ETSI TR 126 935 V7.0.0 (2007-06)

Technical Report

Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Packet Switched (PS) conversational multimedia applications; Performance characterization of default codecs (3GPP TR 26.935 version 7.0.0 Release 7)

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS

Reference
RTR/TSGS-0426935v700

Keywords
GSM, UMTS

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, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Copyright Notification

Intellectual Property Rights

Foreword

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

The present document provides information on the performances of default speech codecs in packet switched conversational multimedia applications. The codecs under test are AMR-NB (Adaptive Multi-Rate Narrowband) and AMR-WB (Adaptive Multi-Rate Wideband). In addition, several ITU-T codecs (G.723.1, G.729, G.722 and G.711) are included in the testing. Experimental test results from the speech quality testing are reported to illustrate the behaviour of these codecs.

The results give information of the performance of PS conversational multimedia applications under various operating and transmission conditions (e.g., considering radio transmission errors, IP packet losses, end-to-end delays, and several types of background noise). The performance results can be used e.g. as guidance for network planning and to appropriately adjust the radio network parameters.

2 References

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

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*

[1] ITU-T Recommendation P.800: "Methods for Subjective Determination of Transmission Quality".

[2] ITU-T Recommendation P.831: "Subjective performance evaluation of network echo cancellers".

[3] ITU-T Recommendation G.711: "Pulse code modulation (PCM) of voice frequencies".

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

[5] ITU-T Recommendation G.723.1: "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s".

[6] ITU-T Recommendation G.722: "7 kHz audio-coding within 64 kbit/s".

[7] IETF RFC 1889: "RTP: A Transport Protocol for Real-Time Applications".

[8] IETF RFC 3267: "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs".

[9] 3GPP TS 34.121: "Terminal Conformance Specification, Radio Transmission and Reception (FDD)" (downlink).

[10] 3GPP TS 25.141: " Base Station (BS) conformance testing (FDD)" (uplink).

[11] 3GPP TR 25.853 "Delay budget within the access stratum".

[12] 3GPP TS 26.235: "Packet switched conversational multimedia applications; Default codecs".

[13] 3GPP TS 26.071: "AMR speech Codec; General description".

[14] 3GPP TS 26.171: "AMR speech codec, wideband; General description".

[15] 3GPP TS 25.322: "Radio Link Control (RLC) protocol specification".

[16] IETF RFC 3095: "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed".

[17] 3GPP TS 34.108 v4.7.0: "Common test environments for User Equipment (UE) conformance testing".

[18] ETSI TR 101 112 v3.1.0 (1997-11): "Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS" (UMTS 30.03 v3.1.0).

3 Abbreviations

3.1 Abbreviations

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

AMR-NB (or AMR)	Adaptive Multi-Rate Narrowband Speech Codec
AMR-WB	Adaptive Multi-Rate Wideband Speech Codec
ANOVA	Analysis of Variance
ASY	ASYmmetric conditions
BLER	Block Error Rate
CMR	Codec Mode Request
COND	Test CONDitions
CN	Core Network
CRC	Cyclic Redundancy Check
DCH	Dedicated Channel
DL	Downlink
DPCH	Dedicated Physical Channel
DTCH	Dedicated Traffic Channel
Eb/No	Ratio of energy per modulating bit to the noise spectral density
FER	Frame Erasure Rate, Frame Error Rate
GAL	Global Analysis Laboratory
GQ	Global Quality (of the conversation)
IA	InterAction (with your partner)
IP	Internet Protocol
ITU-T	International Telecommunication Union - Telecommunications Standardization Sector
LAB	Listening LABoratory
MAC	Medium access control
MANOVA	Multivariate Analysis of Variance
Log-MAP	Logarithmic Maximum A Posteriori
MOS	Mean Opinion Score
NB	Narrowband
PC	PerCeption of impairments (also: Personal Computer)
PDCP	Packet Data Convergence Protocol
PDU	Protocol Data Unit
Pa	Sound Pressure Level (in Pascal)
PL	Packet Loss
plc	Packet Loss Concealment
RC	Radio Conditions
PS	Packet Switched
RB	Radio Bearer
RAB	Radio Access Bearer
RCV	Receive
RLC	Radio Link Control
ROHC	Robust Header Compression
RRM	Radio Resource Management
RTCP	Real-Time Control Protocol
RTP	Real-time Transport Protocol
SYM	SYMmetric conditions
TB size	Transport Block size
TF	Transport Format
ToC	Table of Content

TrCH	Transmission Channel
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UM	Unacknowledged Mode
UMD	Unacknowledged Mode Data
US	difficulty UnderStanding (your partner)
VOIP	Voice over IP
VQ	Voice Quality (of your partner)
WB	Wideband
XMIT	Transmit

4 General Overview

In order to characterize the quality of default speech codecs for packet switched conversational multimedia (AMR-NB and AMR-WB codecs) [12], two series of listening tests were conducted. The testing was carried out from October 2003 until February 2004. The tests were separated into two phases: Phase 1 considered the default speech codecs AMR-NB [13] and AMR-WB [14] in various operating conditions. Phase 2 considered also several other codecs including ITU-T codecs G.723.1 [5], G.729 [4], G.722 [6] and G.711 [3].

In Phase 1, France Telecom R&D acted as host laboratory. The subjective testing laboratories were ARCON for the North American English language, France Telecom R&D for the French language and NTT-AT for the Japanese language. Phase 1 tests consisted of 24 test conditions both for the AMR codec (modes 6.7 and 12.2 kbit/s) and the AMR-WB codec (modes 12.65 and 15.85 kbit/s) with error conditions covering both IP packet loss of 0% and 3% and radio conditions with 10^{-2}, 10^{-3} and 5 10^{-4} BLER (Block Error Rate). End-to-end delays of 300 and 500 ms were covered. Robust Header Compression (ROHC), an optional UMTS functionality, was included for some test cases for AMR-WB. Three types of background noise were used: car, street and cafeteria.

In Phase 2, France Telecom R&D acted as host and listening laboratory. Two languages were used (French and Arabic). The following codecs were tested: AMR-NB (modes 6.7 and 12.2 kbit/s), AMR-WB (modes 12.65 and 15.85 kbit/s), ITU-T G.723.1 (mode 6.4 kbit/s), ITU-T G.729 (mode 8 kbit/s), ITU-T G.722 (mode 64 kbit/s) and ITU-T G.711 (64 kbit/s). Transmission error conditions covered IP packet loss of 0% and 3%.

Siemens provided the real time air interface simulator for the Phase 1. France Telecom provided the IP core network simulator and terminal simulator used in both experiments Phase 1 and Phase 2. IPv6 was employed in the testing. (IPv6 is fully simulated over the radio interface. The CN simulator employs IPv4 but since the only impact is a marginal difference in the end-to-end delay - of the order of ~16 ìs - the use of a particular IP-version in CN part has no impact on the performance results.)

These tests were the first ever conversational tests conducted in any standardization body. Performance evaluation consisted of assessment of 5 aspects: 1) voice quality, 2) difficulty of understanding words, 3) quality of interaction, 4) degree of impairments, and 5) global communication quality. A 5-category rating scale was used for each aspect.

Dynastat performed the global analysis for both phases.

5 Test bed and test plan for Phase 1

This section describes the test plan for the Phase 1 of the conversation test of the AMR-NB (AMR) and AMR-WB in PS networks. All the laboratories participating to this conversation test phase used the same test plan, just the language of the conversation changed. Even if the test rooms or the test equipments are not exactly the same in all the laboratories, the calibration procedures and the tests equipment characteristics and performance guaranteed the similarity of the test conditions.

Annex B contains the instructions for the subjects participating to the conversation tests.

5.1 Test methodology

The protocol described below evaluates the effect of degradation such as delay and dropped packets on the quality of the communications. It corresponds to the conversation-opinion tests recommended by the ITU-T P.800 [1]. First of all, conversation–opinion tests allow subjects passing the test to be in a more realistic situation, close to the actual service conditions experienced by telephone customers. In addition, conversation-opinion tests are suited to assess the effects of impairments that can cause difficulty while conversing (such as delay).

Subjects participate to the test by couple; they are seated in separate sound-proof rooms and are asked to hold a conversation through the transmission chain performed by means of UMTS simulators, and communications are impaired by means of an IP impairments simulator part of the CN simulator and by the air interface simulator, as Figure 1 describes it.

The network configurations (including the terminal equipments) are symmetrical (in the two transmission paths). The only dissymmetry will be due to presence of background noise in one of the test rooms.

5.2 Test arrangement

5.2.1 Description of the testing system

Figure 1 describes the simulation system.

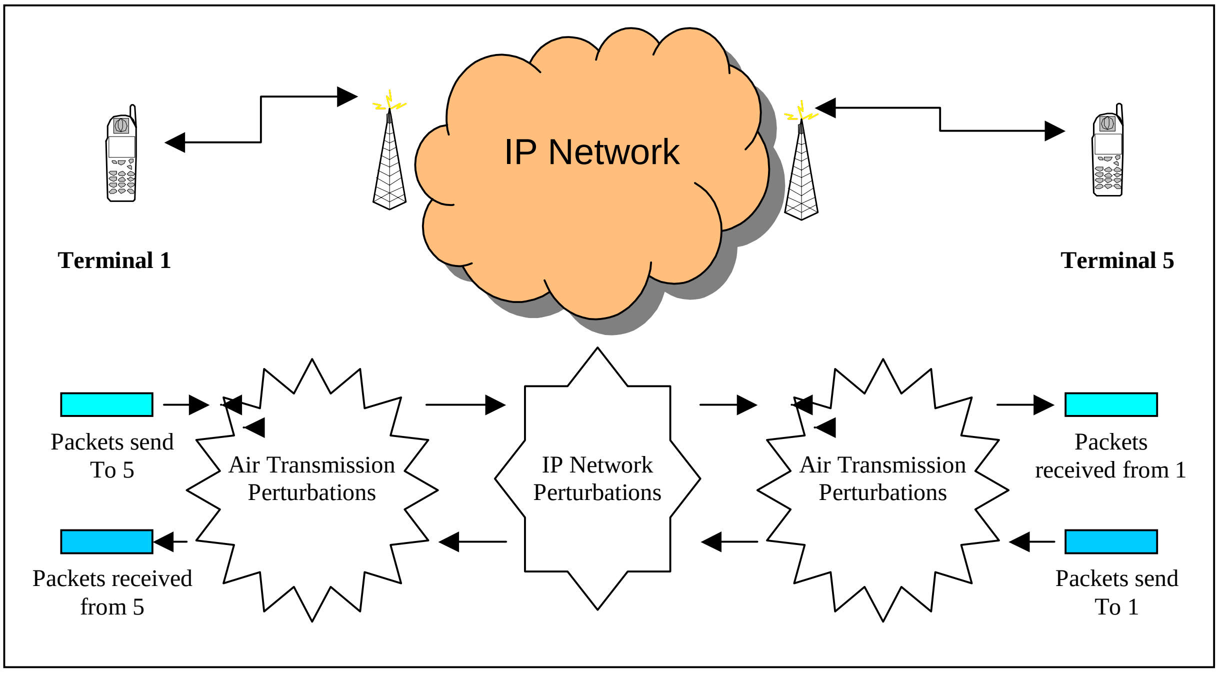

Figure 1: Packet switch audio communication simulator

The PS audio communication has been simulated using 5 PCs as shown in Figure 2.

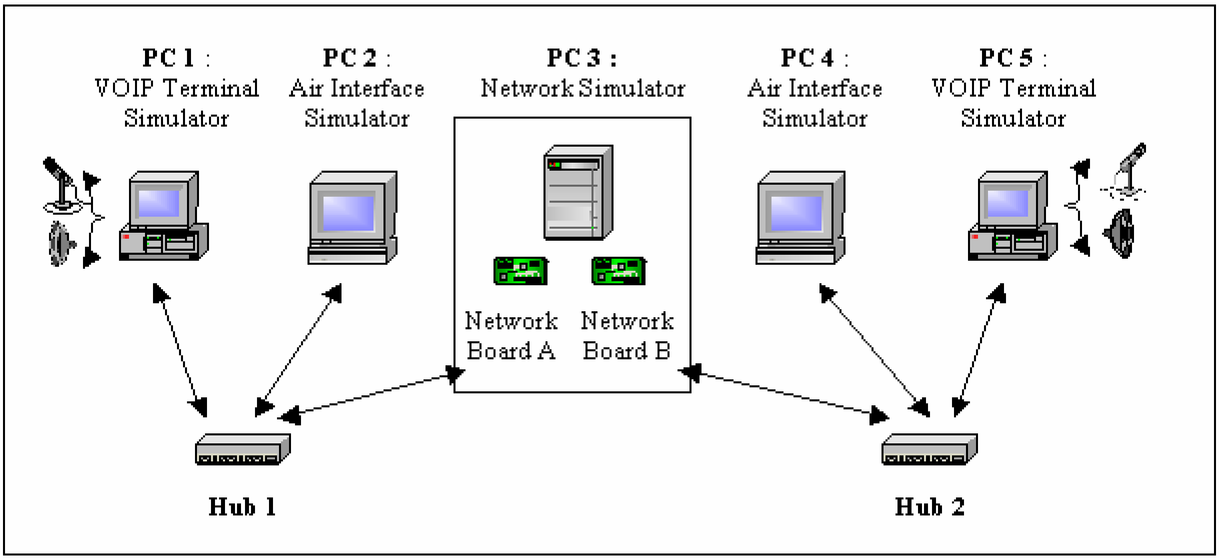

Figure 2: Simulation Platform

PC 1 and PC 5 are running under Windows OS with the VOIP Terminal Simulator Software of France Telecom R&D. PC 2 and PC 4 run under Linux OS with the Air Interface Simulator coming from Siemens AG. And PC 3 runs under WinNT OS with Network Simulator Software (NetDisturb).

The platform simulates a PS interactive communication between two users using PC 1 and PC 5 as their relative VOIP terminals. PC 1 sends AMR (or AMR-WB) encoded packets that are encapsulated using IP/UDP/RTP headers to PC 5. PC 1 receives IP/UDP/RTP audio packets from PC 5.

In fact, the packets created in PC 1 are sent to PC 2. PC 2 simulates the air interface uplink (UL) transmission and then forwards the transmitted packets to PC 4.

In the same way, PC 4 simulates the air interface downlink (DL) transmission and then forwards the packets to PC 5. PC 5 decodes and plays the speech back to the listener.

5.2.2 Network simulator

The core network simulator, as implemented, works under IPv4. However, as the core network simulator acts only on packets (loss, delay,…) the use of IPv4 or IPv6 is equivalent for this test conversation context. Considering the networks perturbations introduced by the simulator and the context of the interactive communications, the simulation using IPv4 perturbation network simulator is adapted to manage and simulate the behaviours of an IPv6 core network.

Figure 3 shows the possible network simulator parameters that can be modified.

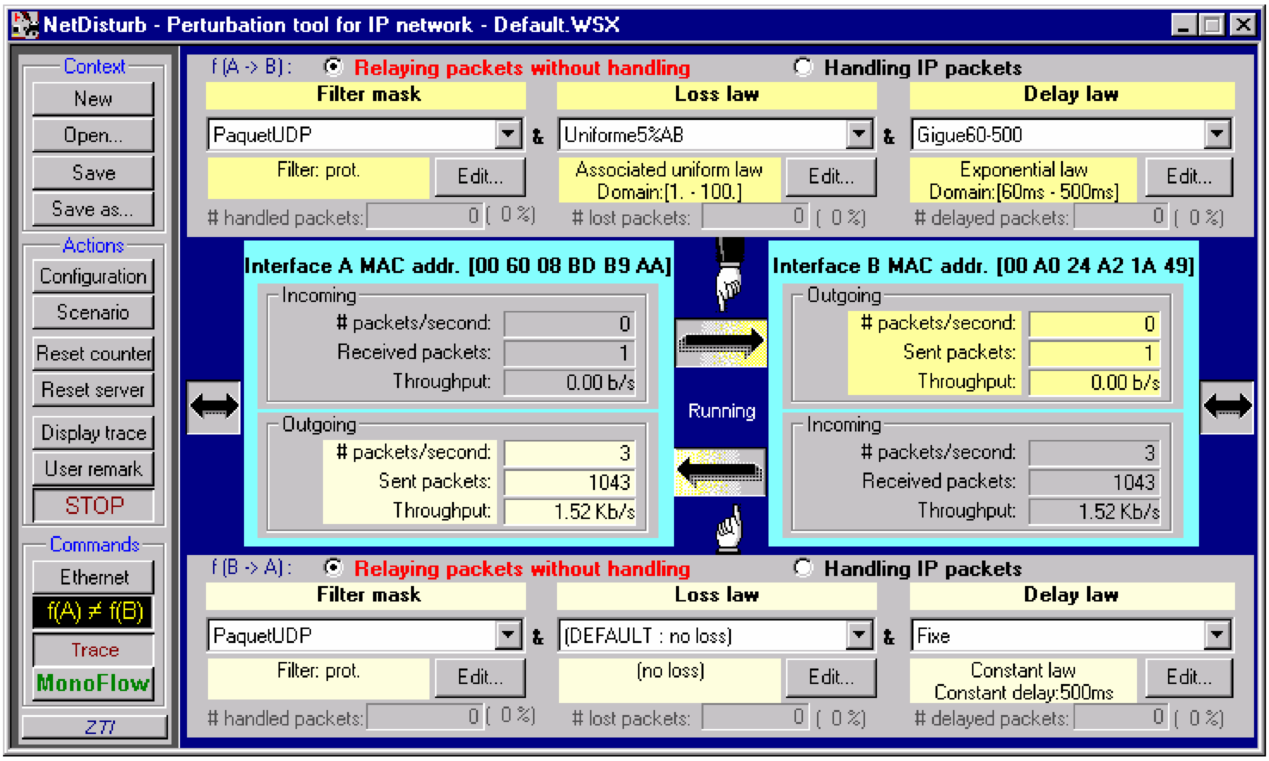

Figure 3: IP simulator interface

On both links, one can choose delay and loss laws. Both links can be treated separately or in the same way. For example, delay can be set to a fixed value but can also be set to another law such as exponential law.

Only loss law and delay law were given values, for delay law the values are 0 or 200 ms and for loss law the possible values: 0% or 3% under bursty law. Both links were treated in the same way.

5.2.3 UMTS simulator choices

The transmission of IP/UDP/RTP/AMR (or AMR-WB) packets over the UMTS air interface is simulated using the RAB described in Section 5.2.3.1. The required functions of the RLC layer are implemented according to [15] and work in real-time. The underlying Physical Layer is simulated offline. Error patterns of block errors (i.e. discarded RLC PDUs) are inserted in the real-time simulation as described in Section 5.2.3.2. For more details on the parameter settings of the Physical Layer simulations see Section 5.2.3.3.

5.2.3.1 RAB and protocols

For the narrowband conversational tests, the AMR is encoded with a maximum of 12.2 kbit/s. The bitstream is encapsulated using IP/UDP/RTP protocols. The air interface simulator receives IPv4 packets from the CN simulator. The RTP packets are extracted and before transmission over the air interface, IPv6/UDP headers are inserted. Finally real IPv6 packets are transmitted over the air interface simulator.

The payload format is the following:

- RTP payload format for AMR-NB (cf. [8]) is used;
- Bandwidth efficient mode is used;
- One speech frame is encapsulated in each RTP packet;
- Interleaving is not used;

The payload header consists of the 4 bits of the CMR (Codec Mode Request). Then 6 bits are added for the ToC (Table of Content). For IPv4, this corresponds to a maximum of 72 bytes per frame that is to say 28.8 kbit/s. This goes up to 92 bytes (36.8 kbit/s) when using IPv6 protocol on the air interface.

RTCP packets are sent. However, in the test conditions defined in the conversation test plans, RTCP is not mandatory, as it is not in a multicast environment (cf. [7]). RTCP reports were sent but not used.

ROHC is an optional functionality in UMTS. In order to reduce the size of the tests and the number of conditions, the ROHC algorithm is not used for the AMR-NB conversation test. This functionality is only tested in the wideband condition.

For the WB conversational tests, the AMR-WB encodes speech at a maximum of 15.85 kbit/s. The bitstream is also encapsulated and transmitted in the same way as for the NB case. For IPv4 a maximum of 81 bytes (41 bytes for the AMR and its payload header plus the 40 bytes of the IP/UDP/RTP headers) per frame are transmitted that is to say 32.4 kbit/s, this goes up to 101 bytes (40.4 kbit/s) when using IPv6 protocol on the air interface.

ROHC algorithm is supported in the AMR-WB conversation test, for the 12.65 kbit/s mode and the 15.85 modes. Header compression is done on the IP/UDP/RTP headers (profile 1). ROHC starts in the unidirectional mode and switches to bi-directional mode as soon as a packet has reached the decompressor and replied with a feedback packet indicating that a mode transition is desired.

The Conversational / Speech / UL:46 DL:46 kbps / PS RAB coming from [17] was used. It is not an optimal RAB for PS conversational test but it was the only one available at the time the test bed and the air interface simulator were designed. The RAB description is given in Table 1.

Table 1: RAB description

Higher layer	RAB/Signalling RB		RAB
PDCP	PDCP header size, bit		8
RLC	Logical channel type		DTCH
	RLC mode		UM
	Payload sizes, bit		920, 304, 96
	Max data rate, bps		46000
	UMD PDU header, bit		8
MAC	MAC header, bit		0
	MAC multiplexing		N/A
Layer 1	TrCH type		DCH
	TB sizes, bit		928, 312, 104
	TFS	TF0, bits	0x928
		TF1, bits	1x104
		TF2, bits	1x312
		TF3, bits	1x928
	TTI, ms		20
	Coding type		TC
	CRC, bit		16
	Max number of bits/TTI after channel coding		2844
	Uplink: Max number of bits/radio frame before rate matching		1422
	RM attribute		180-220

5.2.3.2 Description of the RLC implementation

The UMTS air interface simulator (implemented in PC 2 and 4) receives IP/UDP/RTP/AMR (or AMR-WB) packets on a specified port of the network card (see Figure 4). The IP/UDP/RTP/AMR (or AMR-WB) packets are given to the transmission buffer of the RLC layer, which works in Unacknowledged Mode (UM). The RLC segments or concatenates the IP bitstream in RLC PDUs, adding appropriate RLC headers (sequence number and length indicators). It is assumed that always Transport Format TF 3 is chosen on the physical layer, providing an RLC PDU length including header of 928 bits. In the regular case, one IP packet is placed into an RLC PDU that is filled up with padding bits. Due to delayed packets from the network simulator it may also occur that there are none or no more than one IP packet in the RLC transmission buffer to transmit in the current TTI.

Each TTI of 20ms, an RLC PDU is formed. It is then given to the error insertion block that decides if the RLC PDU is transmitted successfully over the air interface or if it is discarded due to a block error after channel decoding. The physical layer is not simulated in real time, but error pattern files are provided. The error patterns of the air interface transmission are simulated offline according to the settings given in Section 5.2.3.1. They consist of binary decisions for each transmitted RLC PDU, resulting in a certain BLER.

After the error pattern insertion, the RLC of the air interface receiver site receives RLC PDUs in the reception buffer. The sequence numbers of the RLC headers are checked to detect when RLC PDUs have been discarded due to block errors. A discarded RLC PDU can result in one or more lost IP packets, resulting in a certain packet loss rate of the IP packets and thereby in a certain FER of the AMR (or AMR-WB) frames. The IP/UDP/RTP/AMR (or AMR-WB) packets are reassembled and transmitted to the next PC. This PC is either the network simulator (PC 3) in case of uplink transmission, or is one of the terminals (PC 1 or PC 5) in case of downlink transmission.

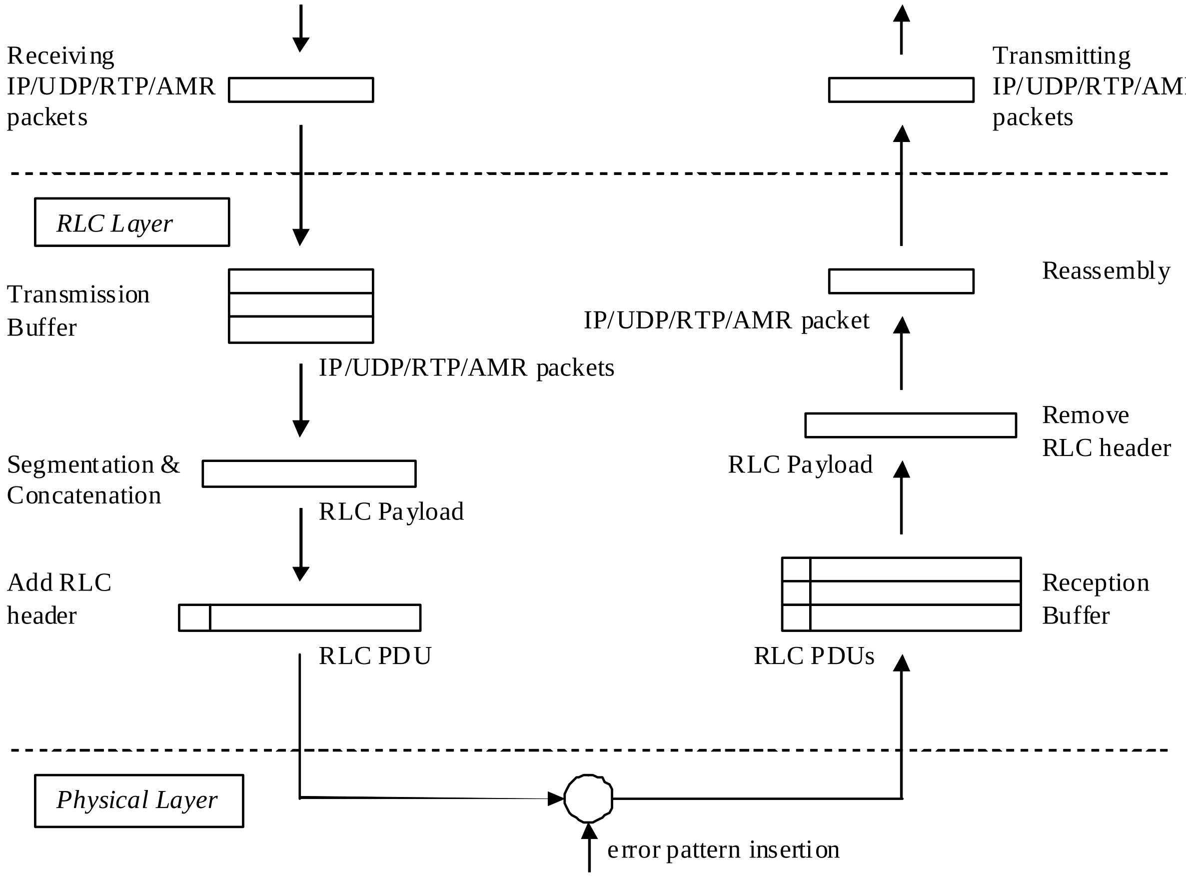

Figure 4: UMTS air interface simulation

5.2.3.3 Physical Layer Implementation

The parameters of the physical layer simulation were set according to the parameters for a DCH in multipath fading conditions given in [9] for the downlink and [10] for the uplink. The TB size is 928 bits and the Turbo decoder uses the Log-MAP algorithm with 4 iterations. The rake receiver has 6 fingers at 60 possible positions.

The different channel conditions given in Tables 2, 3 and 4 were extracted from [18] (Selection procedures for the choice of radio transmission technologies of the UMTS).

Table 2: Indoor Office Test Environment Tapped-Delay-Line Parameters

Tap	Channel A		Doppler
	Rel. Delay (nsec)	Avg. Power (dB)	Spectrum
1	0	0	FLAT
2	50	-3.0	FLAT
3	110	-10.0	FLAT
4	170	-18.0	FLAT
5	290	-26.0	FLAT
6	310	-32.0	FLAT

Table 3: Vehicular Test Environment, High Antenna, Tapped-Delay-Line Parameters

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0.0	CLASSIC
2	310	-1.0	CLASSIC
3	710	-9.0	CLASSIC
4	1090	-10.0	CLASSIC
5	1730	-15.0	CLASSIC
6	2510	-20.0	CLASSIC

Table 4: Outdoor to Indoor and Pedestrian Test Environment Tapped-Delay-Line Parameters

Tap	Channel A		Doppler
	Rel. Delay (nsec)	Avg. Power (dB)	Spectrum
1	0	0	CLASSIC
2	110	-9.7	CLASSIC
3	190	-19.2	CLASSIC
4	410	-22.8	CLASSIC
5	-	-	CLASSIC
6	-	-	CLASSIC

Table 5 (DL) and Table 6 (UL) show approximate results of the air interface simulation for $\frac{DPCH\_E_c}{I_{or}}$ and E_b/N_0 corresponding to the considered BLERs.

Table 5: Downlink performance - approximate $\frac{DPCH\_E_c}{I_{or}}$ for the different channels and BLER

	BLER			
Channel	$5*10^{-2}$	$1*10^{-2}$	$1*10^{-3}$	$5*10^{-4}$
Indoor, 3 km/h ($\hat{I}_{or}/I_{oc}$ = 9 dB)	-13.1 dB	-8.9 dB	-3.4 dB	-2.4 dB
Outdoor to Indoor, 3 km/h ($\hat{I}_{or}/I_{oc}$ = 9 dB)	-13.2 dB	-9.7 dB	-5.9 dB	-5.2 dB
Vehicular, 50 km/h ($\hat{I}_{or}/I_{oc}$ = -3 dB)	-9.35 dB	-8.2 dB	-6.9 dB	-6.55 dB
Vehicular, 120 km/h ($\hat{I}_{or}/I_{oc}$ = -3 dB)	-9.7 dB	-8.95 dB	-7.95 dB	-7.55 dB

Table 6: Uplink performance - approximate E_b/N_0 for the different channels and BLER

Channel	BLER			
	$5*10^{-2}$	$1*10^{-2}$	$1*10^{-3}$	$5*10^{-4}$
Indoor, 3 km/h	3.9 dB	6.4 dB	9.2 dB	9.8 dB
Outdoor to Indoor, 3 km/h	3.7 dB	6.1 dB	8.6 dB	9.2 dB
Vehicular, 50 km/h	-0.9 dB	-0.15 dB	0.55 dB	0.75 dB
Vehicular, 120 km/h	0.2 dB	0.6 dB	1.1 dB	1.3 dB

Outdoor to Indoor channel was used for uplink and downlink in the simulations.

5.2.4 Headsets and Sound Card

To avoid echo problems headsets were used instead of handsets. The monaural headsets are connected to the sound cards of the PCs supporting the speech codec simulators.

The sound level in the earphones can be adjusted, if needed, by the users. But, in practice, the original settings, defined during the preliminary tests, and producing a comfortable listening level, are not modified. The microphones are protected by a foam ball in order to reduce the "pop" effect. It is also suggested to the user to avoid placing the acoustic opening of the microphone in front of the mouth.

5.2.5 Test environment

Each of the two subjects participating to the conversations are installed in a test room. They sit on an armchair, in front of a table. The test rooms are acoustically insulated. All the test equipments are installed in a third room, connected to the test rooms. When needed, the background noise is generated in the appropriate test room through a set of 4 loudspeakers. The background noise level is adjusted and controlled by a sound level meter. The measurement microphone, connected to the Sound level meter is located at the equivalent of the center of the subject's head. The noise level is A weighted.

5.2.6 Calibration and test conditions monitoring

5.2.6.1 Speech level

Before the beginning of a set of experiment, the end-to-end transmission level is checked subjectively, to ensure that there is no problem. If it is necessary to check the speech level following procedure is applied. An artificial mouth placed in front of the microphone of the Headset A, in the LRGP position (see ITU-T Rec. P.64), generates in the artificial ear (according to ITU-T Rec. P57), coupled to the earphone of the Headset B, the nominal level. If necessary, the level is adjusted with the receiving volume control of the headset. Similar calibration is done by inverting headsets A and B.

5.2.6.2 Delay

The overall delay (from the input of sound card A to the output of sound card B) is calculated as shown: On the air interface side, the simulator only receives packets on its network card, processes them and transmits every 20 ms these packets to the following PC. Only processing delay and a possible delay due to a jitter can be added (a packet arrives just after the sending window of the air interface).

The delay is calculated as shown below:

- Encoder side: delay due to account framing, look-ahead, processing and packetization = 45ms
- Uplink delay between UE and Iu: 84.4 ms (see [11])
- Core network delay: a few ms
- Routing through IP: depending on the number of routers.
- Downlink delay between Iu and UE: 71.8 ms (see [11])
- Decoder side, taking into account jitter buffer, de-packetization and processing: 40 ms

The total delay to be considered is at least: 241.2 ms.

5.3 Test conditions for AMR-NB codec

Tables 7 - 9 summarise the test conditions used for AMR-NB testing.

For both AMR-NB and AMR-WB codecs two representative modes were chosen for the testing. The lowest codec modes (such as AMR-NB 4.75) were not included since these are intended to be used mainly temporarily to cope with poor radio conditions. They were expected to provide insufficient quality for conversational applications if used throughout the call (as done in these characterisation tests).

Table 7: Test conditions for AMR-NB

Cond.	Background noise in Room A	Background noise in Room B	Experimental factors		
			Radio cond. (BLER)	IP cond. (Packet loss ratio)	Mode + delay
1	No	No	10^{-2}	0%	6.7 kbit/s (delay 300 ms)
2	No	No	10^{-2}	0%	12.2 kbit/s (delay 500 ms)
3	No	No	10^{-2}	0%	12.2 kbit/s (delay 300 ms)
4	No	No	10^{-2}	3%	6.7 kbit/s (delay 300 ms)
5	No	No	10^{-2}	3%	12.2kbit/s (delay 500 ms)
6	No	No	10^{-2}	3%	12.2 kbit/s (delay 300 ms)
7	No	No	10^{-3}	0%	6.7 kbit/s (delay 300 ms)
8	No	No	10^{-3}	0%	12.2 kbit/s (delay 500 ms)
9	No	No	10^{-3}	0%	12.2 kbit/s (delay 300 ms)
10	No	No	10^{-3}	3%	6.7 kbit/s (delay 300 ms)
11	No	No	10^{-3}	3%	12.2 kbit/s (delay 500 ms)
12	No	No	10^{-3}	3%	12.2 kbit/s (delay 300 ms)
13	No	No	$5\ 10^{-4}$	0%	6.7kbit/s (delay 300 ms)
14	No	No	$5\ 10^{-4}$	0%	12.2kbit/s (delay 500 ms)
15	No	No	$5\ 10^{-4}$	0%	12.2 kbit/s (delay 300 ms)
16	No	No	$5\ 10^{-4}$	3%	6.7kbit/s (delay 300 ms)
17	No	No	$5\ 10^{-4}$	3%	12.2 kbit/s (delay 500 ms)
18	No	No	$5\ 10^{-4}$	3%	12.2 kbit/s (delay 300 ms)
19	Car	No	$5\ 10^{-4}$	3%	12.2 kbit/s (delay 300 ms)
20	No	Car	$5\ 10^{-4}$	3%	12,2 kbit/s (delay 300 ms)
21	Cafeteria	No	$5\ 10^{-4}$	0%	6.7 kbit/s (delay 300 ms)
22	No	Cafeteria	$5\ 10^{-4}$	0%	6.7 kbit/s (delay 300 ms)
23	Street	No	$5\ 10^{-4}$	0%	12.2kbit/s (delay 500 ms)
24	No	Street	$5\ 10^{-4}$	0%	12.2kbit/s (delay 500 ms)

Table 8: Noise types for AMR-NB

Noise type	Level (dB Pa)
Car	60
Street	55
Cafeteria	50

Table 9: Test details for AMR-NB

Listening Level	1	79 dBSPL
Listeners	32	Naïve Listeners
Groups	16	2 subjects/group
Rating Scales	5	
Languages	3	North American English, French, Japanese
Listening System	1	Monaural headset (flat response in the audio bandwidth of interest: 50Hz-7kHz). The other ear is open.
Listening Environment		Room Noise: Hoth Spectrum at 30dBA (as defined by ITU-T Recommendation P.800: Annex A, section A.1.1.2.2.1 Room Noise, with table A.1 and Figure A.1), except when background noise is needed (see Table 8 of this TR).

5.4 Test conditions for AMR-WB codec

Tables 10 - 13 summarise the test conditions used for AMR-WB testing.

Table 10: Test conditions for AMR-WB

Cond.	Experimental factors		
	Radio conditions (BLER)	**IP conditions (Packet loss ratio)**	**Mode**
1	10^{-2}	0%	12,65 kbit/s, ROHC
2	10^{-2}	0%	12,65 kbit/s
3	10^{-2}	0%	15,85 kbit/s, ROHC
4	10^{-2}	3%	12,65 kbit/s, ROHC
5	10^{-2}	3%	12,65 kbit/s
6	10^{-2}	3%	15,85 kbit/s, ROHC
7	10^{-3}	0%	12,65 kbit/s, ROHC
8	10^{-3}	0%	12,65 kbit/s
9	10^{-3}	0%	15,85 kbit/s, ROHC
10	10^{-3}	3%	12,65 kbit/s, ROHC
11	10^{-3}	3%	12,65 kbit/s
12	10^{-3}	3%	15,85 kbit/s, ROHC
13	$5. 10^{-4}$	0%	12,65 kbit/s, ROHC
14	$5. 10^{-4}$	0%	12,65 kbit/s
15	$5. 10^{-4}$	0%	15,85 kbit/s, ROHC
16	$5. 10^{-4}$	3%	12,65 kbit/s, ROHC
17	$5. 10^{-4}$	3%	12,65 kbit/s
18	$5. 10^{-4}$	3%	15,85 kbit/s, ROHC

Table 11: Test conditions with noise for AMR-WB

Cond.	Additional Background noise Room A	Additional Background noise Room B	Experimental factors		
			Radio conditions (BLER)	IP conditions (Packet loss ratio)	Mode
19	Car	No	$5\ 10^{-4}$	3%	12,65 kbit/s, ROHC
20	No	Car	$5\ 10^{-4}$	3%	12,65 kbit/s, ROHC
21	Cafeteria	No	$5\ 10^{-4}$	0%	12,65 kbit/s
22	No	Cafeteria	$5\ 10^{-4}$	0%	12,65 kbit/s
23	Street	No	$5\ 10^{-4}$	0%	15,85 kbit/s, ROHC
24	No	Street	$5\ 10^{-4}$	0%	15,85 kbit/s, ROHC

Table 12: Noise Types for AMR-WB

Noise type	Level (dB Pa)
Car	60
Street	55
Cafeteria	50

Table 13: Test details for AMR-WB

Listening Level	1	79 dBSPL
Listeners	32	Naïve Listeners
Groups	16	2 subjects/group
Rating Scales	5	
Languages	3	North American English, French, Japanese
Listening System	1	Monaural headset (flat response in the audio bandwidth of interest: 50Hz-7kHz). The other ear is open.
Listening Environment		Room Noise: Hoth Spectrum at 30dBA (as defined by ITU-T Recommendation P.800: Annex A, section A.1.1.2.2.1 Room Noise, with table A.1 and Figure A.1),except when background noise is needed (see Table 12 of this TR)

6 Test bed and test plan for Phase 2

The Phase 2 of the listening test was conducted by one listening test laboratory (FT R&D). The different speech coders used in this test are:

- Adaptive Multi-Rate Narrow-Band (AMR-NB), in modes 6.7 kbit/s and 12.2 kbit/s,
- Adaptive Multi-Rate Wide-Band (AMR-WB), in modes 12.65 kbit/s and 15.85 kbit/s,
- ITU-T G.723.1, in mode 6.4 kbit/s,
- ITU-T G.729, in mode 8 kbit/s,
- ITU-T G.722 (wideband codec), in mode 64 kbit/s, with packet loss concealment and,
- ITU-T G.711, with packet loss concealment.

As there is no standardized packet loss concealment for G.711 and G.722, proprietary packet loss concealment algorithms were used for them. The simulated network was tested under two values of IP packet loss (0% and 3%). The testing was done in one test laboratory only, but in two different languages (Arabic and French).

The IP packet contains 20 ms speech frames except for G.723.1 for which IP packet contains 30 ms speech. For G.729 the 20 ms packet consists of two 10 ms frames.

The test methodology was the same as the one applied in Phase 1.

Annex B contains the instructions for the subjects participating to the conversation tests.

6.1 Test arrangement

6.1.1 Description of the proposed testing system

Figure 5 describes the system that was simulated.

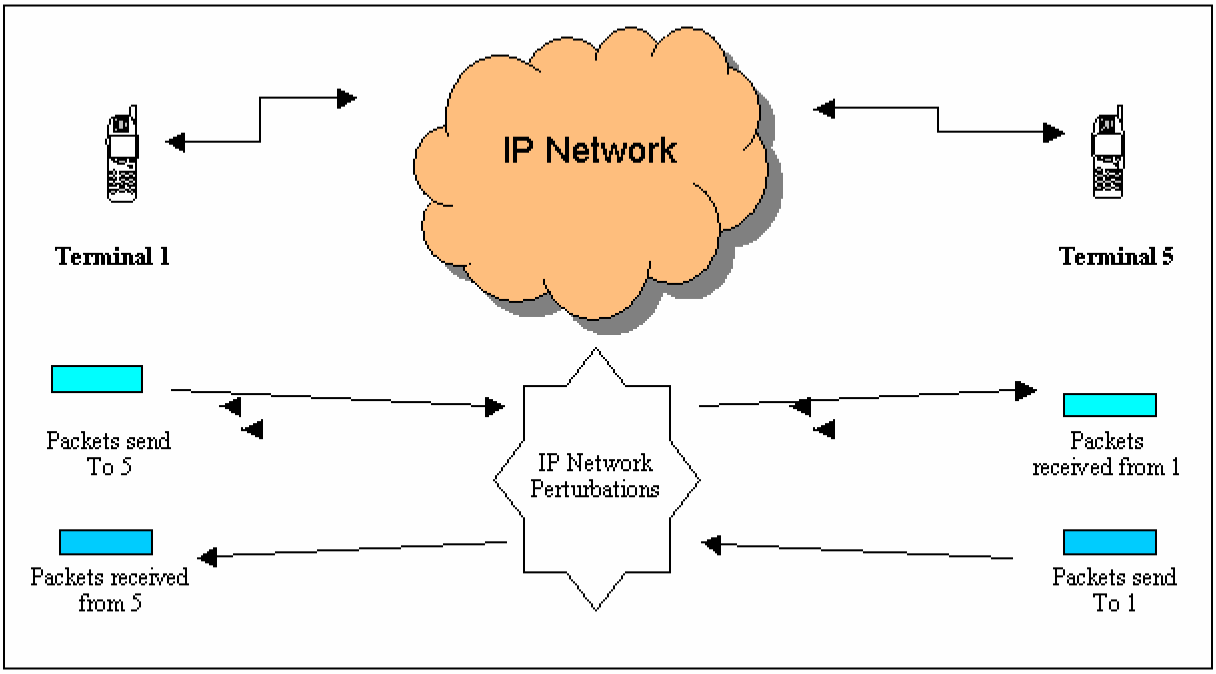

Figure 5: Packet Switched audio communication simulator

This was simulated using 3 PCs as shown in Figure 6.

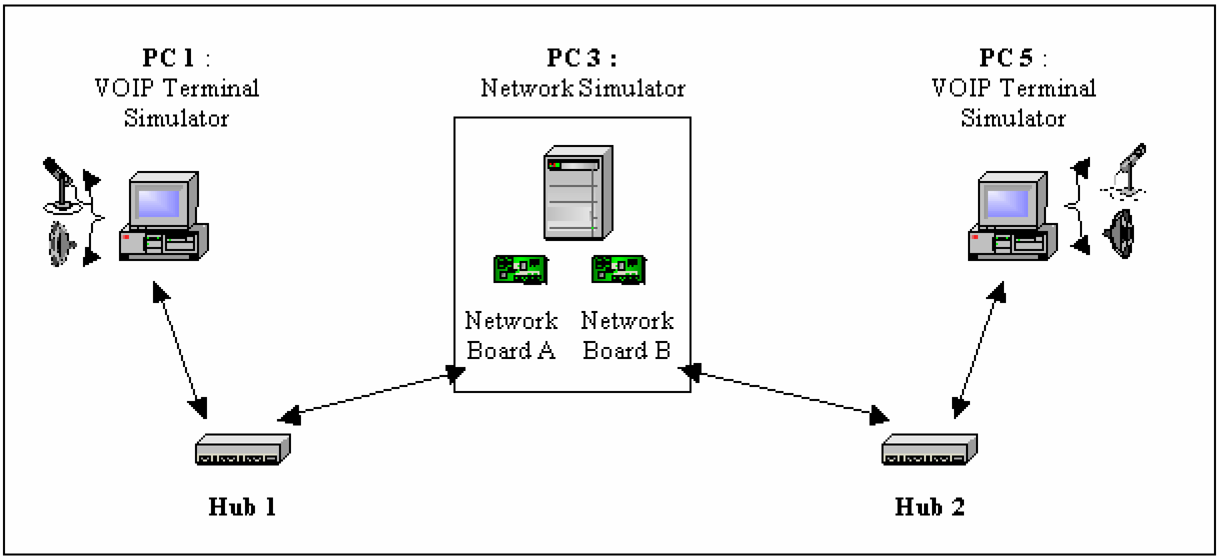

Figure 6: Simulation Platform

PC 1 and PC 5 run under Windows OS with VOIP Terminal Simulator Software of France Telecom R&D. PC 3 run under WinNT OS with Network Simulator Software (NetDisturb).

The platform simulates a packet switched interactive communication between two users using PC 1 and PC 5 as their relative VOIP terminals. PC 1 sends encoded packets that are encapsulated using IP/UDP/RTP headers to PC 5. PC 1 receives these IP/UDP/RTP audio packets from PC 5.

6.1.2 Network simulator

The core network simulator is the same as the one presented in Section 5. The different parameters that can be modified are presented in Figure 3 (Section 5.2.2).

In this test, only "loss law" has two values, all the others settings are fixed. On both links, one can choose delay and loss laws. Both links can be treated separately or in the same way. For example, delay can be set to a fixed value but it can also be set to another law such as exponential law. Only loss law was given values: 0% or 3% under bursty law. Both links were treated in the same way.

Headsets were here also used to reduce echo problems. The monaural headsets are connected to the sound cards of the PCs supporting the different codecs.

The sound level in the earphones can be adjusted, if needed, by the users. But, in practice, the original settings, defined during the preliminary tests, and producing a comfortable listening level, are not modified. The microphones are protected by a foam ball in order to reduce the "pop" effect. It is also suggested to the user to avoid placing the acoustic opening of the microphone in front of the mouth.

The same test environment as in test Phase 1 is used. Each of the two subjects participating to the conversations are installed in a test room. They sit on an armchair, in front of a table. The test rooms are acoustically insulated. All the test equipments are installed in a third room, connected to the test rooms. The background noise level is checked by a sound level meter. The measurement microphone, connected to the Sound level meter is located at the equivalent of the center of the subject's head. The noise level is A weighted.

6.1.3 Calibration and test conditions monitoring

The speech level checking is done in the same way as for Phase 1 (see Section 5.2.6.1).

The overall delay (from the input of sound card A to the output of sound card B) is adjusted for each test condition taking into account the delay of the related codec in order to have a fixed delay around 250ms. This value of 250ms is close to the hypothetical delay computed for AMR-NB and AMR-WB through the UMTS network.

6.2 Test Conditions

The test conditions and details are described in Tables 14 and 15.

Table 14: Test conditions

Cond.	Experimental factors	
	IP conditions (Packet loss ratio)	Mode
1	0%	AMR-NB 6,7kbit/s
2	0%	AMR-NB 12,2 kbit/s
3	0%	AMR-WB 12,65 kbit/s
4	0%	AMR-WB 15,85 kbit/s
5	0%	G. 723.1 6,4 kbit/s
6	0%	G.729 8 kbit/s
7	0%	G.722 64 kbit/s + plc
8	0%	G.711 + plc
9	3%	AMR-NB 6,7kbit/s
10	3%	AMR-NB 12,2 kbit/s
11	3%	AMR-WB 12,65 kbit/s
12	3%	AMR-WB 15,85 kbit/s
13	3%	G. 723.1 6,4 kbit/s
14	3%	G.729 8 kbit/s
15	3%	G.722 64 kbit/s + plc
16	3%	G.711 + plc

Table 15: Test details

Listening Level	1	79 dBSPL
Listeners	32	Naïve Listeners per language
Groups	16	2 subjects/group
Rating Scales	5	
Languages	2	French, Arabic
Listening System	1	Monaural headset (flat response in the audio bandwidth of interest: 50Hz-7kHz). The other ear is open.
Listening Environment		Room Noise: Hoth Spectrum at 30dBA (as defined by ITU-T Recommendation P.800: Annex A, section A.1.1.2.2.1 Room Noise, with table A.1 and Figure A.1)

7 Analysis of test results (for Phase 1 and 2)

This section presents the Global Analysis of the results. The analysis work was performed by Dynastat in its function as the Global Analysis Laboratory (GAL). Annex G presents the GAL Test Plan for characterizing the results of the conversation tests. (Detailed test plans are given in Annexes D and E for Phase 1 and in Annex F for Phase 2).

It should be noted that this is the first instance in any standardisation body of conversation tests being used to characterize the performance of standardized speech codecs, and the first instance of codecs in 3GPP being characterized for packet-switched networks. Moreover, the analyses reported in this document represent a new approach to evaluating the results of conversation tests.

7.1 Conversation Tests

The Phase 1 test plan describes the methodology for conducting the conversation tests. In general, the procedure involved a pair of subjects located in different rooms and communicating over a simulated packet-switched network. The subjects were involved in a task, which required them to communicate in order to solve a specific problem. At the end of their task, each subject was required to rate various aspects of the quality of their conversation. Each of these ratings involved a five-point scale with descriptors appropriate to the aspect of the conversation being rated. Table 16 shows a summary of the five rating scales. (The first row in each column shows the scale abbreviation that will be used throughout this report).

Table 16: Summary of Rating Scales used in the Conversation Tests

VQ		US		IA		PC		GQ	
Voice Quality of your partner		Difficulty Understanding your partner		Interaction with your partner		Perception of impairments		Global Quality of the conversation	
5	Excellent	5	Never	5	Excellent	5	None	5	Excellent
4	Good	4	Rarely	4	Good	4	Not disturbing	4	Good
3	Fair	3	Sometimes	3	Fair	3	Slightly disturbing	3	Fair
2	Poor	2	Often	2	Poor	2	Disturbing	2	Poor
1	Bad	1	All the time	1	Bad	1	Very Disturbing	1	Bad

Since each subject makes five ratings for each condition, there are five dependent variables involved in analyses of the response data. We would expect the ratings on the scales in Table 16 to show some degree of inter-correlation across test conditions. If, in fact, all five were perfectly correlated then we would conclude that they were each measuring the same underlying variable. In this scenario, we could combine them into a single measure (e.g., by averaging them) for purposes of statistical analyses and hypothesis testing. If, on the other hand, the ratings were uncorrelated, we would conclude that each scale is measuring a different underlying variable and should be treated separately in subsequent analyses. In practice, the degree of intercorrelation among such dependent variables usually falls somewhere between these two extremes. Multivariate Analysis of Variance (MANOVA) is a statistical technique designed to evaluate the results of experiments with multiple dependent variables and determine the nature and number of underlying variables. MANOVA was proposed in the GAL test plan for the conversation tests and was used extensively in the analyses presented in this report.

7.2 Experimental Design and Statistical Procedures

The two Phase 1 test plans, AMR Narrowband (AMR-NB) and AMR Wideband (AMR-WB), described similar experimental designs, each experiment involving 24 test conditions (*COND*) and 16 pairs of subjects. The test plans also specified that the experiments would be conducted by three Listening Laboratories (*LAB*), each in a different language: Arcon for North American English, NTT-AT for Japanese, and France Telecom for French.

Of the 24 conditions in both the NB and WB experiments, 18 were described as Symmetrical conditions (SYM), six as Asymmetrical (ASY). In the SYM conditions all subjects were located in a Quiet room, i.e., with no introduced background noise. The six ASY conditions were actually three pairs of conditions where one subject in each conversation-pair was located in a noisy background and the other subject was in the quiet. The data from these sets of paired conditions were sorted to effect a comparison of *sender in noise/receiver in quiet* and s*ender in quiet/receiver in noise* for the three conditions involving noise in the rooms.

The Phase 2 test plan described a single experiment involving 16 conditions conducted by one listening lab (France Telecom) but in two languages, French and Arabic.

For purposes of the GAL, the data from the three experiments, Phase 1-NB, Phase 1-WB, and Phase 2 were separated into five *Sets* of conditions for statistical analyses:

Set 1. Phase 1 - NB/SYM conditions (1-18)

Set 2. Phase 1 - NB/ASY conditions (19-24)

Set 3. Phase 1 - WB/SYM conditions (1-18)

Set 4. Phase 2 - WB/ASY conditions (19-24)

Set 5. Phase 2 - Ph2 conditions (1-16)

For each of these five set of conditions, a three-step statistical process was undertaken to attempt to simplify the final analyses and arrive at the most parsimonious and unambiguous statistical method for characterizing the results of the conversation tests. These procedures involved the following steps:

Step 1) Compute an intercorrelation matrix among the dependent variables for the *Set* of conditions. Substantial inter-correlation among the dependent variables (i.e., correlation coefficients > .50 or < -.50) indicates that the number of dependent variables can be reduced - that there is a reduced set of underlying variables accounting for the variance in the dependent variables.

Step 2) Conduct a MANOVA on the *Set* of scores for the effects of conditions (*COND*) in the *Set*, (18 *COND* for *Set 1*, *6 COND* for *Set 2*, etc.) ignoring other factors. The MANOVA procedure determines the linear combination of the dependent variables that best separates the linear combination of the independent variable, i.e., *COND*. The initial linear combination of dependent variables is the *root* that accounts for maximum variance in the independent variables - it also represents the first underlying variable. A Chi-square test is conducted to determine the significance of the root. Subsequent roots are also extracted from the residual variance and tested with Chi-square for significance with each subsequent root being orthogonal to the preceding root. The number of significant roots indicates the number of significant underlying variables that account for the variance in the dependent variables.

Step 3) If there is only one significant root for the *COND* effect, the *Canonical coefficients* for that root are used to compute a weighted average of the dependent variables to estimate the underlying variable. This composite dependent variable is then used in a univariate ANOVA to test the factors involved in the experiment. Such ANOVA's will produce results that are more parsimonious and less complicated than presenting the results in the multi-dimensional space which would be necessary with multiple dependent variables.

7.3 Narrowband Test - Symmetric conditions (*Set 1*)

Table 18 shows the 1 to 18 test conditions involved in the NB symmetric condition conversation tests. Also shown in the table are the Mean scores for each rating scale by condition and by listening lab. Each score shown in the table is the average of ratings from 32 subjects.

The first step in the process described in the previous section is to examine the inter-correlations among the dependent variables for indications of underlying variables. Table 17 shows the inter-correlation matrix of the five dependent variables for the NB/SYM conditions. Absolute values of correlation above .50 have been bolded in the table. The table shows a high degree of inter-correlation among the dependent variables indicating the presence of a reduced set of underlying variables.

Table 17: Intercorrelations Among the Dependent Variables for the NB/SYM Conditions

NB/S	VQ	US	IA	PC	GQ
VQ	1				
US	**0.65**	1			
IA	0.40	**0.58**	1		
PC	**0.61**	**0.71**	**0.56**	1	
GQ	**0.81**	**0.66**	0.47	**0.69**	1

The second step in the analysis is designed to determine how many underlying variables account for the variance in the five dependent variables. MANOVA for the effects of *COND* was conducted on the NB/SYM data – conditions 1-18. Table 19 summarizes the results of the MANOVA analysis. The table contains two sections. The top section shows the analysis for the main effect of *COND.* It includes the results of univariate ANOVA's for each of the five dependent variables followed by results for the Multivariate-ANOVA (i.e., the MANOVA) for the combination of dependent variables. In Table 19 we can see that the *COND* main effect is highly significant for each of the five individual dependent variables in the univariate ANOVA's as well as for the combination of dependent variables

Table 18: Test Conditions and Mean Scores for each Condition and for each Lab for the Narrowband Experiment

Narrowband - Experimental Parameters							Voice Quality			Understanding			Interaction			Perception			Global Quality		
Cond	**Rm-A**	**Rm-B**	**RC**	**PL**	**Mode**	**Del**	**Arcon**	**FT**	**NTT**	**Arcon**	**FT**	**NTT**	**Arcon**	**FT**	**NTT**	**Arcon**	**FT**	**NTT**	**Arcon**	**FT**	**NTT**
1	Quiet	Quiet	10^{-2}	0	6.7	300	3.47	3.81	3.28	3.94	4.06	4.34	3.78	3.69	4.63	4.00	3.84	4.13	3.56	3.53	3.34
2	Quiet	Quiet	10^{-2}	0	12.2	500	3.50	3.81	3.06	4.16	4.16	4.09	3.59	3.66	4.09	4.06	4.00	3.81	3.66	3.63	3.13
3	Quiet	Quiet	10^{-2}	0	12.2	300	3.81	3.63	3.47	4.16	3.94	4.34	3.88	3.72	4.56	4.19	3.84	4.19	3.88	3.56	3.53
4	Quiet	Quiet	10^{-2}	3	6.7	300	3.25	3.22	2.75	3.66	3.31	3.78	3.66	3.13	4.25	3.66	2.94	3.59	3.28	2.81	2.72
5	Quiet	Quiet	10^{-2}	3	12.2	500	3.44	3.38	2.84	3.69	3.66	3.63	3.72	3.38	4.00	3.84	2.94	3.72	3.50	2.94	2.72
6	Quiet	Quiet	10^{-2}	3	12.2	300	3.41	3.63	3.16	3.88	3.78	4.03	3.88	3.56	4.41	3.88	3.44	4.00	3.41	3.22	3.13
7	Quiet	Quiet	10^{-3}	0	6.7	300	3.91	4.16	3.41	4.19	4.47	4.44	3.94	4.00	4.84	4.34	4.38	4.31	3.78	4.00	3.50
8	Quiet	Quiet	10^{-3}	0	12.2	500	3.72	4.22	3.59	4.22	4.41	4.50	3.72	4.03	4.72	4.09	4.44	4.53	3.97	4.06	3.72
9	Quiet	Quiet	10^{-3}	0	12.2	300	4.00	4.56	3.47	4.38	4.69	4.44	4.03	4.38	4.72	4.44	4.78	4.31	4.16	4.50	3.44
10	Quiet	Quiet	10^{-3}	3	6.7	300	3.28	3.66	3.16	3.72	3.94	4.16	3.78	3.88	4.44	3.91	3.72	4.00	3.31	3.41	3.16
11	Quiet	Quiet	10^{-3}	3	12.2	500	3.75	3.84	3.19	4.13	3.97	4.31	3.81	3.56	4.38	3.94	3.91	4.13	3.66	3.69	3.25
12	Quiet	Quiet	10^{-3}	3	12.2	300	3.50	3.91	3.41	4.00	4.22	4.44	3.97	4.09	4.66	3.88	4.13	4.25	3.53	3.97	3.53
13	Quiet	Quiet	5 x 10^{-4}	0	6.7	300	3.91	4.25	3.59	4.19	4.63	4.47	4.06	4.16	4.72	4.38	4.59	4.44	4.00	4.25	3.59
14	Quiet	Quiet	5 x 10^{-4}	0	12.2	500	3.97	4.34	3.50	4.22	4.47	4.56	3.75	3.97	4.44	4.31	4.53	4.44	3.94	3.97	3.44
15	Quiet	Quiet	5 x 10^{-4}	0	12.2	300	4.03	4.44	4.03	4.53	4.50	4.75	4.09	4.19	4.88	4.47	4.50	4.69	3.97	4.19	3.97
16	Quiet	Quiet	5 x 10^{-4}	3	6.7	300	3.63	3.84	3.19	3.91	3.97	4.25	4.03	3.72	4.63	3.91	3.75	4.06	3.50	3.56	3.34
17	Quiet	Quiet	5 x 10^{-4}	3	12.2	500	3.66	3.88	3.22	4.03	4.22	4.25	3.78	3.78	4.34	4.13	4.13	4.09	3.69	3.78	3.19
18	Quiet	Quiet	5 x 10^{-4}	3	12.2	300	3.56	3.75	3.25	4.03	3.88	4.22	3.69	3.63	4.59	4.09	3.78	4.19	3.72	3.44	3.19
19	Car	Quiet	5 x 10^{-4}	3	12.2	300	3.16	3.63	2.88	3.13	2.97	3.34	3.84	3.06	3.88	3.66	2.72	3.66	3.41	2.53	2.81
20	Quiet	Car	5 x 10^{-4}	3	12.2	300	3.81	3.88	3.50	4.13	3.91	4.44	3.94	3.63	4.44	4.31	3.78	4.25	3.78	3.28	3.53
21	Cafeteria	Quiet	5 x 10^{-4}	0	6.7	300	3.69	4.06	3.13	3.59	3.69	3.88	3.97	3.53	4.38	4.13	3.44	4.00	3.78	3.28	3.16
22	Quiet	Cafeteria	5 x 10^{-4}	0	6.7	300	3.97	4.31	3.53	4.41	4.50	4.50	4.06	4.06	4.66	4.34	4.50	4.38	3.69	4.09	3.56
23	Street	Quiet	5 x 10^{-4}	0	12.2	500	3.66	4.03	3.25	3.53	3.72	4.16	4.00	3.47	4.28	3.94	3.44	4.22	3.81	3.31	3.22
24	Quiet	Street	5 x 10^{-4}	0	12.2	500	3.84	4.19	3.53	4.22	4.38	4.28	4.00	3.91	4.47	4.44	4.22	4.19	3.91	3.91	3.53

Rm-A/Rm-B (Noise environment) RC (Radio Conditions) PL (% Packet Loss) Mode (Bit rate in kbps) Del (Delay in msec)

The bottom section of Table 19 shows the Chi-square tests of the MANOVA roots. It shows only a single significant root (1 through 5), indicating that a single underlying variable accounts for the significant variation in the dependent variables for these conditions. The canonical coefficients for this root are also shown in the table and are used to compute the composite dependent variable that represents the underlying variable for the NB/SYM conditions. The composite dependent variable (**NB/S-CTQ** for **N**arrow**B**and/**S**ymmetric-**C**onversation **T**est **Q**uality) is used to characterize the ratings in the NB/SYM conditions. NB/S-CTQ scores for all conditions and all LAB's in *Set 1* are listed in the Annex A. Equation 1 shows the formula used to compute the composite score for the NB/SYM conditions.

Table 19: Results of MANOVA for *COND* for NB/SYM Conditions

Univariate ANOVA's for Effect *COND* (df = 17, 1710)					
Dep.Var.	VQ	US	IA	PC	GQ
F-Rato	8.25	8.07	5.51	11.80	10.99
Prob.	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
MANOVA for Effect: *COND*					
Statistic	Value	F-Statistic	df	Prob	
Pillai Trace	0.16	3.38	85, 8550	**0.00**	
Test of Residual Roots				Dep.Var.	Canon.Coeff. for Root 1-5
Roots	Chi-Square	df	Prob		
1 through 5	**292.56**	**85**	**0.00**	VQ	0.0382
2 through 5	73.44	64	0.20	US	0.0555
3 through 5	34.14	45	0.88	IA	-0.0013
4 through 5	11.27	28	1.00	PC	0.5073
5 through 5	4.23	13	0.99	GQ	0.4004

Formula used to compute the Conversation Test Quality Score (NB/S-CTQ) for the conditions in Set 1:

$$\text{NB/S-CTQ} = .0426*VQ + .0620*US - .0015 * IA + .5664 * PC + .4470 * GQ \qquad (1)$$

The SYM conditions in the NB experiment are categorized by four experimental factors:

- Radio conditions – 10^{-2}, 10^{-3}, and $5x10^{-4}$
- Packet Loss – 0% and 3%
- AMR-NB mode or bit rate – 6.7 kbps and 12.2 kbps
- Delay – 300 msec and 500 msec

These conditions are assigned to two factorial experimental designs for analysing the effects of three of these factors. Table 20a shows the allocation of the 12 conditions used to evaluate the effects of Radio Conditions, Packet Loss, and Mode – with Delay held constant at 300 msec. Table 20b shows the allocation of the 12 conditions used to evaluate the effects of Radio Conditions, Packet Loss, and Delay – with Mode held constant at 12.2 kbit/s.

Table 20a: NB/SYM: Factorial Design for Effects of Radio Cond., Packet Loss, and Mode

No Noise - 300 msec delay				
6.7kbps / 0% PL			**6.7kbps / 3% PL**	
RC	Cond.#		RC	Cond.#
10^{-2}	1		10^{-2}	4
10^{-3}	7		10^{-3}	10
$5x10^{-4}$	13		$5x10^{-4}$	16
12.2kbps / 0% PL			**12.2kbps / 3% PL**	
RC	Cond.#		RC	Cond.#
10^{-2}	3		10^{-2}	6
10^{-3}	9		10^{-3}	12
$5x10^{-4}$	15		$5x10^{-4}$	18

Table 20b: NB/SYM: Factorial Design for the Effects of Radio Cond., Packet Loss, and Delay

No Noise - 12.2 kbps				
300 msec / 0% PL			**300 msec / 3% PL**	
RC	Cond.#		RC	Cond.#
10^{-2}	3		10^{-2}	6
10^{-3}	9		10^{-3}	12
$5x10^{-4}$	15		$5x10^{-4}$	18
500 msec / 0% PL			**500 msec / 3% PL**	
RC	Cond.#		RC	Cond.#
10^{-2}	2		10^{-2}	5
10^{-3}	8		10^{-3}	11
$5x10^{-4}$	14		$5x10^{-4}$	17

The composite dependent variable, NB/S-CTQ, was computed for the NB/SYM conditions using the equation shown in Eq.1. These composite scores were subjected to factorial ANOVA for the two experimental designs shown in Tables 20a and 20b. The results of those ANOVA's are shown in Tables 21 and 22, respectively.

Table 21: Results of ANOVA of NB/S-CTQ for the Effects of Lab, Radio Conditions (RC), Packet Loss (PL), and Mode

ANOVA for Composite Variable NB/S-CTQ					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	Prob
LAB	1.12	2	0.56	0.79	0.46
RC	39.49	2	19.74	27.61	**0.00**
PL	64.20	1	64.20	89.79	**0.00**
MODE	9.74	1	9.74	13.62	**0.00**
LAB*RC	10.37	4	2.59	3.62	**0.01**
LAB*PL	4.42	2	2.21	3.09	**0.05**
LAB*MODE	0.08	2	0.04	0.06	0.94
RC*PL	0.63	2	0.32	0.44	0.64
RC*MODE	1.76	2	0.88	1.23	0.29
PL*MODE	0.51	1	0.51	0.71	0.40
LAB*RC*PL	2.17	4	0.54	0.76	0.55
LAB*RC*MODE	2.69	4	0.67	0.94	0.44
LAB*PL*MODE	0.43	2	0.22	0.30	0.74
RC*PL*MODE	0.91	2	0.46	0.64	0.53
LAB*RC*PL*MODE	2.36	4	0.59	0.82	0.51
Error	797.99	1116	0.72		
Total	938.88	1151			

Table 21 shows that the main effects for *Radio Conditions*, *Packet Loss*, and *Mode* are significant ($p<.05$) for the NB/S-CTQ composite variable as are the interactions of *LAB x RC* and *LAB x PL*. Figure 7 shows the NB/S-CTQ scores with 95% confidence-interval bars for the factors tested in Table 21. The significant interactions of *RC x LAB* and *PL x LAB* indicate that the pattern of scores for the levels of RC and PL were significantly different across the three LAB's. Figure 9 illustrates the interaction of *LAB x RC*, Fig.10 the interaction of *LAB x PL*.

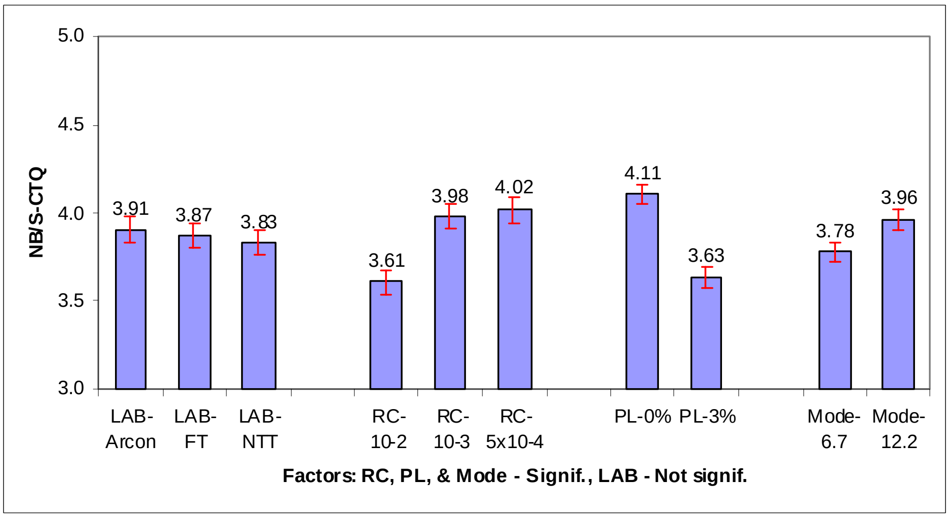

Figure 7: NB/S-CTQ Scores for the Effects of *LAB*, *Radio Conditions*, *Packet Loss*, and *Mode*

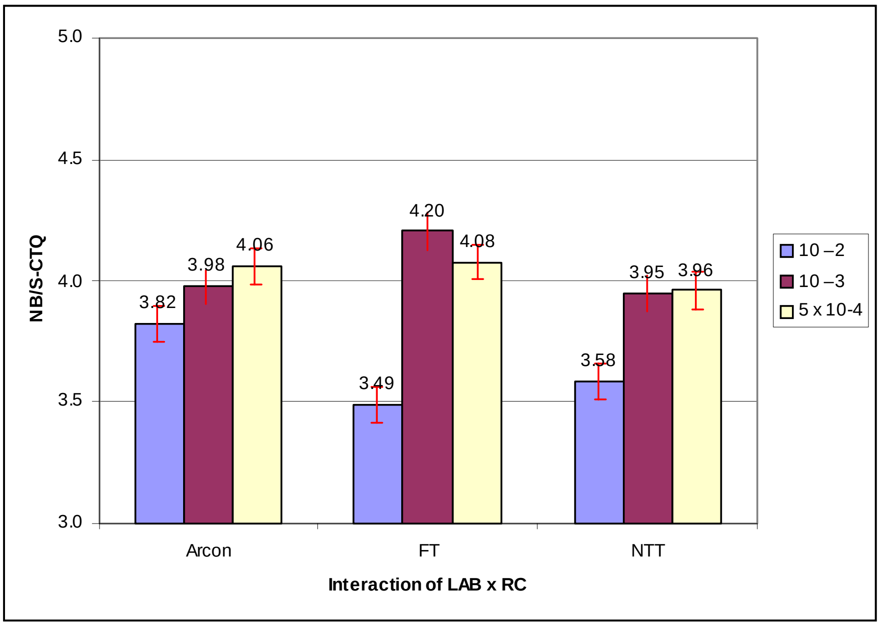

Figure 8: NB/S-CTQ Scores showing the Interaction of *LAB x Radio Conditions*

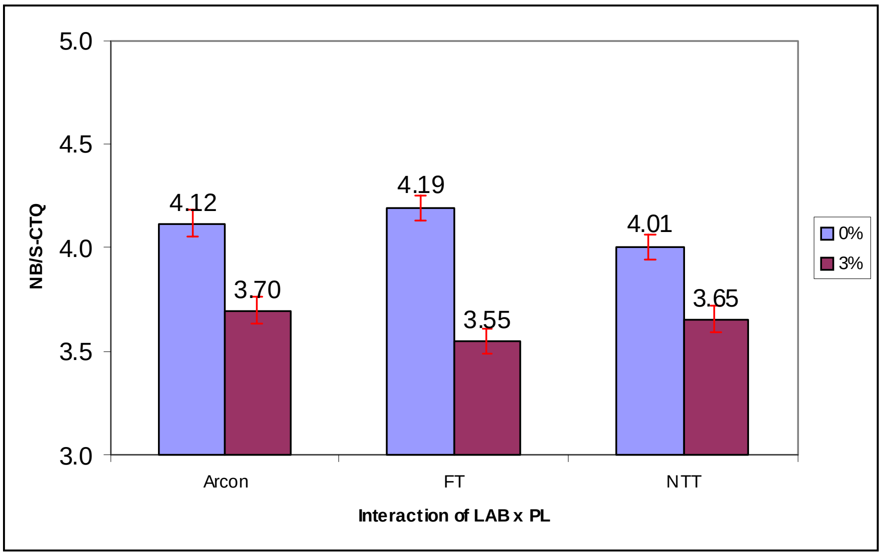

Figure 9: NB/S-CTQ Scores showing the Interaction of *LAB x Packet Loss*

Table 22: Results of ANOVA of NB/S-CTQ for the Effects of *LAB*, *Radio Conditions* (RC), *Packet Loss* (PL), and *Delay*

ANOVA for Composite Variable NB/S-CTQ					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	Prob
LAB	3.10	2	1.55	2.41	0.09
RC	42.54	2	21.27	33.10	**0.00**
PL	44.72	1	44.72	69.61	**0.00**
DELAY	4.06	1	4.06	6.32	**0.01**
LAB*RC	10.47	4	2.62	4.07	**0.00**
LAB*PL	3.52	2	1.76	2.74	0.07
LAB*DELAY	0.64	2	0.32	0.50	0.61
RC*PL	0.10	2	0.05	0.08	0.92
RC*DELAY	1.01	2	0.50	0.79	0.46
PL*DELAY	0.37	1	0.37	0.58	0.45
LAB*RC*PL	1.45	4	0.36	0.57	0.69
LAB*RC*DELAY	4.46	4	1.12	1.74	0.14
LAB*PL*DELAY	0.80	2	0.40	0.62	0.54
RC*PL*DELAY	1.81	2	0.90	1.41	0.25
LAB*RC*PL*DELAY	4.29	4	1.07	1.67	0.15
Error	717.03	1116	0.64		
Total	840.39	1151			

The results in Table 22 show that the main effects for *Radio Conditions*, *Packet Loss*, and *Delay* are significant while only one interaction, *LAB x RC*, is significant. Figure 10 shows the NB/S-CTQ scores with 95% confidence-interval bars for the factors tested in Table 22. Figure 11 illustrates the significant interaction of Lab x RC. The figure shows that the pattern of scores for RC is significantly different across LAB's.

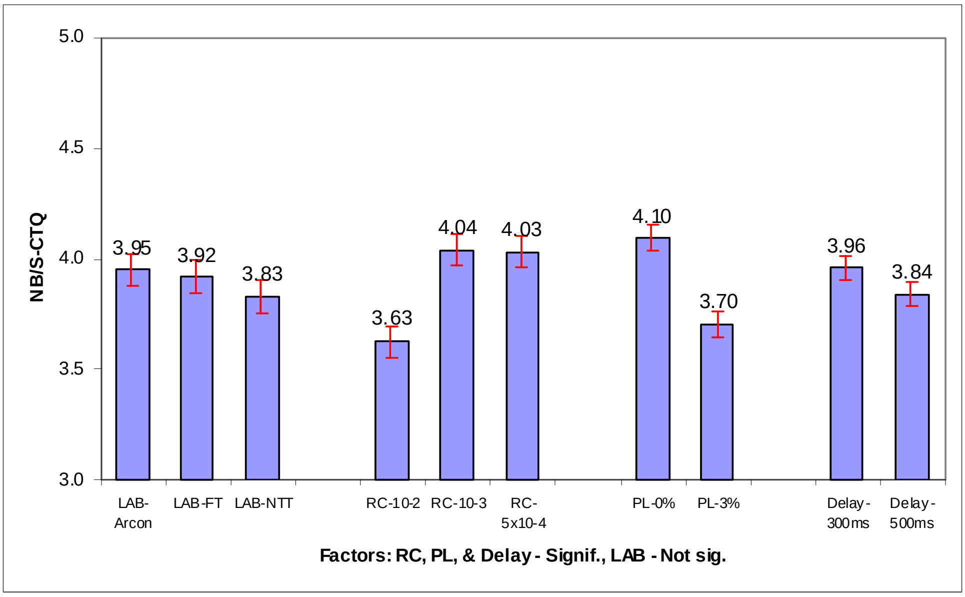

Figure 10: NB/S-CTQ Scores for the Effects of *LAB*, *Radio Conditions*, *Packet Loss*, and *Delay*

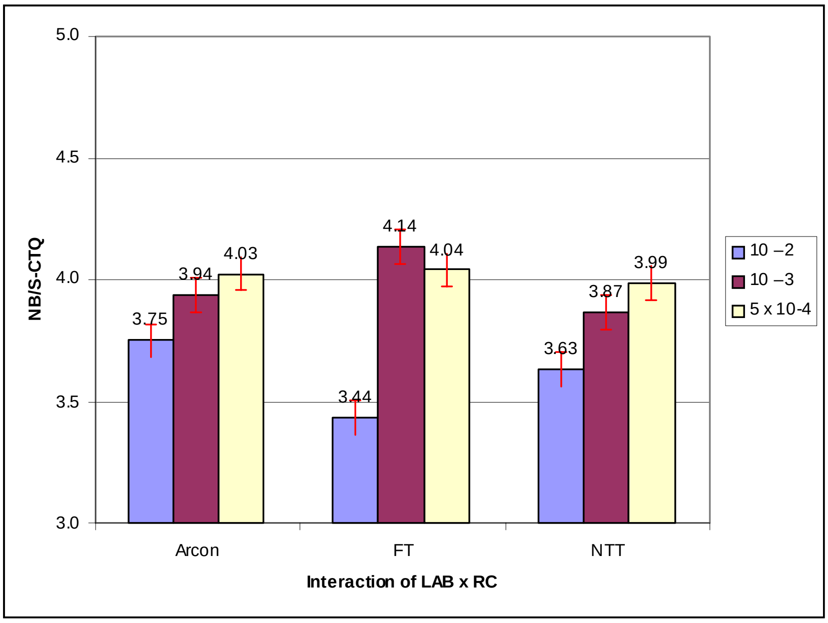

Figure 11: NB/S-CTQ Scores showing the Interaction of *LAB x Radio Conditions*

7.4 Narrowband Test – Asymmetric Conditions (*Set 2*)

Table 18 shows the 6 test conditions involved in the NB asymmetric condition conversation tests (conditions 19 to 24). Also shown in the table are the Mean scores for each rating scale by condition and by listening lab. Each score shown in the table is the average of ratings from 32 subjects.

Table 23 shows the inter-correlation matrix for the dependent variables in the NB/ASY conditions. The degree of inter-correlation among the dependent variables suggests that a reduced set of underlying variables accounts for their variation.

Table 23: Inter-correlations Among the Dependent Variables for the NB/ASY Conditions

WB/A	VQ	US	IA	PC	GQ
VQ	1				
US	**0.60**	1			
IA	0.35	**0.56**	1		
PC	0.44	**0.65**	**0.59**	1	
GQ	**0.65**	**0.64**	**0.56**	**0.68**	1

Table 24 shows the results of MANOVA for the effects of *COND* for the NB/ASY conditions. The analysis shows significant *COND* effects for all the univariate ANOVA's as well as for the MANOVA. The Chi-square tests of the MANOVA roots shows only a single significant root (1 through 5), indicating that a single underlying variable accounts for the significant variation in the dependent variables for these conditions. The canonical coefficients for this root are used to estimate the composite dependent variable that represents the underlying variable for the NB/ASY conditions. The composite dependent variable (**NB/A-CTQ** for **N**arrow**B**and/**A**symmetric-**C**onversation **T**est **Q**uality) is used to characterize the ratings in the NB/ASY conditions. NB/A-CTQ scores for all conditions and all LAB's in *Set 2* are listed in Annex A. Equation 2 shows the formula that was used to compute the values of the composite variable, NB/A-CTQ, for characterizing the NB/ASY conditions.

Table 24: Results of MANOVA for *COND* for NB/ASY Conditions

Univariate ANOVA's for Effect: *COND* (df = 5, 570)					
	VQ	US	IA	PC	GQ
F-Ratio	7.05	22.40	5.99	13.32	10.20
Prob	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
MANOVA for effect: *COND*					
Statistic	Value	F-Ratio	df	Prob	
Pillai Trace	0.18	4.38	25, 2850	**0.00**	
Test of Residual Roots				ependent	Canonical
Roots	Chi-Square	df	Prob	Variable	Coefficient
1 through 5	114.89	25	**0.00**	VQ	0.0894
2 through 5	7.23	16	0.97	US	0.3420
3 through 5	2.70	9	0.98	IA	0.1851
4 through 5	0.31	4	0.99	PC	0.2761
5 through 5	0.04	1	0.84	GQ	0.1074

Formula used to compute the Conversation Test Quality Score (NB/A-CTQ) for the NB/ASY conditions:

$$\text{NB/A-CTQ} = .0894*\text{VQ} + .3420*\text{US} + .1851 * \text{IA} + .2761 * \text{PC} + .1074 * \text{GQ} \quad (2)$$

The six NB/ASY conditions are distinguished by two factors. One factor has three levels with each level differing along a number of dimensions – Noise, Packet Loss, Mode, and Delay. These differences are listed in Table 18, but the factor will be referred to in the following analyses by the factor-name, *Noise*, noting that the conditions differ in more dimensions than noise alone. The second factor relates to the source of the noise. The noise is either in the room of the transmitting subject or in the room of the receiving subject. This factor will be referred to as *Room*. Table 25 shows the results of ANOVA for NB/A for the factors of *LAB*, *Noise*, and *Room*.

Table 25: Results of ANOVA of NB/A-CTQ for the Effects of *LAB*, *Noise*, and *Room*

ANOVA for Composite Variable - NB/A-CTQ					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	Prob
LAB	7.09	2	3.55	5.66	**0.00**
Noise	17.07	2	8.54	13.62	**0.00**
Room	43.76	1	43.76	69.80	**0.00**
LAB x Noise	3.28	4	0.82	1.31	0.27
LAB x Room	2.39	2	1.19	1.90	0.15
NOISE x Room	3.31	2	1.65	2.64	0.07
LAB x Noise x Room	1.19	4	0.30	0.48	0.75
Error	349.80	558	0.63		
Total	427.89	575			

The results of the ANOVA for NB/A-CTQ show that all three factors, *LAB*, *Noise*, and *Room*, are significant, but that none of the interactions are significant. Figure 12 shows the NB/A-CTQ scores with 95% confidence-interval bars for the three factors tested in Table 25.

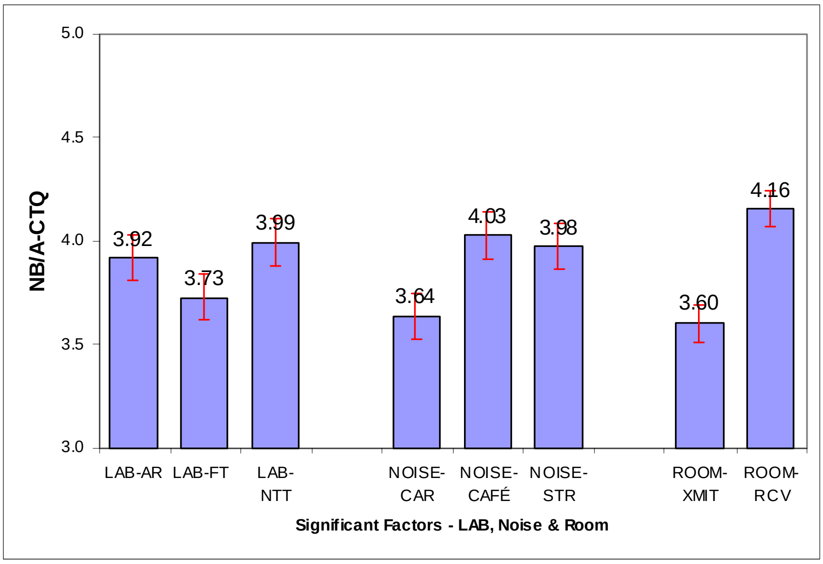

Figure 12: NB/A-CTQ Scores for the Effects of *LAB*, *Noise*, and *Room*

7.5 Wideband Test – Symmetric Conditions (*Set 3*)

Table 27 shows the 18 test conditions involved in the AMR-WB conversation tests (conditions 1 to 18). Also shown in the table are the Mean scores for each rating scale by condition and by listening lab. Each score shown in the table is the average of ratings from 32 subjects.

The initial step in the analysis is to examine the inter-correlation among the dependent variables for indications of underlying variables. Table 26 shows the inter-correlation matrix of the dependent variables for the WB/SYM conditions. Absolute values of correlation above .50 have been bolded in the table. The table shows a high degree of inter-correlation among the dependent variables indicating the presence of a reduced set of significant underlying variables.

Table 26: Intercorrelations Among the Dependent Variables for the WB/SYM Conditions

WB/S	VQ	US	IA	PC	GQ
VQ	1				
US	**0.66**	1			
IA	0.49	**0.51**	1		
PC	**0.59**	**0.59**	**0.51**	1	
GQ	**0.79**	**0.68**	**0.55**	**0.66**	1

The second step in the analysis is designed to determine how many underlying variables account for the variance in the five dependent variables. MANOVA for the effects of *COND* was conducted on the WB/SYM data – conditions 1-18. Table 28 summarizes the results of the analysis. The top section shows the analysis for the main effect of *COND.* This section includes the results of the univariate ANOVA's for each of the five dependent variables followed by the results of the MANOVA. In the table we can see that the *COND* main effect is highly significant for each of the five individual dependent variables in the univariate ANOVA's as well as for the combination of dependent variables in the MANOVA.

The bottom section of the table shows the Chi-square test of the MANOVA roots or underlying variables extracted from the five dependent variables. In Table 28, only the first root (1 through 5) is significant, indicating that a single underlying variable accounts for the significant variation in the dependent variables for these conditions. The canonical coefficients shown in the table are used to estimate the composite dependent variable that represents this root or underlying variable. The composite dependent variable (**WB/S-CTQ** for **W**ide**B**and/**S**ymmetric-**C**onversation **T**est **Q**uality) is computed and used in the third step – ANOVA's to test and characterize the factors of interest in the Wideband/SYM conditions. WB/S-CTQ scores for all conditions and all LAB's for *Set 3* are listed in Annex A. Equation 3 shows the formula that was used to compute the values of the composite variable, WB/S-CTQ, for characterizing the WB/SYM conditions.

Table 27: Test Conditions and Mean Scores for each LAB for the Wideband Experiment

Wideband - Experimental Parameters							Voice Quality			Understanding			Interaction			Perception			Global Quality		
Cond	Rm-A	Rm-B	RC	PL	Mode	RoHC	Arcon	FT	NTT	Arcon	FT	NTT	Arcon	FT	NTT	Arcon	FT	NTT	Arcon	FT	NTT
1	Quiet	Quiet	10^{-2}	0	12.65	RoHC	4.09	4.22	3.84	4.38	4.41	4.34	4.25	4.13	4.53	4.47	4.25	4.31	4.09	4.06	3.75
2	Quiet	Quiet	10^{-2}	0	12.65	-	4.00	4.44	3.97	4.22	4.84	4.53	4.06	4.38	4.72	4.28	4.41	4.31	3.78	4.31	4.00
3	Quiet	Quiet	10^{-2}	0	15.85	RoHC	4.13	4.28	4.13	4.38	4.50	4.69	4.31	4.19	4.66	4.50	4.28	4.59	4.28	4.09	4.22
4	Quiet	Quiet	10^{-2}	3	12.65	RoHC	3.88	3.72	3.72	4.19	4.09	4.03	3.91	4.09	4.28	4.34	3.84	4.06	3.88	3.53	3.59
5	Quiet	Quiet	10^{-2}	3	12.65	-	3.63	3.75	3.72	4.06	3.88	4.06	3.91	3.81	4.38	4.22	3.88	4.16	3.72	3.63	3.69
6	Quiet	Quiet	10^{-2}	3	15.85	RoHC	3.91	3.97	3.84	4.19	4.44	4.28	4.06	4.13	4.53	4.22	4.03	4.28	3.84	3.84	3.81
7	Quiet	Quiet	10^{-3}	0	12.65	RoHC	4.22	4.38	4.00	4.50	4.56	4.69	4.25	4.22	4.75	4.69	4.56	4.63	4.28	4.19	4.00
8	Quiet	Quiet	10^{-3}	0	12.65	-	4.06	4.47	4.06	4.28	4.69	4.72	4.22	4.25	4.69	4.31	4.47	4.69	4.16	4.25	4.22
9	Quiet	Quiet	10^{-3}	0	15.85	RoHC	3.88	4.63	3.94	4.34	4.75	4.53	4.16	4.38	4.75	4.44	4.50	4.53	3.94	4.38	4.06
10	Quiet	Quiet	10^{-3}	3	12.65	RoHC	3.97	4.31	3.97	4.19	4.50	4.41	4.13	4.13	4.66	4.47	4.19	4.53	4.03	3.94	3.97
11	Quiet	Quiet	10^{-3}	3	12.65	-	4.03	4.25	3.75	4.41	4.56	4.34	4.09	4.16	4.50	4.69	4.16	4.28	3.94	3.97	3.81
12	Quiet	Quiet	10^{-3}	3	15.85	RoHC	4.03	4.03	3.91	4.34	4.38	4.47	4.16	4.09	4.66	4.28	4.22	4.38	4.00	3.81	3.91
13	Quiet	Quiet	5×10^{-4}	0	12.65	RoHC	4.09	4.34	4.19	4.34	4.63	4.66	4.16	4.22	4.81	4.59	4.53	4.63	4.00	4.13	4.22
14	Quiet	Quiet	5×10^{-4}	0	12.65	-	4.09	4.59	4.06	4.47	4.81	4.59	4.16	4.44	4.75	4.50	4.56	4.56	4.16	4.38	4.09
15	Quiet	Quiet	5×10^{-4}	0	15.85	RoHC	4.19	4.47	4.03	4.47	4.69	4.66	4.44	4.31	4.78	4.59	4.47	4.59	4.38	4.16	4.06
16	Quiet	Quiet	5×10^{-4}	3	12.65	RoHC	3.94	3.97	3.91	4.25	4.53	4.41	4.00	3.97	4.63	4.25	4.16	4.38	3.84	3.88	4.00
17	Quiet	Quiet	5×10^{-4}	3	12.65	-	4.06	4.19	3.88	4.25	4.47	4.41	4.19	4.13	4.47	4.59	4.28	4.28	4.09	3.94	3.84
18	Quiet	Quiet	5×10^{-4}	3	15.85	RoHC	4.13	4.34	3.81	4.38	4.53	4.56	4.31	4.06	4.59	4.59	4.19	4.44	4.09	3.91	3.81
19	Car	Quiet	5×10^{-4}	3	12.65	RoHC	3.50	4.09	2.97	3.59	3.63	3.00	3.97	3.66	3.47	4.03	3.38	3.19	3.81	3.34	2.78
20	Quiet	Car	5×10^{-4}	3	12.65	RoHC	3.97	4.03	3.78	4.09	4.34	4.38	4.19	3.97	4.50	4.34	3.88	4.31	4.03	3.75	3.84
21	Cafeteria	Quiet	5×10^{-4}	0	12.65	-	3.75	4.38	3.66	3.78	4.38	3.88	3.94	4.09	4.06	4.31	3.97	3.84	3.81	3.81	3.34
22	Quiet	Cafeteria	5×10^{-4}	0	12.65	-	4.16	4.56	4.13	4.47	4.72	4.69	4.25	4.25	4.72	4.59	4.44	4.59	4.13	4.16	4.22
23	Street	Quiet	5×10^{-4}	0	15.85	RoHC	3.81	4.31	3.72	3.63	3.91	4.22	4.13	3.75	4.19	4.41	3.34	4.19	4.13	3.41	3.59
24	Quiet	Street	5×10^{-4}	0	15.85	RoHC	3.94	4.44	4.16	4.31	4.59	4.69	4.19	4.03	4.66	4.56	4.25	4.69	4.03	4.09	4.16

Rm-A/Rm-B (Noise environment) RC (Radio Conditions) PL (% Packet Loss) Mode (Bit rate in kbps) RoHC

Table 28: Results of MANOVA for *COND* for WB/SYM Conditions

Univariate ANOVA's for Effect *COND* (df = 17, 1710)					
Dep.Var.	VQ	US	IA	PC	GQ
F-Rato	3.35	4.36	2.84	3.98	4.14
Prob.	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
MANOVA for Effect: *COND*					
Statistic	Value	F-Statistic	df	Prob	
Pillai Trace	0.08	1.55	85, 8550	**0.00**	
Test of Residual Roots				Dep.Var.	Canon.Coeff. for Root 1-5
Roots	Chi-Square	df	Prob		
1 through 5	**132.56**	**85**	**0.00**	VQ	0.0685
2 through 5	43.32	64	0.98	US	0.3519
3 through 5	25.17	45	0.99	IA	0.1612
4 through 5	8.55	28	1.00	PC	0.2619
5 through 5	2.35	13	1.00	GQ	0.1565

The following formula is used to compute the Conversation Test Quality Score (WB/S-CTQ) for the WB/SYM conditions:

$$\text{WB/S-CTQ} = .0685*\text{VQ} + .3519*\text{US} + .1612*\text{IA} + .2619*\text{PC} + .1565*\text{GQ} \quad (3)$$

The SYM conditions in the WB experiment are categorized by four experimental factors:

- Radio conditions – 10^{-2}, 10^{-3}, and $5x10^{-4}$
- Packet Loss – 0% and 3%
- AMR-WB mode or bit rate – 12.65 kbps and 15.85 kbps
- ROHC

These conditions are assigned to two factorial experimental designs for analysing the effects through ANOVA of three of these factors. Table 29a shows the allocation of the 12 conditions used to evaluate the effects of Radio Conditions, Packet Loss, and Mode – with ROHC held constant. Table 29b shows the allocation of the 12 conditions used to evaluate the effects of Radio Conditions, Packet Loss, and ROHC – Mode held constant at 12.65kbps.

Table 29a: WB/SYM: Factorial Design for the Effects of Radio Cond., Packet Loss, and Mode

No Noise - RoHC

12.65kbps / 0% PL	
RC	Cond.#
10^{-2}	1
10^{-3}	7
$5x10^{-4}$	13

12.65 kbps / 3% PL	
RC	Cond.#
10^{-2}	4
10^{-3}	10
$5x10^{-4}$	16

15.85 kbps / 0% PL	
RC	Cond.#
10^{-2}	3
10^{-3}	9
$5x10^{-4}$	15

15.85 kbps / 3% PL	
RC	Cond.#
10^{-2}	6
10^{-3}	12
$5x10^{-4}$	18

Table 29b: WB/SYM: Factorial Design for the Effects of Radio Cond., Packet Loss, and Mode

No Noise - 12.65 kbps

RoHC / 0% PL	
RC	Cond.#
10^{-2}	1
10^{-3}	7
$5x10^{-4}$	13

RoHc / 3% PL	
RC	Cond.#
10^{-2}	4
10^{-3}	10
$5x10^{-4}$	16

No RoHC / 0% PL	
RC	Cond.#
10^{-2}	2
10^{-3}	8
$5x10^{-4}$	14

No RoHC / 3% PL	
RC	Cond.#
10^{-2}	5
10^{-3}	11
$5x10^{-4}$	17

The composite dependent variable, WB/S-CTQ, was computed for the WB/SYM conditions and subjected to factorial ANOVA for the two experimental designs shown in Tables 29a and 29b. The results of the ANOVA's are shown in Tables 30 and 31, respectively.

Table 30: Results of ANOVA of WB/S-CTQ for the Effects of *Lab*, *Radio Conditions* (RC), *Packet Loss* (PL), and *Mode*

ANOVA for Composite Variable WB/S-CTQ					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	Prob
LAB	6.53	2	3.26	6.52	**0.00**
RC	6.90	2	3.45	6.90	**0.00**
PL	14.33	1	14.33	28.65	**0.00**
MODE	1.41	1	1.41	2.81	0.09
LAB*RC	0.98	4	0.24	0.49	0.75
LAB*PL	0.23	2	0.12	0.23	0.79
LAB*MODE	0.04	2	0.02	0.04	0.96
RC*PL	0.35	2	0.18	0.35	0.70
RC*MODE	1.96	2	0.98	1.96	0.14
PL*MODE	0.09	1	0.09	0.17	0.68
LAB*RC*PL	0.45	4	0.11	0.23	0.92
LAB*RC*MODE	2.25	4	0.56	1.12	0.34
LAB*PL*MODE	0.11	2	0.05	0.11	0.90
RC*PL*MODE	0.01	2	0.01	0.01	0.99
LAB*RC*PL*MODE	1.00	4	0.25	0.50	0.74
Error	558.34	1116	0.50		
Total	594.97	1151			

Table 30 shows that the main effects for *LAB*, *Radio Conditions*, and *Packet Loss* are significant for the WB/S-CTQ composite variable. The factor *Mode* is not significant nor are any of the interactions. Figure 13 shows the WB/S-CTQ scores with 95% confidence-interval bars for the factors tested in Table 30.

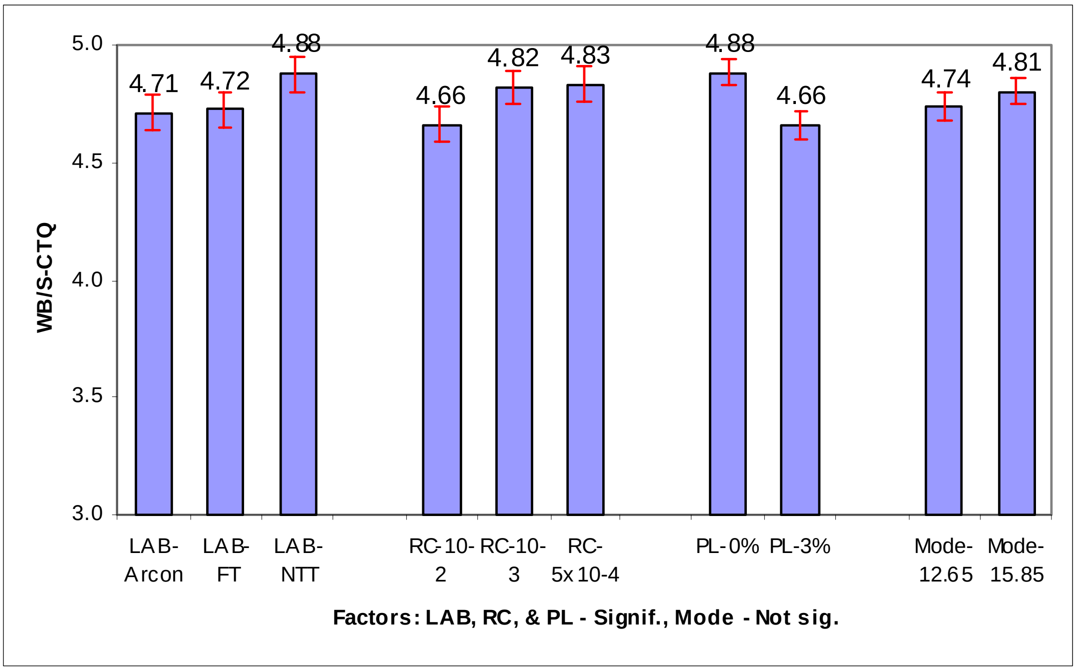

Figure 13: WB/S-CTQ Scores for the Effects of *LAB*, *Radio Conditions*, *Packet Loss*, and *Mode*

Table 31: Results of ANOVA of WB/S-CTQ for the Effects of *LAB*, *Radio Conditions* (RC), *Packet Loss* (PL), and ROHC

ANOVA for Composite Variable WB/S-CTQ					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	Prob
LAB	5.24	2	2.62	5.10	**0.01**
RC	13.59	2	6.80	13.23	**0.00**
PL	19.41	1	19.41	37.79	**0.00**
ROHC	0.07	1	0.07	0.14	0.71
LAB*RC	0.80	4	0.20	0.39	0.82
LAB*PL	2.46	2	1.23	2.39	0.09
LAB*ROHC	0.70	2	0.35	0.68	0.51
RC*PL	1.57	2	0.78	1.52	0.22
RC*ROHC	0.24	2	0.12	0.24	0.79
PL*ROHC	0.11	1	0.11	0.21	0.65
LAB*RC*PL	0.98	4	0.25	0.48	0.75
LAB*RC*ROHC	1.90	4	0.47	0.92	0.45
LAB*PL*ROHC	2.02	2	1.01	1.97	0.14
RC*PL*ROHC	0.50	2	0.25	0.48	0.62
LAB*RC*PL*ROHC	0.85	4	0.21	0.41	0.80
Error	573.40	1116	0.51		
Total	623.84	1151			

The results in Table 31 show that the main effects for *LAB*, *Radio Conditions*, and *Packet Loss* are significant. The factor *ROHC* is not significant nor are any of the interactions. Figure 14 shows the WB/S-CTQ scores with 95% confidence-interval bars for the factors tested in Table 31.

These listening tests were conducted using a fixed size RAB available at this time (size: 46 kbit/s). The test results show that when using ROHC the quality stays the same and the bitrate can be drastically reduced by suppressing the IP/UDP/RTP headers. As a result, a smaller RAB could be used.

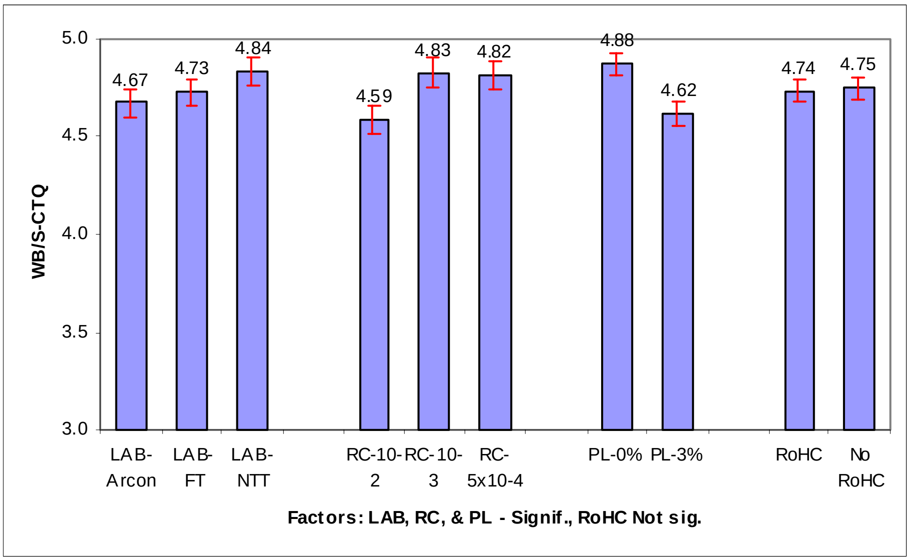

Figure 14: WB/S-CTQ Scores for the Effects of *LAB*, *Radio Conditions*, *Packet Loss*, and *ROHC*

7.6 Wideband Test – Asymmetric Conditions (*Set 4*)

Table 27 shows the 6 test conditions involved in the AMR-WB asymmetric condition conversation tests (condition 19 to 24). Also shown in the table are the Mean scores for each rating scale by condition and by listening lab. Each score shown in the table is the average of ratings from 32 subjects.

Table 32 shows the inter-correlation matrix for the dependent variables in the WB/ASY conditions. The high degree of inter-correlation shown in the table suggests that a reduced set of underlying variables accounts for the variation in the five dependent variables.

Table 32: Inter-correlations Among the Dependent Variables for the WB/ASY Conditions

WB/S	VQ	US	IA	PC	GQ
VQ	1				
US	**0.67**	1			
IA	**0.56**	**0.64**	1		
PC	**0.55**	**0.65**	**0.66**	1	
GQ	**0.72**	**0.73**	**0.69**	**0.73**	1

Table 33 shows the results of MANOVA for the effects of *COND* for the WB/ASY conditions. The analysis shows significant *COND* effects for all the univariate ANOVA's as well as for the MANOVA. The Chi-square tests of the MANOVA roots show only a single significant root (1 through 5), indicating that a single underlying variable accounts for the significant variation in the dependent variables for these conditions. The canonical coefficients for this root were used to compute the composite dependent variable that represents the underlying variable for the WB/Asymmetric conditions. The composite dependent variable (**WB/A-CTQ** for **W**ide**B**and/**A**symmetric-**C**onversation **T**est **Q**uality) is used to characterize the ratings in the WB/ASY conditions. WB/A-CTQ scores for all conditions and all LAB's for *Set 4* are listed Annex A. Equation 4 shows the formula that was used to compute the values of the composite variable, WB/A-CTQ, for characterizing the WB/ASY conditions.

Table 33: Results of MANOVA for *COND* for WB/ASY Conditions

Univariate ANOVA's for Effect: *COND* (df = 5, 570)					
	VQ	US	IA	PC	GQ
F-Ratio	8.38	21.63	8.16	14.10	10.97
Prob	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
MANOVA for effect: *COND*					
Statistic	Value	F-Ratio	df	Prob	
Pillai Trace	0.19	4.53	25, 2850	**0.00**	
Test of Residual Roots				ependent	Canonical
Roots	Chi-Square	df	Prob	Variable	Coefficient
1 through 5	118.45	25	**0.00**	VQ	-0.0970
2 through 5	11.19	16	0.80	US	0.8979
3 through 5	3.80	9	0.92	IA	-0.1103
4 through 5	1.85	4	0.76	PC	0.4136
5 through 5	0.00	1	0.99	GQ	-0.1042

The following formula used to compute the Conversation Test Quality Score (WB/ACTQ) for the WB/ASY conditions.

$$\text{WB/A-CTQ} = -.0970*\text{VQ} + .8979*\text{US} - .1103 * \text{IA} + .4136 * \text{PC} - .1042 * \text{GQ} \quad (4)$$

The six WB/ASY conditions are distinguished by two factors. One factor has three levels with each level differing along a number of dimensions – Noise, Packet Loss, Mode, and ROHC. These differences are listed in Table 27 but the factor will be referred to in the following analyses by the factor-name, *Noise*, noting that the conditions differ in more dimensions than noise alone. The second factor relates to the source of the noise and has two levels. The noise is either in the room of the transmitting subject or in the room of the receiving subject. This factor is referred to as *Room* in the following analyses. Table 34 shows the results of ANOVA for WB/A-CTQ for the factors of *LAB, Noise,* and *Room*.

Table 34: Results of ANOVA of WB/A-CTQ for the Effects of *LAB*, *Noise*, and *Room*

ANOVA for Composite Variable - WB/A-CTQ					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	Prob
LAB	6.06	2	3.03	3.80	**0.02**
NOISE	20.41	2	10.21	12.82	**0.00**
ROOM	63.10	1	63.10	79.24	**0.00**
LAB*NOISE	8.15	4	2.04	2.56	**0.04**
LAB*ROOM	3.16	2	1.58	1.98	0.14
NOISE*ROOM	2.19	2	1.09	1.37	0.25
LAB*NOISE*ROOM	6.20	4	1.55	1.95	0.10
Error	444.37	558	0.80		
Total	553.64	575			

The results of the ANOVA for WB/A-CTQ show that all three factors, *LAB*, *Noise*, and *Room*, are significant but only one of the interactions, *LAB x Noise* is significant. Figure 15 shows the WB/A-CTQ scores with 95% confidence-interval bars for the three factors tested in Table 34. Figure 16 shows how the pattern of scores for the Noise factor is different over the three LAB's resulting in the significant interaction of *Lab x Noise*.

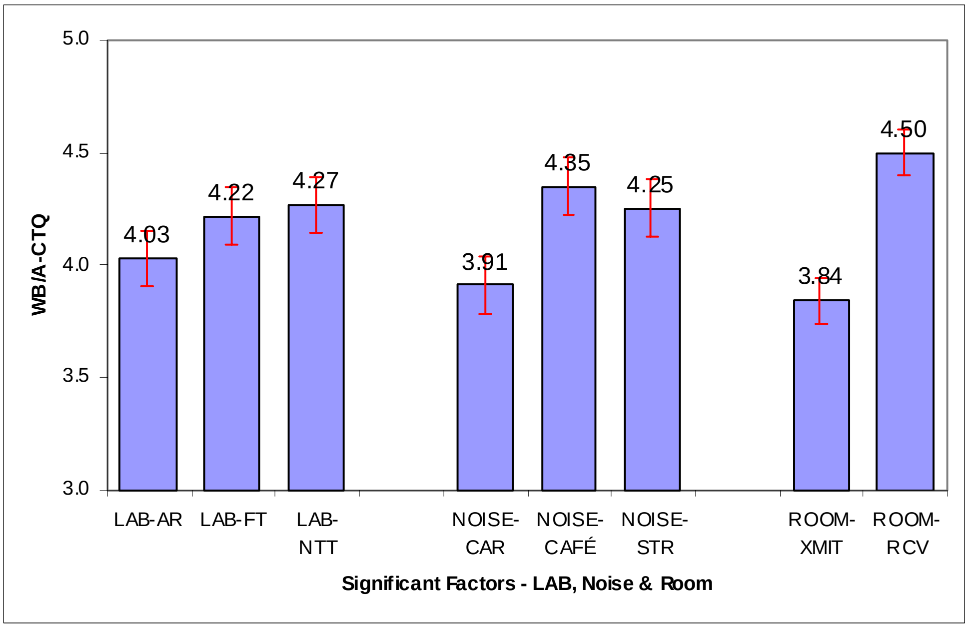

Figure 15: WB/A-CTQ Scores for the Effects of *LAB*, *Noise*, and *Room*

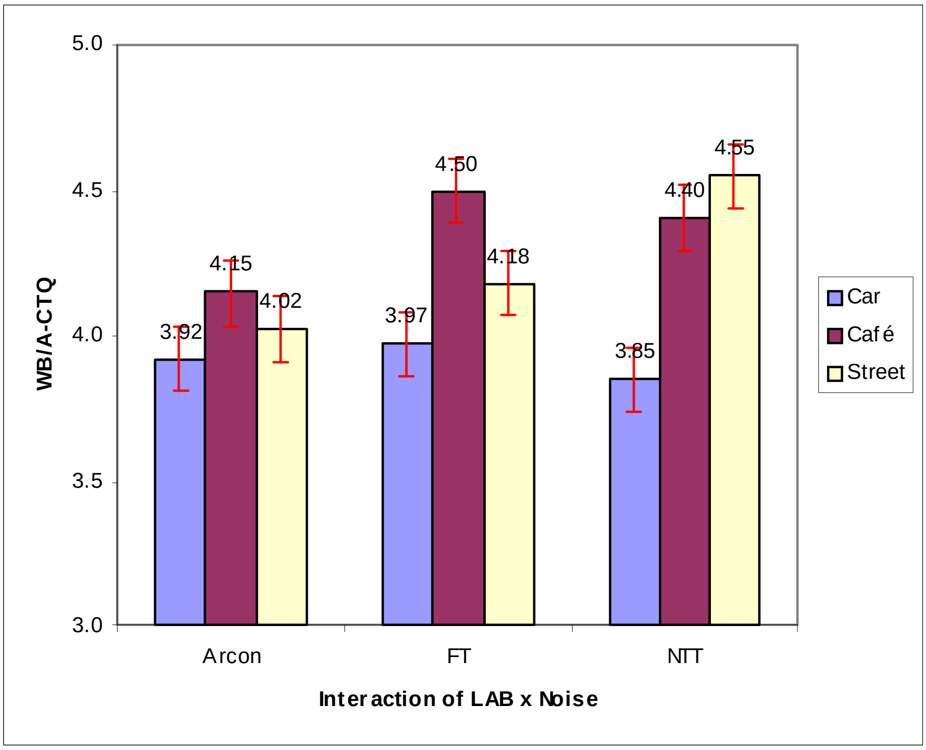

Figure 16: WB/A-CTQ Scores for the Interaction of *LAB x Noise*

7.7 Phase 2 - ITU-T Codec Tests (*Set 5*)

Table 35 shows the test conditions involved in the conversation tests designed to compare the performance of standardized ITU-T codecs in packet switched networks. The test involves eight codecs and two levels of packet loss, 0% and 3%. Scores are shown for each of the five dependent variables by Condition and by Language (Language is referred to by factor-name *LAB* in the following analyses). Each score shown in the table is the average of ratings from 32 listeners.

Table 35: Test Conditions and Scores for each Condition and Lab (Language) for the Codec (Phase 2) Experiment

Set 5 - Phase II Experimental Parameters			Ph2-CTQ Scores		
Cond	PL	Codec, Mode	French	Arabic	Average
1	0	AMR-NB, 6.7kbit/s	4.22	3.94	4.08
2	0	AMR-NB, 12.2kbit/s	4.31	4.05	4.18
3	0	AMR-WB, 12.65kbit/s	4.33	4.30	4.32
4	0	AMR-WB, 15.85kbit/s	4.46	4.31	4.38
5	0	G. 723., 6.4 kbit/s	4.15	3.98	4.07
6	0	G.729, 8kbit/s	4.11	4.18	4.14
7	0	G.722, 64 kbit/s + plc	4.34	4.13	4.24
8	0	G.711 + plc	4.32	4.28	4.30
9	3	AMR-NB, 6.7kbit/s	3.79	3.58	3.68
10	3	AMR-NB, 12.2 kbit/s	4.03	3.88	3.95
11	3	AMR-WB, 12.65kbit/s	4.28	4.04	4.16
12	3	AMR-WB, 15.85kbit/s	4.14	3.99	4.07
13	3	G. 723.1, 6.4 kbit/s	3.87	3.51	3.69
14	3	G.729, 8kbit/s	3.99	3.82	3.90
15	3	G.722, 64 kbit/s + plc	4.33	4.30	4.32
16	3	G.711 + plc	4.34	4.33	4.34

Table 36 shows the inter-correlation matrix for the dependent variables in the Phase 2 experiment. The moderate degree of inter-correlation shown in the table suggests that a reduced set of underlying variables may account for the variation in the five dependent variables.

The following acronyms were used in the tables PL for Packet Loss, FR for French and AB-Arabic.

Table 36: Inter-correlations Among the Dependent Variables for the Codec Conditions.

WB/S	VQ	US	IA	PC	GQ
VQ	1				
US	0.47	1			
IA	**0.50**	**0.54**	1		
PC	0.48	0.42	**0.51**	1	
GQ	**0.60**	**0.53**	**0.62**	**0.61**	1

Table 37 shows the results of MANOVA for the effects of *COND* for the Phase 2 experiment. The analysis shows significant *COND* effects for all the univariate ANOVA's as well as for the MANOVA. The Chi-square tests of the MANOVA roots show only a single significant root (1 through 5), indicating that a single underlying variable accounts for the significant variation in the dependent variables for these conditions. The canonical coefficients for this root were used to compute the composite dependent variable that represents the underlying variable for the Phase 2 conditions. The composite dependent variable (**Ph2-CTQ** for **Ph**ase**2-C**onversation **T**est **Q**uality) is computed and used to characterize the ratings in the Phase 2 experiment. Ph2-CTQ scores for all conditions and all LAB's for *Set 5* are listed in the Appendix. Equation 5 shows the formula that was used to compute the values of the composite variable, Ph2-CTQ, for characterizing the Phase 2 conditions.

Table 37: Results of MANOVA for *COND* for the Phase 2 Conditions

Univariate ANOVA's for Effect: *COND* (df = 15, 1008)					
	VQ	US	IA	PC	GQ
F-Ratio	5.64	2.43	2.68	2.54	4.25
Prob	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
MANOVA for effect: *COND*					
Statistic	Value	F-Ratio	df	Prob	
Pillai Trace	0.12	1.61	75, 5040	**0.00**	
Test of Residual Roots				Dependent	Canonical
Roots	Chi-Square	df	Prob	Variable	Coefficient
1 through 5	**122.26**	**75**	**0.00**	VQ	0.5995
2 through 5	32.44	56	1.00	US	0.0860
3 through 5	19.29	39	1.00	IA	-0.0092
4 through 5	10.45	24	0.99	PC	0.0459
5 through 5	2.58	11	1.00	GQ	0.2778

The following formula was used to compute the Conversation Test Quality Score (Ph2-CTQ) for the Phase 2 conditions:

$$\text{Ph2-CTQ} = .5995*VQ + .0860*US - .0092 * IA + .0459 * PC + .2778 * GQ$$

The 16 Phase 2 conditions are distinguished by two factors, *Codec* and *Packet Loss*. Table 38 shows the results of ANOVA for Ph2-CTQ for these factors.

Table 38: Results of ANOVA of Ph2-CTQ for the Effects of *Codec* and *Packet Loss*

ANOVA for Composite Variable - Ph2-CTQ					
Source	Sum-of-Squares	df	Mean-Square	F-ratio	Prob
LAB	5.71	1	5.71	11.93	**0.00**
CODEC	27.44	7	3.92	8.19	**0.00**
PL	10.33	1	10.33	21.59	**0.00**
LAB*CODEC	1.70	7	0.24	0.51	0.83
LAB*PL	0.07	1	0.07	0.14	0.71
CODEC*PL	7.09	7	1.01	2.12	**0.04**
LAB*CODEC*PL	1.45	7	0.21	0.43	0.88
Error	474.61	992	0.48		
Total	528.38	1023			

The results of the ANOVA for Ph2-CTQ show that all three factors, *LAB*, *Codec*, and *Packet Loss*, are significant as well as the interaction *Codec x Packet Loss*. Figure 17 shows the Ph2-CTQ scores with 95% confidence-interval bars for the factors tested in Table 38. Figure 18 illustrates the interaction of *Codec x Packet Loss*.

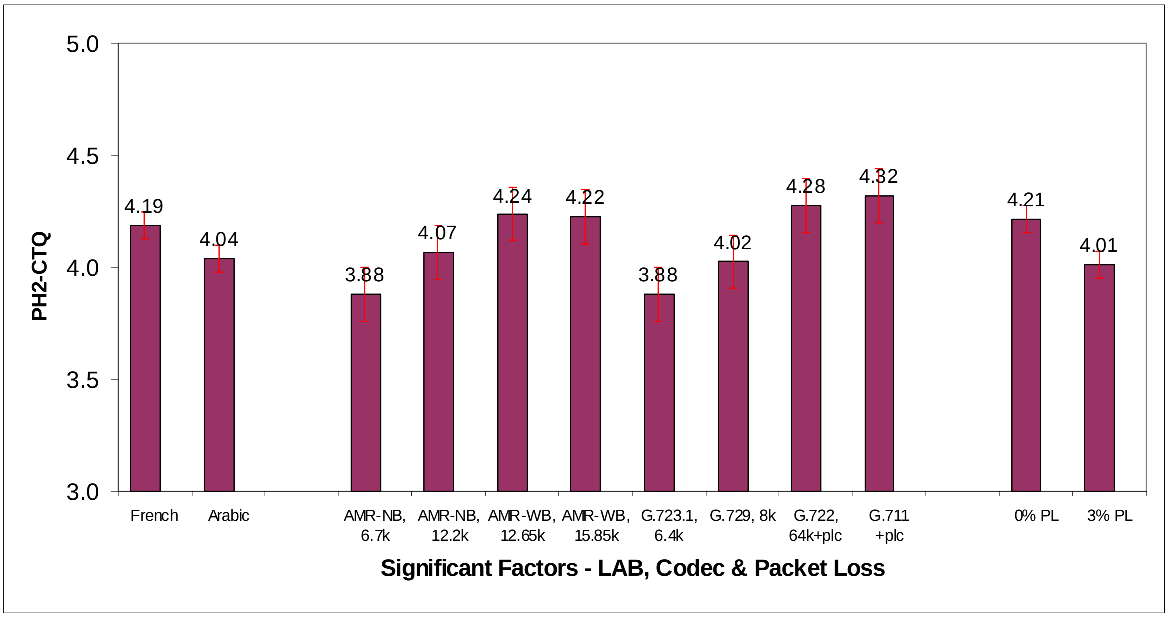

Figure 17: Ph2-CTQ Scores for the Effects of *LAB*, *Codec*, and *Packet Loss*

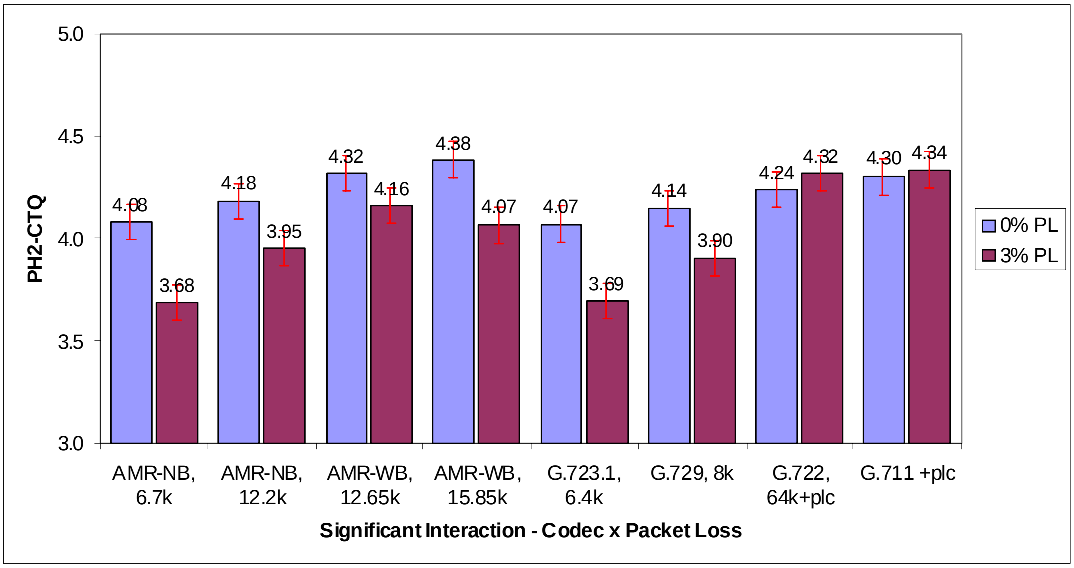

Figure 18: Ph2-CTQ Scores Showing the Interaction of Factors *Codec* and *Packet Loss*

7.8 Summary of Test Result Analysis

For each of the five sets of conditions in the Packet-Switched Conversation Tests, analysis by MANOVA revealed a single underlying variable that accounts for the significant variation in the five opinion rating scales, VQ, US, IA, PC, and GQ. Conversation Test Quality (CTQ) scores were computed for each set of conditions. The CTQ scores were analysed through ANOVA to characterize the conditions involved in the Conversation Tests.

8 Conclusions

The results from conversational tests confirm that the default speech codecs (AMR-NB and AMR-WB) operate well for packet switched conversational multimedia applications over various realistic operating conditions (i.e. packet loss, delay, background noise, radio conditions and ROHC).

The quality is somewhat reduced when packet losses occur and the end-to-end delay is increased, but the overall quality still remains acceptable even with 3% packet loss rate in the terrestrial IP network and up to a maximum of 1% BLER on each radio leg. The results also indicate that users have clear preference for AMR-WB speech over AMR-NB speech.

The performance results can be used as guidance for network planning regarding the QoS parameters for VoIP.

Annex A: Conversation test composite dependent variable scores by condition and Lab

Set 1 - Narrowband/SYM Experimental Parameters							NB/S-CTQ Scores			
Cond	Rm-A	Rm-B	RC	PL	Mode	Del	Arcon	FT	NTT	Average
1	Quiet	Quiet	10^{-2}	0	6.7	300	3.80	3.73	3.79	3.77
2	Quiet	Quiet	10^{-2}	0	12.2	500	3.88	3.85	3.52	3.75
3	Quiet	Quiet	10^{-2}	0	12.2	300	4.05	3.73	3.91	3.89
4	Quiet	Quiet	10^{-2}	3	6.7	300	3.49	2.92	3.22	3.21
5	Quiet	Quiet	10^{-2}	3	12.2	500	3.68	2.99	3.28	3.32
6	Quiet	Quiet	10^{-2}	3	12.2	300	3.67	3.38	3.62	3.55
7	Quiet	Quiet	10^{-3}	0	6.7	300	4.09	4.22	3.96	4.09
8	Quiet	Quiet	10^{-3}	0	12.2	500	4.04	4.28	4.17	4.16
9	Quiet	Quiet	10^{-3}	0	12.2	300	4.31	4.66	3.94	4.30
10	Quiet	Quiet	10^{-3}	3	6.7	300	3.63	3.60	3.64	3.63
11	Quiet	Quiet	10^{-3}	3	12.2	500	3.83	3.82	3.75	3.80
12	Quiet	Quiet	10^{-3}	3	12.2	300	3.73	4.06	3.94	3.91
13	Quiet	Quiet	5×10^{-4}	0	6.7	300	4.20	4.45	4.07	4.24
14	Quiet	Quiet	5×10^{-4}	0	12.2	500	4.14	4.30	4.01	4.15
15	Quiet	Quiet	5×10^{-4}	0	12.2	300	4.26	4.37	4.38	4.34
16	Quiet	Quiet	5×10^{-4}	3	6.7	300	3.73	3.69	3.75	3.72
17	Quiet	Quiet	5×10^{-4}	3	12.2	500	3.93	3.98	3.71	3.87
18	Quiet	Quiet	5×10^{-4}	3	12.2	300	3.92	3.65	3.75	3.77

Set 2 - Narrowband/ASY Experimental Parameters							NB/A-CTQ Scores			
Cond	Rm-A	Rm-B	RC	PL	Mode	Del	Arcon	FT	NTT	Average
19	Car	Quiet	5×10^{-4}	3	12.2	300	3.44	2.93	3.43	3.27
20	Quiet	Car	5×10^{-4}	3	12.2	300	4.08	3.75	4.20	4.01
21	Cafeteria	Quiet	5×10^{-4}	0	6.7	300	3.84	3.58	3.86	3.76
22	Quiet	Cafeteria	5×10^{-4}	0	6.7	300	4.21	4.36	4.31	4.29
23	Street	Quiet	5×10^{-4}	0	12.2	500	3.77	3.58	4.01	3.79
24	Quiet	Street	5×10^{-4}	0	12.2	500	4.17	4.18	4.14	4.16

Set 3 - Wideband/SYM - Experimental Parameters							WB/S-CTQ Scores			
Cond	Rm-A	Rm-B	RC	PL	Mode	RoHC	Arcon	FT	NTT	Average
1	Quiet	Quiet	10^{-2}	0	12.65	RoHC	4.76	4.68	4.73	4.72
2	Quiet	Quiet	10^{-2}	0	12.65	-	4.55	5.01	4.90	4.82
3	Quiet	Quiet	10^{-2}	0	15.85	RoHC	4.82	4.75	5.05	4.87
4	Quiet	Quiet	10^{-2}	3	12.65	RoHC	4.53	4.35	4.44	4.44
5	Quiet	Quiet	10^{-2}	3	12.65	-	4.42	4.21	4.52	4.38
6	Quiet	Quiet	10^{-2}	3	15.85	RoHC	4.53	4.60	4.70	4.61
7	Quiet	Quiet	10^{-3}	0	12.65	RoHC	4.90	4.87	5.05	4.94
8	Quiet	Quiet	10^{-3}	0	12.65	-	4.68	4.92	5.10	4.90
9	Quiet	Quiet	10^{-3}	0	15.85	RoHC	4.69	5.01	4.97	4.89
10	Quiet	Quiet	10^{-3}	3	12.65	RoHC	4.64	4.69	4.88	4.74
11	Quiet	Quiet	10^{-3}	3	12.65	-	4.77	4.72	4.72	4.74
12	Quiet	Quiet	10^{-3}	3	15.85	RoHC	4.66	4.61	4.86	4.71
13	Quiet	Quiet	5×10^{-4}	0	12.65	RoHC	4.74	4.88	5.09	4.91
14	Quiet	Quiet	5×10^{-4}	0	12.65	-	4.80	5.07	5.01	4.96
15	Quiet	Quiet	5×10^{-4}	0	15.85	RoHC	4.93	4.92	5.05	4.97
16	Quiet	Quiet	5×10^{-4}	3	12.65	RoHC	4.55	4.64	4.84	4.67
17	Quiet	Quiet	5×10^{-4}	3	12.65	-	4.73	4.70	4.75	4.72
18	Quiet	Quiet	5×10^{-4}	3	15.85	RoHC	4.81	4.68	4.88	4.79

Set 4 - Wideband/ASY - Experimental Parameters							WB/A-CTQ Scores			
Cond	Rm-A	Rm-B	RC	PL	Mode	RoHC	Arcon	FT	NTT	Average
19	Car	Quiet	5×10^{-4}	3	12.65	RoHC	3.69	3.62	3.17	3.49
20	Quiet	Car	5×10^{-4}	3	12.65	RoHC	4.14	4.32	4.53	4.33
21	Cafeteria	Quiet	5×10^{-4}	0	12.65	-	3.83	4.35	4.01	4.06
22	Quiet	Cafeteria	5×10^{-4}	0	12.65	-	4.47	4.65	4.80	4.64
23	Street	Quiet	5×10^{-4}	0	15.85	RoHC	3.71	3.87	4.32	3.97
24	Quiet	Street	5×10^{-4}	0	15.85	RoHC	4.34	4.49	4.78	4.54

Set 5 - Phase II Experimental Parameters			Ph2-CTQ Scores		
Cond	PL	Codec, Mode	French	Arabic	Average
1	0	AMR-NB, 6.7kbit/s	4.22	3.94	4.08
2	0	AMR-NB, 12.2kbit/s	4.31	4.05	4.18
3	0	AMR-WB, 12.65kbit/s	4.33	4.30	4.32
4	0	AMR-WB, 15.85kbit/s	4.46	4.31	4.38
5	0	G. 723., 6.4 kbit/s	4.15	3.98	4.07
6	0	G.729, 8kbit/s	4.11	4.18	4.14
7	0	G.722, 64 kbit/s + plc	4.34	4.13	4.24
8	0	G.711 + plc	4.32	4.28	4.30
9	3	AMR-NB, 6.7kbit/s	3.79	3.58	3.68
10	3	AMR-NB, 12.2 kbit/s	4.03	3.88	3.95
11	3	AMR-WB, 12.65kbit/s	4.28	4.04	4.16
12	3	AMR-WB, 15.85kbit/s	4.14	3.99	4.07
13	3	G. 723.1, 6.4 kbit/s	3.87	3.51	3.69
14	3	G.729, 8kbit/s	3.99	3.82	3.90
15	3	G.722, 64 kbit/s + plc	4.33	4.30	4.32
16	3	G.711 + plc	4.34	4.33	4.34

Annex B: Instructions to subjects

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

You are going to have a conversation with another user. The test situation is simulating communications between two mobile phones. The most of the situations will correspond to silent environment conditions, but some other will simulate more specific situations, as in a car, or in a railway station or in an office environment, when other people are discussing in the background.

After the completion of each call conversation, you will have to give your opinions on the quality, by answering to the following questions that will be displayed on the screen of the black box in front of you. Your judgment will be stored. You have 8 seconds to answer to each question. After "pressing" the button on the screen, another question will be displayed. You continue the procedure for the 5 following questions.

Question 1: How do you judge the quality of the voice of your partner?

Excellent	Good	Fair	Poor	Bad

Question 2: Do you have difficulties to understand some words?

All the time	Often	Sometimes	Rarely	Never

Question 3: How did you judge the conversation when you interacted with your partner?

Excellent interactivity (similar to face-to-face situation)	Good interactivity (in few moments, you were talking simultaneously, and you had to interrupt yourself)	Fair interactivity (sometimes, you were talking simultaneously, and you had to interrupt yourself)	Poor interactivity (often, you were talking simultaneously, and you had to interrupt yourself)	Bad interactivity (it was impossible to have an interactive conversation)

Question 4: Did you perceive any impairment (noises, cuts,…)? In that case, was it:

No impairment	Slight impairment, but not disturbing	Impairment slightly disturbing	Impairment disturbing	Very disturbing Impairment

Question 5: How do you judge the global quality of the communication?

Excellent	Good	Fair	Poor	Bad

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

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

Annex C: Example Scenarios for the conversation test

The pretexts used for conversation test are those developed by the Ruhr University (Bochum, Germany) within the context of ITU-T SG12. These scenarios have been elaborated to allow a well-balanced conversation within both participants and lasting approximately 2’30 or 3’, and to stimulate the discussion between persons that know each other to facilitate the naturalness of the conversation. They are derived from typical situations of every day life: railways inquiries, rent a car or an apartment, etc. Each condition should be given a different scenario.

Examples coming from ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

Scenario 1: Pizza service

Subject 1:

Your Name:	Clemence
Reason for the call	1 large Pizza
Condition which should be applied to the exchange of information	For 2 people, Vegetarian pizza preferred
Information you want to receive from your partner	Topping Price
Information that your partner requires	Delivery address : 41 industry street, Oxford Phone : 7 34 20
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	How long will it take?

Subject 2:

Your Name :	Pizzeria Roma			
Information from which you should select the details which your partner requires	Pizzas	1 person	2 persons	4 persons
	Toscana (ham, mushrooms, tomatoes, cheese)	3.2£	5.95£	10.5£
	Tonno (Tuna, onions, tomatoes, cheese)	3.95£	7.5£	13.95£
	Fabrizio (salami, ham, tomatoes, cheese)	4.2£	7.95£	14.95£
	Vegetarian (spinach, mushrooms, tomatoes, cheese)	4.5£	8.5£	15.95£
Information you want to receive from your partner	Name address telephone number			
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.				

Scenario 2 : Information on flights

Subject 1:

Your Name:	Parker
Reason for the call	Intended journey: London Heathrow → Düsseldorf
Condition which should be applied to the exchange of information	On June 23rd, Morning flight, Direct flight preferred
Information you want to receive from your partner	Departure: Arrival Flight number
Information that your partner requires	Reservation: 1 seat, Economy class Address: 66 middle street, Sheffield Phone: 21 08 33
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	From which airport is it easier to get into Cologne center : Düsseldorf or Cologne/Bonn

Subject 2:

Your Name :	Heathrow flight information			
Information from which you should select the details which your partner requires	Flight schedule	Lufthansa	British Airways	Lufthansa
	Flight number	LH 2615	BA 381	LH 413
	London Heathrow departure	6:30	6:35	8:20
	Brussels arrival Brussels departure		7:35 8:00	
	Düsseldorf arrival	7:35	9:05	9:25
Information you want to receive from your partner	Name address telephone number number of seats Class: Business or Economy			
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.				

Annex D: Test Plan for the AMR Narrow-Band Packet Switched Conversation Test

Source:	**Siemens1, France Telecom2**
Title:	**Test Plan for the AMR Narrow-Band Packet switched Conversation Test**
Document for:	**Approval**
Agenda Item:	**14.1**

1. Introduction

This document contains the test plan of one conversation test for the Adaptive Multi-Rate Narrow-Band (AMR-NB) in Packet Switched networks.

All the laboratories participating to this conversation test phase will use the same test plan, just the language of the conversation would change.

Even if the test rooms or the test equipments are not exactly the same in all the laboratories, the calibration procedures and the tests equipment characteristics and performance (as defined in this document) will guarantee the similarity of the test conditions.

1 contact: ImreVarga

Imre.Varga @siemens.com

Tel: +49 89 722 47537

Siemens AG, ICM MP

Grillparzerstrasse 10a, 81675 Munich, Germany

2 contacts:

Catherine Quinquis catherine.quinquis@francetelecom.com Tel: +33 2 96 05 14 93 France Telecom T&I/R&D 2 avenue Pierre Marzin, 22397 Lannion, France	Jean-Yves Monfort Jeanyves.monfort@francetelecom.com Tel : +33296053171 France Telecom T&I/R&D 2 avenue Pierre Marzin, 22397 Lannion, France

Section 2 gives references, conventions and contacts, section 3 details the test methodology, including test arrangement and test procedure, and section 4 defines the financial considerations.

Annex A contains the instructions for the subjects participating to the conversation tests.

Annex B contains the description of results to be provided to the Analysis Laboratory (if any) by the testing laboratories.

Annex C contains the list of statistical comparisons to be performed.

Considerations about IPV6 versus IPV4 are given in section 3.2.

RoHC is not implemented in AMR-NB conversation test. The effect of RoHC should be extrapolated from the results observed in AMR-WB conversation test.

2. References, Conventions, and Contacts

2.1Permanent Documents

ITU-T Rec.P.800	Methods for Subjective Determination of Transmission Quality	
ITU-T Rec. P.831	Subjective performance evaluation of network echo cancellers	This Recommendation defines conversation test procedures based on handset telephones, and gives inputs for the calibration.

2.2 Key Acronyms

AMR-NB	Adaptive Multi-Rate Narrowband Speech Codec
AMR-WB	Adaptive Multi-Rate Wide-band Speech Codec
MOS	Mean Opinion Score

2.3 Contact Names

The following persons should be contacted for questions related to the test plan.

Section	Contact Person/Email	Organisation	Address	Telephone/Fax
Experiments and results analysis	J-Y Monfort	France Telecom R&D	2, Avenue P. Marzin, 22307 Lannion Cédex France	Tel : +33296053171 Fax : +33296051316
AOB	Paolo Usai paolo.usai@etsi.fr	ETSI MCC	650 Route des Lucioles 06921 Sophia Antipolis Cedex France	Tel: 33 (0)4 92 94 42 36 Fax: 33 (0)4 93 65 28 17

2.4 Responsibilities

Each test laboratory has the responsibility to organize its conversation tests.

The list of Test laboratories participating to the conversation test phase.

Lab	Company	Language	Statistical analysis	Reporting
1	LAB1			
2	LAB2			

3. Test methodology

3.1 Introduction

The protocol described below evaluates the effect of degradation such as delay and dropped packets on the quality of the communications. It corresponds to the conversation-opinion tests recommended by the ITU-T P.800 [1]. First of all, conversation–opinion tests allow subjects passing the test to be in a more realistic situation, close to the actual service conditions experienced by telephone customers. In addition, conversation-opinion tests are suited to assess the effects of impairments that can cause difficulty while conversing (such as delay).

Subjects participate to the test by couple; they are seated in separate sound-proof rooms and are asked to hold a conversation through the transmission chain performed by means of UMTS simulators and communications are impaired by means of an IP impairments simulator part of the CN simulator and by the air interface simulator, as the figure below describes it.

The network configurations (including the terminal equipments) will be symmetrical (in the two transmission paths). The only dissymmetry will be due to presence of background noise in one of the test rooms.

3.2 Test arrangement

3.2.1 Description of the proposed testing system

This contribution describes a UMTS simulator for the characterization of the AMR speech codecs when the bitstream is transmitted over a PS network. The procedure to do the conversational listening test has been earlier described in [1].

Figure 1 describes the system that is going to be simulated:

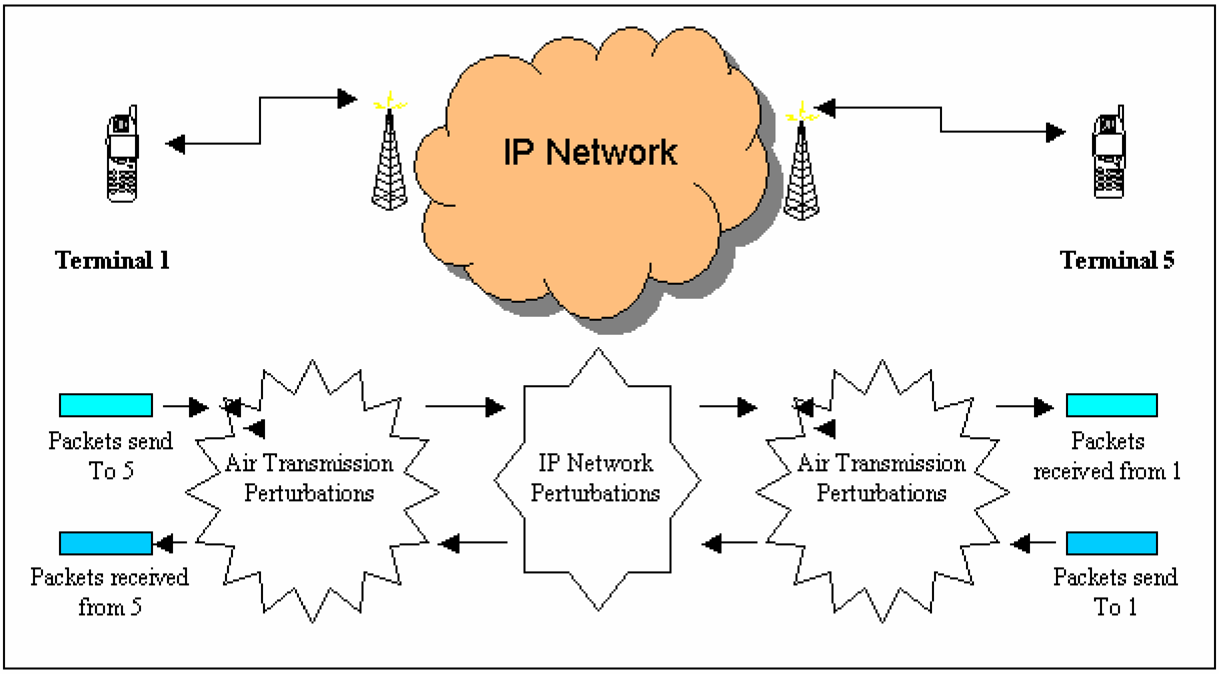

Figure 1: Packet switch audio communication simulator

This will be simulated using 5 PCs as shown in Figure 2.

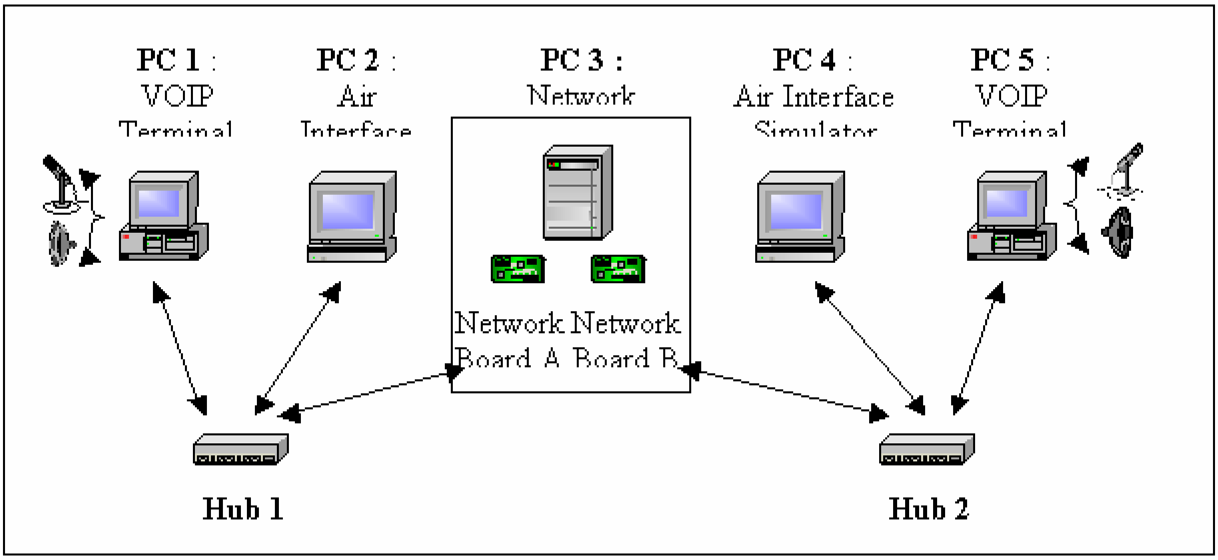

Figure 2: Simulation Platform

PC 1 and PC 5: PCs under Windows OS with VOIP Terminal Simulator Software of France Telecom R&D.

PC 2 and PC 4: PCs under Linux OS with Air Interface Simulator of Siemens AG.

PC 3: PCs under WinNT OS with Network Simulator Software (NetDisturb).

<u>Basic Principles:</u>

The platform simulates a packet switch interactive communication between two users using PC1 and PC5 as their relatives VOIP terminals. PC1 sends AMR encoded packets that are encapsulated using IP/UDP/RTP headers to PC5. PC1 receives these IP/UDP/RTP audio packets from PC5.

In fact, the packets created in PC1 are sent to PC2. PC2 simulates the air interface Up Link transmission and then forwards the transmitted packets to PC4.

In the same way, PC4 simulates the air interface Down Link transmission and then forwards the packets to PC5. PC5 decodes and plays the speech back to the listener.

3.2.2 France Telecom Network simulator

The core network simulator, as implemented, works under IPv4.

However, as the core network simulator acts only on packets (loss, delay,…) the use of Ipv4 or Ipv6 is equivalent for this test conversation context. Considering the networks perturbations introduced by the simulator and the context of the interactive communications, the simulation using IPv4 perturbation network simulator is adapted to manage and simulate the behaviours of an IPv6 core network.

Figure 3 shows the possible parameters that can be modified.

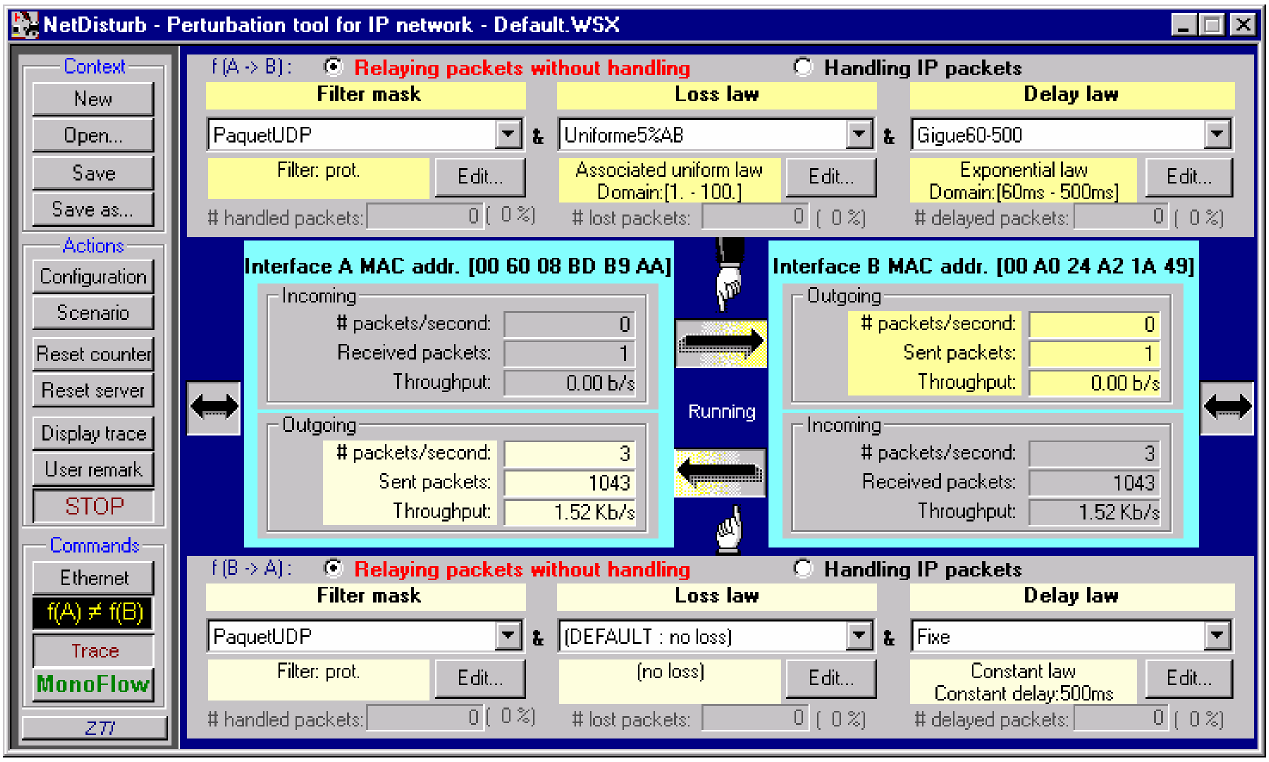

Figure 3: IP simulator interface

On both links, one can choose delay and loss laws. Both links can be treated separately or on the same way. For example, delay can be set to a fixed value but can also be set to another law such as exponential law.

3.2.3 UMTS simulator choices

The transmission of IP/UDP/RTP/AMR packets over the UMTS air interface is simulated using the RAB described in Section 3.2.3.1 The required functions of the RLC layer are implemented according to TS 25.322 and work in real-time. The underlying Physical Layer is simulated offline. Error patterns of block errors (i.e. discarded RLC PDUs) are inserted in the real-time simulation as described in Section 3.2.3.2 For more details on the parameter settings of the Physical Layer simulations see Section3.2.3.3

3.2.3.1 RAB and protocols

For our conversational tests, the AMR will encode speech at a maximum of 12.2 kbit/s. The bitstream will be encapsulated using IP/UDP/RTP protocols. The air interface simulator will receive IPv4 (or IPv6) packets from the CN simulator. The RTP packets will be extracted and before transmission over the air interface, IPv6 headers will be inserted. Finally real IPv6 packets are transmitted over the air interface simulator.

The payload Format should be the following:

- RTP Payload Format for AMR-NB (RFC 3267) will be used;
- Bandwidth efficient mode will be used;
- One speech frame shall be encapsulated in each RTP packet;
- Interleaving will not be used;

The payload header will then consist of the 4 bits of the CMR (Codec Mode Request). Then 6 bits is added for the ToC (Table of Content). For IPv4, this corresponds to a maximum of 72 bytes per frame that is to say 28.8 kbit/s, this goes up to 92 bytes (36.8 kbit/s) when using IPv6 protocol on the air interface.

RTCP packets will be sent. However, in the test conditions defined in the conversation test plans, RTCP is not mandatory, as it is not in a multicast environment (see IETF rfc 1889) we are not going to make use of the RTCP reports.

ROHC is an optional functionality in UMTS. In order to reduce the size of the tests and the number of condition ROHC algorithm will not be used for AMR-NB conversation test. This functionality will only be tested in the wideband condition. The Conversational / Speech / UL:42.8 DL:42.8 kbps / PS RAB RAB coming from TS 34.108 v4.7.0 will be used:

Here is the RAB description:

Higher layer	RAB/Signalling RB		**RAB**
PDCP	PDCP header size, bit		8
RLC	Logical channel type		DTCH
	RLC mode		UM
	Payload sizes, bit		920, 304, 96
	Max data rate, bps		46000
	UMD PDU header, bit		8
MAC	MAC header, bit		0
	MAC multiplexing		N/A
Layer 1	TrCH type		DCH
	TB sizes, bit		928, 312, 104
	TFS	TF0, bits	0x928
		TF1, bits	1x104
		TF2, bits	1x312
		TF3, bits	1x928
	TTI, ms		20
	Coding type		TC
	CRC, bit		16
	Max number of bits/TTI after channel coding		2844
	Uplink: Max number of bits/radio frame before rate matching		1422
	RM attribute		180-220

3.2.3.2 Description of the RLC implementation

The UMTS air interface simulator (PC 2 and 4) receives IP/UDP/RTP/AMR packets on a specified port of the network card (see Figure 4). The IP/UDP/RTP/AMR packets are given to the transmission buffer of the RLC layer, which works in UM. The RLC will segment or concatenate the IP bitstream in RLC PDUs, adding appropriate RLC headers (sequence number and length indicators). It is assumed that always Transport Format TF 3 is chosen on the physical layer, providing an RLC PDU length including header of 928 bits. In the regular case, one IP packet is placed into an RLC PDU that is filled up with padding bits. Due to delayed packets from the network simulator it may also occur that there are more than one IP packets in the RLC transmission buffer to transmit in the current TTI.

Each TTI of 20ms, an RLC PDU is formed. It is then given to the error insertion block that decides if the RLC PDU is transmitted successfully over the air interface or if it is discarded due to a block error after channel decoding. The physical layer will not be simulated in real time, but error pattern files will be provided. The error patterns of the air interface transmission will be simulated according to the settings given in Section 0. They consist of binary decisions for each transmitted RLC PDU, resulting in a certain BLER.

After the error pattern insertion, the RLC of the air interface receiver site receives RLC PDUs in the reception buffer. The sequence numbers of the RLC headers are checked to detect when RLC PDUs have been discarded due to block errors. A discarded RLC PDU will result in one or more lost IP packets, resulting in a certain packet loss rate of the IP packets and thereby in a certain FER of the AMR frames. The IP/UDP/RTP/AMR packets are reassembled and transmitted to the next PC. This PC is either the network simulator (PC 3) in case of uplink transmission, or it is one of the terminals (PC 1 or 5) in case of downlink transmission.

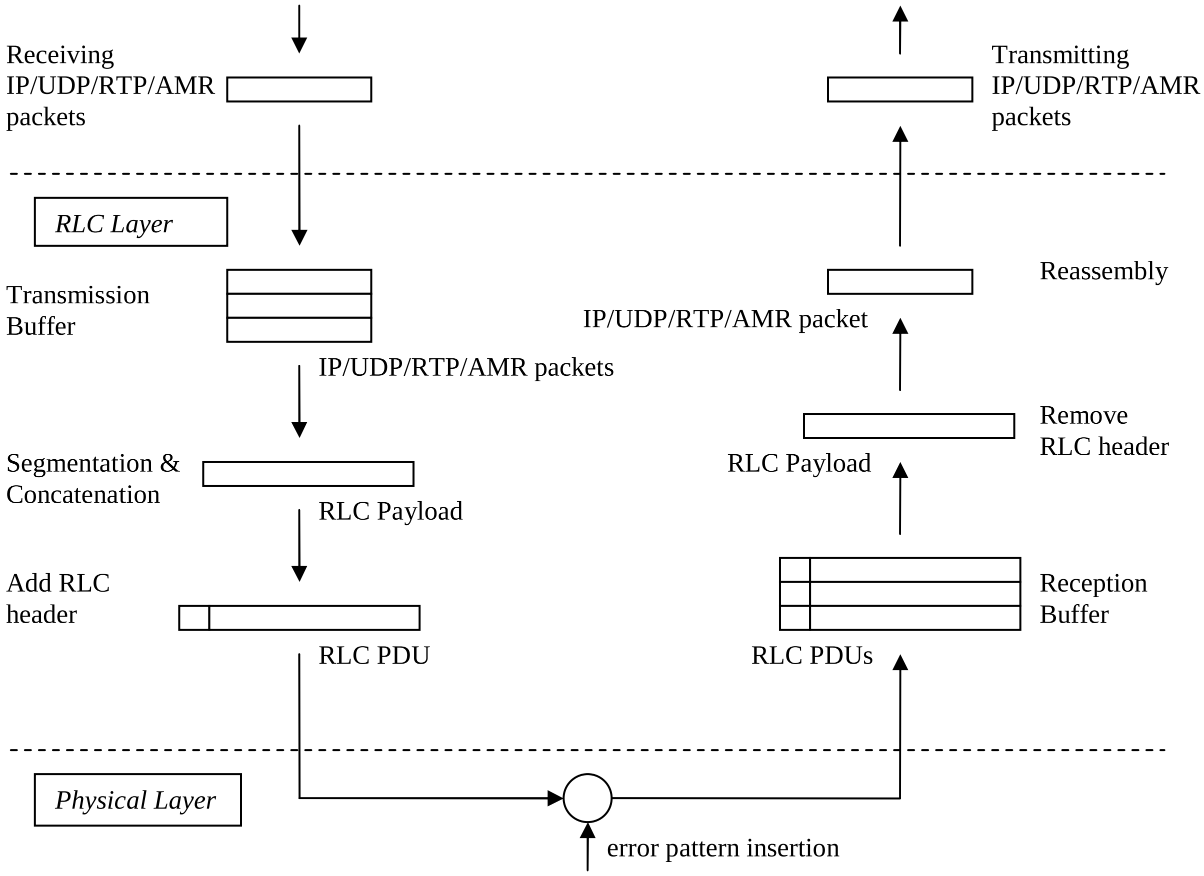

Figure 4: UMTS air interface simulation

3.2.3.3 Physical Layer Implementation

The parameters of the physical layer simulation were set according to the parameters for a DCH in multipath fading conditions given in TS 34.121 (downlink) and TS 25.141 (uplink). The TB size is 928 bits and the Turbo decoder uses the Log-MAP algorithm with 4 iterations. The rake receiver has 6 fingers at 60 possible positions.

The different channel conditions given in **Table 1**, **Table 2**, and **Table 3** were extracted from TR 101 112 (Selection procedures for the choice of radio transmission technologies of the UMTS) and also mentioned in the annex of the document S4-020680.

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0	FLAT
2	50	-3.0	FLAT
3	110	-10.0	FLAT
4	170	-18.0	FLAT
5	290	-26.0	FLAT
6	310	-32.0	FLAT

Table 1: Indoor Office Test Environment Tapped-Delay-Line Parameters

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0.0	CLASSIC
2	310	-1.0	CLASSIC
3	710	-9.0	CLASSIC
4	1090	-10.0	CLASSIC
5	1730	-15.0	CLASSIC
6	2510	-20.0	CLASSIC

Table 2: Vehicular Test Environment, High Antenna, Tapped-Delay-Line Parameters

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0	CLASSIC
2	110	-9.7	CLASSIC
3	190	-19.2	CLASSIC
4	410	-22.8	CLASSIC
5	-	-	CLASSIC
6	-	-	CLASSIC

Table 3: Outdoor to Indoor and Pedestrian Test Environment Tapped-Delay-Line Parameters

Table 4 (DL) and **Table 5** (UL) show approximate results of the air interface simulation for $\frac{DPCH\_E_c}{I_{or}}$ and E_b/N_0 corresponding to the considered BLERs.

	BLER			
Channel	**5*10^{-2}**	**1*10^{-2}**	**1*10^{-3}**	**5*10^{-4}**
Indoor, 3 km/h ($\hat{I}_{or}/I_{oc}$ = 9 dB)	-13.1 dB	-8.9 dB	-3.4 dB	-2.4 dB
Outdoor to Indoor, 3 km/h ($\hat{I}_{or}/I_{oc}$ = 9 dB)	-13.2 dB	-9.7 dB	-5.9 dB	-5.2 dB
Vehicular, 50 km/h ($\hat{I}_{or}/I_{oc}$ = -3 dB)	-9.35 dB	-8.2 dB	-6.9 dB	-6.55 dB
Vehicular, 120 km/h ($\hat{I}_{or}/I_{oc}$ = -3 dB)	-9.7 dB	-8.95 dB	-7.95 dB	-7.55 dB

Table 4: Downlink performance - approximately $\frac{DPCH\_E_c}{I_{or}}$ for the different channels and BLER

	BLER			
Channel	**5*10^{-2}**	**1*10^{-2}**	**1*10^{-3}**	**5*10^{-4}**
Indoor, 3 km/h	3.9 dB	6.4 dB	9.2 dB	9.8 dB
Outdoor to Indoor, 3 km/h	3.7 dB	6.1 dB	8.6 dB	9.2 dB
Vehicular, 50 km/h	-0.9 dB	-0.15 dB	0.55 dB	0.75 dB
Vehicular, 120 km/h	0.2 dB	0.6 dB	1.1 dB	1.3 dB

Table 5: Uplink performance - approximately E_b/N_0 for the different channels and BLER

3.2.4 Headsets and Sound Card

To avoid echo problems, it has been decided to use headsets, instead of handsets. The monaural headsets are connected to the sound cards of the PCs supporting the AMR simulators.

The sound level in the earphones can be adjusted, if needed, by the users. But, in practice, the original settings, defined during the preliminary tests, and producing a comfortable listening level, will not be modified. The microphones are protected by a foam ball in order to reduce the "pop" effect. It is also suggested to the user to avoid to place the acoustic opening of the microphone in front of the mouth.

3.2.5 Test environment

Each of the two subjects participating to the conversations is installed in a test room. They sit on an armchair, in front of a table. The test rooms are acoustically insulated. All the test equipments are installed in a third room, connected to the test rooms. When needed, the background noise is generated in the appropriate test room through a set of 4 loudspeakers. The background noise level is adjusted and controlled by a sound level meter. The measurement microphone, connected to the Sound level meter is located at the equivalent of the center of the subject's head. The noise level is A weighted.

3.2.6 Calibration and test conditions monitoring

Speech level

Before the beginning of a set of experiment, the end to end transmission level is checked subjectively, to ensure that there is no problem. If it is necessary to check the speech level following procedure will apply. An artificial mouth placed in front of the microphone of the Headset A, in the LRGP position -See ITU-T Rec. P.64-, generates in the artificial ear (according to ITU-T Rec. P57) coupled to the earphone of the Head set B the nominal level defined in section 4.3. If necessary, the level is adjusted with the receiving volume control of the headset. The similar calibration is done by inverting headsets A and B.

Delay

The overall delay (from the input of sound card A to the output of sound card B) will be evaluated for each test condition.

The hypothetical delay is calculated as shown :

On the air interface side, the simulator only receives packets on its network card, process them and transmits every 20 ms these packets to the following PC. Only processing delay and a possible delay due to a jitter can be added (a packet arrives just after the sending window of the air interface).

The hypothetical delay is calculated as shown :

On encoder side, delay have to take into account framing, look-ahead, processing and packetization: 45ms

Uplink delay between UE and Iu: 84.4 ms (see TR25.853)

Core network delay: a few ms

Routing through IP: depending on the number of routers.

Downlink delay between Iu and Ue: 71.8 ms (see TR25.853)

And delay on decoder side, taking into account jitter buffer, de-packetization and processing, 40 ms

The total delay to be considered is at least: 241.2 ms

3.3 Test Conditions

Based on circuit switched testing experiments, SA4 expects AMR 4.75 kb/s to provide insufficient quality for conversational applications. SA4 does not recommend testing AMR 4.75kb/s, this mode is considered as fall back solution in case of poor radio conditions.

Condition	**Additional Background noise** **Room A**	**Additional Background noise** **Room B**	**Experimental actors**		
			Radio conditions	IP conditions (Packet loss ratio)	Mode + delay
1	No	No	10^{-2}	0%	6,7kbit/s (delay 300 ms)
2	No	No	10^{-2}	0%	12.2 kbit/s (delay 500 ms)

Condition	**Additional Background noise** **Room A**	**Additional Background noise** **Room B**	**Experimental actors**		
			Radio conditions	IP conditions (Packet loss ratio)	Mode + delay
3	No	No	10^{-2}	0%	12,2 kbit/s (delay 300 ms)
4	No	No	10^{-2}	3%	6,7kbit/s (delay 300 ms)
5	No	No	10^{-2}	3%	12.2kbit/s(delay 500 ms)
6	No	No	10^{-2}	3%	12,2 kbit/s (delay 300 ms)
7	No	No	10^{-3}	0%	6,7kbit/s (delay 300 ms)
8	No	No	10^{-3}	0%	12.2kbit/s(delay 500 ms)
9	No	No	10^{-3}	0%	12,2 kbit/s (delay 300 ms)
10	No	No	10^{-3}	3%	6,7kbit/s (delay 300 ms)
11	No	No	10^{-3}	3%	12.2kbit/s(delay 500 ms)
12	No	No	10^{-3}	3%	12,2 kbit/s (delay 300 ms)
13	No	No	$5\ 10^{-4}$	0%	6,7kbit/s (delay 300 ms)
14	No	No	$5\ 10^{-4}$	0%	12.2kbit/s(delay 500 ms)
15	No	No	$5\ 10^{-4}$	0%	12,2 kbit/s (delay 300 ms)
16	No	No	$5\ 10^{-4}$	3%	6,7kbit/s (delay 300 ms)
17	No	No	$5\ 10^{-4}$	3%	12.ékbit/s(delay 500 ms)
18	No	No	$5\ 10^{-4}$	3%	12,2 kbit/s (delay 300 ms)
19	Car	No	$5\ 10^{-4}$	3%	12,2 kbit/s (delay 300 ms)
20	No	Car	$5\ 10^{-4}$	3%	12,2 kbit/s (delay 300 ms)
21	Cafeteria	No	$5\ 10^{-4}$	0%	6,7 kbit/s (delay 300 ms)
22	No	Cafeteria	$5\ 10^{-4}$	0%	6,7 kbit/s (delay 300 ms)
23	Street	No	$5\ 10^{-4}$	0%	12.2kbit/s(delay 500 ms)
24	No	Street	$5\ 10^{-4}$	0%	12.2kbit/s(delay 500 ms)

Noise types

Noise type	**Level (dB Pa A**
Car	60
Street	55
Babble	50

Listening Level	1	79 dBSPL
Listeners	32	Naïve Listeners
Groups	16	2 subjects/group
Rating Scales	5	
Languages	1	See table
Listening System	1	Monaural headset (flat response in the audio bandwidth of interest: 50Hz-7kHz). The other ear is open.
Listening Environment		Room Noise: Hoth Spectrum at 30dBA (as defined by ITU-T, Recommendation P.800, Annex A, section A.1.1.2.2.1 Room Noise, with table A.1 and Figure A.1), except when background noise is needed (see table)

Annex A Example Instructions for the conversation test

Table : Instructions to subjects.

INSTRUCTIONS TO SUBJECTS

In this experiment we are evaluating systems that might be used for telecommunication services.
You are going to have a conversation with another user. The test situation is simulating communications between two mobile phones. The most of the situations will correspond to silent environment conditions, but some other will simulate more specific situations, as in a car, or in a railway station or in an office environment, when other people are discussing in the background.

After the completion of each call conversation, you will have to give your opinions on the quality, by answering to the following questions that will be displayed on the screen of the black box in front of you. Your judgment will be stored. You have 8 seconds to answer to each question. After "pressing" the button on the screen, another question will be displayed. You continue the procedure for the 5 following questions.

Question 1: How do you judge the quality of the voice of your partner?

Excellent	Good	Fair	Poor	Bad

Question 2: Do you have difficulties to understand some words?

All the time	Often	Some time to time	Rarely	Never

Question 3: How did you judge the conversation when you interacted with your partner?

Excellent interactivity (similar to face-to-face situation)	Good interactivity (in few moments, you were talking simultaneously, and you had to interrupt yourself)	Fair interactivity (sometimes, you were talking simultaneously, and you had to interrupt yourself)	Poor interactivity (often, you were talking simultaneously, and you had to interrupt yourself)	Bad interactivity (it was impossible to have an interactive conversation)

Question 4: Did you perceive any impairment (noises, cuts,…)? In that case, was it:

No impairment	Slight impairment, but not disturbing	Impairment slightly disturbing	Impairment disturbing	Very disturbing Impairment

Question 5: How do you judge the global quality of the communication?

Excellent	Good	Fair	Poor	Bad

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

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

Annex B: Example Scenarios for the conversation test

The pretexts used for conversation test are those developed by the Rurh University (Bochum, Germany) within the context of ITU-T SG12 . These scenarios have been elaborated to allow a conversation well balanced within both participants and lasting approximately 2’30 or 3’, and to stimulate the discussion between persons that know each other to facilitate the naturalness of the conversation. They are derived from typical situations of every day life: railways inquiries, rent a car or an apartment, etc. Each condition should be given a different scenario.

Examples coming from ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

- Scenario 1 : Pizza service

Subject 1:

<table>
<tr><td>Your Name :</td><td>Clemence</td></tr>
<tr><td>Reason for the call</td><td>1 large Pizza</td></tr>
<tr><td>Condition which should be applied to the exchange of information</td><td>For 2 people,
Vegetarian pizza prefered</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Topping
Price</td></tr>
<tr><td>Information that your partner requires</td><td>Delivery address : 41 industry street,Oxford
Phone : 7 34 20</td></tr>
<tr><td>Question to which neither you nor your partner will have information.
You should discuss and find a solution that is acceptable to both of you.</td><td>How long will it take?</td></tr>
</table>

Subject 2:

<table>
<tr><td>Your Name :</td><td>Pizzeria Roma</td></tr>
<tr><td>Information from which you should select the details which your partner requires</td><td>
<table>
<tr><td>Pizzas</td><td>1 person</td><td>2 persons</td><td>4 persons</td></tr>
<tr><td>Toscana
(ham, mushrooms, tomatoes,cheese)</td><td>3.2£</td><td>5.95£</td><td>10.5£</td></tr>
<tr><td>Tonno
(Tuna, onions, tomatoes, cheese)</td><td>3.95£</td><td>7.5£</td><td>13.95£</td></tr>
<tr><td>Fabrizio
(salami, ham, tomatoes, heese)</td><td>4.2£</td><td>7.95£</td><td>14.95£</td></tr>
<tr><td>Vegetaria
(spinach, mushrooms, tomatoes,cheese)</td><td>4.5£</td><td>8.5£</td><td>15.95£</td></tr>
</table>
</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Name
address
telephone number</td></tr>
<tr><td>Question to which neither you nor your partner will have information.
You should discuss and find a solution that is acceptable to both of you.</td><td></td></tr>
</table>

- Scenario 2 : Information on flights

Subject 1:

Your Name :	Parker
Reason for the call	Intended journey: London Heathrow → Düsseldorf
Condition which should be applied to the exchange of information	On June 23th, Morning flight, Direct flight preferred
Information you want to receive from your partner	Departure : Arrival Flight number
Information that your partner requires	Reservation : 1 seat, Economy class Address: 66 middle street, Sheffield Phone: 21 08 33
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	From which airport is it easier to get into Cologne center : Düsseldorf or Cologne/Bonn

Subject 2:

<table>
<tr><td>Your Name :</td><td>Heathrow flight information</td></tr>
<tr><td>Information from which you should select the details which your partner requires</td><td>
<table>
<tr><td>Flight schedule</td><td>Lufthansa</td><td>British Airways</td><td>Lufthansa</td></tr>
<tr><td>Flight number</td><td>LH 2615</td><td>BA 381</td><td>LH 413</td></tr>
<tr><td>London Heathrow departure</td><td>6:30</td><td>6:35</td><td>8:20</td></tr>
<tr><td>Brussels arrival
Brussels departure</td><td></td><td>7:35
8:00</td><td></td></tr>
<tr><td>Düsseldorf arrival</td><td>7:35</td><td>9:05</td><td>9:25</td></tr>
</table>
</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Name
address
telephone number
number of seats
Class : Business or Economy</td></tr>
<tr><td>Question to which neither you nor your partner will have information.
You should discuss and find a solution that is acceptable to both of you.</td><td></td></tr>
</table>

ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

Annex C: Results to be provided

For contractual purposes, the information which needs to be provided is defined here.

The information required from each test Laboratory is a table containing the following information for each of the conditions in the experiment:

The "Mean Opinion Score (MOS)" obtained for all the subjects.
When the conditions are symmetrical, the mean value is calculated from all the result for the two test rooms..

For the dissymmetric conditions, the mean is calculated on the two test conditions, each result cumulating the results obtained in each condition of background noise.

The Standard Deviation of the "MOS" obtained for all the subjects, for each test condition.

The specific statistical comparisons are specified in Annex C.

Annex D: Data analysis and presentation of results

D.1 Calculation of MOS and Standard Deviation

The (overall) MOS/DMOS for confounded subjects for condition C (Yc) can then be obtained from:

$$Y_c = \frac{1}{T}\sum_{t=1}^{T} Y_{c,t}$$

The standard deviation (S) for condition C, denoted as Sc, can be calculated as:

$$S_c = \sqrt{\frac{1}{L \times T - 1}\sum_{t=1}^{T}\sum_{l=1}^{L}(X_{c,l,t} - Y_c)^2}$$

Finally, the confidence interval (CI) at the (1-α) level can be calculated for $N = L \times T$ as:

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

D.2 Presentation of Basic Statistical Results

The test results should be reported by the test Laboratory and the Global Analysis Laboratory as follows:

Calculate and tabulate "Mean Opinion Scores" for the (opinion scales, Standard Deviations and Confidence Intervals as shown in Table E.1.

Table C.1 - Layout for presentation of test results.

D.3 Thorough analysis

Two statistical analyses should be conducted on the data obtained with these subjective scales. The first analysis consists in a Multiple ANalysis OF VAriance (MANOVA), which globally indicates the possible effect of the experimental factors (*i.e.*, different conditions). Then, a specific ANOVA should be run on each dependent variable (the five scales) to test if there is an effect of a specific experimental factor for a given subjective variable. In other words, these statistical analyses indicate if the differences observed between the MOS obtained for the different conditions are significant, for one given dependant variable (ANOVA) or for the whole of dependant variables (MANOVA). Finally, Pearson's linear correlations should be computed between the results of all subjective variables, to see which are those preponderant or dependent on others.

Annex E: Test Plan for the AMR Wide-Band Packet Switched Conversation Test

Source:	**Siemens1, France Telecom2**
Title:	**Test Plan for the AMR Wide-Band Packet Switched Conversation Test**
Document for:	**Approval**
Agenda Item:	**14.1**

1. Introduction

This document contains the test plan of a conversation test for the Adaptive Multi-Rate Wide-Band (AMR-WB) in Packet Switched network.

All the laboratories participating to this conversation test phase will use the same test plan just the language of the conversation would change.

Even if the test rooms or the test equipments are not exactly the same in all the laboratories, the calibration procedures and the tests equipment characteristics and performance (as defined in this document) will guarantee the similarity of the test conditions.

Section 2 gives references, conventions and contacts, section 3 details the test methodology, including test arrangement and test procedure, and section 4 defines the financial considerations.

Annex A contains the instructions for the subjects participating to the conversation tests.

1 contact: ImreVarga

Imre.Varga @siemens.com

Tel: +49 89 722 47537

Siemens AG, ICM MP

Grillparzerstrasse 10a, 81675 Munich, Germany

2 contacts:

Catherine Quinquis	Jean-Yves Monfort
Catherine.quinquis@francetelecom.com	Jeanyves.monfort@francetelecom.com
Tel: +33 2 96 05 14 93	Tel : +33296053171
France Telecom T&I/R&D	France Telecom T&I/R&D
2 avenue Pierre Marzin, 22397 Lannion, France	2 avenue Pierre Marzin, 22397 Lannion, France

Annex B contains the description of results to be provided to the Analysis Laboratory (if any) by the testing laboratories.

Annex C contains the list of statistical comparisons to be performed.

Considerations about IPV6 versus IPV4 are given in section 3.2.

RoHC is implemented for AMR-WB conversation test, but only for the AMR-WB mode at 12,65 kbit/s

2. References, Conventions, and Contacts

2.1Permanent Documents

ITU-T Rec.P.800	Methods for Subjective Determination of Transmission Quality	
ITU-T Rec. P.831	Subjective performance evaluation of network echo cancellers	This Recommendation defines conversation test procedures based on handset telephones, and gives inputs for the calibration.

2.2 Key Acronyms

AMR-NB Adaptive Multi-Rate Narrowband Speech Codec

AMR WB Adaptive Multi-Rate Wide-band Speech Codec

MOS Mean Opinion Score

2.3 Contact Names

The following persons should be contacted for questions related to the test plan.

Section	Contact Person/Email	Organisation	Address	Telephone/Fax
Experiments and results analysis	J-Y Monfort	France Telecom R&D	2, Avenue P. Marzin, 22307 Lannion Cédex France	Tel : +33296053171 Fax : +33296051316
AOB	Paolo Usai paolo.usai@etsi.fr	ETSI MCC	650 Route des Lucioles 06921 Sophia Antipolis Cedex France	Tel: 33 (0)4 92 94 42 36 Fax: 33 (0)4 93 65 28 17

2.4 Responsibilities

Each test laboratory has the responsibility to organize its conversation tests.

The list of Test laboratories participating to the conversation test phase.

Lab	Company	Language	Statistical analysis	Reporting
1	Lab1			
2	Lab2			

3. Test methodology

3.1 Introduction

The protocol described below evaluates the effect of degradation such as delay and dropped packets on the quality of the communications. It corresponds to the conversation-opinion tests recommended by the ITU-T P.800 [1]. First of all, conversation–opinion tests allow subjects passing the test to be in a more realistic situation, close to the actual service conditions experienced by telephone customers. In addition, conversation-opinion tests are suited to assess the effects of impairments that can cause difficulty while conversing (such as delay).

Subjects participate to the test by couple; they are seated in separate sound-proof rooms and are asked to hold a conversation through the transmission chain of the UMTS simulator Communications are impaired by means of an IP impairments simulator simulator part of the CN simulator and by the air interface simulator, as the figure below describes it.

The network configurations (including the terminal equipments) will be symmetrical (in the two transmission paths). The only dissymmetry will be due to presence of background noise in one of the test rooms.

3.2 Test arrangement

3.2.1 Description of the proposed testing system

This contribution describes a UMTS simulator for the characterization of the AMR speech codecs when the bitstream is transmitted over a PS network. The procedure to do the conversational listening test has been earlier described in [1].

Figure 1 describes the system that is going to be simulated:

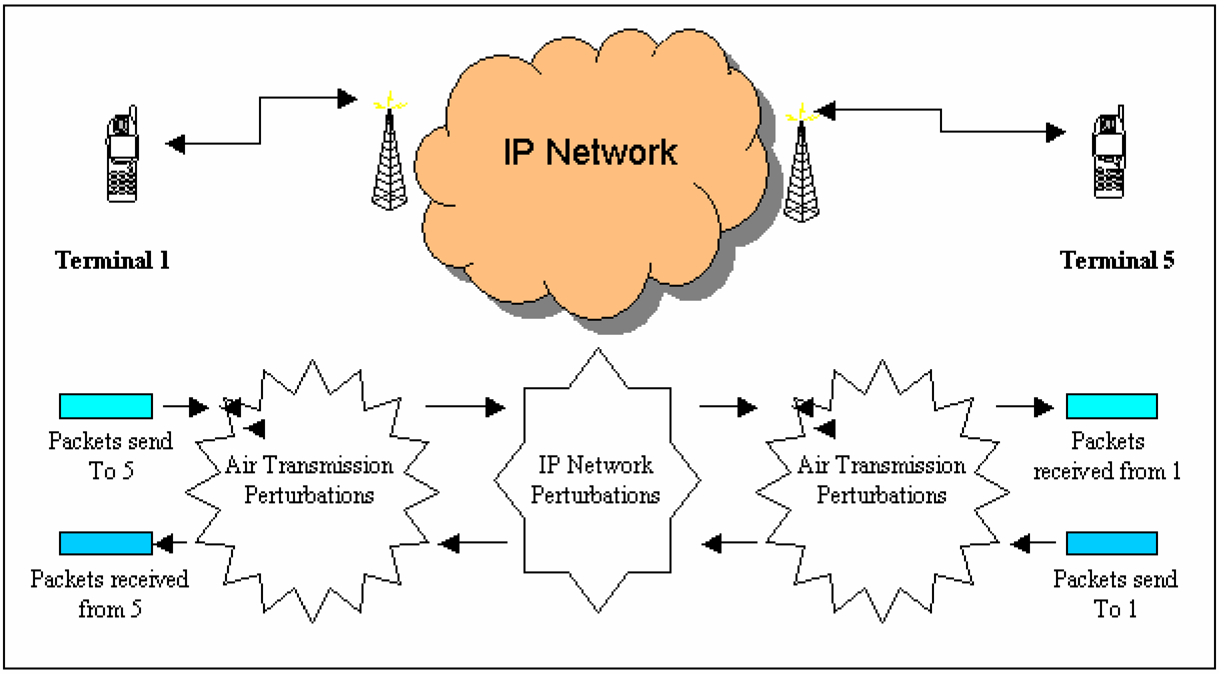

Figure 1: Packet switch audio communication simulator

This will be simulated using 5 PCs as shown in Figure 2.

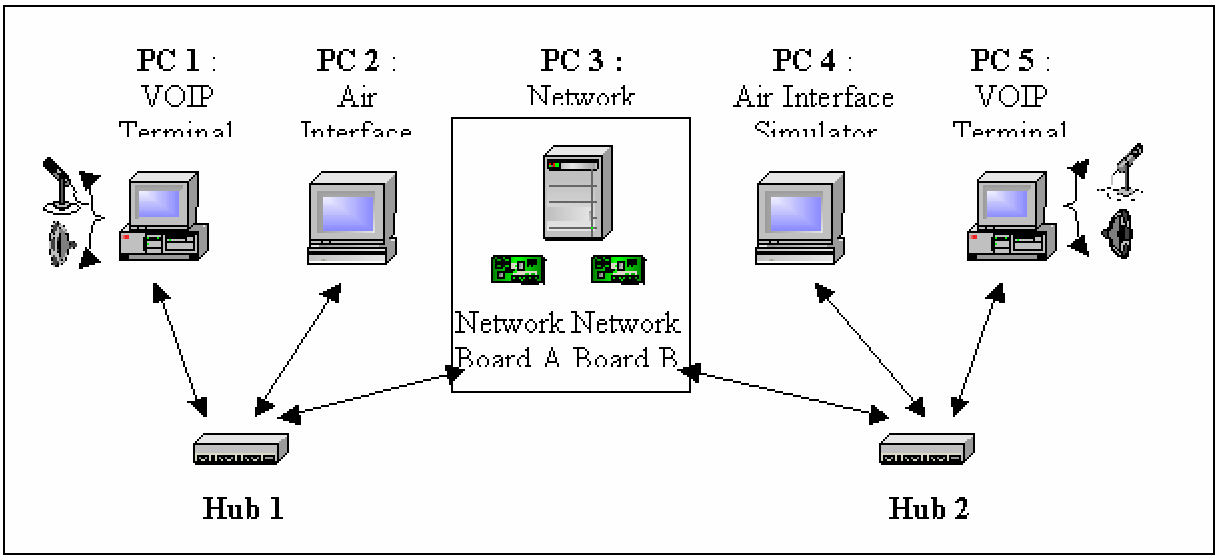

Figure 2: Simulation Platform

PC 1 and PC 5: PCs under Windows OS with VOIP Terminal Simulator Software of France Telecom R&D.

PC 2 and PC 4: PCs under Linux OS with Air Interface Simulator of Siemens AG.

PC 3: PCs under WinNT OS with Network Simulator Software (NetDisturb).

Basic Principles :

The platform simulates a packet switch interactive communication between two users using PC1 and PC5 as their relatives VOIP terminals. PC1 sends AMR encoded packets that are encapsulated using IP/UDP/RTP headers to PC5. PC1 receives these IP/UDP/RTP audio packets from PC5.

In fact, the packets created in PC1 are sent to PC2. PC2 simulates the air interface Up Link transmission and then forwards the transmitted packets to PC4.

In the same way, PC4 simulates the air interface Down Link transmission and then forwards the packets to PC5. PC5 decodes and plays the speech back to the listener.

3.2.2 France Telecom Network simulator

The core network simulator, as implemented, works under IPv4.

However, as the core network simulator acts only on packets (loss, delay,…) the use of IPv4 or IPv6 is equivalent for this test conversation context. Considering the networks perturbations introduced by the simulator and the context of the interactive communications, the simulation using IPv4 perturbation network simulator is adapted to manage and simulate the behaviours of an IPv6 core network.

. Figure 3 shows the possible parameters that can be modified.

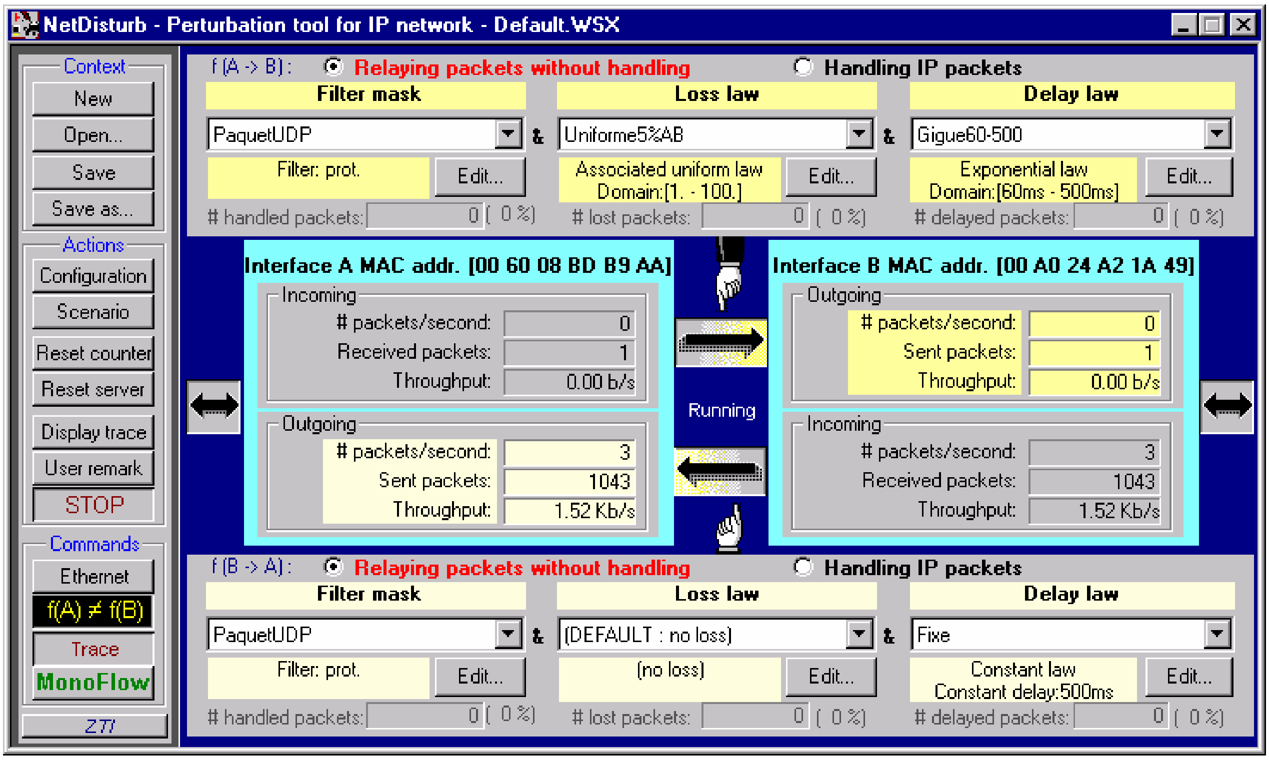

Figure 3: IP simulator interface

On both links, one can choose delay and loss laws. Both links can be treated separately or on the same way. For example, delay can be set to a fixed value but can also be set to another law such as exponential law.

3.2.3 UMTS simulator choices

The transmission of IP/UDP/RTP/AMR packets over the UMTS air interface is simulated using the RAB described in Section 3.2.3.1. The required functions of the RLC layer are implemented according to TS 25.322 and work in real-time. The underlying Physical Layer is simulated offline. Error patterns of block errors (i.e. discarded RLC PDUs) are inserted in the real-time simulation as described in Section 3.2.3.2. For more details on the parameter settings of the Physical Layer simulations see Section 3.2.3.3.

3.2.3.1 RAB and protocols

For our conversational tests, the AMR-WB will encode speech at a maximum of 15.85 kbit/s. The bitstream will be encapsulated using IP/UDP/RTP protocols. The air interface simulator will receive IPv4 packets from the IP network simulator. The RTP packets will be extracted and before transmission over the air interface, IPv6 headers will be inserted. Then a new IP/UDP/RTP packet will be transmitted through the air interface simulator.

The payload Format should be the following:

- RTP Payload Format for AMR-WB (RFC 3267) will be used;
- Bandwidth efficient mode will be used;
- One speech frame shall be encapsulated in each RTP packet;
- Interleaving will not be used;

The payload header will then consist of the 4 bits of the CMR (Codec Mode Request). Then 6 bits are added for the ToC (Table of Content). For IPv4 a maximum of 81 bytes (41 bytes for the AMR and its payload header plus the 40 bytes of the IP/UDP/RTP headers) per frame will be transmitted that is to say 32.4 kbit/s, this will go up to 101 bytes (40.4 kbit/s) when using IPv6 protocol on the air interface.

ROHC algorithm will be supported for AMR-WB conversation test, for the 12.65 kbit/s mode and the 15.85 mode. Header compression will be done on the IP/UDP/RTP headers. ROHC will start in the unidirectional mode and switch to bidirectional mode as soon as a packet has reached the decompressor and it has replied with a feedback packet indicating that a mode transition is desired.

The Conversational / Speech / UL:42.8 DL:42.8 kbps / PS RAB RAB coming from TS 34.108 v4.7.0 will be used:

Here is the RAB description:

Higher layer	RAB/Signalling RB		**RAB**
PDCP	PDCP header size, bit		8
RLC	Logical channel type		DTCH
	RLC mode		UM
	Payload sizes, bit		920, 304, 96
	Max data rate, bps		46000
	UMD PDU header, bit		8
MAC	MAC header, bit		0
	MAC multiplexing		N/A
Layer 1	TrCH type		DCH
	TB sizes, bit		928, 312, 104
	TFS	TF0, bits	0x928
		TF1, bits	1x104
		TF2, bits	1x312
		TF3, bits	1x928
	TTI, ms		20
	Coding type		TC
	CRC, bit		16
	Max number of bits/TTI after channel coding		2844
	Uplink: Max number of bits/radio frame before rate matching		1422
	RM attribute		180-220

3.2.3.2 Description of the RLC implementation

The UMTS air interface simulator (PC 2 and 4) receives IP/UDP/RTP/AMR packets on a specified port of the network card (see Figure 4). The IP/UDP/RTP/AMR packets are given to the transmission buffer of the RLC layer, which works in UM. The RLC will segment or concatenate the IP bitstream in RLC PDUs, adding appropriate RLC headers (sequence number and length indicators). It is assumed that always Transport Format TF 3 is chosen on the physical layer, providing an RLC PDU length including header of 928 bits. In the regular case, one IP packet is placed into an RLC PDU that is filled up with padding bits. Due to delayed packets from the network simulator it may also occur that there are more than one IP packets in the RLC transmission buffer to transmit in the current TTI.

Each TTI of 20ms, an RLC PDU is formed. It is then given to the error insertion block that decides if the RLC PDU is transmitted successfully over the air interface or if it is discarded due to a block error after channel decoding. The physical layer will not be simulated in real time, but error pattern files will be provided. The error patterns of the air interface transmission will be simulated according to the settings given in Section 0. They consist of binary decisions for each transmitted RLC PDU, resulting in a certain BLER.

After the error pattern insertion, the RLC of the air interface receiver site receives RLC PDUs in the reception buffer. The sequence numbers of the RLC headers are checked to detect when RLC PDUs have been discarded due to block errors. A discarded RLC PDU will result in one or more lost IP packets, resulting in a certain packet loss rate of the IP packets and thereby in a certain FER of the AMR frames. The IP/UDP/RTP/AMR packets are reassembled and transmitted to the next PC. This PC is either the network simulator (PC 3) in case of uplink transmission, or it is one of the terminals (PC 1 or 5) in case of downlink transmission.

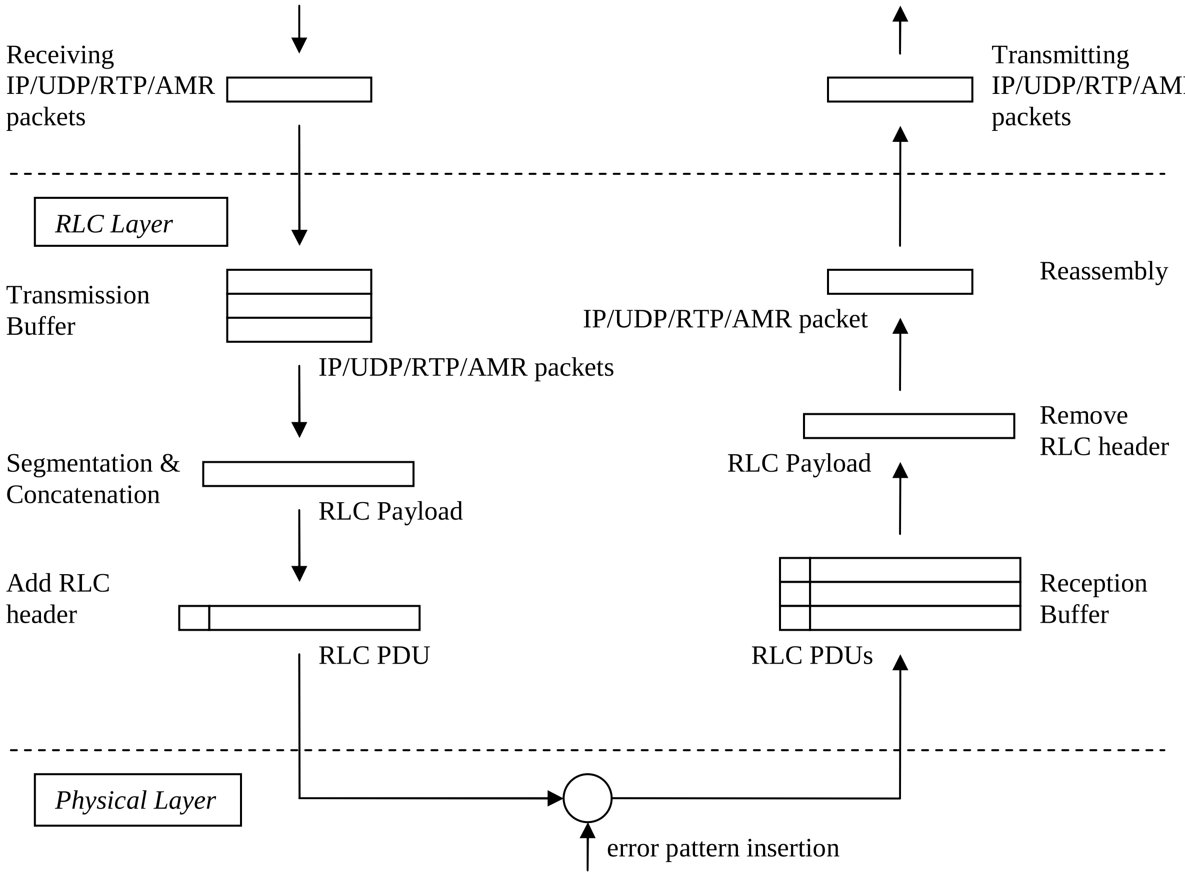

Figure 4: UMTS air interface simulation

3.2.3.3 Physical Layer Implementation

The parameters of the physical layer simulation were set according to the parameters for a DCH in multipath fading conditions given in TS 34.121 (downlink) and TS 25.141 (uplink). The TB size is 928 bits and the Turbo decoder uses the Log-MAP algorithm with 4 iterations. The rake receiver has 6 fingers at 60 possible positions.

The different channel conditions given in **Table 1**, **Table 2**, and **Table 3** were extracted from TR 101 112 (Selection procedures for the choice of radio transmission technologies of the UMTS) and also mentioned in the annex of the document S4-020680.

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0	FLAT
2	50	-3.0	FLAT
3	110	-10.0	FLAT
4	170	-18.0	FLAT
5	290	-26.0	FLAT
6	310	-32.0	FLAT

Table 1: Indoor Office Test Environment Tapped-Delay-Line Parameters

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0.0	CLASSIC
2	310	-1.0	CLASSIC
3	710	-9.0	CLASSIC
4	1090	-10.0	CLASSIC
5	1730	-15.0	CLASSIC
6	2510	-20.0	CLASSIC

Table 2: Vehicular Test Environment, High Antenna, Tapped-Delay-Line Parameters

Tap	Channel A		Doppler Spectrum
	Rel. Delay (nsec)	Avg. Power (dB)	
1	0	0	CLASSIC
2	110	-9.7	CLASSIC
3	190	-19.2	CLASSIC
4	410	-22.8	CLASSIC
5	-	-	CLASSIC
6	-	-	CLASSIC

Table 3: Outdoor to Indoor and Pedestrian Test Environment Tapped-Delay-Line Parameters

Table 4 (DL) and **Table 5** (UL) show approximate results of the air interface simulation for $\frac{DPCH\_E_c}{I_{or}}$ and E_b/N_0 corresponding to the considered BLERs.

	BLER			
Channel	**5*10^{-2}**	**1*10^{-2}**	**1*10^{-3}**	**5*10^{-4}**
Indoor, 3 km/h ($\hat{I}_{or}/I_{oc}$ = 9 dB)	-13.1 dB	-8.9 dB	-3.4 dB	-2.4 dB
Outdoor to Indoor, 3 km/h ($\hat{I}_{or}/I_{oc}$ = 9 dB)	-13.2 dB	-9.7 dB	-5.9 dB	-5.2 dB
Vehicular, 50 km/h ($\hat{I}_{or}/I_{oc}$ = -3 dB)	-9.35 dB	-8.2 dB	-6.9 dB	-6.55 dB
Vehicular, 120 km/h ($\hat{I}_{or}/I_{oc}$ = -3 dB)	-9.7 dB	-8.95 dB	-7.95 dB	-7.55 dB

Table 4: **Downlink performance - approximately** $\frac{DPCH\_E_c}{I_{or}}$ **for the different channels and BLER**

	BLER			
Channel	**5*10^{-2}**	**1*10^{-2}**	**1*10^{-3}**	**5*10^{-4}**
Indoor, 3 km/h	3.9 dB	6.4 dB	9.2 dB	9.8 dB
Outdoor to Indoor, 3 km/h	3.7 dB	6.1 dB	8.6 dB	9.2 dB
Vehicular, 50 km/h	-0.9 dB	-0.15 dB	0.55 dB	0.75 dB
Vehicular, 120 km/h	0.2 dB	0.6 dB	1.1 dB	1.3 dB

Table 5: **Uplink performance - approximately E_b/N_0 for the different channels and BLER**

3.2.4Headsets and Sound Card

To avoid echo problems, it has been decided to use headsets, instead of handsets. The monaural headsets are connected to the sound cards of the PCs supporting the AMR simulators.

The sound level in the earphones can be adjusted, if needed, by the users. But, in practice, the original settings, defined during the preliminary tests, and producing a comfortable listening level, will not be modified. The microphones are protected by a foam ball in order to reduce the "pop" effect. It is also suggested to the user to avoid to place the acoustic opening of the microphone in front of the mouth.

3.2.5 Test environment

Each of the two subjects participating to the conversations is installed in a test room. They sit on an armchair, in front of a table. The test rooms are acoustically insulated. All the test equipments are installed in a third room, connected to the test rooms. When needed, the background noise is generated in the appropriate test room through a set of 4 loudspeakers. The background noise level is adjusted and controlled by a sound level meter. The measurement microphone, connected to the Sound level meter is located at the equivalent of the center of the subject's head. The noise level is A weighted.

3.2.6 Calibration and test conditions monitoring

Speech level

Before the beginning of a set of experiment, the end to end transmission level is checked subjectively, to ensure that there is no problem. If it is necessary to check the speech level following procedure will apply.

An artificial mouth placed in front of the microphone of the Headset A, in the LRGP position -See ITU-T Rec. P.64-, generates in the artificial ear (according to ITU-T Rec. P57) coupled to the earphone of the Head set B the nominal level defined in section 4.3. If necessary, the level is adjusted with the receiving volume control of the headset. The similar calibration is done by inverting headsets A and B.

Delay

The overall delay (from the input of sound card A to the output of sound card B) will be evaluated for each test condition.

The hypothetical delay is calculated as shown :

On the air interface side, the simulator only receives packets on its network card, process them and transmits every 20 ms these packets to the following PC. Only processing delay and a possible delay due to a jitter can be added (a packet arrives just after the sending window of the air interface).

The hypothetical delay is calculated as shown :

On encoder side, delay have to take into account framing, look-ahead, processing and packetization: 45ms

Uplink delay between UE and Iu: 84.4 ms (see TR25.853)

Core network delay: a few ms

Routing through IP: depending on the number of routers.

Downlink delay between Iu and Ue: 71.8 ms (see TR25.853)

And delay on decoder side, taking into account jitter buffer, de-packetization and processing, 40 ms

The total delay to be considered is at least: 241.2 ms.

Note : The actual delay will be measured on the test equipment.

3.3 Test Conditions

The 24 test conditions are :

Condition	**Experimental actors**		
	Radio conditions	IP conditions (Packet loss ratio)	Mode
1	10^{-2}	0%	12,65 kbit/s, RoHC
2	10^{-2}	0%	12,65 kbit/s
3	10^{-2}	0%	15,85 kbit/s, RoHC
4	10^{-2}	3%	12,65 kbit/s, RoHC
5	10^{-2}	3%	12,65 kbit/s
6	10^{-2}	3%	15,85 kbit/s, RoHC
7	10^{-3}	0%	12,65 kbit/s, RoHC
8	10^{-3}	0%	12,65 kbit/s
9	10^{-3}	0%	15,85 kbit/s, RoHC
10	10^{-3}	3%	12,65 kbit/s, RoHC
11	10^{-3}	3%	12,65 kbit/s
12	10^{-3}	3%	15,85 kbit/s, RoHC
13	$5\ 10^{-4}$	0%	12,65 kbit/s, RoHC
14	$5\ 10^{-4}$	0%	12,65 kbit/s
15	$5\ 10^{-4}$	0%	15,85 kbit/s, RoHC
16	$5\ 10^{-4}$	3%	12,65 kbit/s, RoHC
17	$5\ 10^{-4}$	3%	12,65 kbit/s
18	$5\ 10^{-4}$	3%	15,85 kbit/s, RoHC

Condition	**Additional Background noise Room A**	**Additional Background noise Room B**	**Experimental actors**		
			Radio conditions	IP conditions (Packet loss ratio)	Mode
19	Car	No	$5\ 10^{-4}$	3%	12,65 kbit/s, RoHC

20	No	Car	$5\ 10^{-4}$	3%	12,65 kbit/s, RoHC
21	Cafeteria	No	$5\ 10^{-4}$	0%	12,65 kbit/s
22	No	Cafeteria	$5\ 10^{-4}$	0%	12,65 kbit/s
23	Street	No	$5\ 10^{-4}$	0%	15,85 kbit/s, RoHC
24	No	Street	$5\ 10^{-4}$	0%	15,85 kbit/s, RoHC

Noise types

Noise type	**Level (dB Pa A**
Car	60
Street	55
Bable	50

Listening Level	1	79 dBSPL
Listeners	32	Naïve Listeners
Groups	16	2 subjects/group
Rating Scales	5	
Languages	1	See table
Listening System	1	Monaural headset (flat response in the audio bandwidth of interest: 50Hz-7kHz). The other ear is open.
Listening Environment		Room Noise: Hoth Spectrum at 30dBA (as defined by ITU-T, Recommendation P.800, Annex A, section A.1.1.2.2.1 Room Noise, with table A.1 and Figure A.1),except when background noise is needed (see table)

Annex A Example Instructions for the conversation test

Table : Instructions to subjects.

INSTRUCTIONS TO SUBJECTS

In this experiment we are evaluating systems that might be used for telecommunication services.
You are going to have a conversation with another user. The test situation is simulating communications between two mobile phones. The most of the situations will correspond to silent environment conditions, but some other will simulate more specific situations, as in a car, or in a railway station or in an office environment, when other people are discussing in the background.

After the completion of each call conversation, you will have to give your opinions on the quality, by answering to the following questions that will be displayed on the screen of the black box in front of you. Your judgment will be stored. You have 8 seconds to answer to each question. After "pressing" the button on the screen, another question will be displayed. You continue the procedure for the 5 following questions.

Question 1: How do you judge the quality of the voice of your partner?

Excellent	Good	Fair	Poor	Bad

Question 2: Do you have difficulties to understand some words?

All the time	Often	Some time to time	Rarely	Never

Question 3: How did you judge the conversation when you interacted with your partner?

Excellent interactivity (similar to face-to-face situation)	Good interactivity (in few moments, you were talking simultaneously, and you had to interrupt yourself)	Fair interactivity (sometimes, you were talking simultaneously, and you had to interrupt yourself)	Poor interactivity (often, you were talking simultaneously, and you had to interrupt yourself)	Bad interactivity (it was impossible to have an interactive conversation)

Question 4: Did you perceive any impairment (noises, cuts,…)? In that case, was it:

No impairment	Slight impairment, but not disturbing	Impairment slightly disturbing	Impairment disturbing	Very disturbing Impairment

Question 5: How do you judge the global quality of the communication?

Excellent	Good	Fair	Poor	Bad

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

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

Annex B: Example Scenarios for the conversation test

The pretexts used for conversation test are those developed by the Rurh University (Bochum, Germany) within the context of ITU-T SG12 . These scenarios have been elaborated to allow a conversation well balanced within both participants and lasting approximately 2'30 or 3', and to stimulate the discussion between persons that know each other to facilitate the naturalness of the conversation. They are derived from typical situations of every day life: railways inquiries, rent a car or an apartment, etc. Each condition should be given a different scenario.

Examples coming from ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

- Scenario 1 : Pizza service

Subject 1:

Your Name :	Clemence
Reason for the call	1 large Pizza
Condition which should be applied to the exchange of information	For 2 people, Vegetarian pizza prefered
Information you want to receive from your partner	Topping Price
Information that your partner requires	Delivery address : 41 industry street,Oxford Phone : 7 34 20
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	How long will it take?

Subject 2:

<table>
<tr><td>Your Name :</td><td>Pizzeria Roma</td></tr>
<tr><td>Information from which you should select the details which your partner requires</td><td>
<table>
<tr><td>Pizzas</td><td>1 person</td><td>2 persons</td><td>4 persons</td></tr>
<tr><td>Toscana
(ham, mushrooms, tomatoes,cheese)</td><td>3.2£</td><td>5.95£</td><td>10.5£</td></tr>
<tr><td>Tonno
(Tuna, onions, tomatoes, cheese)</td><td>3.95£</td><td>7.5£</td><td>13.95£</td></tr>
<tr><td>Fabrizio
(salami, ham, tomatoes, heese)</td><td>4.2£</td><td>7.95£</td><td>14.95£</td></tr>
<tr><td>Vegetaria
(spinach, mushrooms, tomatoes,cheese)</td><td>4.5£</td><td>8.5£</td><td>15.95£</td></tr>
</table>
</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Name
address
telephone number</td></tr>
<tr><td>Question to which neither you nor your partner will have information.
You should discuss and find a solution that is acceptable to both of you.</td><td></td></tr>
</table>

- Scenario 2 : Information on flights

Subject 1:

Your Name :	Parker
Reason for the call	Intended journey: London Heathrow → Düsseldorf
Condition which should be applied to the exchange of information	On June 23th, Morning flight, Direct flight preferred
Information you want to receive from your partner	Departure : Arrival Flight number
Information that your partner requires	Reservation : 1 seat, Economy class Address: 66 middle street, Sheffield Phone: 21 08 33
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	From which airport is it easier to get into Cologne center : Düsseldorf or Cologne/Bonn

Subject 2:

<table>
<tr><td>Your Name :</td><td>Heathrow flight information</td></tr>
<tr><td>Information from which you should select the details which your partner requires</td><td>
<table>
<tr><td>Flight schedule</td><td>Lufthansa</td><td>British Airways</td><td>Lufthansa</td></tr>
<tr><td>Flight number</td><td>LH 2615</td><td>BA 381</td><td>LH 413</td></tr>
<tr><td>London Heathrow departure</td><td>6:30</td><td>6:35</td><td>8:20</td></tr>
<tr><td>Brussels arrival
Brussels departure</td><td></td><td>7:35
8:00</td><td></td></tr>
<tr><td>Düsseldorf arrival</td><td>7:35</td><td>9:05</td><td>9:25</td></tr>
</table>
</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Name
address
telephone number
number of seats
Class : Business or Economy</td></tr>
<tr><td>Question to which neither you nor your partner will have information.
You should discuss and find a solution that is acceptable to both of you.</td><td></td></tr>
</table>

Annex C: Results to be provided

For contractual purposes, the information which needs to be provided is defined here.

The information required from each test Laboratory is a table containing the following information for each of the conditions in the experiment:

The "Mean Opinion Score (MOS)" obtained for all the subjects.
When the conditions are symmetrical, the mean value is calculated from all the result for the two test rooms..

For the dissymmetric conditions, the mean is calculated on the two test conditions, each result cumulating the results obtained in each condition of background noise.

The Standard Deviation of the "MOS" obtained for all the subjects, for each test condition.

The specific statistical comparisons are specified in Annex C.

Annex D: Data analysis and presentation of results

D.1 Calculation of MOS and Standard Deviation

The (overall) MOS/DMOS for confounded subjects for condition C (Yc) can then be obtained from:

$$Y_c = \frac{1}{T}\sum_{t=1}^{T} Y_{c,t}$$

The standard deviation (S) for condition C, denoted as Sc, can be calculated as:

$$S_c = \sqrt{\frac{1}{L \times T - 1}\sum_{t=1}^{T}\sum_{l=1}^{L}(X_{c,l,t} - Y_c)^2}$$

Finally, the confidence interval (CI) at the (1-α) level can be calculated for $N = L \times T$ as:

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

D.2 Presentation of Basic Statistical Results

The test results should be reported by the test Laboratory and the Global Analysis Laboratory as follows:

Calculate and tabulate "Mean Opinion Scores" for the (opinion scales, Standard Deviations and Confidence Intervals as shown in Table E.1.

Table C.1 - Layout for presentation of test results.

D.3 Thorough analysis

Two statistical analyses should be conducted on the data obtained with these subjective scales. The first analysis consists in a Multiple ANalysis OF VAriance (MANOVA), which globally indicates the possible effect of the experimental factors (*i.e.*, different conditions). Then, a specific ANOVA should be run on each dependent variable (the five scales) to test if there is an effect of a specific experimental factor for a given subjective variable. In other words, these statistical analyses indicate if the differences observed between the MOS obtained for the different conditions are significant, for one given dependant variable (ANOVA) or for the whole of dependant variables (MANOVA). Finally, Pearson's linear correlations should be computed between the results of all subjective variables, to see which are those preponderant or dependent on others.

Annex F: Test plan for Packet Switched Conversation Tests for Comparison of Quality Offered by Different Speech Coders

Source:	**France Telecom R&D**
Title:	**Test plan for packet switched conversation test. Comparison of quality offered by different speech coders.**
Document For:	**Discussion and Approval**
Agenda Item:	

Introduction

This document proposes a conversation test plan to compare the quality obtained with several different speech coders, over packet switched networks.

The different speech coders used in this test are

Adaptive Multi-Rate Narrow-Band (AMR-NB), in modes 6.7 kbit/s and 12.2 kbit/s,

Adaptive Multi-Rate Wide-Band (AMR-WB), in modes 12.65 kbit/s and 15.85 kbit/s,

ITU-T G.723.1, in mode 6.4 kbit/s,

ITU-T G.729, in mode 8 kbit/s,

ITU-T G.722, in mode 64 kbit/s, with packet loss concealment and,

ITU-T G.711, with packet loss concealment.

As there is no standardized packet loss concealment, plc for G.711 and G.722 are proprietary algorithms.

The simulated network will include two values of IP packet loss.

The test will be done in one test laboratory, only, but in two different languages.

This discussion gives references, conventions and contacts, section 3 details the test methodology, including test arrangement and test procedure,

Annex A contains the instructions for the subjects participating to the conversation tests.

Annex B contains the description of results to be provided to the Analysis Laboratory (if any) by the testing laboratories.

Annex C contains the list of statistical comparisons to be performed.

2. References, Conventions, and Contacts

2.1Permanent Documents

	ITU-T Rec.P.800	Methods for Subjective Determination of Transmission Quality
	ITU-T Rec. P.831	Subjective performance evaluation of network echo cancellers
	ITU-T Rec. G.711	Pulse code modulation (PCM) of voice frequencies
	ITU-T Rec. G.729	Coding of speech at8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)
	ITU-T Rec. G.723.1	Speech coders : Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s
	ITU-T Rec. G.722	7 kHz audio-coding within 64 kbit/s

2.2 Key Acronyms

AMR-NB Adaptive Multi-Rate Narrowband Speech Codec

AMR-WB Adaptive Multi-Rate Wide-band Speech Codec

MOS Mean Opinion Score

2.3 Contact Names

The following persons should be contacted for questions related to the test plan.

Section	Contact Person/Email	Organisation	Address	Telephone/Fax
Experiments and results analysis	L. Gros Laeticia.gros@francetelecom.com	France Telecom R&D	2, Avenue P. Marzin, 22307 Lannion Cédex France	Tel : +3329605 0720 Fax : +33296051316
AOB	Paolo Usai paolo.usai@etsi.fr	ETSI MCC	650 Route des Lucioles 06921 Sophia Antipolis Cedex France	Tel: 33 (0)4 92 94 42 36 Fax: 33 (0)4 93 65 28 17

2.4 Responsibilities

Each test laboratory has the responsibility to organize its conversation tests.

The list of Test laboratories participating to the conversation test phase.

Lab	Company	Language
1	France Telecom R&D	French
	France Telecom R&D	Arabic

3. Test methodology

3.1 Introduction

The protocol described below evaluates the effect of degradation such as delay and dropped packets on the quality of the communications. It corresponds to the conversation-opinion tests recommended by the ITU-T P.800 [1]. First of all, conversation–opinion tests allow subjects passing the test to be in a more realistic situation, close to the actual service conditions experienced by telephone customers. In addition, conversation-opinion tests are suited to assess the effects of impairments that can cause difficulty while conversing (such as delay).
Subjects participate to the test by couple; they are seated in separate sound-proof rooms and are asked to hold a conversation through the transmission chain performed by means of networks simulators and communications are impaired by means of an IP impairments simulator part of the CN simulator, as the figure below describes it.

3.2 Test arrangement

3.2.1 Description of the proposed testing system

This contribution describes a networks simulator for the characterization of the different speech codecs when the bitstream is transmitted over a PS network. The procedure to do the conversational listening test has been earlier described in [1].

Figure 1 describes the system that is going to be simulated:

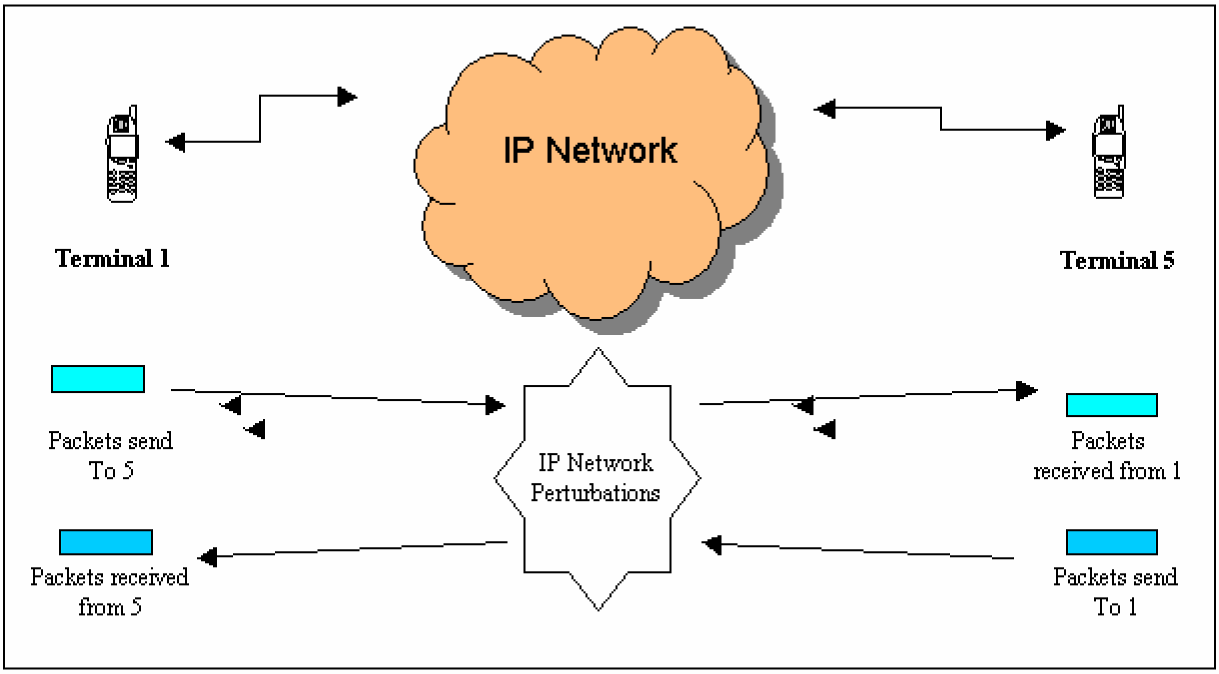

Figure 1: Packet switch audio communication simulator

This will be simulated using 5 PCs as shown in Figure 2.

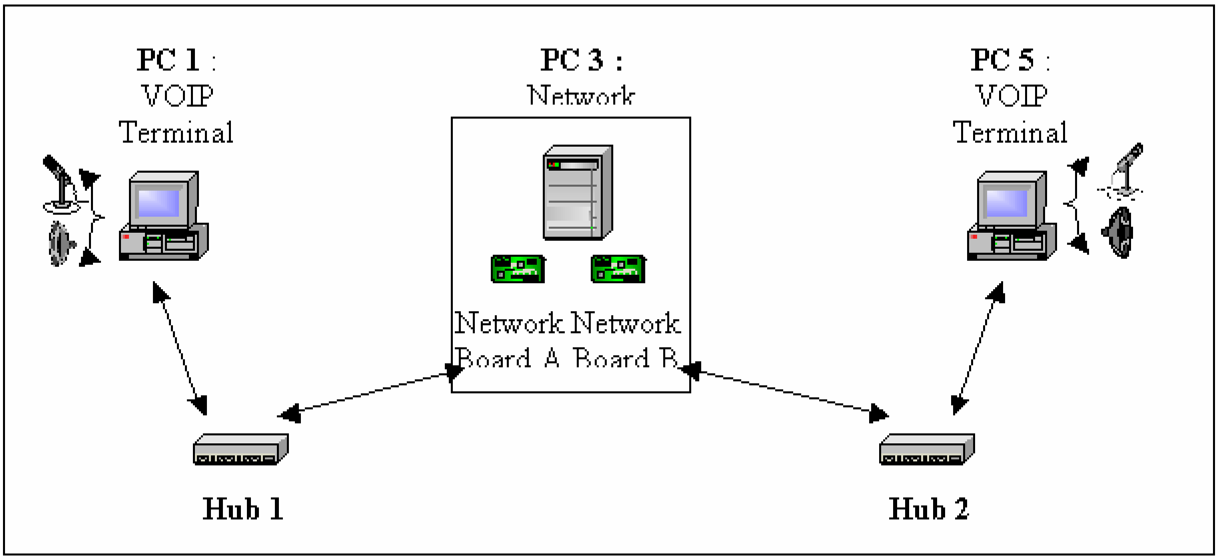

Figure 2: Simulation Platform

PC 1 and PC 5: PCs under Windows OS with VOIP Terminal Simulator Software of France Telecom R&D.

PC 3: PCs under WinNT OS with Network Simulator Software (NetDisturb).

<u>Basic Principles:</u>

The platform simulates a packet switch interactive communication between two users using PC1 and PC5 as their relatives VOIP terminals. PC1 sends encoded packets that are encapsulated using IP/UDP/RTP headers to PC5. PC1 receives these IP/UDP/RTP audio packets from PC5.

3.2.2 France Telecom Network simulator

The core network simulator, as implemented, works under IPv4.

Figure 3 shows the possible parameters that can be modified, but, in this test, only "loss Law" will have two values, all the others settings being fixed.

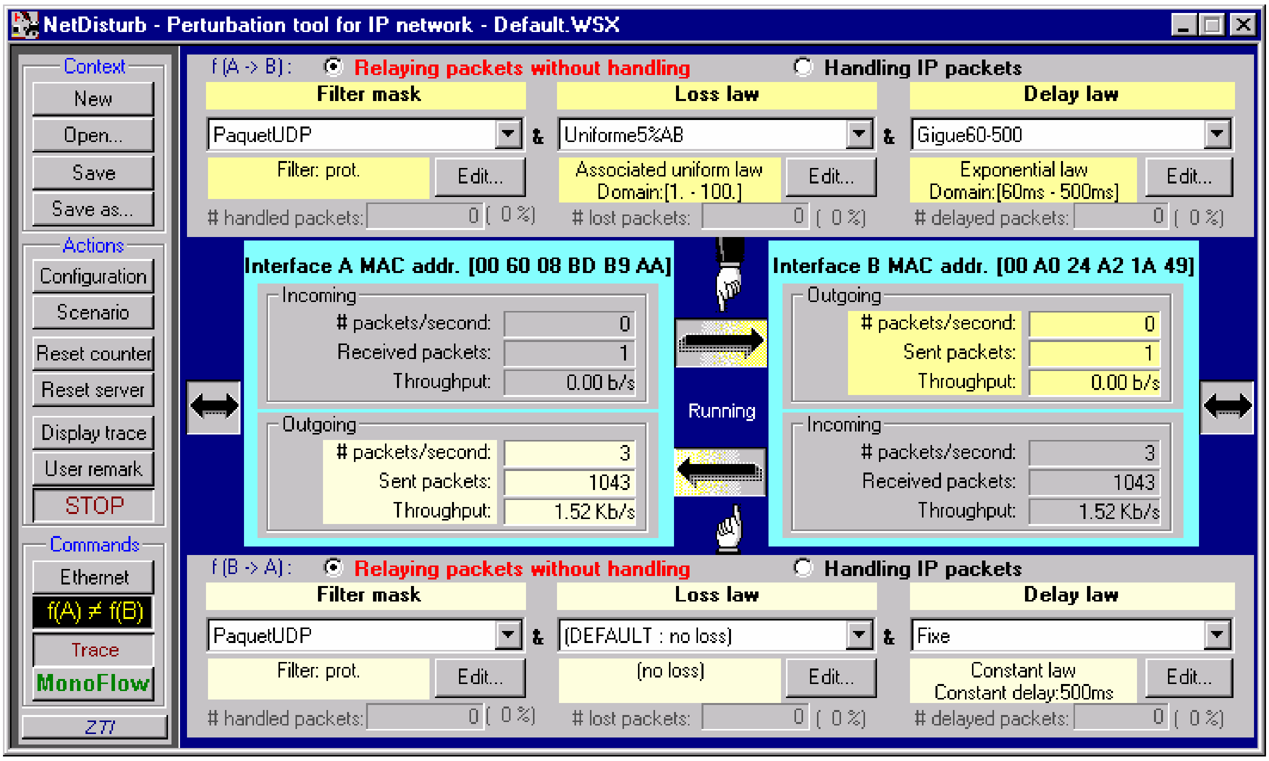

Figure 3: IP simulator interface

On both links, one can choose delay and loss laws. Both links can be treated separately or on the same way. For example, delay can be set to a fixed value but can also be set to another law such as exponential law.

3.2.3 Headsets and Sound Card

To avoid echo problems, it has been decided to use headsets, instead of handsets. The monaural headsets are connected to the sound cards of the PCs supporting the AMR simulators.

The sound level in the earphones can be adjusted, if needed, by the users. But, in practice, the original settings, defined during the preliminary tests, and producing a comfortable listening level, will not be modified. The microphones are protected by a foam ball in order to reduce the "pop" effect. It is also suggested to the user to avoid to place the acoustic opening of the microphone in front of the mouth.

3.2.4 Test environment

Each of the two subjects participating to the conversations is installed in a test room. They sit on an armchair, in front of a table. The test rooms are acoustically insulated. All the test equipments are installed in a third room, connected to the test rooms. The background noise level is checked by a sound level meter. The measurement microphone, connected to the Sound level meter is located at the equivalent of the center of the subject's head. The noise level is A weighted.

3.2.5 Calibration and test conditions monitoring

Speech level

Before the beginning of a set of experiment, the end to end transmission level is checked subjectively, to ensure that there is no problem. If it is necessary to check the speech level following procedure will apply. An artificial mouth placed in front of the microphone of the Headset A, in the LRGP position -See ITU-T Rec. P.64-, generates in the artificial ear (according to ITU-T Rec. P57) coupled to the earphone of the Head set B the nominal level defined in section 4.3. If necessary, the level is adjusted with the receiving volume control of the headset. The similar calibration is done by inverting headsets A and B.

Delay

The overall delay (from the input of sound card A to the output of sound card B) will be adjusted for each test condition taking into account the delay of the related codec in order to have a fixed delay around 250ms. This value of 250ms is close to the hypothetical delay computed for AMR and AMRWB through the UMTS network.

3.3 Test Conditions

Condition	**Experimental actors**		
		IP conditions (Packet loss ratio)	Mode
1		0%	AMR NB 6,7kbit/s
2		0%	AMR-NB 12,2 kbit/s
3		0%	AMR-WB 12,65 kbit/s
4		0%	AMR-WB 15,85 kbit/s
5		0%	G. 723.1 6,4 kbit/s
6		0%	G.729 8 kbit/s
7		0%	G.722 64 kbit/s + plc
8		0%	G.711 + plc
9		3%	AMR NB 6,7kbit/s
10		3%	AMR-NB 12,2 kbit/s (delay 300 ms)
11		3%	AMR-WB 12,65 kbit/s
12		3%	AMR-WB 15,85 kbit/s
13		3%	G. 723.1 6,4 kbit/s
14		3%	G.729 8 kbit/s
15		3%	G.722 64 kbit/s + plc
16		3%	G.711 + plc

Listening Level	1	79 dBSPL
Listeners	32	Naïve Listeners per langage
Groups	16	2 subjects/group
Rating Scales	5	
Languages	1	See table
Listening System	1	Monaural headset (flat response in the audio bandwidth of interest: 50Hz-7kHz). The other ear is open.
Listening Environment		Room Noise: Hoth Spectrum at 30dBA (as defined by ITU-T, Recommendation P.800, Annex A, section A.1.1.2.2.1 Room Noise, with table A.1 and Figure A.1),

References

Tdoc S4-030564- Test Plan for the AMR Narrow-Band Packet switched Conversation test

Tdoc S4-030565- Test Plan for the AMR Wide-Band Packet switched Conversation test

END

Annex A Example Instructions for the conversation test

Table : Instructions to subjects.

INSTRUCTIONS TO SUBJECTS

In this experiment we are evaluating systems that might be used for telecommunication services.
You are going to have a conversation with another user. The test situation is simulating communications between two mobile phones. All the situations will correspond to silent environment condition

After the completion of each call conversation, you will have to give your opinions on the quality, by answering to the following questions that will be displayed on the screen of the black box in front of you. Your judgment will be stored. You have 8 seconds to answer to each question. After "pressing" the button on the screen, another question will be displayed. You continue the procedure for the 5 following questions.

Question 1: How do you judge the quality of the voice of your partner?

Excellent	Good	Fair	Poor	Bad

Question 2: Do you have difficulties to understand some words?

All the time	Often	Some time to time	Rarely	Never

Question 3: How did you judge the conversation when you interacted with your partner?

Excellent interactivity (similar to face-to-face situation)	Good interactivity (in few moments, you were talking simultaneously, and you had to interrupt yourself)	Fair interactivity (sometimes, you were talking simultaneously, and you had to interrupt yourself)	Poor interactivity (often, you were talking simultaneously, and you had to interrupt yourself)	Bad interactivity (it was impossible to have an interactive conversation)

Question 4: Did you perceive any impairment (noises, cuts,…)? In that case, was it:

No impairment	Slight impairment, but not disturbing	Impairment slightly disturbing	Impairment disturbing	Very disturbing Impairment

Question 5: How do you judge the global quality of the communication?

Excellent	Good	Fair	Poor	Bad

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

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

Annex B: Example Scenarios for the conversation test

The pretexts used for conversation test are those developed by the Rurh University (Bochum, Germany) within the context of ITU-T SG12 . These scenarios have been elaborated to allow a conversation well balanced within both participants and lasting approximately 2’30 or 3’, and to stimulate the discussion between persons that know each other to facilitate the naturalness of the conversation. They are derived from typical situations of every day life: railways inquiries, rent a car or an apartment, etc. Each condition should be given a different scenario.

Examples coming from ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

- Scenario 1 : Pizza service

Subject 1:

Your Name :	Clemence
Reason for the call	1 large Pizza
Condition which should be applied to the exchange of information	For 2 people, Vegetarian pizza prefered
Information you want to receive from your partner	Topping Price
Information that your partner requires	Delivery address : 41 industry street,Oxford Phone : 7 34 20
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	How long will it take?

Subject 2:

<table>
<tr><td>Your Name :</td><td>Pizzeria Roma</td></tr>
<tr><td>Information from which you should select the details which your partner requires</td><td>
<table>
<tr><td>Pizzas</td><td>1 person</td><td>2 persons</td><td>4 persons</td></tr>
<tr><td>Toscana
(ham, mushrooms, tomatoes,cheese)</td><td>3.2£</td><td>5.95£</td><td>10.5£</td></tr>
<tr><td>Tonno
(Tuna, onions, tomatoes, cheese)</td><td>3.95£</td><td>7.5£</td><td>13.95£</td></tr>
<tr><td>Fabrizio
(salami, ham, tomatoes, heese)</td><td>4.2£</td><td>7.95£</td><td>14.95£</td></tr>
<tr><td>Vegetaria
(spinach, mushrooms, tomatoes,cheese)</td><td>4.5£</td><td>8.5£</td><td>15.95£</td></tr>
</table>
</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Name
address
telephone number</td></tr>
<tr><td>Question to which neither you nor your partner will have information.
You should discuss and find a solution that is acceptable to both of you.</td><td></td></tr>
</table>

- Scenario 2 : Information on flights

Subject 1:

Your Name :	Parker
Reason for the call	Intended journey: London Heathrow → Düsseldorf
Condition which should be applied to the exchange of information	On June 23th, Morning flight, Direct flight preferred
Information you want to receive from your partner	Departure : Arrival Flight number
Information that your partner requires	Reservation : 1 seat, Economy class Address: 66 middle street, Sheffield Phone: 21 08 33
Question to which neither you nor your partner will have information. You should discuss and find a solution that is acceptable to both of you.	From which airport is it easier to get into Cologne center : Düsseldorf or Cologne/Bonn

Subject 2:

<table>
<tr><td>Your Name :</td><td>Heathrow flight information</td></tr>
<tr><td>Information from which you should select the details which your partner requires</td><td>
<table>
<tr><td>Flight schedule</td><td>Lufthansa</td><td>British Airways</td><td>Lufthansa</td></tr>
<tr><td>Flight number</td><td>LH 2615</td><td>BA 381</td><td>LH 413</td></tr>
<tr><td>London Heathrow departure</td><td>6:30</td><td>6:35</td><td>8:20</td></tr>
<tr><td>Brussels arrival
Brussels departure</td><td></td><td>7:35
8:00</td><td></td></tr>
<tr><td>Düsseldorf arrival</td><td>7:35</td><td>9:05</td><td>9:25</td></tr>
</table>
</td></tr>
<tr><td>Information you want to receive from your partner</td><td>Name
address
telephone number
number of seats
Class : Business or Economy</td></tr>
<tr><td>Question to which neither you nor your partner will have information.
You should discuss and find a solution that is acceptable to both of you.</td><td></td></tr>
</table>

ITU-T SG 12 COM12-35 "Development of scenarios for short conversation test", 1997

Annex C: Results to be provided

For contractual purposes, the information which needs to be provided is defined here.

The information required from each test Laboratory is a table containing the following information for each of the conditions in the experiment:

The "Mean Opinion Score (MOS)" obtained for all the subjects.
When the conditions are symmetrical, the mean value is calculated from all the result for the two test rooms..

For the dissymmetric conditions, the mean is calculated on the two test conditions, each result cumulating the results obtained in each condition of background noise.

The Standard Deviation of the "MOS" obtained for all the subjects, for each test condition.

The specific statistical comparisons are specified in Annex C.

Annex D: Data analysis and presentation of results

D.1 Calculation of MOS and Standard Deviation

The (overall) MOS/DMOS for confounded subjects for condition C (Yc) can then be obtained from:

$$Y_c = \frac{1}{T}\sum_{t=1}^{T} Y_{c,t}$$

The standard deviation (S) for condition C, denoted as Sc, can be calculated as:

$$S_c = \sqrt{\frac{1}{L \times T - 1}\sum_{t=1}^{T}\sum_{l=1}^{L}(X_{c,l,t} - Y_c)^2}$$

Finally, the confidence interval (CI) at the (1-α) level can be calculated for $N = L \times T$ as:

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

D.2 Presentation of Basic Statistical Results

The test results should be reported by the test Laboratory and the Global Analysis Laboratory as follows:

Calculate and tabulate "Mean Opinion Scores" for the (opinion scales, Standard Deviations and Confidence Intervals as shown in Table E.1.

Table C.1 - Layout for presentation of test results.

D.3 Thorough analysis

Two statistical analyses should be conducted on the data obtained with these subjective scales. The first analysis consists in a Multiple ANalysis OF VAriance (MANOVA), which globally indicates the possible effect of the experimental factors (*i.e.*, different conditions). Then, a specific ANOVA should be run on each dependent variable (the five scales) to test if there is an effect of a specific experimental factor for a given subjective variable. In other words, these statistical analyses indicate if the differences observed between the MOS obtained for the different conditions are significant, for one given dependant variable (ANOVA) or for the whole of dependant variables (MANOVA). Finally, Pearson's linear correlations should be computed between the results of all subjective variables, to see which are those preponderant or dependent on others.

Annex G: Test Plan for Global Analysis of PSS Conversation Tests

Source:	**Dynastat[1]**
Title:	**Proposed Test Plan for Global Analysis of PSS Conversation Tests (R1)**
Document for:	**Discussion and Approval**
Agenda Item:	**7, 13.1**

1. Introduction

This contribution presents a proposal for conducting a Global Analysis of the results derived from the 3GPP Conversation Tests for Packet Switched (PS) networks. Phase I of these tests are described in two test plans -- S4-030564 for conversation tests using the Adaptive Multi-Rate Narrow-Band (AMR-NB) codec, S4-030565 for conversation tests using the Adaptive Multi-Rate Wide-Band (AMR-WB) codec. The test plan for the Phase II tests are described in S4-030747 for conversation tests comparing various ITU-T standardized speech codecs. The Phase I test plans specify similar experimental designs involving 24 test conditions and 16 pairs of subjects. They also specify that three Listening Laboratories (LL) will conduct the tests in different languages: Arcon for North American English (NAE), NTT-AT for Japanese, and France Telecom for French. The Phase II test plan involves 16 conditions and a single Listening Lab (France Telecom) conducting the test in two languages (French and Arabic).

2. Phase I - AMR-NB Tests

Table 1 shows the 24 test conditions involved in the AMR-NB conversation tests.

Table 1. Test Conditions in the PS Conversation Tests for AMR-NB

Condition	Room A	Room B	Radio conditions	Packet loss (%)	Mode (kbps)	Delay (msec)
1	No	No	10^{-2}	0	6.7	300
2	No	No	10^{-2}	0	12.2	500
3	No	No	10^{-2}	0	12.2	300
4	No	No	10^{-2}	3	6.7	300
5	No	No	10^{-2}	3	12.2	500
6	No	No	10^{-2}	3	12.2	300
7	No	No	10^{-3}	0	6.7	300
8	No	No	10^{-3}	0	12.2	500
9	No	No	10^{-3}	0	12.2	300
10	No	No	10^{-3}	3	6.7	300
11	No	No	10^{-3}	3	12.2	500
12	No	No	10^{-3}	3	12.2	300
13	No	No	5×10^{-4}	0	6.7	300
14	No	No	5×10^{-4}	0	12.2	500
15	No	No	5×10^{-4}	0	12.2	300
16	No	No	5×10^{-4}	3	6.7	300
17	No	No	5×10^{-4}	3	12.2	500
18	No	No	5×10^{-4}	3	12.2	300
19	Car	No	5×10^{-4}	3	12.2	300
20	No	Car	5×10^{-4}	3	12.2	300
21	Cafeteria	No	5×10^{-4}	0	6.7	300
22	No	Cafeteria	5×10^{-4}	0	6.7	300
23	Street	No	5×10^{-4}	0	12.2	500
24	No	Street	5×10^{-4}	0	12.2	500

Test conditions 1-18 are symmetrical in that both subjects in a conversation-pair are listening in quiet (i.e., no noise) rooms. Conditions 19-24, on the other hand, are asymmetrical, one subject is listening in a quiet room, the other in a noisy room. Conditions 1-18 are categorized by four experimental factors:

- Delay – 300 msec and 500 msec
- AMR-NB mode (rate) – 6.7 kbps and 12.2 kbps
- Packet Loss – 0% and 3%
- Radio conditions – 10^{-2}, 10^{-3}, and $5x10^{-4}$

These conditions can be assigned to two factorial designs for analysing the effects of three of these factors. Table 2 shows the conditions involved in the two three-factor analyses for the AMR-NB experiments. Using the 12 conditions shown in Table 2a, the effects of Rate, Radio Conditions, and Packet Loss can be evaluated (Delay held constant at 300 msec). Using the 12 conditions shown in Table 2b, the effects of Delay, Radio Conditions, and Packet Loss can be evaluated (Rate held constant at 12.2 kbps).

Table 2a AMR-NB: Factorial Design for the Effects of Rate, Radio Cond., and Packet Loss

No Noise - 300 msec delay			
6.7kbps / 0% PL		**6.7kbps / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	1	10^{-2}	4
10^{-3}	7	10^{-3}	10
$5x10^{-4}$	13	$5x10^{-4}$	16
12.2kbps / 0% PL		**12.2kbps / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	3	10^{-2}	6
10^{-3}	9	10^{-3}	12
$5x10^{-4}$	15	$5x10^{-4}$	18

Table 2b – AMR-NB: Factorial Design for the Effects of Delay, Radio Cond., and Packet Loss

No Noise - 12.2 kbps			
300 msec / 0% PL		**300 msec / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	3	10^{-2}	6
10^{-3}	9	10^{-3}	12
$5x10^{-4}$	15	$5x10^{-4}$	18
500 msec / 0% PL		**500 msec / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	2	10^{-2}	5
10^{-3}	8	10^{-3}	11
$5x10^{-4}$	14	$5x10^{-4}$	17

The three sets of paired conditions involving noise (i.e., conditions 19/20, 21/22, and 23/24) can be used to compare the effects of *sender in noise/receiver in quiet* with those for *sender in quiet/receiver in noise* for the three noise environments.

3. Phase I - AMR-WB Tests

Table 3 shows the test conditions involved in the AMR-WB conversation tests. As in the AMR-NB tests, conditions 1-18 are symmetrical and conditions 19-24 are asymmetrical. Conditions 1-18 are categorized by four experimental factors:

- o RoHC – present and absent
- o AMR-NB mode (rate) – 6.7 kbps and 12.2 kbps
- o Packet Loss – 0% and 3%
- o Radio conditions – 10^{-2}, 10^{-3}, and $5x10^{-4}$

Table 3. Test Conditions in the PS Conversation Tests for AMR-WB

Condition	Room A Noise	Room B Noise	Radio conditions	Packet loss (%)	Mode (kbps)	RoHC
1	No	No	10^{-2}	0	12.65	RoHC
2	No	No	10^{-2}	0	12.65	
3	No	No	10^{-2}	0	15.85	RoHC
4	No	No	10^{-2}	3	12.65	RoHC
5	No	No	10^{-2}	3	12.65	
6	No	No	10^{-2}	3	15.85	RoHC
7	No	No	10^{-3}	0	12.65	RoHC
8	No	No	10^{-3}	0	12.65	
9	No	No	10^{-3}	0	15.85	RoHC
10	No	No	10^{-3}	3	12.65	RoHC
11	No	No	10^{-3}	3	12.65	
12	No	No	10^{-3}	3	15.85	RoHC
13	No	No	5×10^{-4}	0	12.65	RoHC
14	No	No	5×10^{-4}	0	12.65	
15	No	No	5×10^{-4}	0	15.85	RoHC
16	No	No	5×10^{-4}	3	12.65	RoHC
17	No	No	5×10^{-4}	3	12.65	
18	No	No	5×10^{-4}	3	15.85	RoHC
19	Car	No	5×10^{-4}	3	12.65	RoHC
20	No	Car	5×10^{-4}	3	12.65	RoHC
21	Cafeteria	No	5×10^{-4}	0	12.65	
22	No	Cafeteria	5×10^{-4}	0	12.65	
23	Street	No	5×10^{-4}	0	15.85	RoHC
24	No	Street	5×10^{-4}	0	15.85	RoHC

Consistent with the AMR-NB tests, conditions 1-18 can be assigned to two factorial designs for analysing the effects of three of these factors. Table 4 shows the conditions involved in the two three-factor analyses for the AMR-WB experiments. Using the 12 conditions shown in Table 4a, the effects of Rate, Radio Conditions, and Packet Loss can be evaluated (RoHC present in all conditions). Using the 12 conditions shown in Table 4b, the effects of RoHC, Radio Conditions, and Packet Loss can be evaluated (Rate held constant at 12.65 kbps).

Table 4a AMR-NB: Factorial Design for the Effects of Rate, Radio Cond., and Packet Loss

No Noise - RoHC			
12.65kbps / 0% PL		**12.65 kbps / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	1	10^{-2}	4
10^{-3}	7	10^{-3}	10
$5x10^{-4}$	13	$5x10^{-4}$	16
15.85 kbps / 0% PL		**15.85 kbps / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	3	10^{-2}	6
10^{-3}	9	10^{-3}	12
$5x10^{-4}$	15	$5x10^{-4}$	18

Table 4b – AMR-NB: Factorial Design for the Effects of RoHC, Radio Cond., and Packet Loss

No Noise - 12.65 kbps			
RoHC / 0% PL		**RoHc / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	3	10^{-2}	6
10^{-3}	9	10^{-3}	12
$5x10^{-4}$	15	$5x10^{-4}$	18
No RoHC / 0% PL		**No RoHC / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	2	10^{-2}	5
10^{-3}	8	10^{-3}	11
$5x10^{-4}$	14	$5x10^{-4}$	17

Again, consistent with the tests for AMR-NB, the three sets of paired conditions involving noise (i.e., conditions 19/20, 21/22, and 23/24) can be used to compare the effects of *sender in noise/receiver in quiet* with those for *sender in quiet/receiver in noise* for the three noise environments.

4. Phase II - ITU-T Codec Tests

Table 5 shows the test conditions involved in the conversation tests designed to compare the performance of standardized ITU-T codecs in packet switched networks. The test involves eight codecs and two levels of packet loss, 0% and 3%.

Condition	IP conditions (Packet loss)	Codec, Mode
1	0%	AMR-NB, 6.7kbit/s
2	0%	AMR-NB, 12.2kbit/s
3	0%	AMR-WB, 12.65kbit/s
4	0%	AMR-WB, 15.85kbit/s
5	0%	G. 723., 6.4 kbit/s
6	0%	G.729, 8kbit/s
7	0%	G.722, 64 kbit/s + plc
8	0%	G.711 + plc
9	3%	AMR-NB, 6.7kbit/s
10	3%	AMR-NB, 12.2 kbit/s
11	3%	AMR-WB, 12.65kbit/s
12	3%	AMR-WB, 15.85kbit/s
13	3%	G. 723.1, 6.4 kbit/s
14	3%	G.729, 8kbit/s
15	3%	G.722, 64 kbit/s + plc
16	3%	G.711 + plc

Table 5. Test Conditions in the PS Conversation Tests for ITU-T Codecs

5. Global Analyses

The purpose of the Global Analysis task is to bring together the results from the different Listening Labs/languages (Phase I - NAE, French, Japanese; Phase II – French, Arabic) and combine them, where appropriate, such that conclusions may be drawn about the performance of the AMR-NB and AMR-WB codecs in packet switched networks. This task is complicated by the fact that in the conversation tests multiple criterion measures are collected for each condition. In the tests involved here, listeners are required to rate each condition on five aspects of the communication situation:

- Quality of the voice of their partner
- Difficulty of understanding words
- Quality of interaction with their partner
- Degree of impairments
- Global communication quality

Each of these criteria is measured using ratings on five-category rating scales. Each criterion also represents a separate independent variable which must be evaluated in a Global Analysis. The appropriate analysis for this situation is a Multivariate Analysis of Variance (MANOVA). The first step in MANOVA involves an omnibus test for the combination of all independent variables. A number of statistical techniques may be employed in MANOVA to determine whether the independent variables are measuring different or the same underlying variable. Other techniques, discriminant analysis in particular, determine the contribution provided by each independent variable to a composite variable that maximally separates the data on the dependent variables. The omnibus MANOVA test is then followed by separate Analyses of Variance (ANOVA) for each independent variable. The F-ratios for the individual ANOVA's are adjusted (Bonferroni) to account for the fact that multiple tests are being performed. It is proposed here to perform MANOVA's and the associated univariate ANOVA's separately for each of the six experiments (AMR-NB and AMR-WB from each of the three listening labs). Examination of the results of these analyses will determine if there is a single composite independent variable for each experiment and whether these composites are similar across experiments and across listening labs. The results of these analyses will determine whether it is appropriate to combine the results across listening labs.

Pearson's correlation coefficients will be computed to identify and illustrate the inter-relationships among the dependent variables.

If the results can be legitimately combined across listening labs, a nested ANOVA for *Conditions* and *Listening Labs* will be conducted separately for each codec, AMR-NB and AMR-WB. Table 5 shows a generalized Source Table for the appropriate ANOVA with the effects of *Listening Labs* nested within the effects of *Subjects*.

One task of the Global Analysis exercise will be to provide an Excel spreadsheet to the individual Listening Labs for delivery of the raw ratings. The Global Analysis task will also include a comprehensive report containing the results of the various statistical analyses described above. Dynastat will present the final report at the February 2004 meeting of 3GPP-SA4.

Effect (Source of Variation)	**F Ratio**
Conditions	$MS_{Cond} / MS_{Cond\ x\ SwLL}$
Subjects	- - -
Listening Labs (LL)	MS_{LL} / MS_{SwLL}
Subjects within LL (SwLL)	- - -
Conditions x Subjects	- - -
Conditions x LL	$MS_{Cond\ x\ LL} / MS_{Cond\ x\ SwLL}$
Conditions x SwLL	- - -
Total	- - -

Table 6. Generalized ANOVA Source Table for Combining Results across Listening Labs.

6. References

S4-030564 Test Plan for the AMR Narrow-Band Packet Switched Conversation Test

S4-030565 Test Plan for the AMR Wide-Band Packet Switched Conversation Test

S4-030747 Test plan for Packet Switched Conversation Test. Comparison of quality offered by different speech coders.

Annex H: Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2004-06	SP-24	S4-040342			Version 6.0.0 approved at 3GPP TSG SA#24	2.0.0	6.0.0
2007-06	SP-36				Version for Release 7	6.0.0	7.0.0

History

Document history		
V7.0.0	June 2007	Publication