Technical Specification Group Services and System Aspects **TSGS#23(04)0063**

Meeting #23, Phoenix, USA, 15 - 18 March 2004

Source: **TSG-SA WG4**

Title: 3GPP TS 26.935 version 1.0.0 "Packet switched conversational multimedia applications; Default codecs; Performance characterization" (Release 6)

Agenda Item: **7.4.3**

Presentation of Specification to TSG SA Plenary

Presentation to: **TSG SA Meeting #23**

Document for presentation: **TR 26.935 "Packet switched conversational multimedia applications; Default codecs; Performance characterization", Version 1.0.0**

Presented for: **Information**

Abstract of document:

The present document contains the results of internationally co-ordinated conversational tests conducted in French, North-American, Japanese and Arabic language to characterise the performance of AMR-NB and AMR-WB codecs for packet switched conversational multimedia applications. An example of VoIP communication has been simulated complying with a potentially realistic scenario and fulfilling the state-of-the-art 3GPP specifications. Other speech codecs than AMR-NB and AMR-WB have been included as well in the Phase 2 of testing. The results from the conversational tests confirm that the speech codecs AMR-NB and AMR-WB operate well for packet switched conversational multimedia applications over various operating conditions. The performance results can be used as guidance for network planning regarding the QoS parameters for VoIP.

Changes since last presentation:

This is version 1.0.0, provided for information to TSG-SA#23 Plenary.

Outstanding Issues: none

The TS should be ready for approval at next TSA SA#24.

Contentious Issues:

None.

Comment(s):

This TR has been drafted after the very recent provision (at SA4#30 meeting) of the Phase 2 test results and of the global analysis of the whole exercise, and it has been approved by SA4 by correspondence, during the week immediately preceding SA#23 Plenary, to be provided "for information" to SA#23 Plenary. However, the document has been already commented over the reflector SA4, and changes have been requested to the actual text, in order to improve the accuracy of the document, before the final version is provided for approval at SA#24 Plenary. The comments/remarks already received concern the following items:

1) the Title of the TR has been requested to be changed (to replace the word "default" by the 3GPP codecs actually tested during the exercise, i.e. "AMR-NB and AMR-WB");

2) the Conclusions of the TR have been requested to be re-drafted, e.g. by removing actual values of delay and/or packet losses that, if not interpreted correctly, could contradict existing ITU-T Recommendations G.114 and G.109;

3) a clean-up work has been requested to take place to remove a number of editorial mistakes already noticed in the text, tables and/or figures of the TR (c/o Editor taking into account the available input from collaborators to this action).

3GPP TR 26.935 V1.0.0 (2004-03)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
Packet Switched Conversational Multimedia Applications;
Performance Characterisation of Default Codecs
(Release 6)

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

Keywords

<keyword[, keyword]>

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

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. Experimental test results from the speech quality testing are reported to illustrate the behaviour of these codecs. 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.

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 adjusting 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 Rec. P.800: “Methods for Subjective Determination of Transmission Quality”

[2] ITU-T Rec. P.831: “Subjective performance evaluation of network echo cancellers”

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

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

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

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

[7] IETF RFC 1889

[8] IETF RFC 3267

[9] TS 34.121 (downlink)

[10] TS 25.141 (uplink)

[11] S4-030564: Test Plan for the AMR Narrow-Band Packet Switched Conversation Test

[12] S4-030565: Test Plan for the AMR Wide-Band Packet Switched Conversation Test

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

[14] S4-030818: Proposed Test Plan for Global Analysis of PSS Conversation Tests

[15] 3GPP TR 25.853

[16] 3GPP TS 26.235

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
ASY	ASYmmetric conditions
BLER	Block Error Rate
COND	Test CONDitions
CN	Core Network
CRC	Cyclic Redundancy Check
DL	Downlink
DPCH	Dedicated Physical Channel
Ec/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
LAB	Listening LABoratory
MAC	Medium access control
MANOVA	Multivariate Analysis of Variance
MOS	Mean Opinion Score
NB	Narrowband
PC	PerCeption of impairments (also: Personal Computer)
PDCP	Packet Data Convergence Protocol
PDU	Protocol Data Unit
Pa	Soumd Pressure Level (in Pascal)
PL	Packet Loss
Plc	Packet Loss Concealment
RC	Radio Conditions
PS	Packet Switched
RB	Radio Bearer
RAB	Radio Access Bearer
RLC	Radio Link Control
RoHC	Robust Header Compression
RRM	Radio Resource Management
RTP	Real-time Transport Protocol
SYM	SYMmetric conditions
TB size	Transport Block size
TF	Transport Format
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

4 General Overview

In order to characterize the quality of default speech codecs for packet switched conversational multimedia (AMR-NB and AMR-WB codecs) [16], 3GPP conducted two series of listening tests. 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 and AMR-WB in various operating conditions. Phase 2 considered also several other codecs including ITU-T codecs G.723.1, G.729, G.722 and G.711.

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 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 included. 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. 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 µs) the use of a particular IP-version in CN part has no impact on the performance results.

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.

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 the figure below 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.

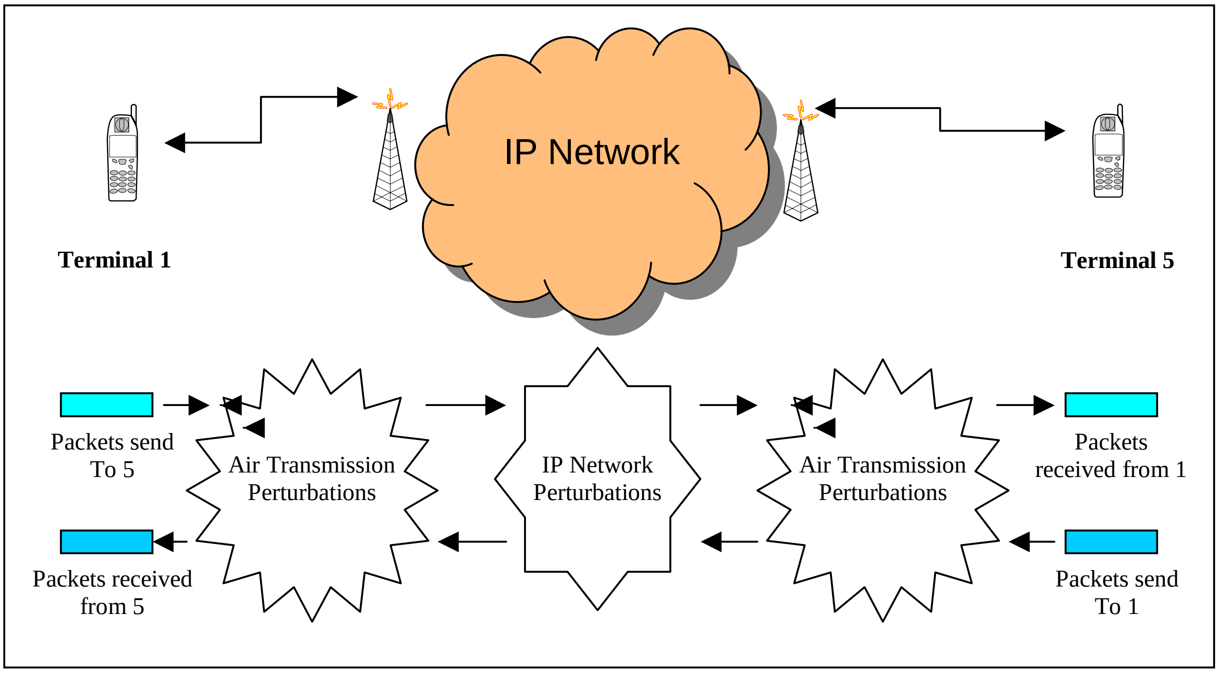


Figure 1: Packet switch audio communication simulator

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

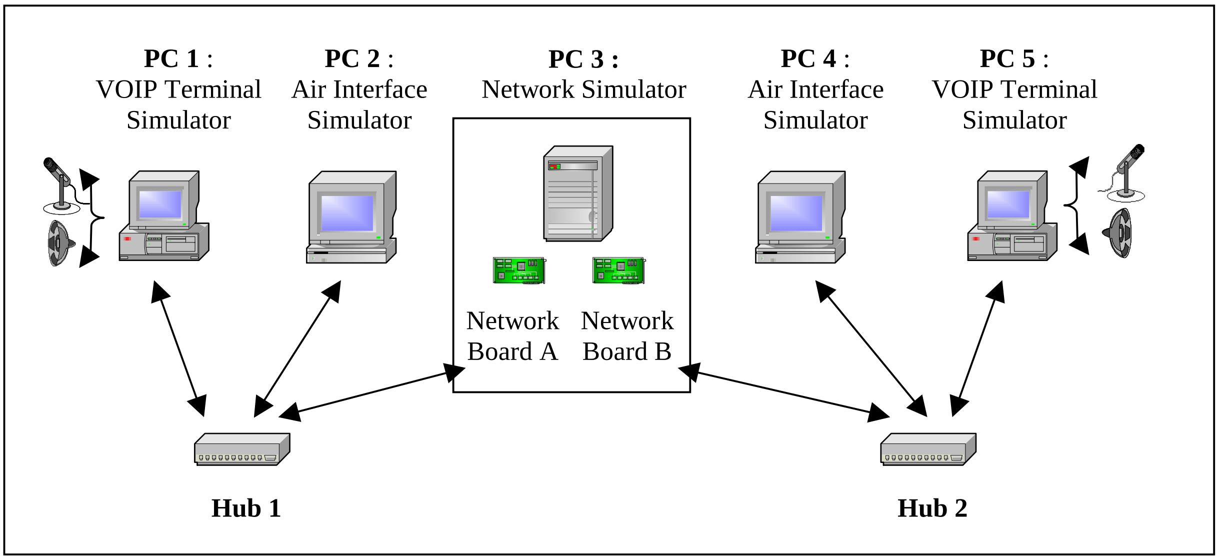


Figure 2: Simulation Platform

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 PC3 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 relatives 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 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 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.

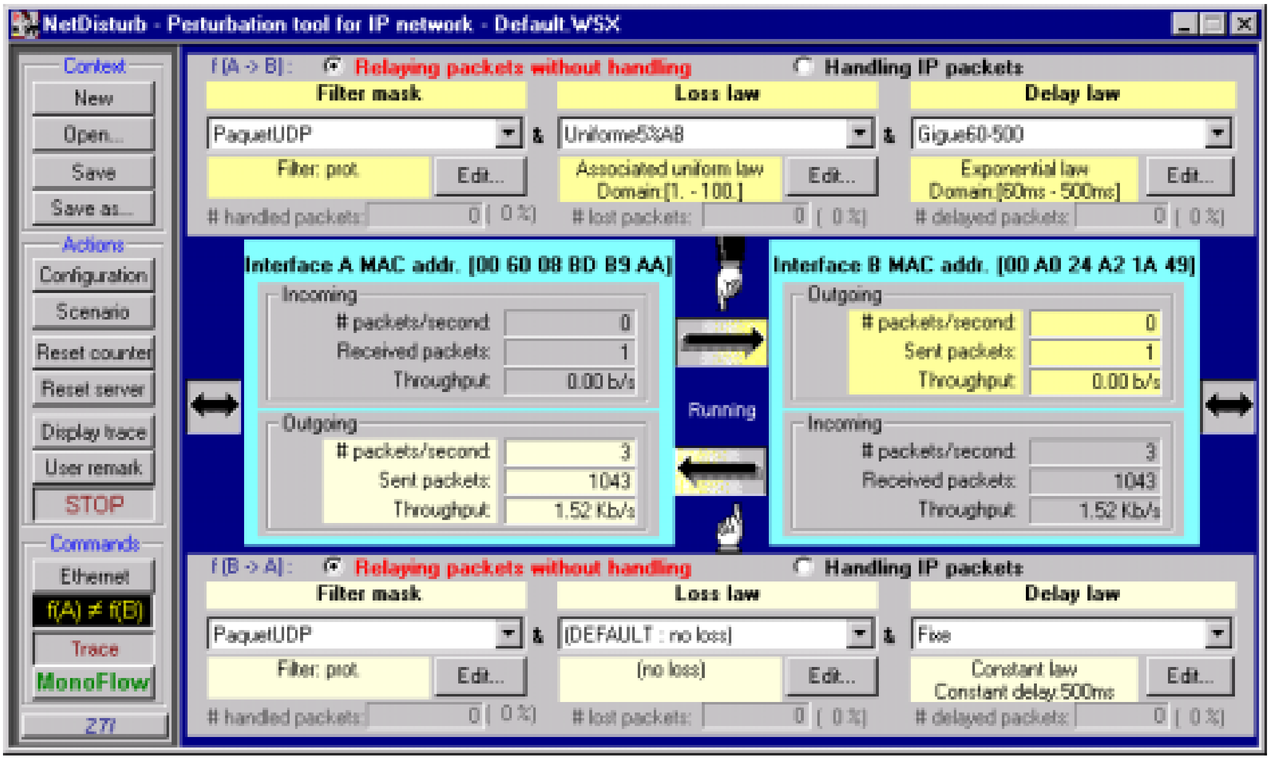


Figure 3: IP simulator interface

On both links, one can choose delay and loss laws. Both links can be treated separately or 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.

5.2.3 UMTS simulator choices

The transmission of IP/UDP/RTP/AMR packets over the UMTS air interface is simulated using the RAB described in Section 5.2.3.1. 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 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 Narrow Band conversational tests, the AMR was encoded with 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 was the following:

- RTP Payload Format for AMR-NB (cf. [8]) 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 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 were 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]) 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 was 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 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 is supported in the AMR-WB conversation test, for the 12.65 kbit/s mode and the 15.85 modes. Header compression will be done on the IP/UDP/RTP headers. ROHC will start in the unidirectional mode and switch to bi-directional 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:46 DL:46 kbps / PS RAB coming from TS 34.108 v4.7.0 was used. It is not an optimal RAB to do PS conversational test but it was the only one available at the time the test bed and the air interface simulator were designed. 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

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 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 are simulated in real time, but error pattern files are provided. The error patterns of the air interface transmission are simulated 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 (PC3) in case of uplink transmission, or it is one of the terminals (PC1 or PC5) in case of downlink transmission.

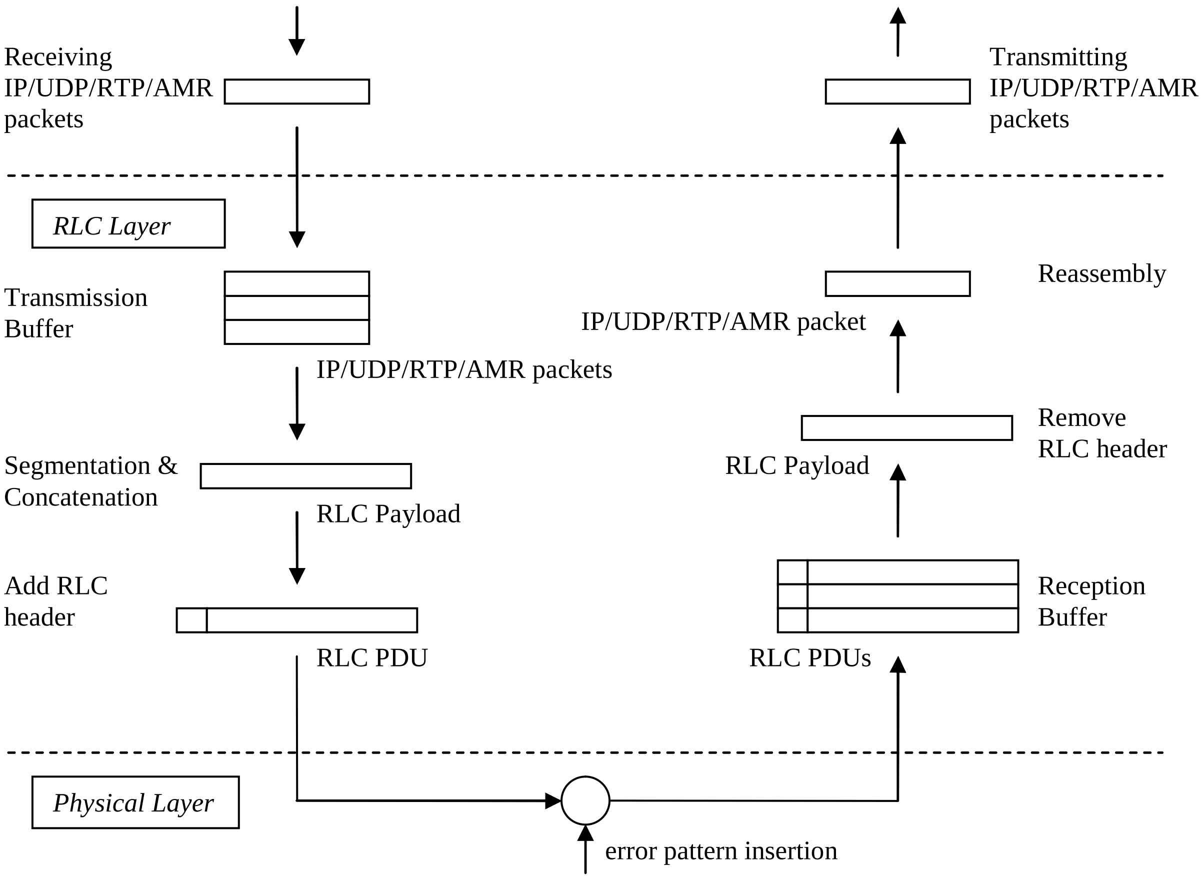


Figure 4: UMTS air interface simulation

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 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).

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

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

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

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 to place 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 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.

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 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.

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, 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 delay is calculated as shown:

- Encoder side: delay due to account framing, look-ahead, processing and packetization = 45ms
- Uplink delay between UE and Iu: 84.4 ms (see [15])
- 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 [15])
- 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 NB

The AMR 4.75 kb/s mode is intended to be used only temporarily to cope with poor radio conditions. It was expected to provide insufficient quality for conversational applications if used throughout the call. Therefore, a higher mode of 6.7 kbit/s was selected as the “low-rate mode” for these conversational tests (where mode adoption during call is not allowed).

Cond.	Background noise in Room A	Background noise in Room B	Experimental actors		
			Radio cond.	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)

Noise types

Noise type	Level (dB Pa)
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)

5.4 Test Conditions for WB

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)
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)

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 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:

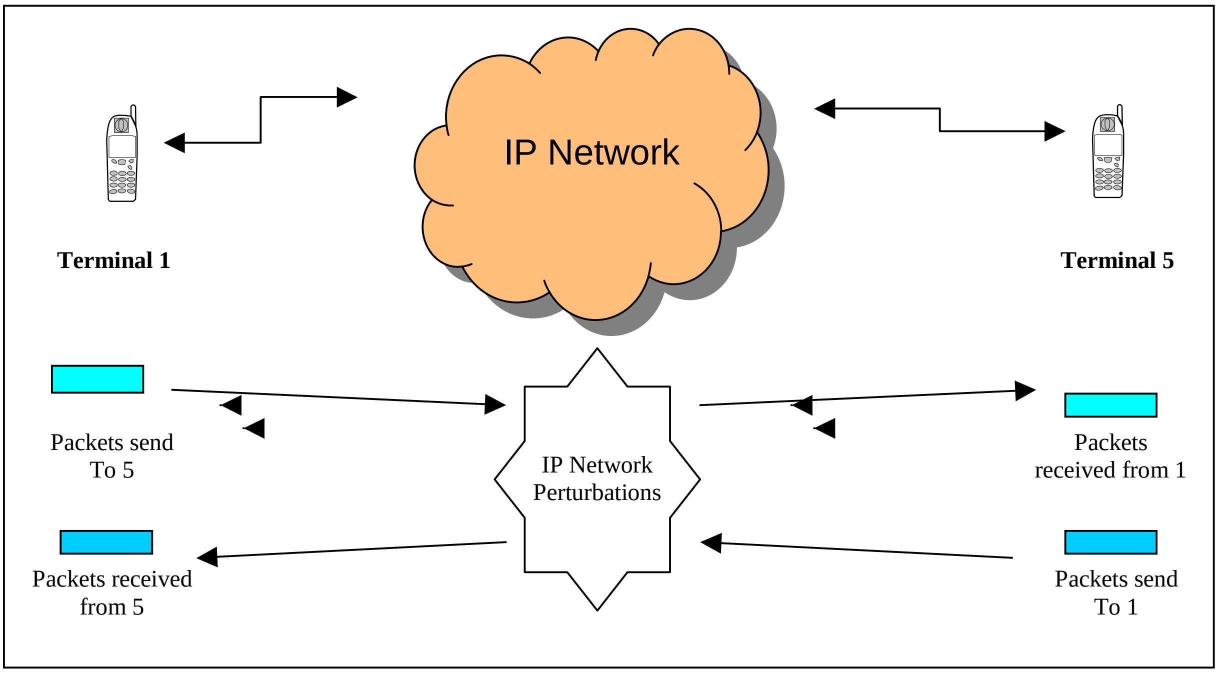


Figure 5: Packet switch audio communication simulator

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

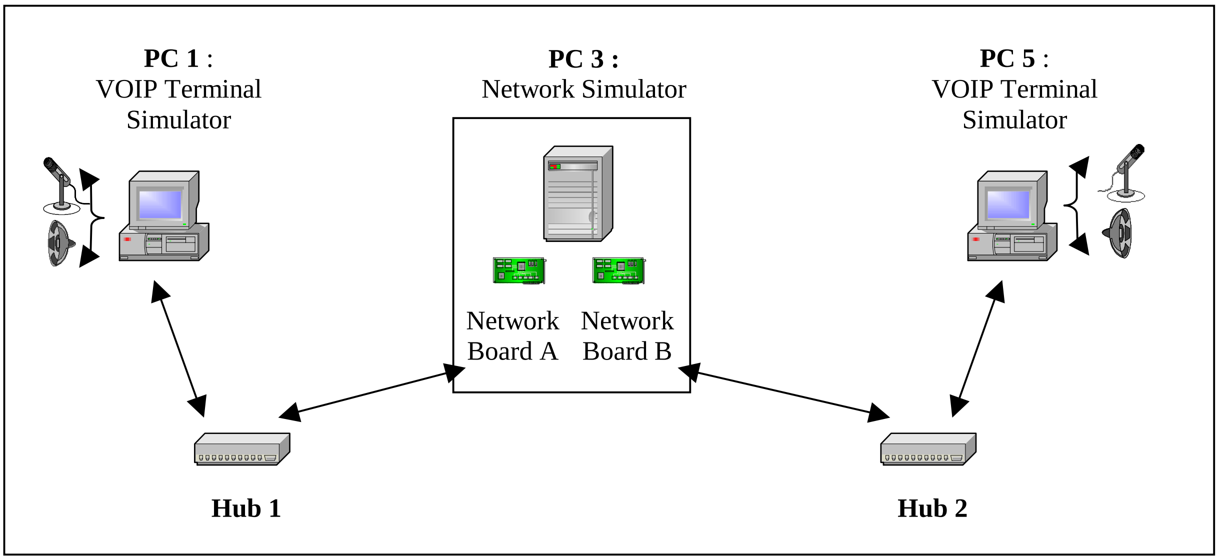


Figure 6: Simulation Platform

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 switch interactive communication between two users using PC 1 and PC 5 as their relatives VOIP terminals. PC 1 sends encoded packets that are encapsulated using IP/UDP/RTP headers to PC5. PC1 receives these IP/UDP/RTP audio packets from PC 5.

6.1.2 Network simulator

The core network simulator was used 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" will have two values, all the others settings being fixed.

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.

Headsets were here also used to reduce echo problems. 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.

The same test environment as in test Phase 1 was used. 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.

6.1.3 Calibration and test conditions monitoring

The speech level checking is done in the same way as previously.

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-NB and AMR-WB through the UMTS network.

6.2 Test Conditions

The test conditions are described in the 2 tables below.

Cond.	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 language
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),

7 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). Document [14] presents the GAL Test Plan for characterizing the results of the conversation tests. (Detailed test plans are given in [11] and [12] for Phase 1 and in [13] 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 described 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 6 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 6: 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 6 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 8 on the following page 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 7 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 7: 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 9 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 9 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 (MANOVA), i.e., the Pillai Trace and the associated F-statistic is highly significant in the MANOVA[1].

[1] For MANOVA, there is no single universally accepted procedure for hypothesis testing but rather a number of different methods. For the analyses that follow, we have chosen Pillai Trace and the associated F-statistic as the criterion for significance, primarily because of its robustness to violations of MANOVA assumptions.

Table 8: 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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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 9 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 9: 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:

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

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 10a 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 10b 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 10a: 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 10b: 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 10a and 10b. The results of those ANOVA's are shown in Tables 11 and 12, respectively.

Table 11: 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 11 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 11. 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*.

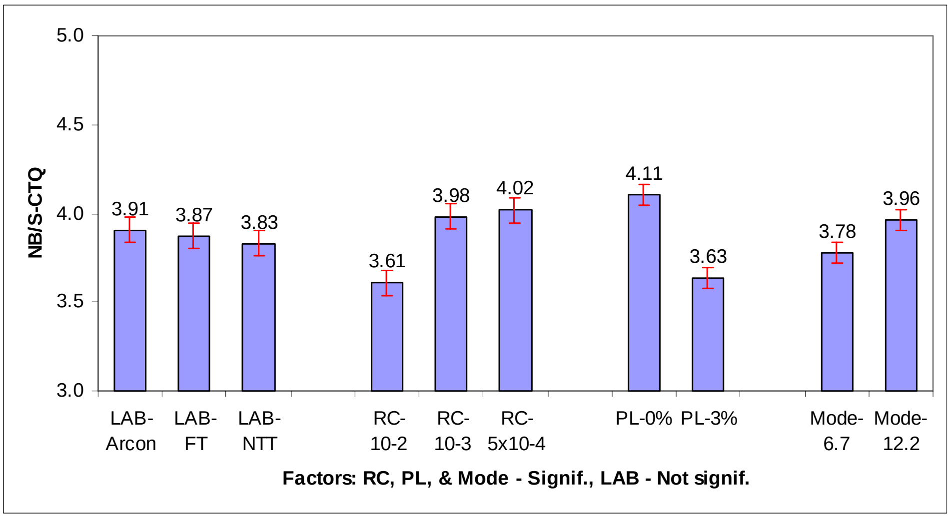


Figure 7: NB/S-CTQ Scores for the Effects of *LAB*, *Radio Conditions*, *Packet Loss*, and *Mode*

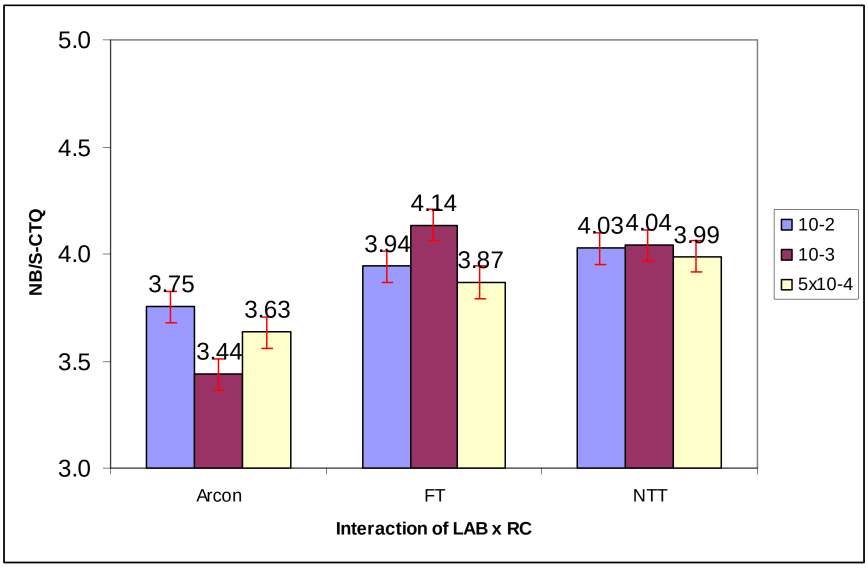


Figure 8: NB/S-CTQ Scores showing the Interaction of *LAB x Radio Conditions*

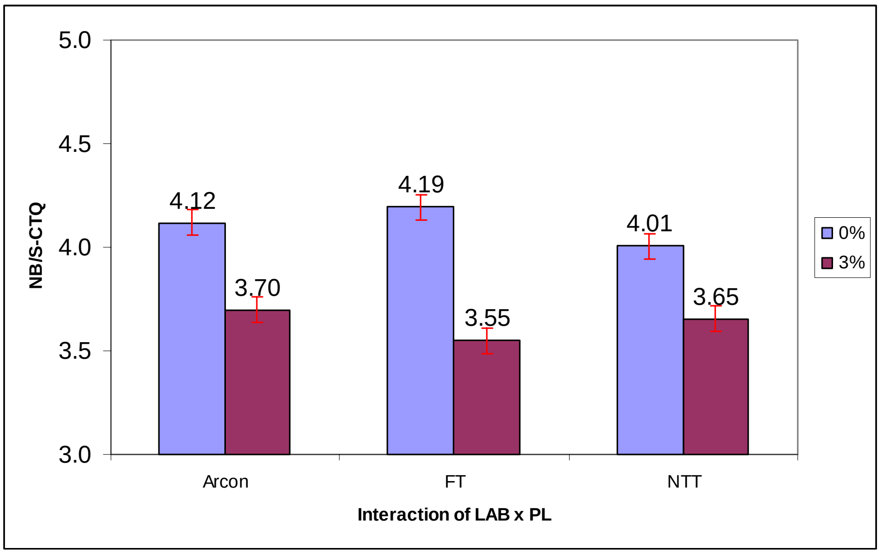


Figure 9: NB/S-CTQ Scores showing the Interaction of *LAB x Packet Loss*

Table 12: 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 12 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 12. 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.

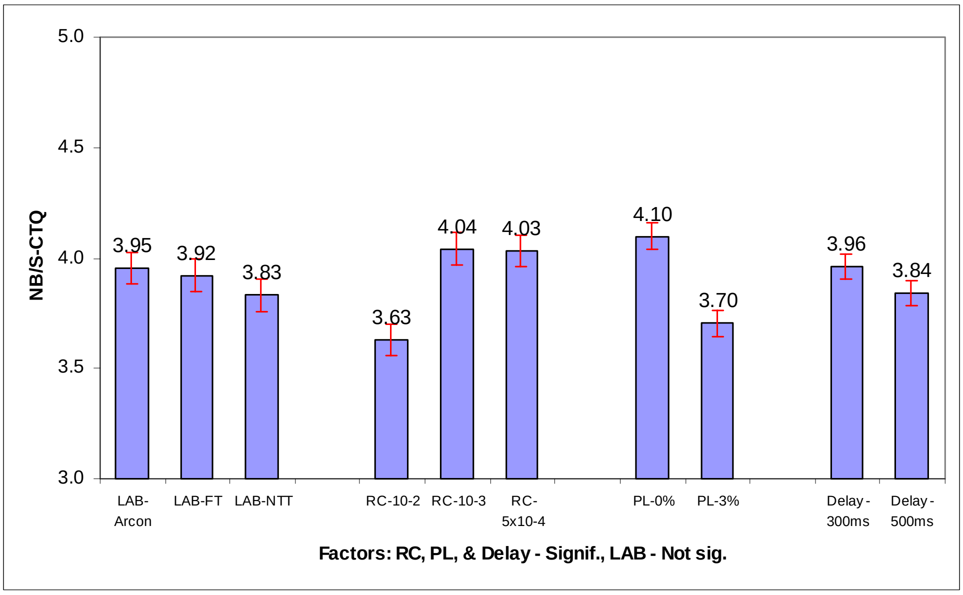
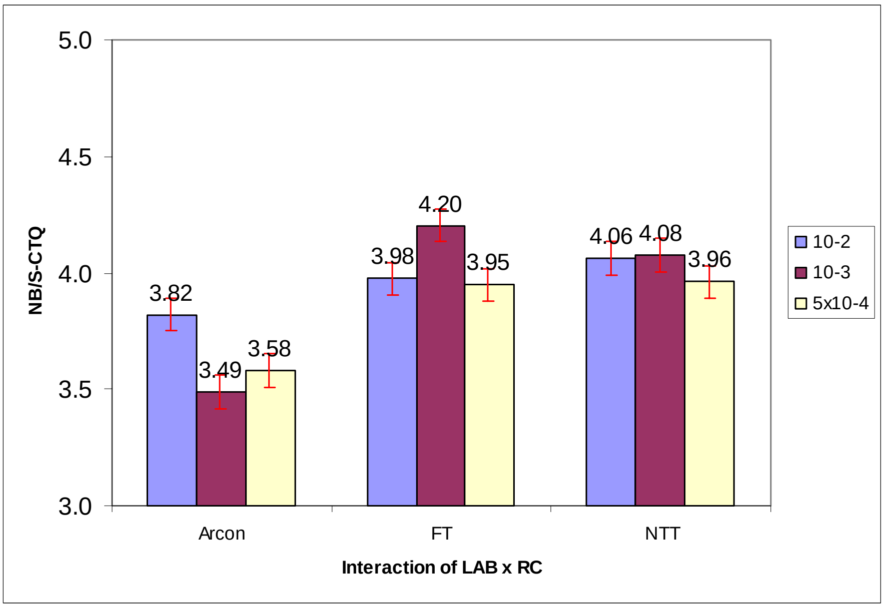


Figure 10: NB/S-CTQ Scores for the Effects of *LAB*, *Radio Conditions*, *Packet Loss*, and *Delay*

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



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

Table 3 shows the 6 test conditions involved in the NB 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 13 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 13: 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 14 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 the Appendix. 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 14: 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:

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

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 3, 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 15 shows the results of ANOVA for NB/A for the factors of *LAB*, *Noise*, and *Room*.

Table 15: 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 15.

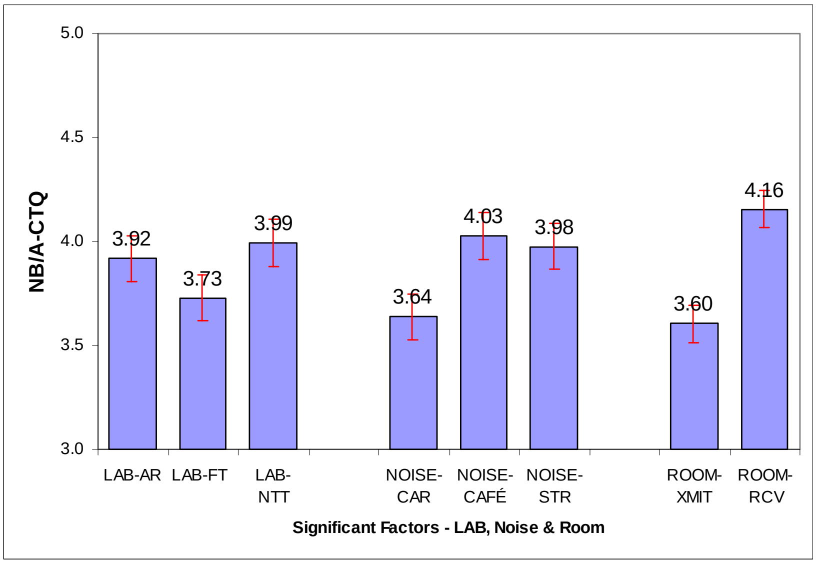


Figure 12: NB/A-CTQ Scores for the Effects of *LAB*, *Noise*, and *Room*

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

Table 17 on the next page shows the 18 test conditions involved in the AMR-WB conversation tests (condition 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 analyses is to examine the inter-correlation among the dependent variables for indications of underlying variables. Table 16 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 16: 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 18 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 18, 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 the Appendix. 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 17: Test Conditions and Mean Scores for each LAB for the Wideband Experiment

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

Wideband - 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	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

Table 18: 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

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

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 19a shows the allocation of the 12 conditions used to evaluate the effects of Radio Conditions, Packet Loss, and Mode – with RoHC held constant. Table 19b 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 19a: WB/SYM: Factorial Design for the Effects of Radio Cond., Packet Loss, and Mode

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 19b: WB/SYM: Factorial Design for the Effects of Radio Cond., Packet Loss, and Mode

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

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 19a and 19b. The results of the ANOVA's are shown in Tables 20 and 21, respectively.

Table 20: 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 20 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 20.

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

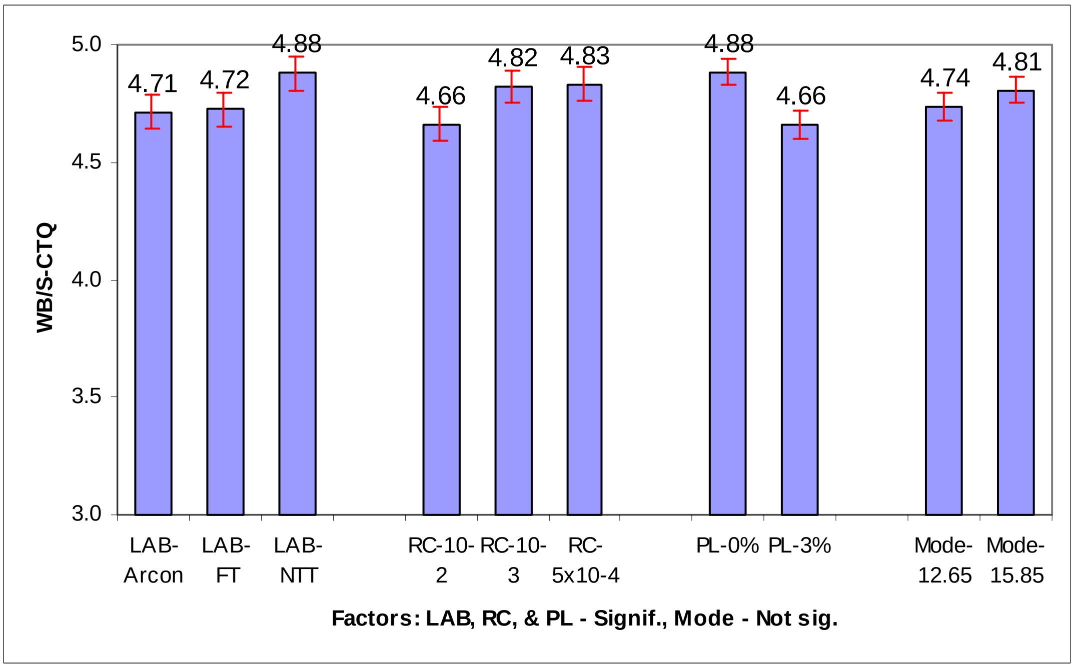


Table 21: Results of ANOVA of WB/S-CTQ for the Effects of *LAB*, *Radio Conditions* (RC), *Packet Loss* (PL), and *RoHC*

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 21 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 21.

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.

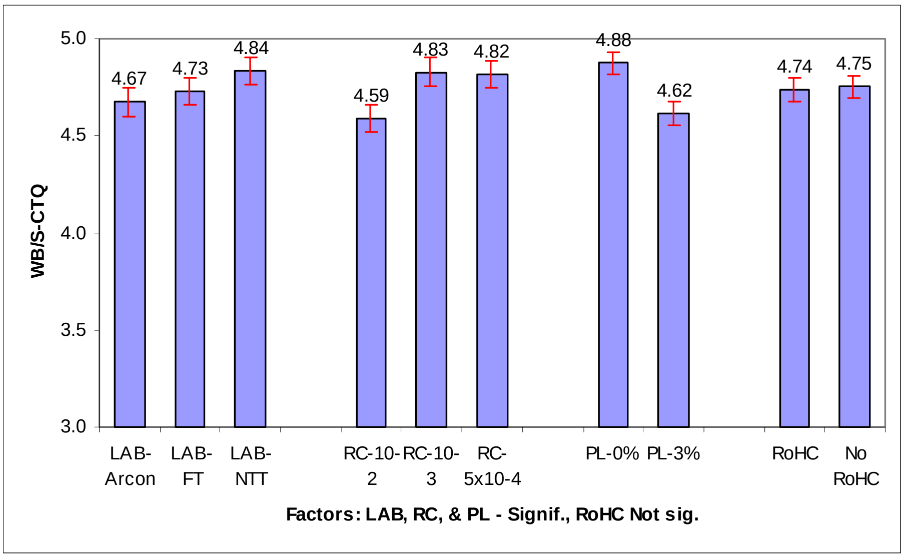


Figure 14: WB/S-CTQ Scores for the Effects of *LAB*, *Radio Conditions*, *Packet Loss*, and *RoHC*

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

Table 17 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 22 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 22: 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 23 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 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 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 in the Appendix. 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 23: 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.

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

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 17 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 24 shows the results of ANOVA for WB/A-CTQ for the factors of *LAB*, *Noise*, and *Room*.

Table 24: 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 24. 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*.

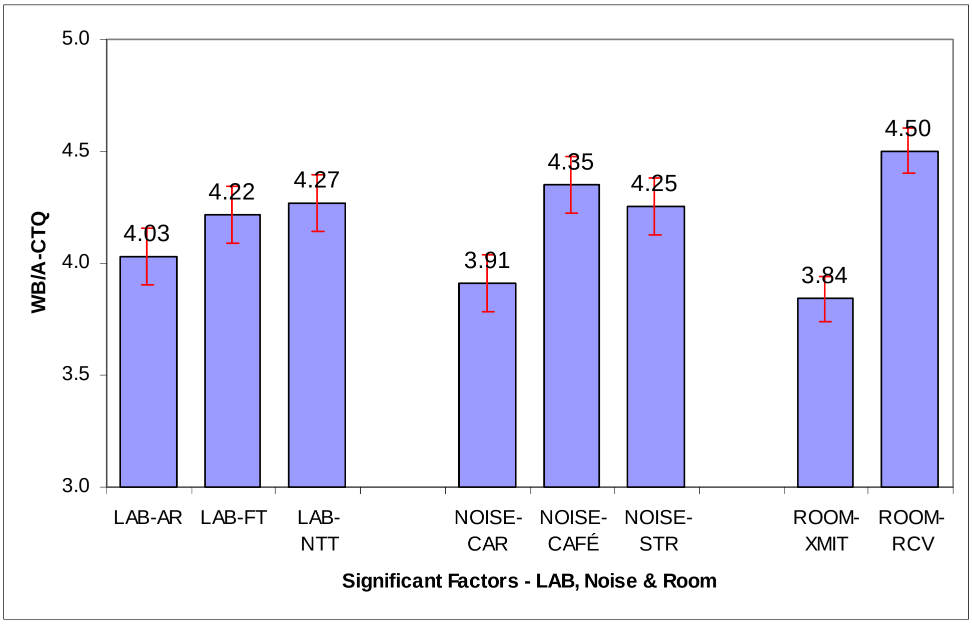


Figure 15: WB/A-CTQ Scores for the Effects of *LAB*, *Noise*, and *Room*

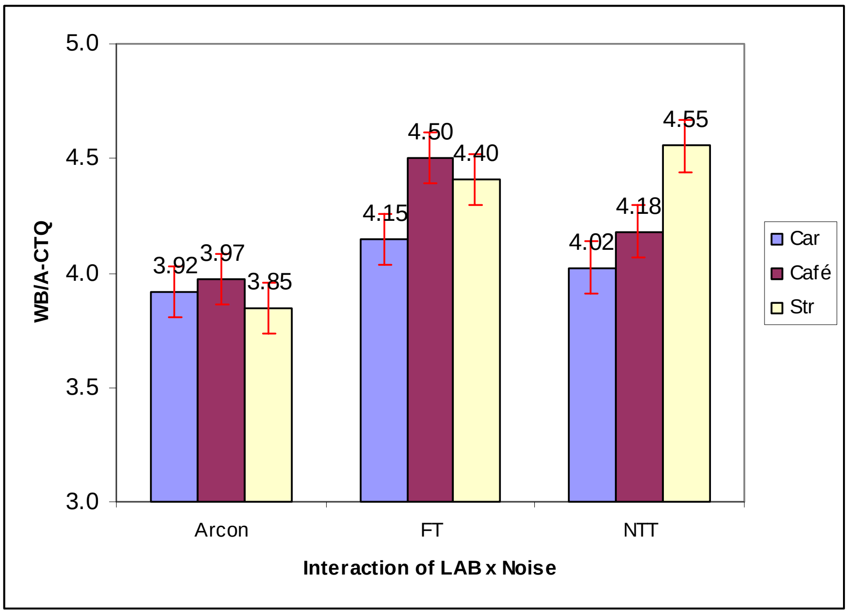


Figure 16: WB/A-CTQ Scores for the Interaction of *LAB x Noise*

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

Table 25 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 25: 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 26 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 26: 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 27 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 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 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 27: 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 28 shows the results of ANOVA for Ph2-CTQ for these factors.

Table 28: 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 28. Figure 18 illustrates the interaction of *Codec x Packet Loss*.

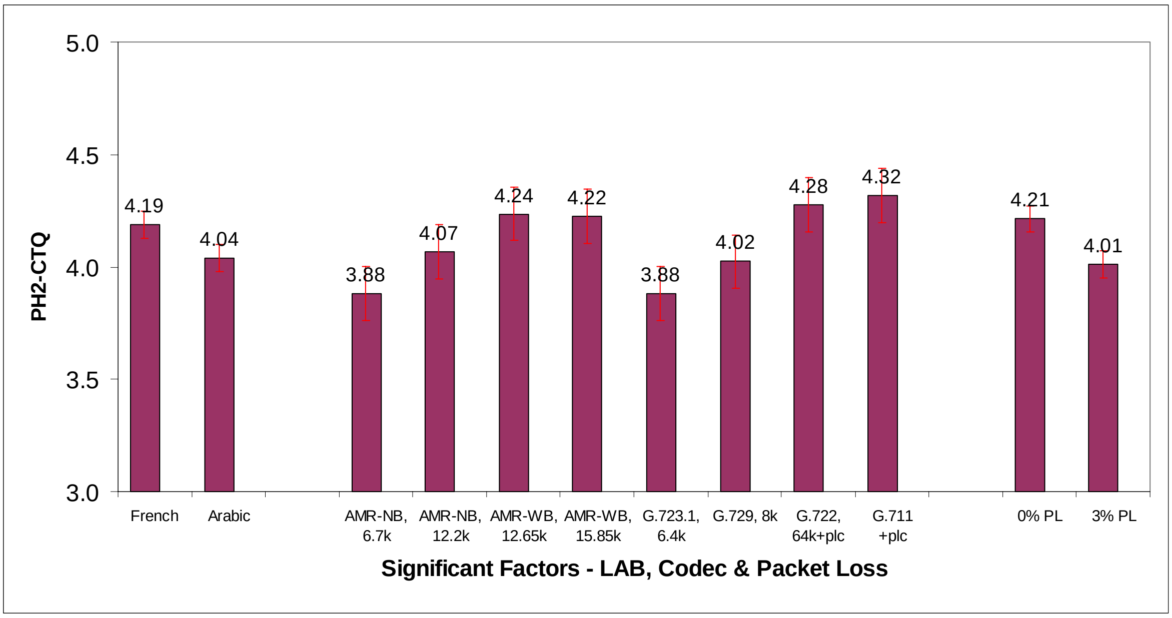


Figure 17: Ph2-CTQ Scores for the Effects of *LAB*, *Codec*, and *Packet Loss*

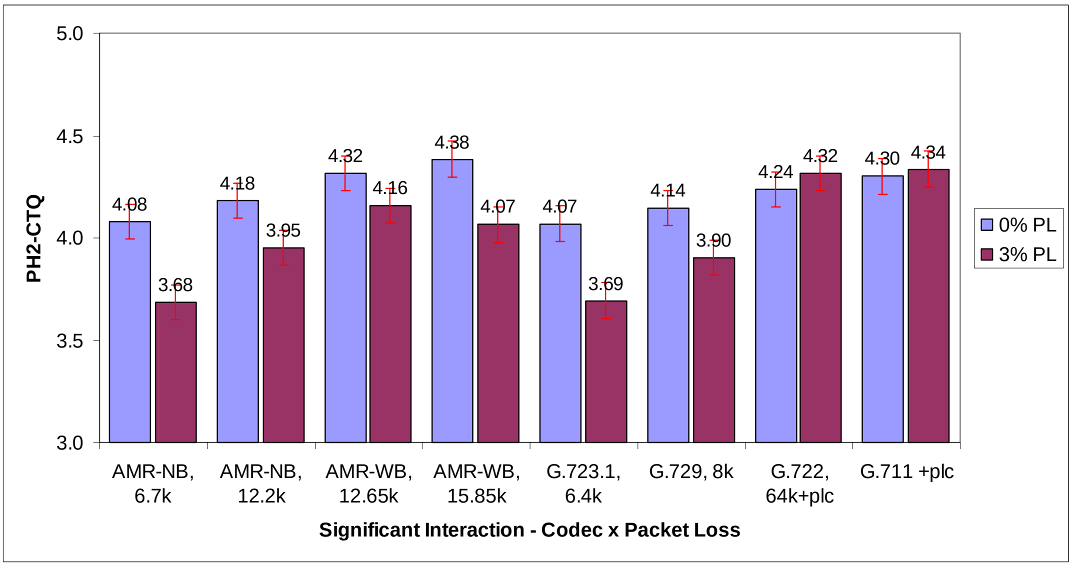


Figure 18: Ph2-CTQ Scores Showing the Interaction of Factors *Codec* and *Packet Loss*

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 operating conditions.

The quality is somewhat reduced when packet losses occur and the end-to-end delay is increased, but the overall quality still remains good even with 3% packet loss rate and 500 ms end-to-end transmission delay. The results also indicate that users have clear preference to wideband speech over narrowband.

The performance results can be used e.g. as guidance for network planning regarding the QoS parameters for VoIP (on end-to-end delay and target packet loss rates).

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

Phase 1 - 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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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	5×10^{-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

Phase 1 - Wideband - 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	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

Phase 2 Experiment			VQ Voice Quality		US Understand		IA Interaction		PC Perception		GQ Global Quality	
Cond	**PL**	**Codec, Mode**	**FR**	**AB**	**FR**	**AB**	**FR**	**AB**	**FR**	**AB**	**FR**	**AB**
1	0	AMR-NB, 6.7kbit/s	4.25	3.94	4.44	4.28	4.13	4.06	4.13	4.28	4.09	3.78
2	0	AMR-NB, 12.2kbit/s	4.41	4.13	4.56	4.38	4.28	3.91	4.16	4.09	4.06	3.78
3	0	AMR-WB, 12.65kbit/s	4.34	4.41	4.59	4.50	4.34	4.13	4.31	4.38	4.22	4.00
4	0	AMR-WB, 15.85kbit/s	4.50	4.44	4.72	4.47	4.34	4.13	4.22	4.25	4.31	4.00
5	0	G. 723., 6.4 kbit/s	4.22	4.09	4.47	4.09	4.06	3.88	4.03	4.06	3.94	3.69
6	0	G.729, 8kbit/s	4.16	4.34	4.47	4.28	3.97	3.94	4.19	4.06	3.88	3.81
7	0	G.722, 64 kbit/s + plc	4.41	4.28	4.63	4.44	4.19	4.09	4.25	4.03	4.13	3.72
8	0	G.711 + plc	4.41	4.44	4.56	4.44	4.13	4.03	4.13	4.31	4.09	3.88
9	3	AMR-NB, 6.7kbit/s	3.78	3.66	4.00	4.03	3.81	3.56	3.78	3.66	3.75	3.25
10	3	AMR-NB, 12.2 kbit/s	4.16	3.94	4.38	4.28	4.00	3.72	3.94	3.97	3.66	3.59
11	3	AMR-WB, 12.65kbit/s	4.38	4.09	4.38	4.34	4.06	3.84	4.09	4.00	4.06	3.81
12	3	AMR-WB, 15.85kbit/s	4.13	4.09	4.53	4.34	4.03	3.94	4.03	4.00	4.06	3.66
13	3	G. 723.1, 6.4 kbit/s	3.91	3.53	4.44	3.91	3.91	3.75	3.66	3.63	3.66	3.34
14	3	G.729, 8kbit/s	4.06	3.91	4.34	4.03	4.03	3.78	3.94	4.00	3.72	3.53
15	3	G.722, 64 kbit/s + plc	4.44	4.44	4.53	4.50	4.13	4.09	4.16	4.13	4.06	3.97
16	3	G.711 + plc	4.44	4.44	4.56	4.50	4.25	4.03	4.13	4.38	4.09	4.03

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	Bas

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 C: 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 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

Information that your partner requires	Flight number 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 colspan="4">Heathrow flight information</td></tr>
<tr><td rowspan="5">Information from which you should select the details which your partner requires</td><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>
<tr><td>Information you want to receive from your partner</td><td colspan="4">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 colspan="4"></td></tr>
</table>

Annex D:
Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2004-03	*23*	*S4-040063*			*Version 1.0.0 presented for information*		