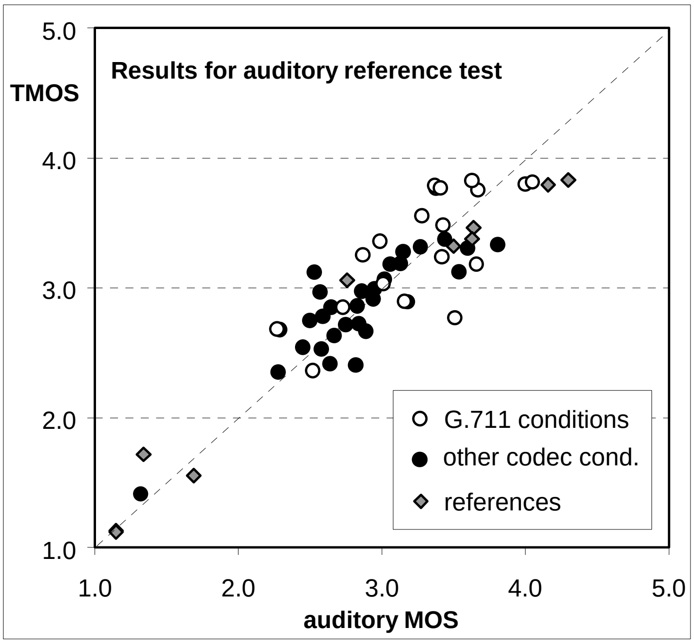


Figure 14: MOS versus TMOS for VoIP conditions and references, detailed plot

TOSQA calculates not only a speech quality value which is focused on quality during speech activity, but also an additional value called "connection quality: CQ". This algorithm is still under test and therefore not referenced in ITU-T. Here besides other values, a special weighting of intervals without speech activity is taken into account for the connection quality result. If the current version of this algorithm of TOSQA "connection quality" will be used for assessment, the correlation increases further to 0,935.

During the competition for a new standard for objective speech quality measurement within ITU-T Q13/12, also a VoIP database was assessed. This database was recorded during the ETSI Tiphon SuperOp, Hawaii in September 1999. Here a correlation of 0,949 between TOSQA and auditory MOS could be reached. In case of TOSQA "connection quality" this result is increasing to 0,965.

These higher correlations are caused by the smaller range of included conditions in this test. This so called "SuperOp-test" contains only G.711 [16] and G.723.1 [18] conditions under different rates of packet loss and delay jitter. There only two coding algorithm were tested and only one manufacturer was involved in the test.

5.2.2.4 TOSQA Results

Voice samples were processed across the VoIP scenario for the estimation of the speech quality. Because of the random behaviour of the IP network emulation (NIST Net) 16 voice samples (8 seconds voice each) have been processed to average the random influence for each condition. Nevertheless, there may be still some cases where this random behaviour influences the results.

For analysis of the conditions under test (C01...C06), the TMOS values of all 16 speech samples, German and English, were averaged.

Tables 7, 8 and 9 show the speech quality estimation from the codec types which were mostly used in the participants VoIP equipment.

The tables give also the minimum and maximum TMOS values for each condition. In some cases there is a large difference between TMOS min and TMOS max. This is caused by the random influence of packet loss, the point of time when the packet loss occurs, the influence of VAD and the PLC algorithm itself.

5.2.2.4.1 G.711 codec

Table 7 shows the average test results from C1 to C6 for the G.711 [16] codec.

Table 7: Average Test Results for G.711 [16], VAD on, PLC on/off, PL = 20 ms to 30 ms

Condition	TMOS average	TMOS min	TMOS max
C1: Drop = 0 %, Delay = 0 ms, Jitter = 0 ms	4,2	4,2	4,2
C2: Drop = 1 %, Delay = 0 ms, Jitter = 0 ms	3,7	3,3	4,0
C3: Drop = 2 %, Delay = 0 ms, Jitter = 0 ms	3,4	3,1	3,8
C4: Drop = 3 %, Delay = 0 ms, Jitter = 0 ms	3,4	2,8	3,7
C5: Drop = 5 %, Delay = 0 ms, Jitter = 0 ms	3,1	2,7	3,5
C6: Drop = 1 %, Delay = 50 ms, Jitter = 20 ms	3	2,7	3,3

As expected the TMOS values are decreasing with higher packet loss values. In general the results show that for the G.711 [16] codec the reached TMOS values, especially for C1, are fully comparable with the simulated reference values.

5.2.2.4.2 G.723 codec

Table 8 shows the test results from C1 to C6 for the G.723 codec.

Table 8: Average Test Results for G.723, VAD on, PLC on/off, PL = 30 ms

Condition	TMOS average	TMOS min	TMOS max
C1: Drop = 0 %, Delay = 0 ms, Jitter = 0 ms	3,3	3,2	3,5
C2: Drop = 1 %, Delay = 0 ms, Jitter = 0 ms	3,2	3,1	3,4
C3: Drop = 2 %, Delay = 0 ms, Jitter = 0 ms	3,1	3,0	3,2
C4: Drop = 3 %, Delay = 0 ms, Jitter = 0 ms	3,0	2,9	3,1
C5: Drop = 5 %, Delay = 0 ms, Jitter = 0 ms	2,8	2,7	2,9
C6: Drop = 1 %, Delay = 50 ms, Jitter = 20 ms	2,7	2,0	3,1

Normally one may expect better values, especially for condition C1. The result of condition 1 is mainly caused by some bad voice samples where the codec seemed to be desynchronized.

5.2.2.4.3 G.729 codec

Table 9 shows the average test results from C1 to C6 for all vendors for the G.729 [22] codec.

Table 9: Average Test Results for G.729 [22], VAD on, PLC on/off, PL = 10 ms to 30 ms

Condition	TMOS average	TMOS min	TMOS max
C1: Drop = 0 %, Delay = 0 ms, Jitter = 0 ms	3,3	3,1	3,5
C2: Drop = 1 %, Delay = 0 ms, Jitter = 0 ms	3,1	2,9	3,3
C3: Drop = 2 %, Delay = 0 ms, Jitter = 0 ms	3,0	2,9	3,2
C4: Drop = 3 %, Delay = 0 ms, Jitter = 0 ms	2,9	2,9	3,0
C5: Drop = 5 %, Delay = 0 ms, Jitter = 0 ms	2,7	2,6	2,9
C6: Drop = 1 %, Delay = 50 ms, Jitter = 20 ms	2,3	1,8	2,7

5.2.2.4.4 Summary and conclusion

Figure 15 again give a summarized overview of all TMOS results. It displays also the references used in this test to compare the results of the real processed voice samples against codec simulations.

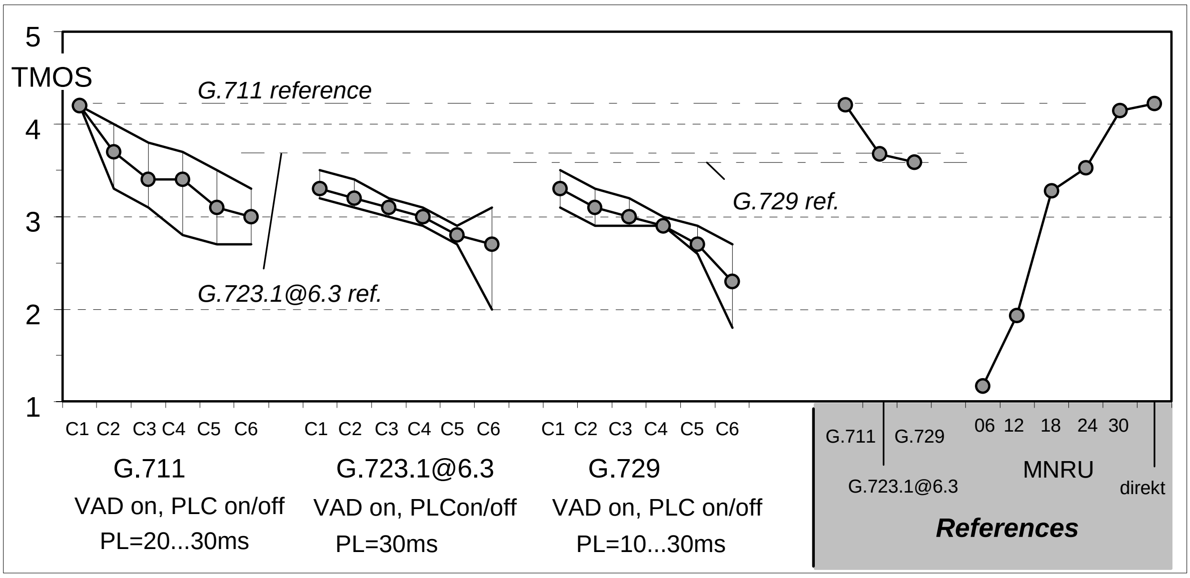


Figure 15: Summary of all conditions (incl. reference conditions)

5.3 Delay

5.3.1 Delay between two analogue PBX subscribers

Source: Robert Bosch GmbH: ETSI TIPHON 14TD081.

5.3.1.1 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.

5.3.1.2 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 signalling; 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.

Figure 16 shows the principal test Set Up.

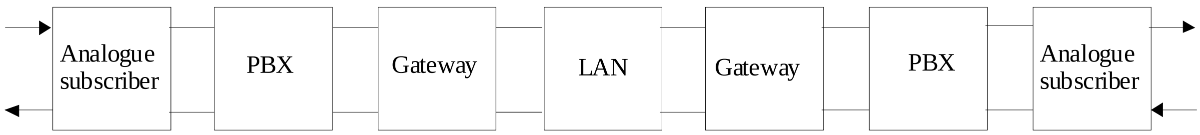


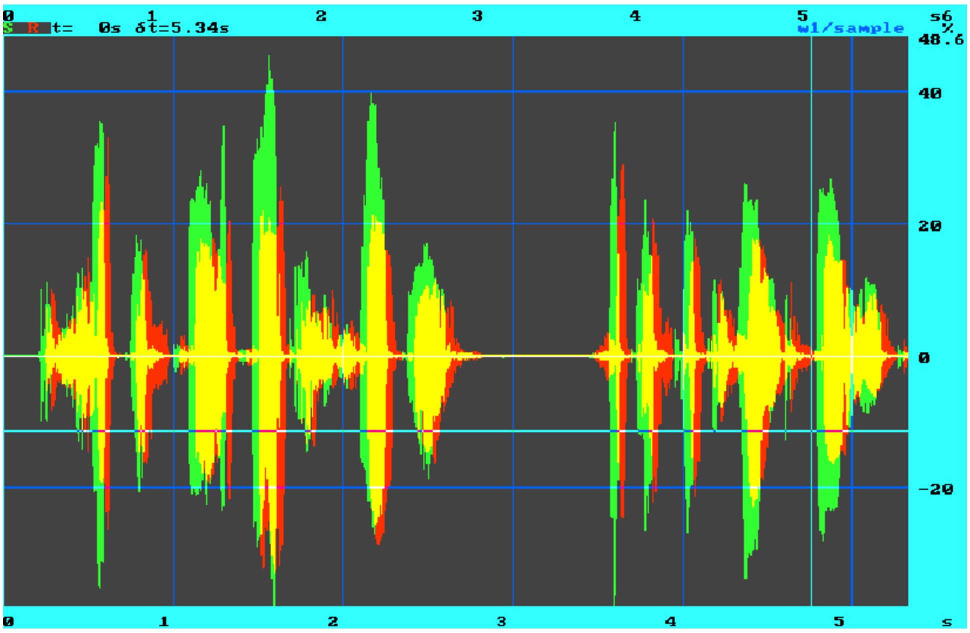
Figure 16: Set Up between two analogue PBX subscribers

5.3.1.3 Results

The results of Delay measurements are summarized in table 10.

Table 10: Measurement results of delay measurements of two different gateways

Gateway A	Gateway B
The mean one-way delay for both subscribers is variable between 125 ms and 250 ms, with an arithmetical mean value at 160 ms.	The mean one-way delay is constant with a value between 90 ms and 110 ms (see figure 17).



NOTE: Red: original send signal;
Green: receive signal (transmitted via the test Set Up).

Figure 17: Delay behaviour of the test Set Up for two analogue PBX subscribers

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.

5.3.1.4 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).

5.3.2 Delay between two PABX systems with analogue subscribers

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

5.3.2.1 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 Ω 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.

5.3.2.2 Result

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

5.3.3 Delay between two PC SW clients

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

5.3.3.1 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 Ω 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.

5.3.3.2 Results

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

5.3.4 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.

5.3.4.1 Measurement Set Up

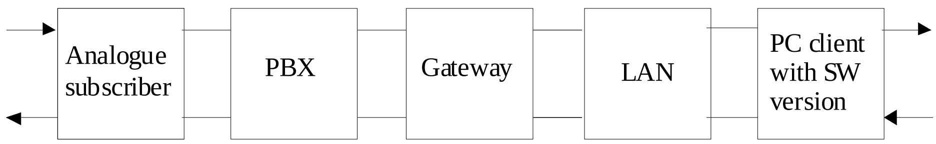
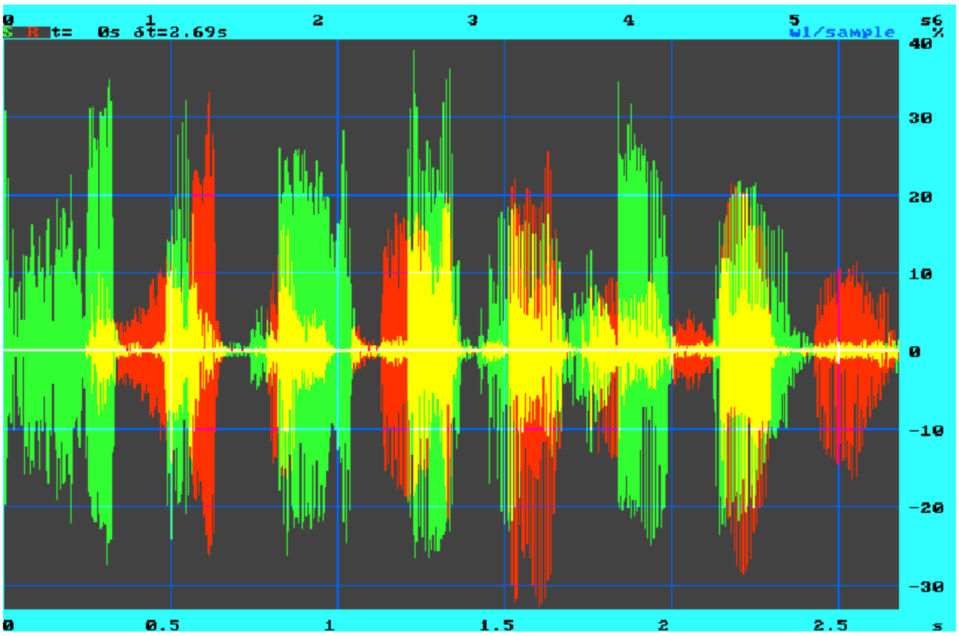


Figure 18: Set Up between an analogue PBX subscriber and a PC client

5.3.4.2 Results

Figure 19 shows the delay behaviour of the test Set Up for the transmission direction from PC client to analogue subscriber.



NOTE: Red: original send signal;
Green: receive signal (transmitted via the test Set Up).

Figure 19: Delay behaviour of the test Set Up for the transmission direction from PC client to analogue subscriber

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.

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 19 shows that no significant delay jitter could be observed.

5.3.4.3 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.

5.4 Echo

5.4.1 Echo between two analogue PBX subscribers

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

5.4.1.1 Measurement Set Up

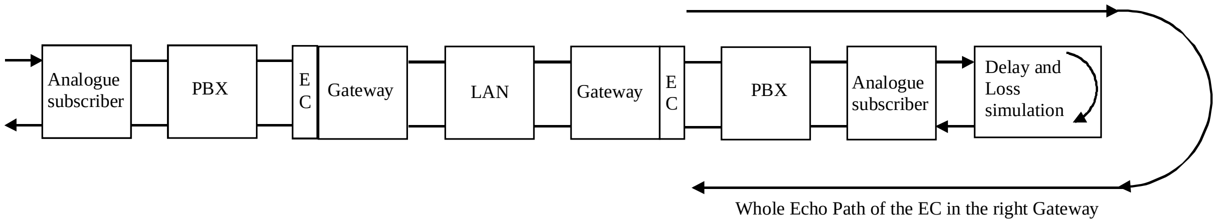


Figure 20: Measurement Set up for Echo measurements between two analogue PBX subscribers

Two different cases of loss were simulated.

- In case of simulated loss in the whole echo path of 7 dB:

  The delay and loss simulation device (see figure 20) 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 4 dB:

  The delay and loss simulation device (see figure 20) 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).

5.4.1.2 Results

Table 11 shows the results of simulated loss in the whole echo path.

Table 11: Results of Echo loss simulation

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.	The adaptation of the EC is very slow; 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, 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.	

5.4.1.3 Conclusion

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

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 ms to 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.

5.5 Impairment Factors

5.5.1 Transmission Impairments according to ITU-T G.113

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

ITU-T has provided tables of transmission impairment factors for a couple of speech codecs. These impairment factors are marked as Ie values in every table.

Table 12 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 13 and 15, and for propagation error patterns EP1 and EP2 in table 15. These values are provisional only as they were determined in single or a few experiments.

Table 12: Provisional Planning Values for the Equipment Impairment Factor Ie

Codec Type	Reference	Operating Rate kbit/s	Ie Value
PCM	G.711 [16]	64	0
ADPCM	G.726 [19], G.727 [20]	40	2
	G.721 [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	ETS 300 581-2 [37], Half-rate	5,6	23
ACELP	EN 300 726 [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 13: Provisional planning values for the equipment impairment factor Ie under conditions of random packet loss, codecs G.729A [23] + VAD, G.723.1-A [41] + VAD and GSM EFR

% Packet Loss	G.729A [23] + VAD	G.723.1-A [41] + 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 [41] + VAD: 1;• GSM EFR: 1.			

Table 14: 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 15: 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 to 32
	EP2	31 to 42
GSM-FR	EP1	32 to 39
	EP2	40 to 45
GSM-EFR	EP1	15 to 22
	EP2	26 to 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.		

5.6 R-Values

5.6.1 Analogue PBX- and SW Client scenarios

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

5.6.1.1 Introduction

The results given in clause 5.5.1 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 to 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]. An Advantage Factor A which is sometimes discussed for Internet-Telephony does not apply for business applications.

5.6.1.2 Results

R values were calculated for different scenarios.

5.6.1.2.1 Two Gateways type A with analogue telephone sets

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

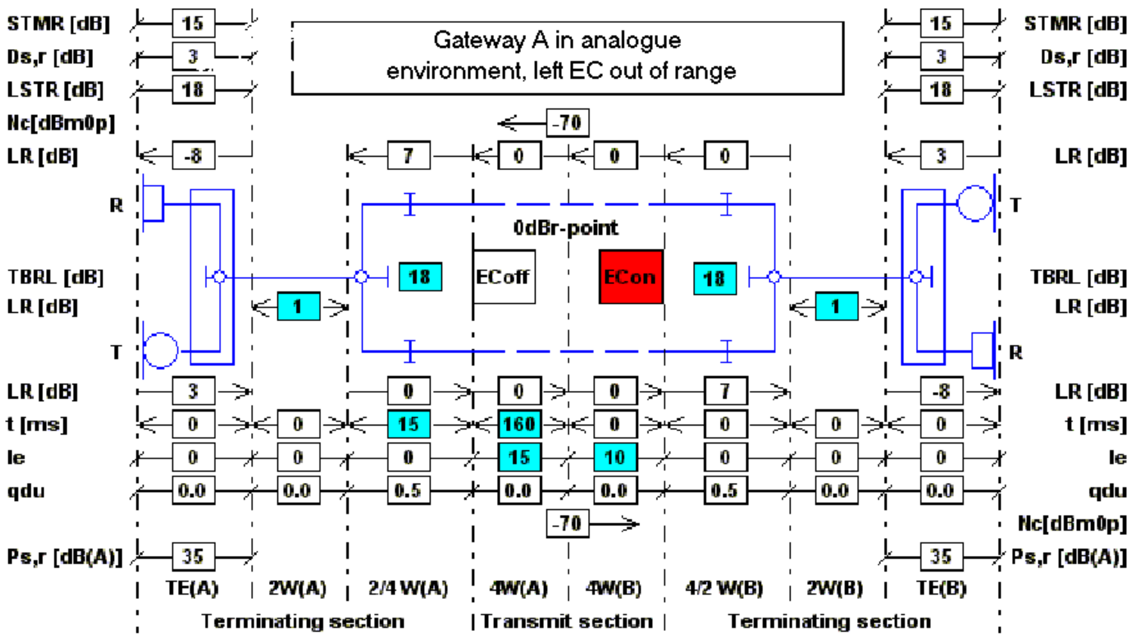


Figure 21: Speech transmission quality as perceived by the user at the left terminal where the EC is out of his operational range

Table 16: 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

Table 17: 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

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

5.6.1.2.2 Gateway type A with analogue subscriber behind PBX

In this scenario the analogue subscriber behind the PBX is on the left side and the PC client is on the right side.

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

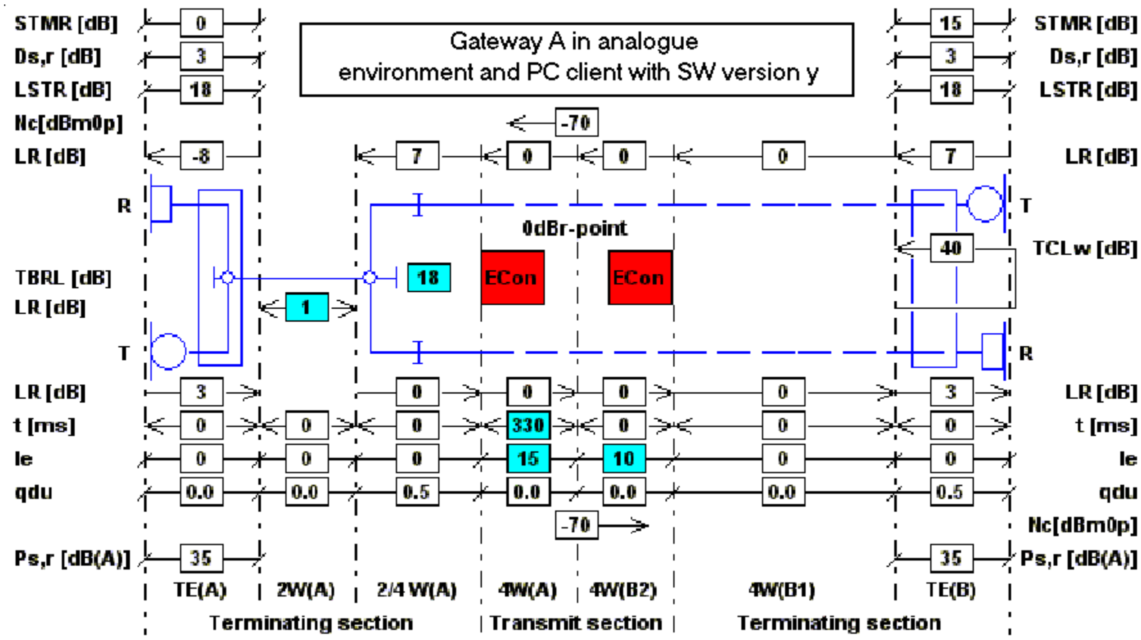


Figure 22: Speech transmission quality as perceived by the user at the left terminal (analogue telephone set)

Table 18: 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

Table 19: 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

5.6.1.2.3 Gateway type B with analogue telephone sets on both sides

The analogue telephone sets are used 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).

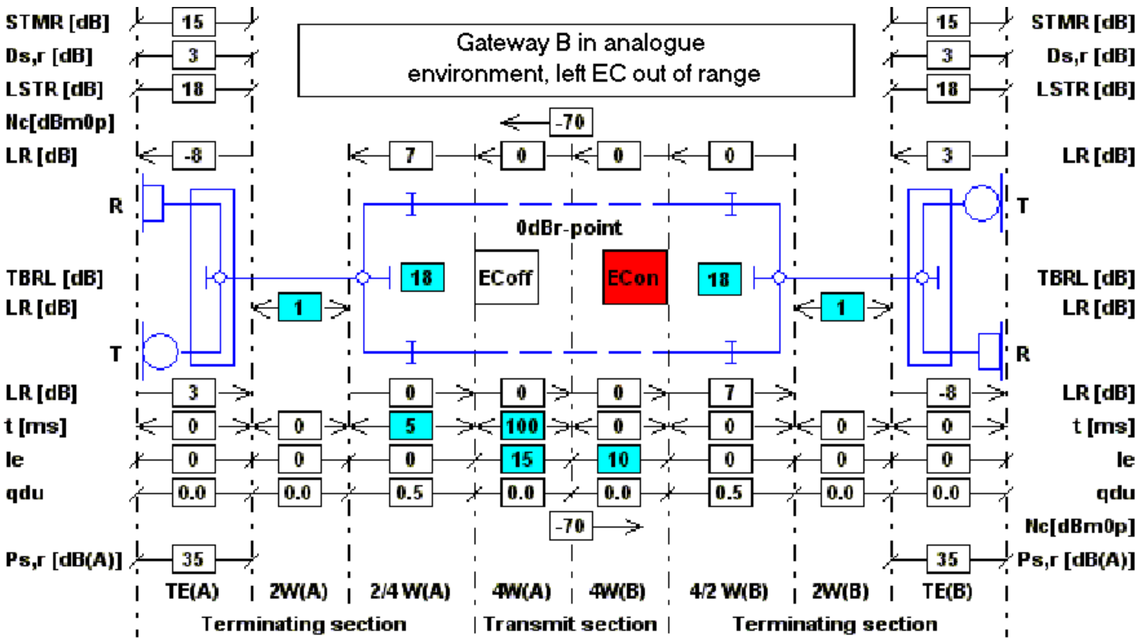


Figure 23: Speech transmission quality as perceived by the user at the left terminal where the EC is out of his operational range

Table 20: 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

Table 21: 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

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

5.6.1.3 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.

5.7 Advanced Measurement Techniques

5.7.1 QoS Measurements of IP-Configurations

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

5.7.1.1 Configuration and Measurement Set Up

For the test 2 IP configurations were available which can be found in figures 24 and 25. 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 Ω 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.

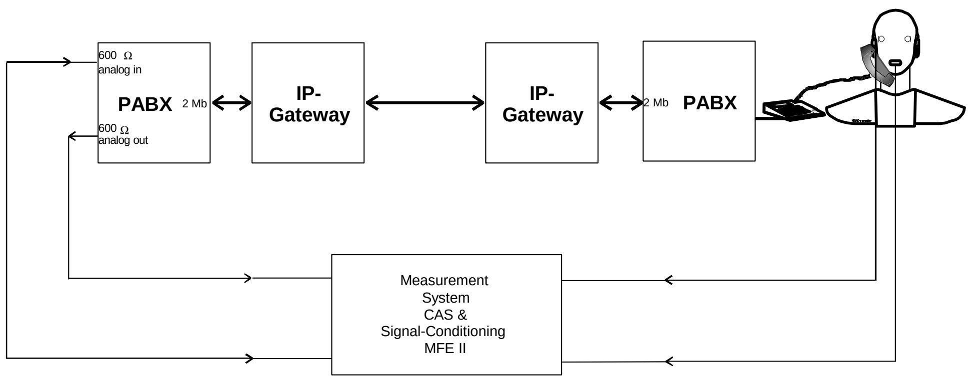


Figure 24: Configuration 1 - One way transmission delay = 70 ms, Coding ITU-T Recommendation G.729 [22], "Standard" - handset telephone

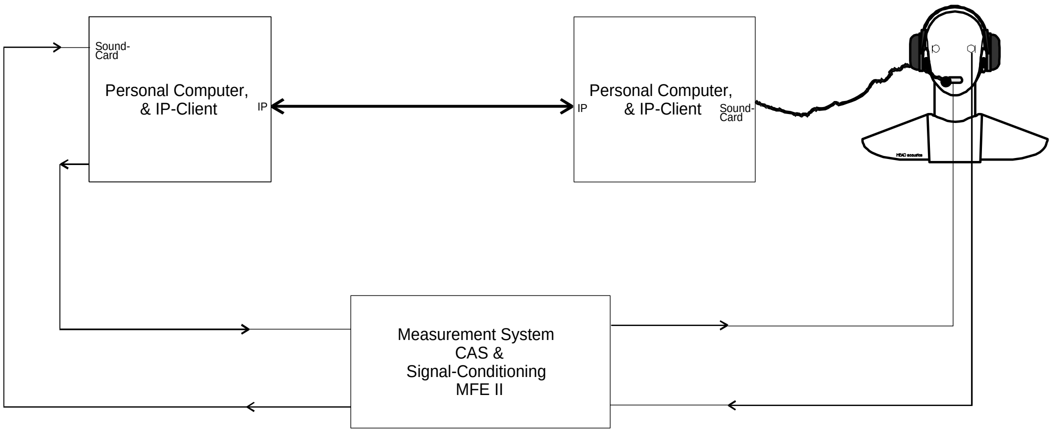


Figure 25: Configuration 2 - One way transmission time = 400 ms to 530 ms, variable, Coding ITU-T Recommendation G.723 [40], PC with soundcard and headset

Figure 26: Test set up with head set and hand set

5.7.1.2 Results

5.7.1.2.1 Parameters in Single Talk Conditions

The following parameters are described as "Traditional" Parameters.

Figures 27 and 28 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.

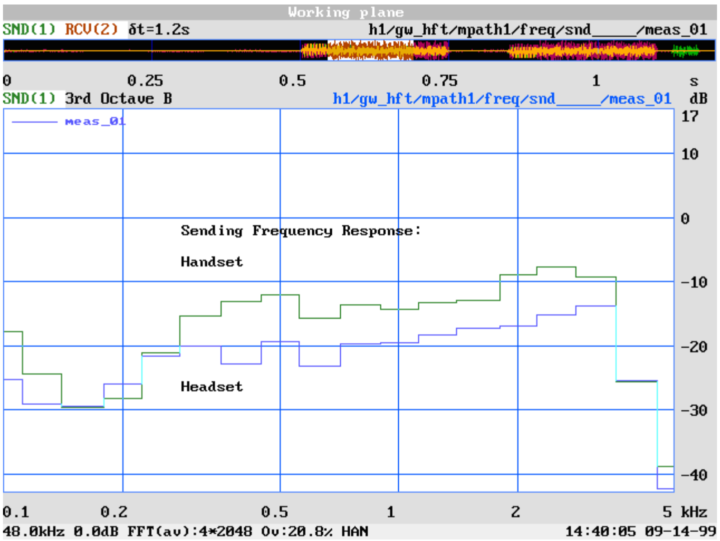


Figure 27: Sending frequency responses for configuration 1 (handset) and 2 (headset)

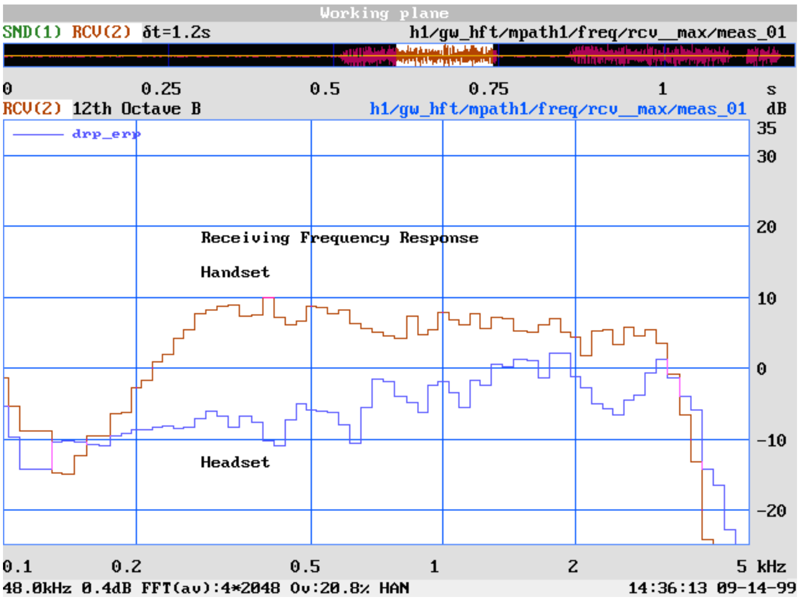


Figure 28: Receiving frequency responses for configuration 1 (handset) and 2 (headset)

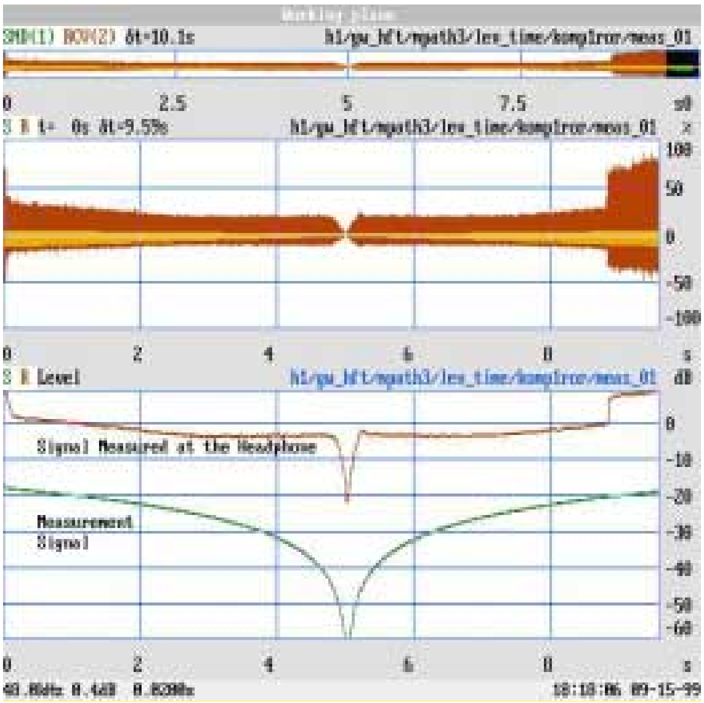
In the following clauses more detail results in single talk conditions especially for switching, AGC characteristics and echo loss will be given.

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

5.7.1.2.2 Level dependant input-output characteristics

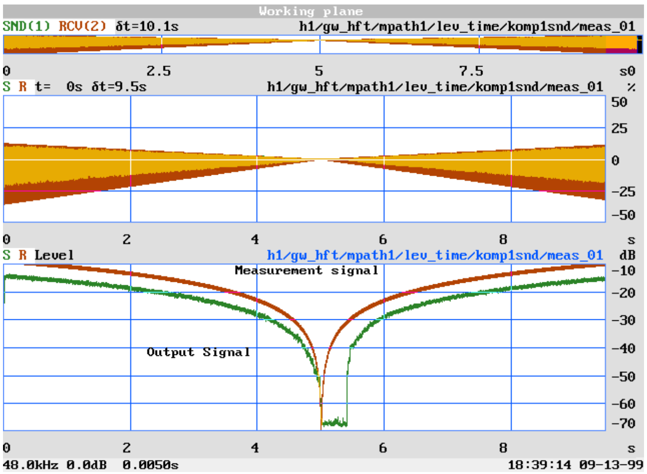
In sending and receiving direction a VAD is detectable which for both configurations indicate no or only minor distortions caused by syllable clipping. The measurement was made using an increasing voiced sound of artificial voice (ITU-T Recommendation P.50 [42]), monitoring the output signal and comparing 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 30. In addition –depending on the input signal level- strong level variations can be seen (beginning and end of the test sequence in figure 29).



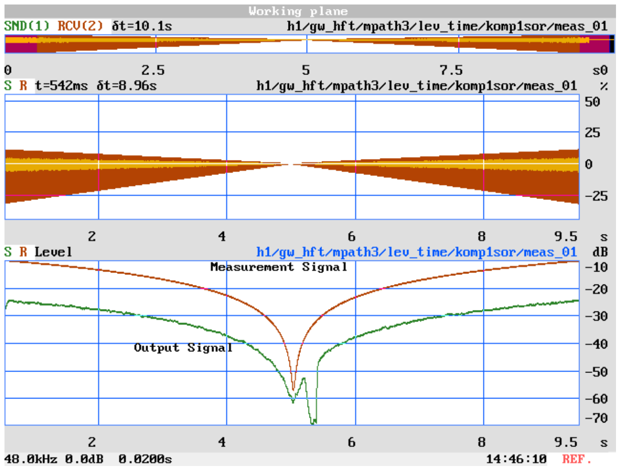
NOTE: upper: time signal;
lower: level vs. time.

Figure 29: Level depending input - output characteristics in receiving, configuration2



NOTE: upper: time signal;
lower: level vs. time.

Figure 30: Level dependant input-output characteristics in sending, configuration 1



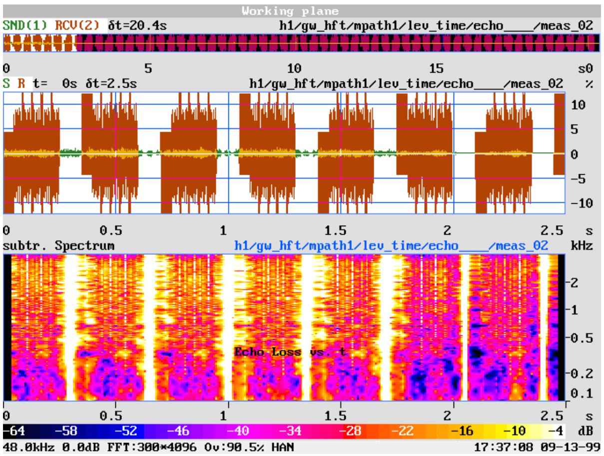
NOTE: upper: time signal;
lower: level vs. time.

Figure 31: Level dependant input-output characteristics in sending, configuration 2

5.7.1.2.3 Echo loss and convergence

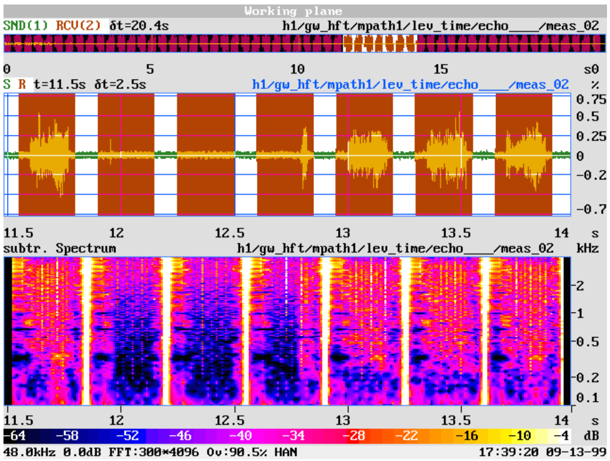
The echo loss was measured always from the electrical input to the electrical output. Certainly in single talk the characteristics of the echo canceller(s) are quite important for the subjective impression of a connection. Figure 32 and 33 show measurements using the composite source signal CSS as defined in ITU-T Recommendation P.501 [29] and ITU-T Recommendation G.168 [15] at configuration 1. Figures 32 and 33 show the convergence as a function of time, displayed in the spectral domain. Figure 32 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 33 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 condition is 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 32 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 29.



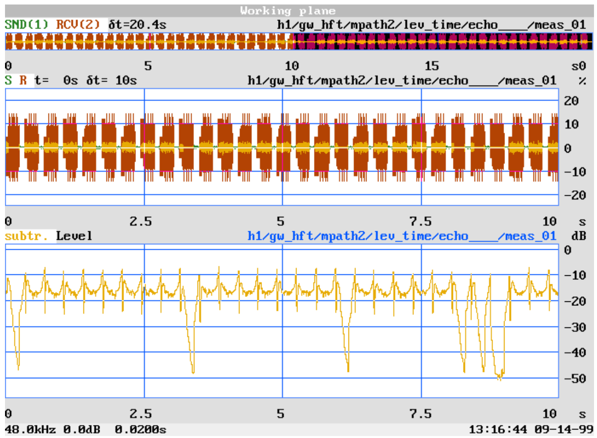
NOTE: red: measurement signal;
green: echo signal;
upper: time signal;
lower: echo loss vs. time.

Figure 32: Convergence vs. time; configuration 1



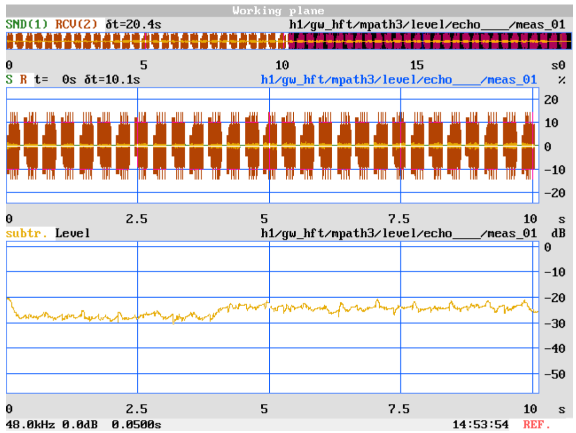
NOTE: red: measurement signal;
green: echo signal;
upper: time signal;
lower: echo loss vs. time.

Figure 33: Echo loss vs. time, EC converged after 11 s; configuration 2



NOTE: red: measurement signal;
green: echo signal;
upper: time signal;
lower: echo loss vs. time.

Figure 34: Echo loss vs. time when exceeding the processing window of the echo canceller, configuration 1



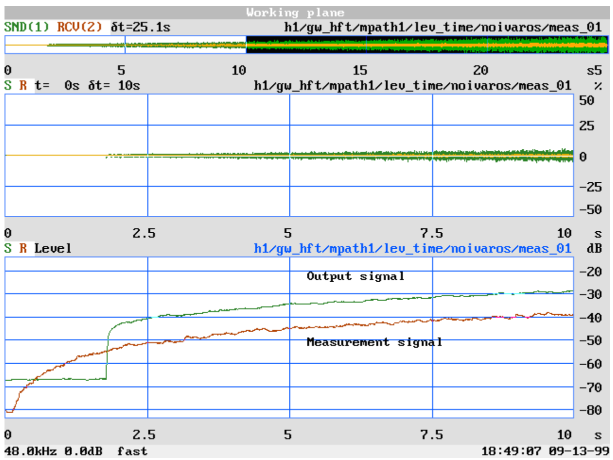
NOTE: red: measurement signal;
green: echo signal;
upper: time signal;
lower: echo loss vs. time.

Figure 35: Echo loss vs. time, configuration 2

5.7.1.2.4 Performance in sending direction in the presence of background noise

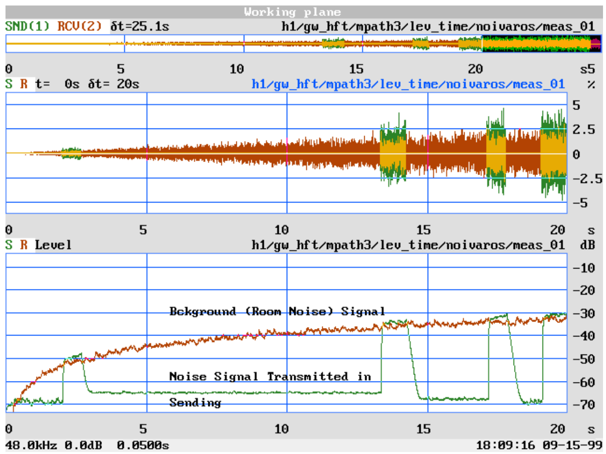
A very importance performance parameter is the system behaviour in the presence of background noise. Some measurement examples are shown in figures 36 and 37. For configuration 1 figure 36 shows, that the background noise (indicated in red) is transmitted only when a certain level is reached. After reaching 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 37, the background noise is constantly interrupted for various levels.



NOTE: red: measurement signal;
green: output signal at the line interface;
upper: time signal;
lower: level vs. time.

Figure 36: Background noise transmission in sending level versus time, configuration 1

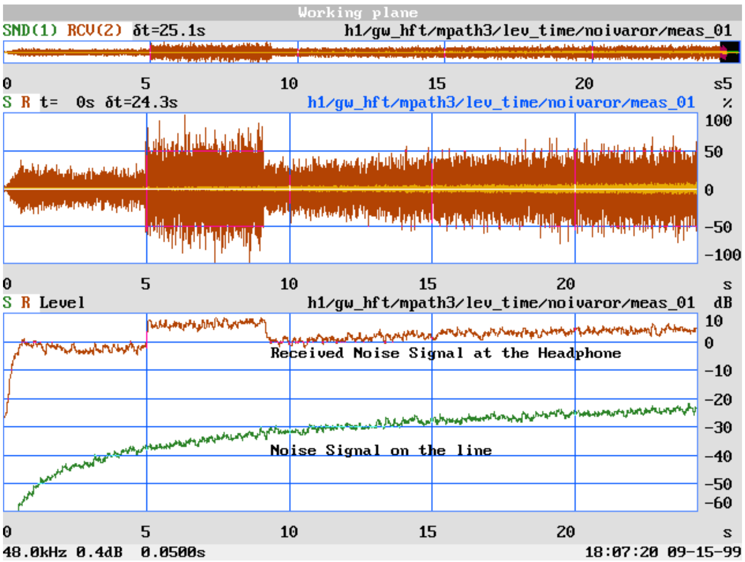


NOTE: red: measurement signal;
green: output signal at the line interface;
upper: time signal;
lower: level vs. time.

Figure 37: Background noise transmission in sending, level versus time, configuration 2

5.7.1.2.5 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 characteristic as it can be seen in figure 38.



NOTE: green: measurement signal;
red: output signal at the headphone);
upper: time signal;
lower: level vs. time.

Figure 38: Background noise transmission in receiving, level versus time, configuration 2

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 38. This will result in a very bad background noise performance, especially if this system is connected with a terminal providing the intermittent transmission behaviour as shown in figure 37.

5.7.1.2.6 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 39 and 40 and is inserted simultaneously in sending and receiving direction.

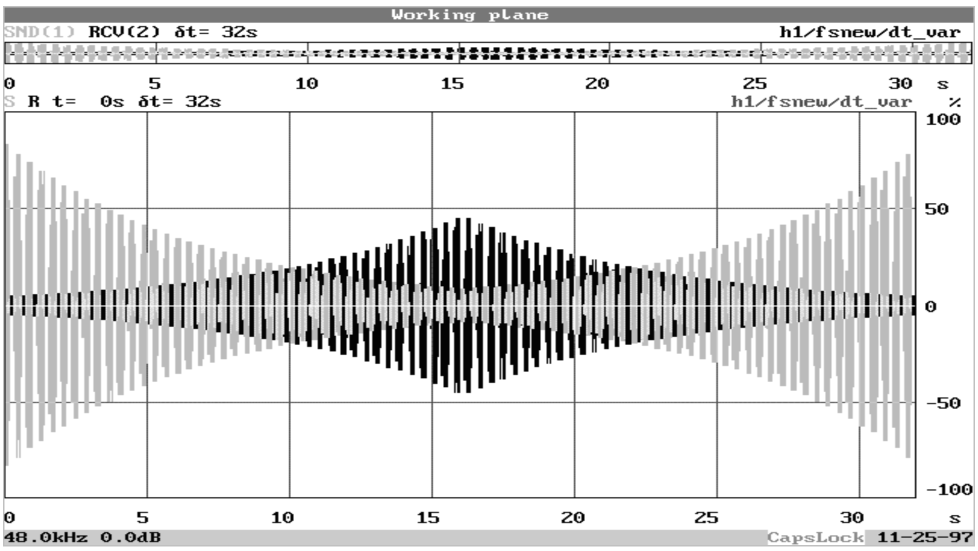


Figure 39: Overview of double talk test signal

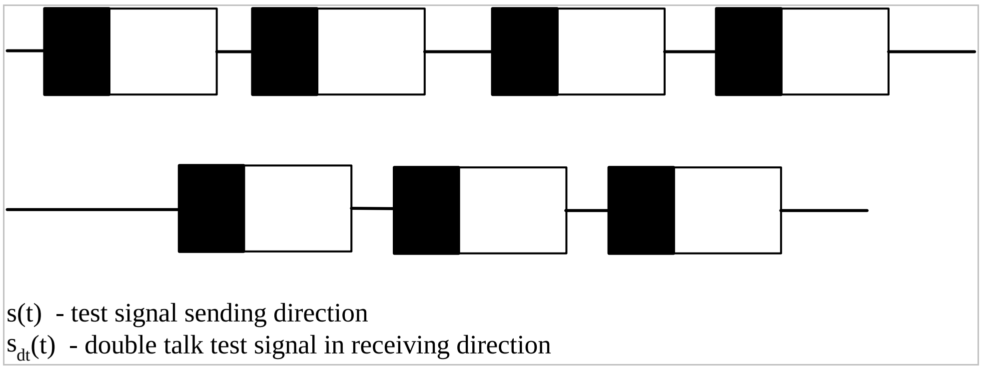
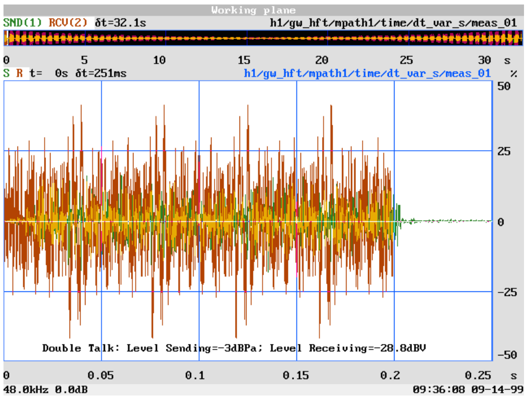


Figure 40: Cut out of the complete measurement sequence with detailed view on the overlap of sending and receiving direction signal, principle arrangement

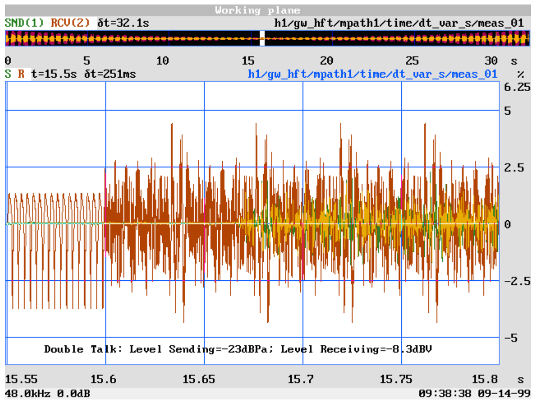
By using this test signal the speech-like CSS Sequences are fed into sending and receiving simultaneously with varying level. The level is increasing (or decreasing) by 0,5 dB from CSS package to CSS package. Thus the level variation during double talk is 20 dB in both directions.

Some measurement results for configuration 1 are given below.



NOTE: red: measurement signal;
green: output signal at the line interface

Figure 41: Transmission during double talk in sending for high sending level, configuration 1



NOTE: red: measurement signal;
green: output signal at the line interface

Figure 42: Transmission during double talk in sending for low sending level, configuration 1

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.

5.7.1.2.7 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 could 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.

5.7.1.3 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 well 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.

5.7.2 QoS Measurements of ETSI Speech Quality Test Event 2000

5.7.2.1 Introduction

This clause describes the test methodologies and the results of the measurements which were carried out during the 1st ETSI Speech Quality Test Event. The tests were conducted by T-Nova Deutsche Telekom Innovationsgesellschaft mbH Berkom (T-Nova) in collaboration with HEAD acoustics GmbH (HEAD acoustics).

The aim of the test event was to determine the speech quality of various voices over IP equipment under certain IP network conditions. During the test event, speech material as well as measurement data were collected by transferring voice samples and artificial signals across the Voice over IP Set Up.

5.7.2.2 Measurement Set Up

Measurements at the acoustical interface were conducted with two measurement scenarios:

- Acoustical transmitter interface – Acoustical receiver interface ("Acoustical-Acoustical");
- Electrical transmitter interface – Acoustical receiver interface ("Electrical-Acoustical").

Measurements were executed using mainly with artificial test signals (according to ITU-T Recommendation P.501 [29]).

The artificial test signals were used for measuring of transmission parameters according the ITU-T Recommendation P.500 series.

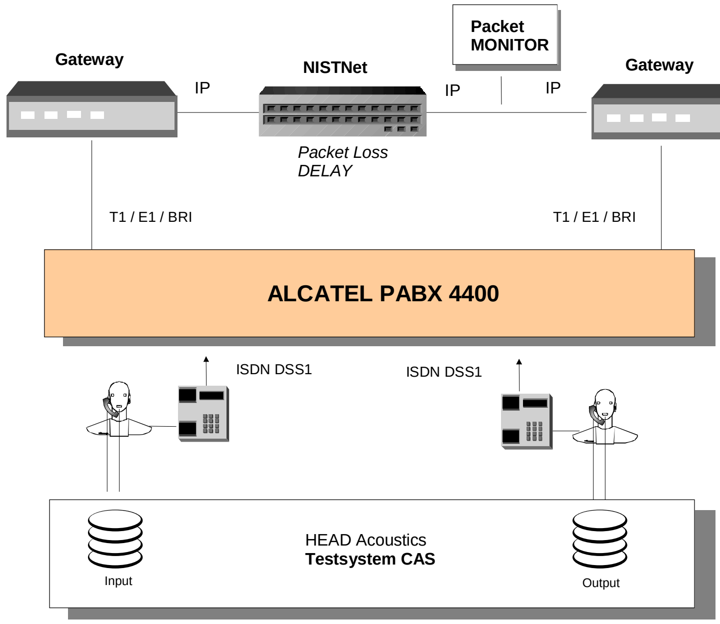


Figure 43: "Acoustical – Acoustical" Measurement Set Up with Reference ISDN Terminals

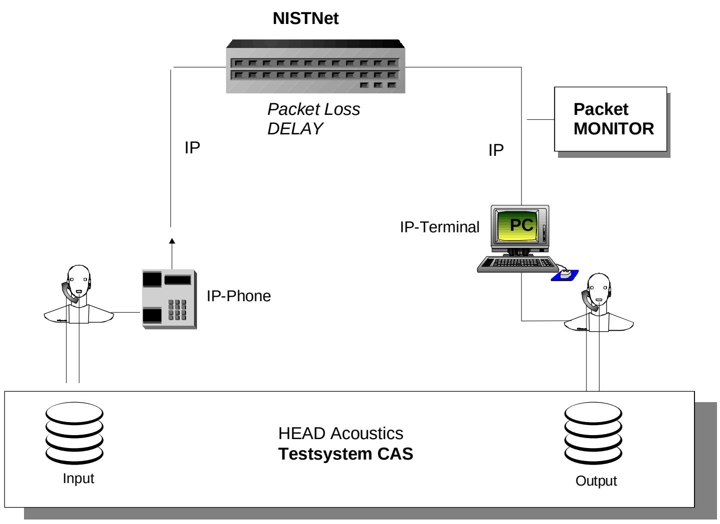


Figure 44: "Acoustical – Acoustical" Measurement setup with IP Terminals (handsets or headsets)

The reference terminals provided were standard digital handset terminals according to ETSI TBR 8. For all kind of measurements a packet loss generator and a packet loss monitor were included in the Set Up.

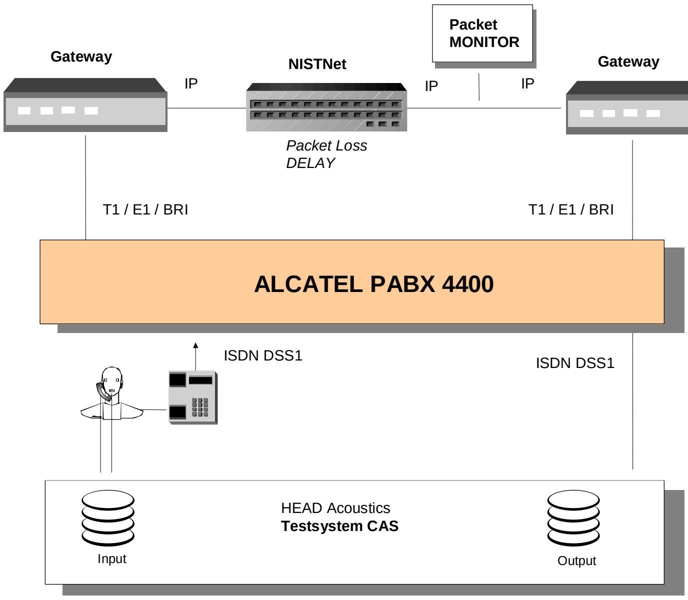


Figure 45: Measurement Set Up "Acoustical – Electrical" for Gateway to Gateway Configuration

For the tests the handsets of the terminals are applied to the HATS using the positioning as described in ITU-T Recommendation P.64 [28] with defined pressure force. The test sequences, natural speech as well as the artificial sequences were automatically introduced and recorded by the test system CAS and stored on hard disc.

Instead of the PABX telephone an IP telephone was used if provided in combination with a gateway which interfaces to the PABX. This is shown in figure 46.

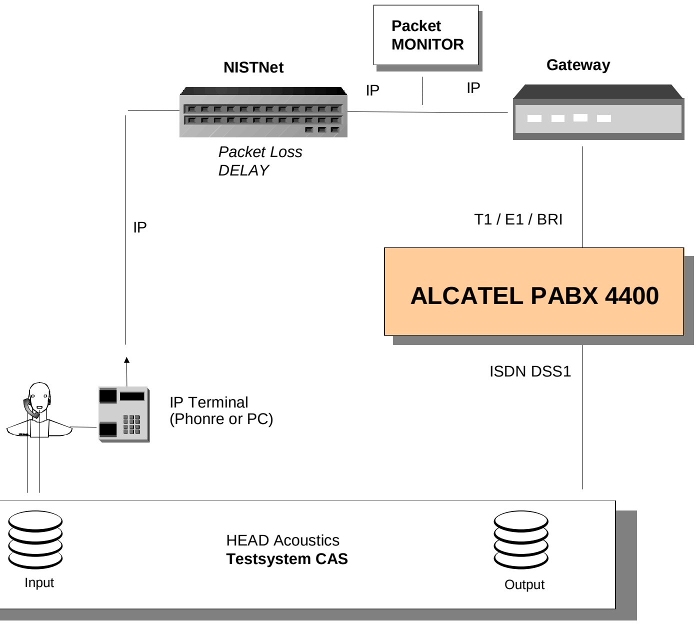


Figure 46: Measurement Set Up "Acoustical – Electrical" for IP-Terminal to Gateway Configuration

The IP network conditions for all kinds of acoustical measurements ("electrical – acoustical" and "acoustical – acoustical") were:

Table 22: Network Conditions for All Kinds of Acoustical Measurements

Condition	Packet Loss (Equal)	Additional Delay[1)]	Delay Variation
1	0	No	No
2	0	100 ms	No
3	0	100 ms	20 ms
4	1 %	100 ms	No
5	1 %	100 ms	20 ms
6	3 %	100 ms	No
NOTE: Additional IP network delay was introduced using NIST Net.			

Also for all kinds of acoustical measurements it was recommended to use a packet length of at least 60 ms for audio frames (cf. electrical-electrical measurements). This packet length had no influence on auditive or instrumental assessment methods.

5.7.2.3 Results

5.7.2.3.1 Packet loss

Figures 47 and 48 demonstrate two analysis examples obtained from recordings under network conditions with simulated packet loss. The figures show an enlarged time sequence from the recorded signal using the periodical repetition of a voiced sound as test signal.

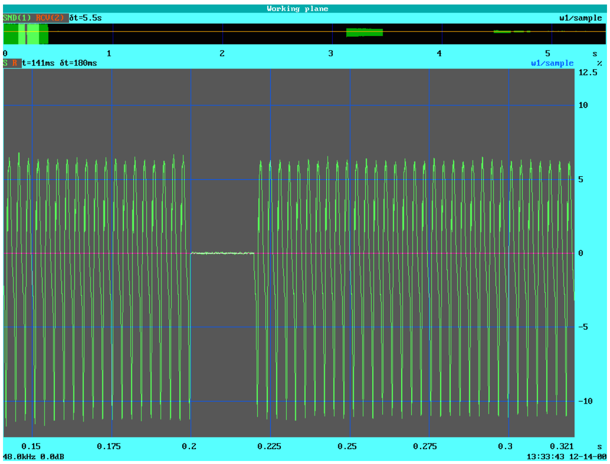


Figure 47: Occurrence of packet loss

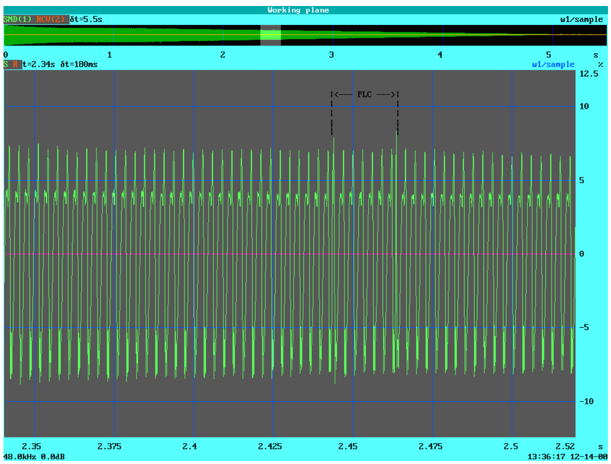


Figure 48: Packet loss concealment

A resulting signal gap due to one lost IP packet can be analysed in figure 47. Packet loss concealment (PLC) is obviously not implemented in the equipment which was under test. A typical PLC implementation can be seen in figure 48. The lost IP packet obviously is substituted by a previous frame, but signal irregularities at the beginning and the end of this substituted frame can be analysed. These resulting peaks are audible and annoying.

This result reflects one kind of PLC implementation which can be used to cover the influence of lost packets in the network.

5.7.2.3.2 Transmission characteristics for background noise

The test signal consists of the spectral shaped noise signal (Hoth spectrum) with increasing test signal level vs. time. The measurement result derived from the test via one ISDN line and the PBX is shown in figure 49 for comparison. The red signal in the upper windows represents the test signal, the green signal the measured signal (transmitted signal). Note, that the overlapped red and green colour results in yellow. The analysis curves in the lower window show the calculated level vs. time (calculated in time domain using a 35 ms time constant). The transmission is linear, no level variations can be observed, the two curves (red for the original test signal level, green for the transmitted signal) are in parallel. Typical results derived from tests during the event are shown in figure 50.

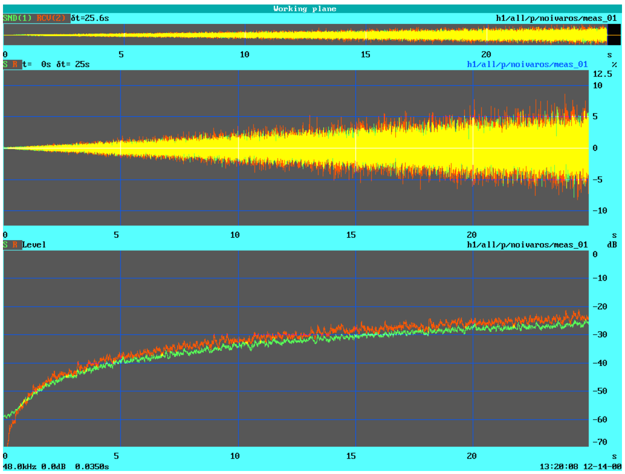
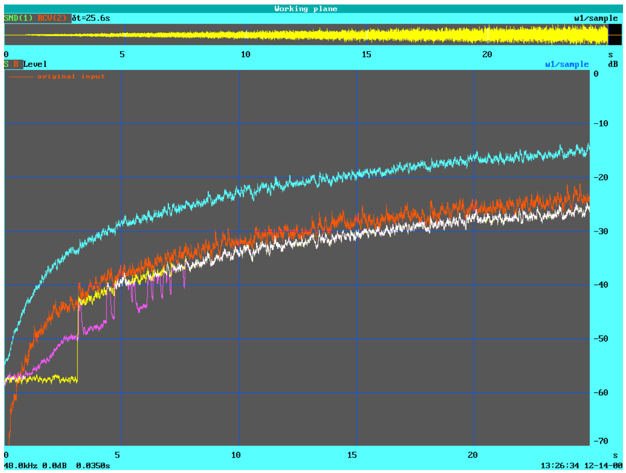


Figure 49: Transmission characteristic for background noise (PBX reference connection)



NOTE: red: original test signal.

Figure 50: Typical analysis results derived from tests with 3 manufacturers

The results displayed in yellow, magenta and cyan represent different kinds of implementations. An activation level threshold (yellow), the influence of automatic gain control (AGC, cyan) and an implementation with adaptive comfort noise injection can be analysed clearly.

5.7.2.3.3 Transmission performance under double talk performance

The transmission characteristics under double talk conditions were determined using the test signal consisting of two decorrelated composite source signals which are periodically repeated and applied with different signal levels to both sides of the connections. The recordings were carried out in one direction and consequently the signal which is fed at the other end of the connection should be transmitted and no additional signal components should occur. The following 4 figures demonstrate typical measurement results which were obtained during the event. In each figure the transmitted and recorded signal is shown in green and the original test signal is displayed in red. In order not to confuse the reader the double talk signal which is fed in the opposite transmission path is not given in these figures.

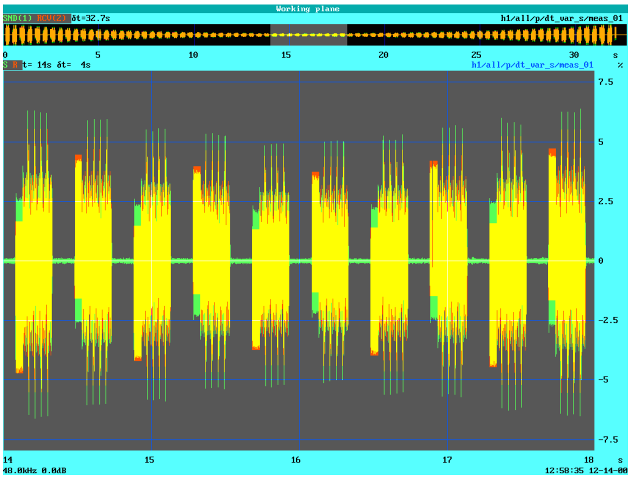


Figure 51: Result for the PBX connection

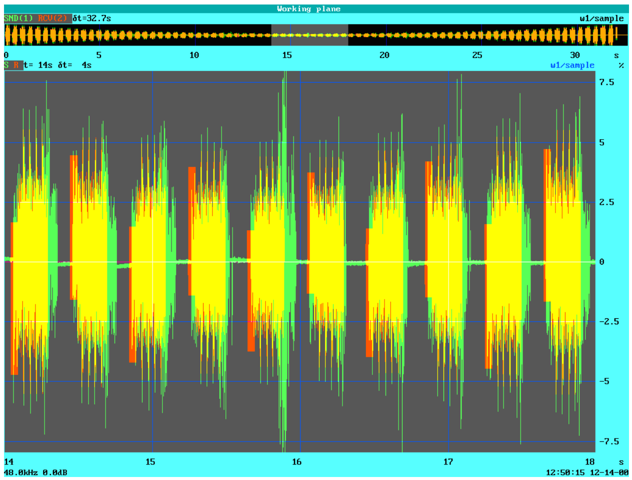


Figure 52: Typical result demonstrating echo components and short term clipping

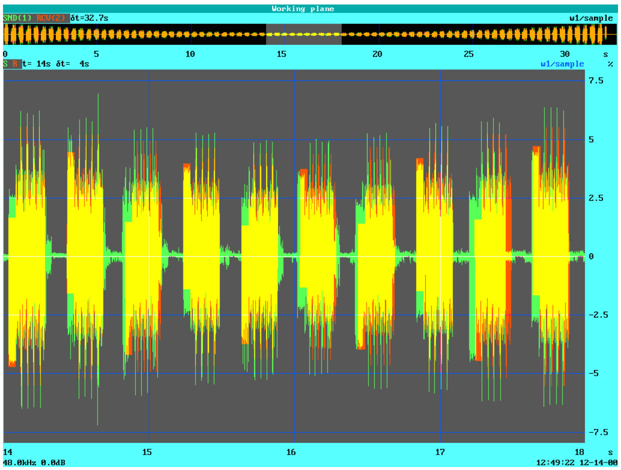


Figure 53: Typical result demonstrating echo components

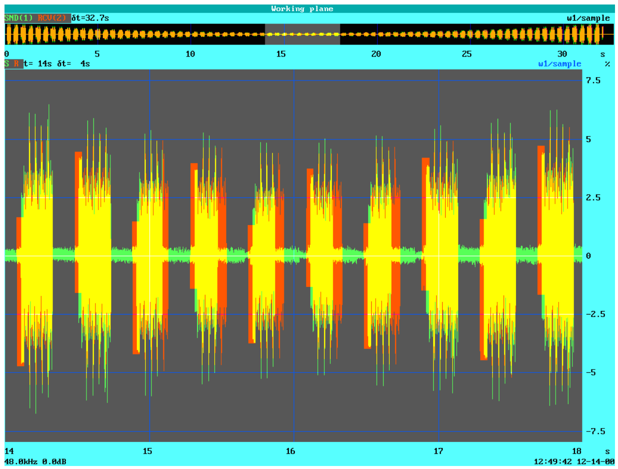


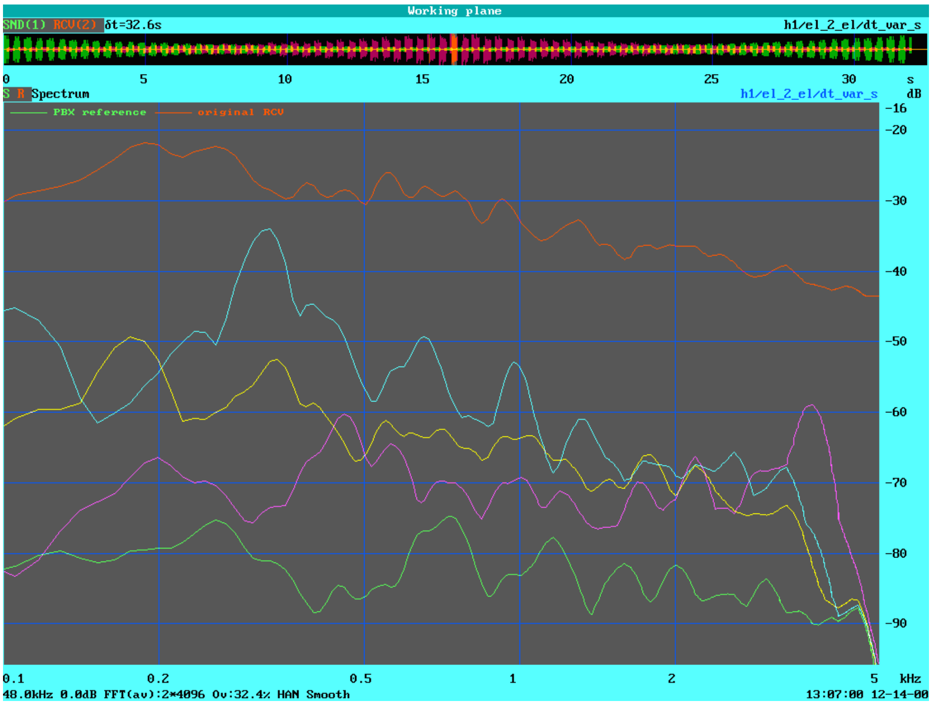
Figure 54: Typical result demonstrating echo components and clipping

Figure 51 shows the example for the analysis obtained for the PBX reference connection and points out that all signal bursts are completely transmitted. Clipping does not occur. During the pauses between 2 signal bursts the channel noise floor is recorded. No additional signal components which could be identified as disturbances (like echoes or others) occur.

The complete transmission of all signal bursts can also be seen in figure 52 except a short frontend clipping. Additional high level echo components (green) can be determined during the pauses between the signal bursts. Basically the same result can be analysed in figure 53 for another tested connection. One example for the occurrence of signal clipping (parts of the green signals are missing) is demonstrated in figure 54. An additional high level noise floor is recorded between the signal bursts in this example.

These three measurement results demonstrate some disturbances which are introduced by the implemented echo cancellers and the non-linear processors.

Figure 55 compares the power density spectra analysed during the signal pauses under double talk conditions. The red curve represents the curve for the original test signal which was applied in the opposite transmission path. The green curve was derived from the measurement of the PBX connection. The other curves in cyan, yellow and magenta represent the results for 3 different test connections. Obviously significant differences compared to the PBX connection occurred for the connections under test during the event.



NOTE: red: original test signal;
green: PBX connection;
others: typical results measured during the event).

Figure 55: Comparison of power density spectra analysed in the pauses between two signals bursts during the double talk period

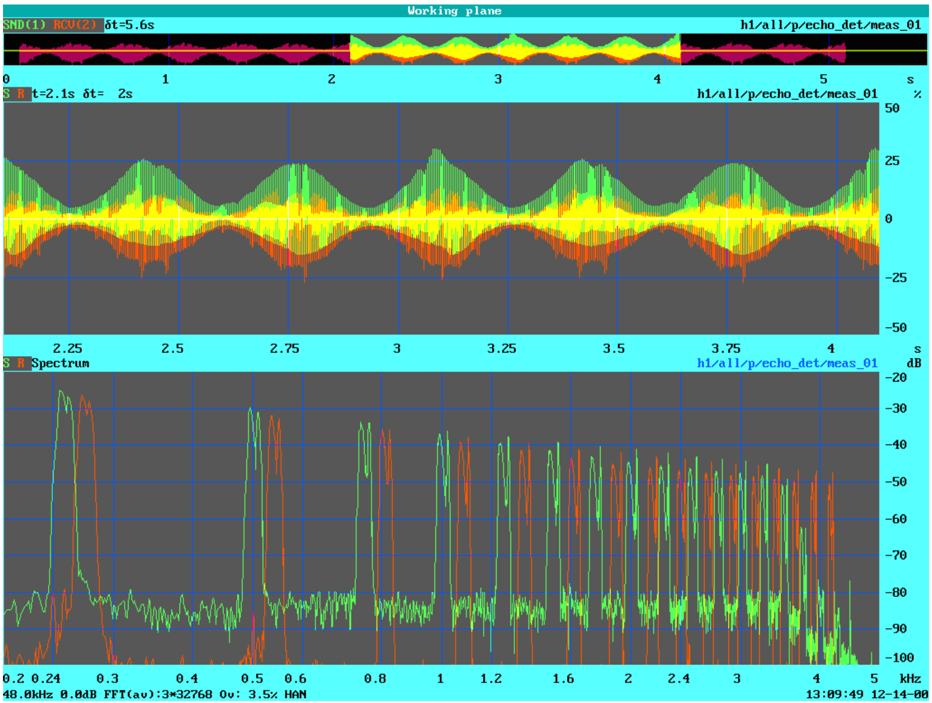
Further analysis demonstrates that these signal components are caused by echoes which only occur during double talk periods. It should be noted that the test connection was completely 4-wire and therefore physically did not produce any echo.

5.7.2.3.4 Detailed analysis of echo during double talk

A specific test signal according to ITU-T Recommendation P.501 [29] consisting of a two channel signal with a comb filter structure was used for a more detailed analysis of the echo disturbances during double talk. This signal is suited to determine and measure the echo more precisely.

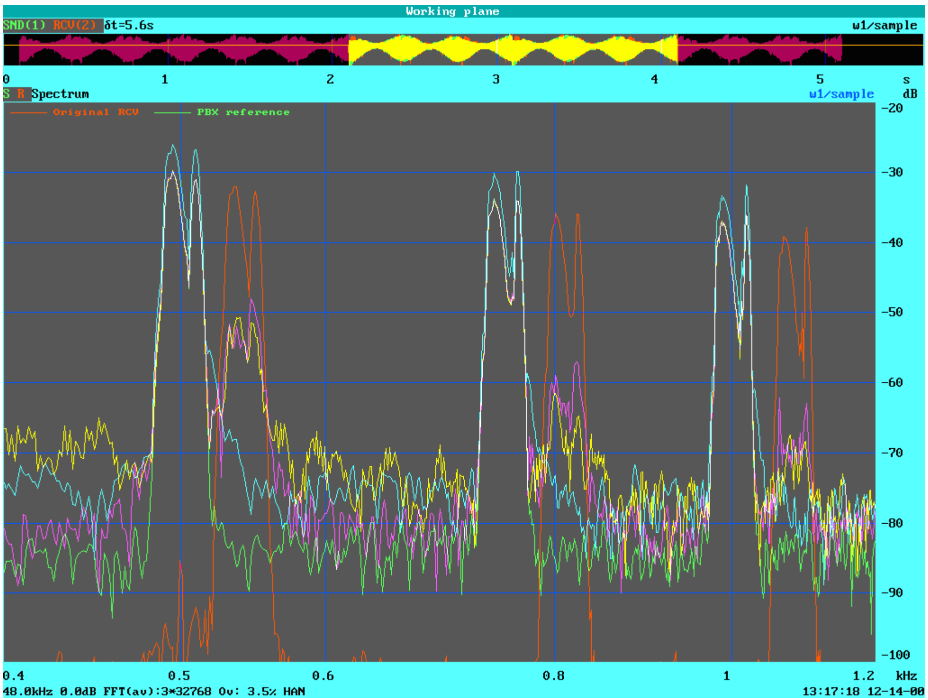
The test result which was obtained for the PBX connection is analysed in the example in figure 56. The time signals are shown in the upper window. The analysis window (lower window) shows the power density spectrum calculated by Fourier transform for the measured signal (green) and the original test signal in receiving direction (red). In principle signal components in the measured (green) signal which correlate to the excitation frequencies of the original signal (red, note that this was the original one which may cause echoes) can be analysed as echoes.

The analysis in figure 56 shows that the PBX connection was echo-free under double talk conditions. A more detailed analysis example for the enlarged frequency range between 400 Hz and 1,2 kHz is given in figure 57. Again typical measurement results obtained for 3 different manufacturers are analysed together with the PBX connection (green). The figure shows significant echo components for some of the tested connections. These components can be detected in the frequency range which was excited by the original test signal (red).



NOTE: green: measured signal containing double talk signal and echo;
red: original test signal.

Figure 56: Determination of echo during double talk test result for the PBX reference connection upper window: time sequences, lower window: power density spectra



NOTE: green: PBX reference connection;
red: original test signal;
others: typical results measured during the event.

Figure 57: Enlarged frequency range with a comparison of typical test results

5.8 Jitter Buffer

Source: KPN; Problems with the behaviour of jitter buffers and their influence on the end-to-end speech quality; ESTI TIPHON 22TD47.

5.8.1 Simulation

All results presented in this clause are only based on simulations, all the way from source to receiver. No measurements or real IP packets are involved, but all experiments were done within a discrete event driven simulator environment.

5.8.1.1 Static Jitter Buffer

First, it is necessary to determine what experiments are useful. Many experiments can be carried out by changing the link capacities, the number of sources, the codec, etc. For the purpose of comparing static and dynamic jitter buffers, and for recommendations on what buffer size to use, changing the jitter buffer size while keeping all other parameters the same, is the most interesting experiment, and therefore the one performed.

The following series of plots, figures 58, 59, 60 and 61, give the jitter buffer loss percentage for a jitter buffer delay of 10 ms, 15 ms, 20 ms, and 30 ms respectively. At 40 ms, no jitter buffer loss occurred.

This series of plots not only shows the dependence of the jitter buffer loss on the jitter buffer size, but also an important other phenomenon: the dependence of the jitter buffer loss on the network delay of the *first* packet of a voice connection. This can be understood intuitively: if the first packet is "early", i.e. it has a delay lower than the average delay, more packets will be considered "late". If the first packet is "late", more (sometimes all) packets will be considered "on time". This is clearly visible in for instance figure 58: the lower the "first delay", the higher the loss, although some degree of variation in the jitter buffer loss for the same "first delay" does exist.

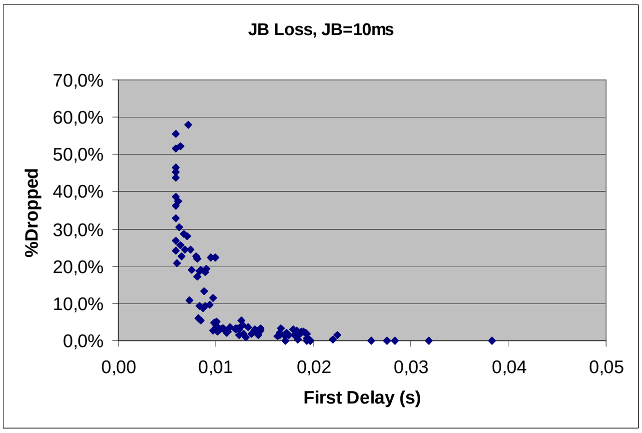


Figure 58: Jitter buffer loss for 10 ms jitter buffer

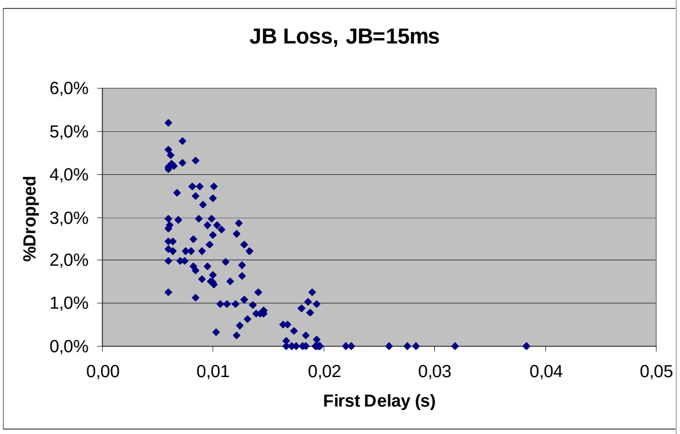


Figure 59: Jitter buffer loss for 15 ms jitter buffer

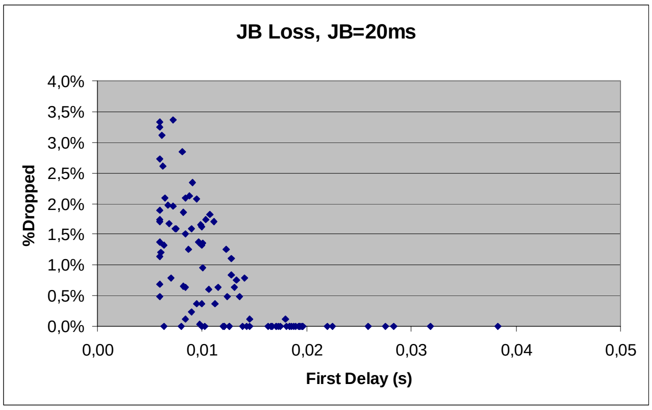


Figure 60: Jitter buffer loss for 20 ms jitter buffer

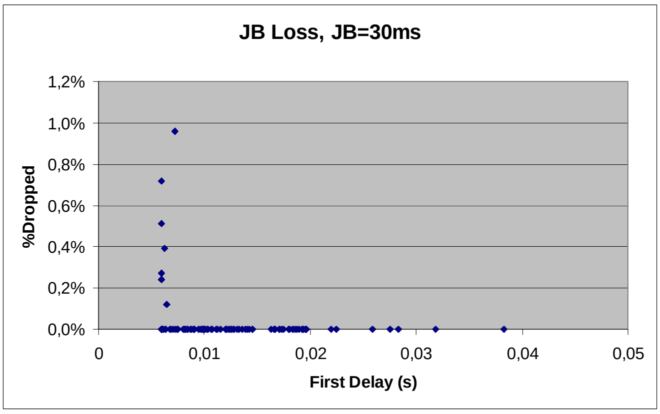


Figure 61: Jitter buffer loss for 30 ms jitter buffer

Figure 62 summarizes the previous four plots by giving the maximum jitter buffer loss for the four different jitter buffer sizes. Clearly, the higher the jitter buffer size, the lower the jitter buffer loss.

Figure 62: Maximum jitter buffer loss depending on jitter buffer size

5.8.1.2 Dynamic Jitter Buffer

The results of simulating (100 times) a dynamic jitter buffer of 10 ms, 20 ms and 25 ms are given in figures 63, 64 and 65 respectively. For 30 ms, no jitter buffer loss occurred.

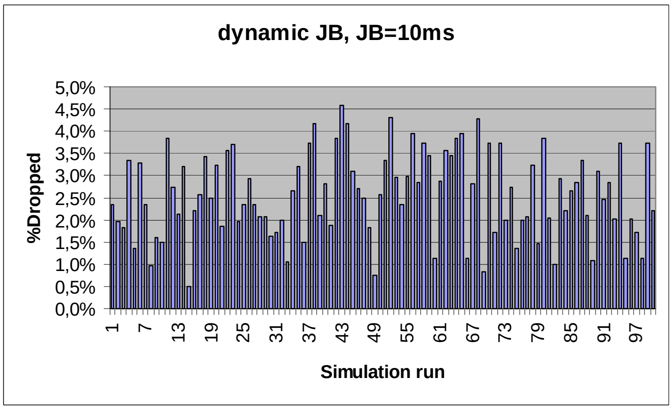


Figure 63: Jitter buffer loss for 10 ms dynamic jitter buffer

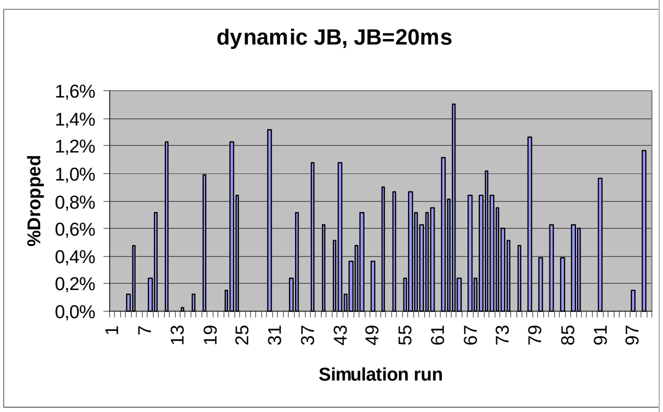


Figure 64: Jitter buffer loss for 20 ms dynamic jitter buffer

Figure 66 summarizes the previous three plots by giving the maximum jitter buffer loss for the three different dynamic jitter buffer sizes.

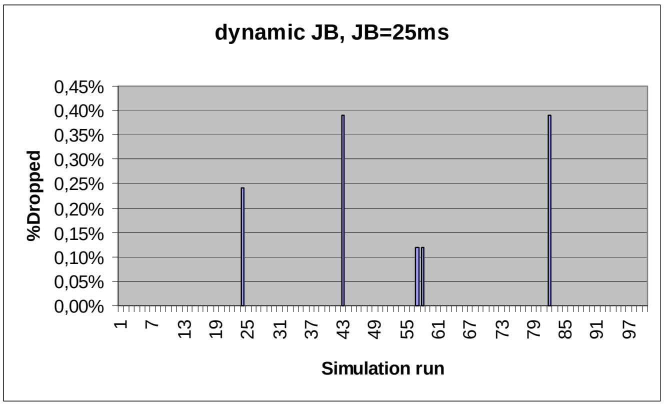


Figure 65: Jitter buffer loss for 25 ms dynamic jitter buffer

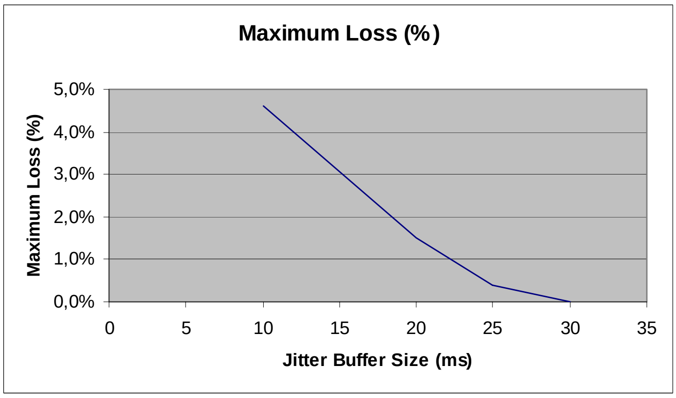


Figure 66: Maximum jitter buffer loss depending on dynamic jitter buffer size

5.8.1.3 Comparison between Static and Dynamic Jitter Buffer Simulations

Figure 67 compares the maximum loss of a static and a dynamic jitter buffer, as given earlier in figures 62 and 66, except for the high value for the 10 ms static jitter buffer.

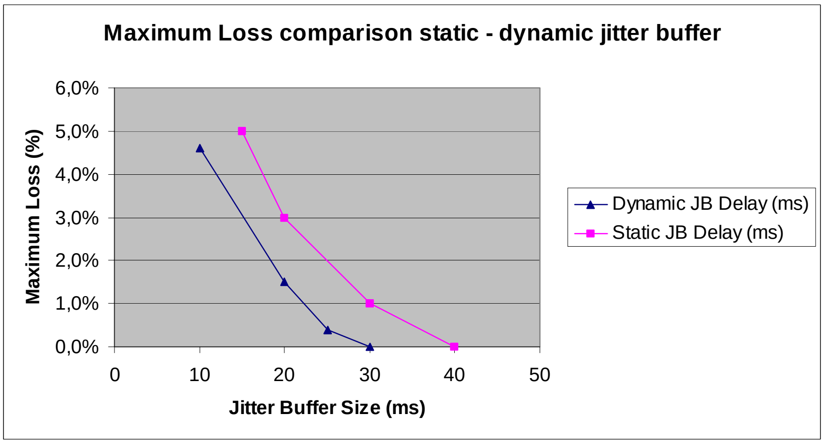


Figure 67: Comparing loss between a static and a dynamic jitter buffer

On the basis of the simulation experiments leading to this result, one can conclude that the particular dynamic jitter buffer simulated can be around 10 ms smaller than the static jitter buffer, for the same jitter buffer loss.

5.8.2 Practical Experiments

5.8.2.1 Introduction

These theoretical findings and evaluation results were checked in practical experiments with prototype implementations. By testing the equipment in terms of speech quality in an experimental Set Up and comparing the results with success criteria for a realistic telephony network, it can be determined how well the equipment functions.

The equipment in the experimental environment has to meet the success criteria for speech quality in a realistic telephony network, i.e. each telephony connection is considered to be successful ("toll quality") if it fits each of the following requirements.

Table 23: The success criteria for speech quality in a realistic telephony network

No	Success criteria	Value	
1	One-way delay	<= 150 [ms]	ITU-T Recommendation G.114 [43]
2	Listening quality	>= 4,0 [MOS]	PESQ (draft ITU-T Recommendation P.862 [39])
3	Conversational quality	>= 3,5 [MOS]	Expert listener events

The measurements following in this clause are the results of tests with an Internet simulator that can cause packet loss, delay and delay jitter. In these tests the packet loss was varied from 0 to 10 %.

5.8.2.2 Results

The influence of packet loss using an Internet simulator for an example implementation is given in figures 68 and 69. The results show a dramatic decrease of the speech quality when one percent packet loss was introduced. In addition, with one percent packet loss, the one-way delay increases from 100 ms to 400 ms.

The expected behaviour would be the opposite of what was measured. When proper jitter buffers are implemented (and packet loss concealment is used) one would expect that for low percentages of packet loss the listening quality would not drop that much compared to 0 % loss. Also (low values of) packet loss in itself is no reason for the excessive increase in end-to-end delay.

These effects can possibly be explained due to the poor working of jitter buffer implementations. In practice it appeared very difficult to find out the explicit algorithms that were used by vendors in implementations of jitter buffers.

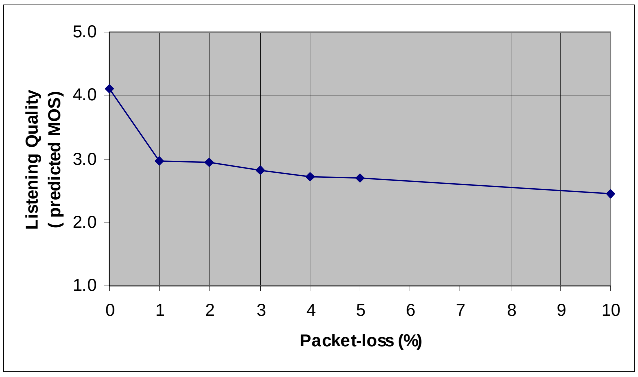


Figure 68: The packet-loss adjustment of the internet simulator versus the listening quality measured with PESQ

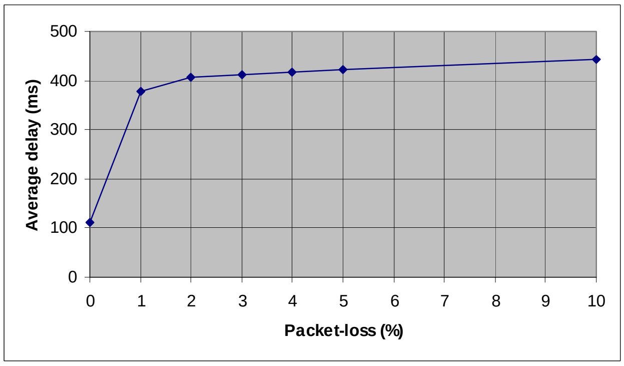


Figure 69: The packet-loss adjustment of the internet simulator versus the average delay

5.8.3 Conclusions

The end-to-end speech quality is very unpredictable when packet loss occurs, possibly due to poor jitter buffer implementations. Dramatic decreases in conversational quality were observed in various tests under realistic network conditions with prototype implementations.

Annex A: Bibliography

ETSI TS 102 024: "Telecommunications and Internet protocol Harmonization Over Networks (TIPHON) Release 4; End to End Quality of Service in TIPHON Systems".

ETSI TS 102 025: "Telecommunications and Internet protocol Harmonization Over Networks (TIPHON) Release 5; End to End Quality of Service in TIPHON Systems".

History

Document history		
V1.1.1	July 2000	Publication
V2.1.1	February 2002	Publication