Technical Specification Group Services and System Aspects
Meeting #22, Maui, Hawaii, USA, 15-18 December 2003

TSGS#22(03)0680
Agenda Item: 7.4.3

3GPP TSG-SA4 Meeting #29
Tampere, November 24 – 28, 2003

Tdoc S4-030832

Source: **TSG SA WG4** (Dynastat)

Title: **Test plan for 3G packet switched conversation tests: Global Analysis of Phase 1 & Phase 2 Conversation Test results**

Document for: **Approval**

Agenda Item: **7.4.3**

1. Introduction

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

2. Phase I - AMR-NB Tests

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

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

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

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

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

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

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

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

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

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

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

3. Phase I - AMR-WB Tests

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

- RoHC – present and absent
- AMR-WB mode (rate) – 12.65 kbps and 15.85 kbps
- Packet Loss – 0% and 3%
- Radio conditions – 10^{-2}, 10^{-3}, and $5x10^{-4}$

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

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

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

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

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

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

No Noise - 12.65 kbps			
RoHC / 0% PL		**RoHc / 3% PL**	
RC	Cond.#	RC	Cond.#
10^{-2}	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

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

4. Phase II - ITU-T Codec Tests

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

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

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

5. Global Analyses

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

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

Each of these criteria is measured using ratings on five-category rating scales. Each criterion also represents a separate independent variable which must be evaluated in a Global Analysis. The appropriate analysis for this situation is a Multivariate Analysis of Variance (MANOVA). The first step in MANOVA involves an omnibus test for the combination of all independent variables. A number of statistical techniques may be employed in MANOVA to determine whether the

independent variables are measuring different or the same underlying variable. Other techniques, discriminant analysis in particular, determine the contribution provided by each independent variable to a composite variable that maximally separates the data on the dependent variables. The omnibus MANOVA test is then followed by separate Analyses of Variance (ANOVA) for each independent variable. The F-ratios for the individual ANOVA's are adjusted (Bonferroni) to account for the fact that multiple tests are being performed. It is proposed here to perform MANOVA's and the associated univariate ANOVA's separately for each of the six experiments (AMR-NB and AMR-WB from each of the three listening labs). Examination of the results of these analyses will determine if there is a single composite independent variable for each experiment and whether these composites are similar across experiments and across listening labs. The results of these analyses will determine whether it is appropriate to combine the results across listening labs.

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

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

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

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

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

6. References[1]

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

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

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

[1] 1 contact: Alan Sharpley
sharpley@dynastat.com
Tel: +1 512-476-4797
Dynastat, Inc. - 2704 Rio Grande
Austin, Texas 78705 USA