ETSI SMG11 **Tdoc SMG11/NS 422/99**

Kyoto, Japan 6-10 December 1999

Title: **ETSI/AMR NS Selection Experiment 10 - COMSAT Results**

Source*: **COMSAT Laboratories**

Summary

This document presents a summary for ETSI Adaptive Multi Rate (AMR) Noise Selection (NS) Selection Phase Experiment 10 conducted by COMSAT in the Japanese language.

1. Introduction

A listening laboratory evaluation was performed in Japanese by COMSAT Laboratories in accordance with the AMR NS Selection Phase Experiment 10. This experiment was designed to evaluate the performance of the noise suppression algorithms for the AMR codec in the presence of a variety of background noise types and in the presence of speech input level variation and tandem. The test design is defined in Section 13 of the AMR NS Selection Subjective Test Plan [1]. COMSAT performed Experiment 10 using a subset of the Japanese speech material available in the NTT Speech Database. Twenty-four native speakers of the Japanese language performed as subjects in the test, which was nominally balanced for gender. The raw data collected was used to derive gender-wise and combined-gender MOS and standard deviation statistics. Additionally, a rank-order analysis was performed for the different impairments in the experiment.

2. Source Material

Seven sentence-triplets were selected for two male and two female talkers from the NTT Speech Database, for a total of 28 different source speech stimuli. Six sentence-triplets per talker were allocated for the main assessment sessions, and one sentence-triplet per talker was allocated for the practice session. All files had an exact duration of 12 seconds. The source material was provided to COMSAT, the designated Host Laboratory, which was responsible for all pre- and post-processing according to [2].

3. Experimental Design

The test design followed the specification in the AMR NS Selection Test Plan, and is summarized in Tables 1 and 2.

Contact: Simão F. Campos-Neto
COMSAT Labs
22300 COMSAT Drive
Clarksburg, MD 20871
USA

Tel: +1-301-428-4516
Fax: +1-301-428-9287
E-mail: simao.campos@comsat.com

Table 1:
Factors and Conditions for Experiment 10.

Main Codec Conditions	**#**	**Notes**
Noise Suppresser Candidates	6	
Codec	1	AMR
Codec Modes (FR/HR)	FR	12.2 kbit/s rate
BERs	0	
Input level	3	Nominal (-26dBov), +10dB, -10dB
Noise	4	Static Car, Dynamic Car, Music, Background Talkers
Tandeming	0	
Input Characteristic	1	GSM Filtered
Codec references	**#**	**Notes**
All Experiments	1	AMR without NS
Other references	**#**	**Notes**
Direct	1	Nominal Level, Static Car, GSM Filtered
MNRU	5	Nominal level, with background noise, GSM Filtered, Q=6, 12, 18, 24, 30dB
Common Conditions	**#**	**Notes**
GSM Channel	0	No channel model
Number of primary talkers	4	2 male + 2 female (plus one sentence-triplet for 4 additional talkers, for use in the interfering talker conditions)
Number of speech samples	28	6/ talker for the main test + 1/ talker for the Practice session
Listening Level	1	-15dBPa (79dB SPL) at ERP
Listeners	24	Naive Listeners
Randomizations	6	6 groups of 4 listeners
Rating Scale	1	Modified ACR Instructions
Replications	1	Original Presentation Only

4. Processed Material

The host laboratory provided a CDROM with 1168 processed speech files, which corresponds to the processing of (6+1) sentence-triplets per talker for four talkers through 48 test conditions. See [3] for details on the source speech processing.

5. Listening Sessions

5.1 Presentation Sequence Material

COMSAT used the grouping and randomization sequences specified in the AMR NS Selection Test Plan for Experiment 10.

Table 2:
Experiment 10: Influence of Input level and Tandeming in Clean Speech

Cond	Codec	Noise Type	SNR (dB)	Input Level
1	Direct	Static Car	6 dB	-26dBov
2	MNRU-30	StaticCar	6 dB	-26dBov
3	MNRU-24	Static Car	6 dB	-26dBov
4	MNRU-18	Static Car	6 dB	-26dBov
5	MNRU-12	Static Car	6 dB	-26dBov
6	MNRU-6	Static Car	6 dB	-26dBov
7	AMR@12.2	Static Car	6 dB	-16dBov
8	AMR@12.2	Static Car	6 dB	-26dBov
9	AMR@12.2	Static Car	6 dB	-36dBov
10	AMR@12.2	Dynamic Car	6 dB	-26dBov
11	AMR@12.2	Music	15 dB	-26dBov
12	AMR@12.2	IT	15 dB	-26dBov
13	AMR/NS1@12.2	Static Car	6 dB	-16dBov
14	AMR/NS2@12.2	Static Car	6 dB	-16dBov
15	AMR/NS3@12.2	Static Car	6 dB	-16dBov
16	AMR/NS4@12.2	Static Car	6 dB	-16dBov
17	AMR/NS5@12.2	Static Car	6 dB	-16dBov
18	AMR/NS6@12.2	Static Car	6 dB	-16dBov
19	AMR/NS1@12.2	Static Car	6 dB	-26dBov
20	AMR/NS2@12.2	Static Car	6 dB	-26dBov
21	AMR/NS3@12.2	Static Car	6 dB	-26dBov
22	AMR/NS4@12.2	Static Car	6 dB	-26dBov
23	AMR/NS5@12.2	Static Car	6 dB	-26dBov
24	AMR/NS6@12.2	Static Car	6 dB	-26dBov
25	AMR/NS1@12.2	Static Car	6 dB	-36dBov
26	AMR/NS2@12.2	Static Car	6 dB	-36dBov
27	AMR/NS3@12.2	Static Car	6 dB	-36dBov
28	AMR/NS4@12.2	Static Car	6 dB	-36dBov
29	AMR/NS5@12.2	Static Car	6 dB	-36dBov
30	AMR/NS6@12.2	Static Car	6 dB	-36dBov
31	AMR/NS1@12.2	Dynamic Car	6 dB	-26dBov
32	AMR/NS2@12.2	Dynamic Car	6 dB	-26dBov
33	AMR/NS3@12.2	Dynamic Car	6 dB	-26dBov
34	AMR/NS4@12.2	Dynamic Car	6 dB	-26dBov
35	AMR/NS5@12.2	Dynamic Car	6 dB	-26dBov
36	AMR/NS6@12.2	Dynamic Car	6 dB	-26dBov
37	AMR/NS1@12.2	Music	15 dB	-26dBov
38	AMR/NS2@12.2	Music	15 dB	-26dBov
39	AMR/NS3@12.2	Music	15 dB	-26dBov
40	AMR/NS4@12.2	Music	15 dB	-26dBov
41	AMR/NS5@12.2	Music	15 dB	-26dBov
42	AMR/NS6@12.2	Music	15 dB	-26dBov
43	AMR/NS1@12.2	IT	15 dB	-26dBov
44	AMR/NS2@12.2	IT	15 dB	-26dBov
45	AMR/NS3@12.2	IT	15 dB	-26dBov
46	AMR/NS4@12.2	IT	15 dB	-26dBov
47	AMR/NS5@12.2	IT	15 dB	-26dBov
48	AMR/NS6@12.2	IT	15 dB	-26dBov

5.2 Listeners

The subjective assessment was performed using 24 listeners (nominally balanced between male and female), divided into six groups of four listeners each.

The listener selection criteria were compliant with the AMR NS Selection Test Plan, noting that Audiometric testing was not performed on the listeners, for legal reasons. Test subjects were selected from an existing pool of native Japanese language listeners for which past assessment performance data exists indicating their general hearing integrity. Subject performance within this experiment was compared to the overall performance of all listeners used in each experiment as a check on the hearing integrity of each listener at the time of testing. The pre-test listener orientation used by COMSAT conformed to that specified in the Test Plan.

5.3 Audio Presentation

The processed speech material was presented to groups of listeners, seated at separate, visually screened listening stations contained within an acoustically conditioned sound room meeting the requirements for an NC 20 acoustic facility. Presentation was made monaurally using a telephone handset, driven by a distribution amplifier set to deliver monophonic speech to the listener's preferred listening ear at an active level of -15 dBPa (79 dB SPL), using a B&K 4153 Artificial Ear with circumaural headphone adapter, 4134 Microphone element and 2610 Measurement Amplifier.

The processed speech files were stored within the main facility computer and presented to the listeners under program control as 16 kHz samples through a 16-bit, D/A coupled to the input of the distribution amplifier through a Frequency Devices 9002 Eight-pole Elliptic Filter, set for a bandpass of 200-Hz to 3.4-kHz. Auxiliary filtering was performed to achieve an overall modified-IRS receive characteristic.

The listener responses were registered on auxiliary computers. One of these voting terminals is contained within each voting station. Voting was permitted following the completed presentation of each voting stimulus (in this experiment, each sentence-triplet). All seated listeners were required to register responses prior to the subsequent presentation of a new stimulus. Once a group of listeners was conditioned to the dynamics of the voting procedure, the voting response time for each presented stimulus was nominally three seconds for each presented stimulus.

5.4 Scoring

Within experiments using a Mean-Opinion-Score (MOS) method of assessment, the presented sentence-triplets were scored by the listeners using a five-point perceived quality scale as either Excellent, Good, Fair, Poor, or Bad. The quality designations were presented on the screen of the voting terminals and selected through the use of a pointing device. The voting screen was rendered neutral during the presentation of each new stimulus.

As all seated listeners completed their voting, the votes for all stations were transferred to the main facility computer prior to the presentation of subsequent new material. The votes of each group of listeners for each presentation set of speech material were stored as ASCII files within the main facility computer for subsequent analysis and presentation.

Upon completion of the listening sessions, all raw data were de-scrambled and consolidated into a single ASCII file, which was used for the statistical analysis.

6. Statistical Analysis

Table 3 presents the basic statistical analysis data produced by COMSAT for AMR NS Selection Experiment 10, similar to the data provided to the Global Analysis Laboratory. Each test condition received a total of 96 votes. In the table, *Condition* represents the test condition number, *Factor* is the circuit impairment, *MOS* is the Mean Opinion Score, *SD* is the standard deviation, *Se* is the standard error, and *+95%* and *-95%* represent the upper and lower 95% confidence interval, respectively..

Table 3

MOS, standard deviation, standard error, and 95% confidence intervals for COMSAT's Experiment 10

Cnd	Codec	Factor	N	MOS	SD	Se	+95%	-95%
1	Direct	StaCar/Nom/06dB	96	2.375	0.811	0.083	2.537	2.213
2	MNRU-30	StaCar/Nom/06dB/Q30	96	2.208	0.794	0.081	2.367	2.050
3	MNRU-24	StaCar/Nom/06dB/Q24	96	1.927	0.684	0.070	2.064	1.790
4	MNRU-18	StaCar/Nom/06dB/Q18	96	1.635	0.682	0.070	1.772	1.499
5	MNRU-12	StaCar/Nom/06dB/Q12	96	1.229	0.492	0.050	1.328	1.131
6	MNRU-06	StaCar/Nom/06dB/Q06	96	1.052	0.266	0.027	1.105	0.999
7	AMR@12.2	StaCar/Hi/06dB	96	2.438	0.779	0.079	2.593	2.282
8	AMR@12.2	StaCar/Nom/06dB	96	2.365	0.822	0.084	2.529	2.200
9	AMR@12.2	StaCar/Lo/06dB	96	2.354	0.754	0.077	2.505	2.203
10	AMR@12.2	DynCar/Nom/06dB	96	2.448	0.857	0.087	2.619	2.277
11	AMR@12.2	Mus/Nom/15dB	96	4.250	0.834	0.085	4.417	4.083
12	AMR@12.2	IT/Nom/15dB	96	3.073	1.225	0.125	3.318	2.828
13	AMR/NS1@12.2	StaCar/Hi/06dB	96	2.260	0.757	0.077	2.412	2.109
14	AMR/NS2@12.2	StaCar/Hi/06dB	96	2.865	0.720	0.073	3.009	2.721
15	AMR/NS3@12.2	StaCar/Hi/06dB	96	2.729	0.814	0.083	2.892	2.566
16	AMR/NS4@12.2	StaCar/Hi/06dB	96	2.844	0.670	0.068	2.978	2.710
17	AMR/NS5@12.2	StaCar/Hi/06dB	96	2.979	0.680	0.069	3.115	2.843
18	AMR/NS6@12.2	StaCar/Hi/06dB	96	2.875	0.743	0.076	3.024	2.726
19	AMR/NS1@12.2	StaCar/Nom/06dB	96	2.781	0.743	0.076	2.930	2.633
20	AMR/NS2@12.2	StaCar/Nom/06dB	96	2.969	0.852	0.087	3.139	2.798
21	AMR/NS3@12.2	StaCar/Nom/06dB	96	2.698	0.727	0.074	2.843	2.552
22	AMR/NS4@12.2	StaCar/Nom/06dB	96	2.896	0.801	0.082	3.056	2.736
23	AMR/NS5@12.2	StaCar/Nom/06dB	96	2.875	0.743	0.076	3.024	2.726
24	AMR/NS6@12.2	StaCar/Nom/06dB	96	2.813	0.772	0.079	2.967	2.658
25	AMR/NS1@12.2	StaCar/Lo/06dB	96	2.760	0.692	0.071	2.899	2.622
26	AMR/NS2@12.2	StaCar/Lo/06dB	96	2.885	0.819	0.084	3.049	2.722
27	AMR/NS3@12.2	StaCar/Lo/06dB	96	2.688	0.744	0.076	2.836	2.539
28	AMR/NS4@12.2	StaCar/Lo/06dB	96	2.833	0.735	0.075	2.980	2.686
29	AMR/NS5@12.2	StaCar/Lo/06dB	96	2.917	0.691	0.071	3.055	2.778
30	AMR/NS6@12.2	StaCar/Lo/06dB	96	2.906	0.769	0.078	3.060	2.752
31	AMR/NS1@12.2	DynCar/Nom/06dB	96	3.083	0.937	0.096	3.271	2.896
32	AMR/NS2@12.2	DynCar/Nom/06dB	96	3.063	0.892	0.091	3.241	2.884
33	AMR/NS3@12.2	DynCar/Nom/06dB	96	2.927	0.861	0.088	3.099	2.755
34	AMR/NS4@12.2	DynCar/Nom/06dB	96	3.000	0.871	0.089	3.174	2.826
35	AMR/NS5@12.2	DynCar/Nom/06dB	96	3.021	0.808	0.082	3.182	2.859
36	AMR/NS6@12.2	DynCar/Nom/06dB	96	3.135	0.841	0.086	3.304	2.967
37	AMR/NS1@12.2	Mus/Nom/15dB	96	4.094	0.907	0.093	4.275	3.912
38	AMR/NS2@12.2	Mus/Nom/15dB	96	4.104	0.900	0.092	4.284	3.924
39	AMR/NS3@12.2	Mus/Nom/15dB	96	4.198	0.878	0.090	4.374	4.022
40	AMR/NS4@12.2	Mus/Nom/15dB	96	4.208	0.928	0.095	4.394	4.023
41	AMR/NS5@12.2	Mus/Nom/15dB	96	4.219	0.908	0.093	4.400	4.037
42	AMR/NS6@12.2	Mus/Nom/15dB	96	4.198	0.947	0.097	4.387	4.008
43	AMR/NS1@12.2	IT/Nom/15dB	96	3.188	1.276	0.130	3.443	2.932
44	AMR/NS2@12.2	IT/Nom/15dB	96	3.167	1.303	0.133	3.427	2.906
45	AMR/NS3@12.2	IT/Nom/15dB	96	3.063	1.255	0.128	3.314	2.811
46	AMR/NS4@12.2	IT/Nom/15dB	96	3.115	1.305	0.133	3.376	2.854
47	AMR/NS5@12.2	IT/Nom/15dB	96	3.010	1.269	0.130	3.264	2.757
48	AMR/NS6@12.2	IT/Nom/15dB	96	3.198	1.211	0.124	3.440	2.956

Table 4

Rank-order presentation grouped by impairment for combined talkers in Experiment 10

Cnd	Codec	Factor	N	MOS	+95%	-95%	t	HSD	D
1	Direct	StaCar/Nom/06dB	96	2.375	2.537	2.213			
2	MNRU-30	StaCar/Nom/06dB/Q30	96	2.208	2.367	2.050			
3	MNRU-24	StaCar/Nom/06dB/Q24	96	1.927	2.064	1.790			N/A
4	MNRU-18	StaCar/Nom/06dB/Q18	96	1.635	1.772	1.499			
5	MNRU-12	StaCar/Nom/06dB/Q12	96	1.229	1.328	1.131			
6	MNRU-06	StaCar/Nom/06dB/Q06	96	1.052	1.105	0.999			
17	AMR/NS5@12.2	StaCar/Hi/06dB	96	2.979	3.115	2.843			>
18	AMR/NS6@12.2	StaCar/Hi/06dB	96	2.875	3.024	2.726			>
14	AMR/NS2@12.2	StaCar/Hi/06dB	96	2.865	3.009	2.721			>
16	AMR/NS4@12.2	StaCar/Hi/06dB	96	2.844	2.978	2.710			>
15	AMR/NS3@12.2	StaCar/Hi/06dB	96	2.729	2.892	2.566			>
7	AMR@12.2	StaCar/Hi/06dB	96	2.438	2.593	2.282			-
13	AMR/NS1@12.2	StaCar/Hi/06dB	96	2.260	2.412	2.109			=
20	AMR/NS2@12.2	StaCar/Nom/06dB	96	2.969	3.139	2.798			>
22	AMR/NS4@12.2	StaCar/Nom/06dB	96	2.896	3.056	2.736			>
23	AMR/NS5@12.2	StaCar/Nom/06dB	96	2.875	3.024	2.726			>
24	AMR/NS6@12.2	StaCar/Nom/06dB	96	2.813	2.967	2.658			>
19	AMR/NS1@12.2	StaCar/Nom/06dB	96	2.781	2.930	2.633			>
21	AMR/NS3@12.2	StaCar/Nom/06dB	96	2.698	2.843	2.552			>
8	AMR@12.2	StaCar/Nom/06dB	96	2.365	2.529	2.200			-
29	AMR/NS5@12.2	StaCar/Lo/06dB	96	2.917	3.055	2.778			>
30	AMR/NS6@12.2	StaCar/Lo/06dB	96	2.906	3.060	2.752			>
26	AMR/NS2@12.2	StaCar/Lo/06dB	96	2.885	3.049	2.722			>
28	AMR/NS4@12.2	StaCar/Lo/06dB	96	2.833	2.980	2.686			>
25	AMR/NS1@12.2	StaCar/Lo/06dB	96	2.760	2.899	2.622			>
27	AMR/NS3@12.2	StaCar/Lo/06dB	96	2.688	2.836	2.539			>
9	AMR@12.2	StaCar/Lo/06dB	96	2.354	2.505	2.203			-
36	AMR/NS6@12.2	DynCar/Nom/06dB	96	3.135	3.304	2.967			>
31	AMR/NS1@12.2	DynCar/Nom/06dB	96	3.083	3.271	2.896			>
32	AMR/NS2@12.2	DynCar/Nom/06dB	96	3.063	3.241	2.884			>
35	AMR/NS5@12.2	DynCar/Nom/06dB	96	3.021	3.182	2.859			>
34	AMR/NS4@12.2	DynCar/Nom/06dB	96	3.000	3.174	2.826			>
33	AMR/NS3@12.2	DynCar/Nom/06dB	96	2.927	3.099	2.755			>
10	AMR@12.2	DynCar/Nom/06dB	96	2.448	2.619	2.277			-
11	AMR@12.2	Mus/Nom/15dB	96	4.250	4.417	4.083			-
41	AMR/NS5@12.2	Mus/Nom/15dB	96	4.219	4.400	4.037			=
40	AMR/NS4@12.2	Mus/Nom/15dB	96	4.208	4.394	4.023			=
39	AMR/NS3@12.2	Mus/Nom/15dB	96	4.198	4.374	4.022			=
42	AMR/NS6@12.2	Mus/Nom/15dB	96	4.198	4.387	4.008			=
38	AMR/NS2@12.2	Mus/Nom/15dB	96	4.104	4.284	3.924			=
37	AMR/NS1@12.2	Mus/Nom/15dB	96	4.094	4.275	3.912			=
48	AMR/NS6@12.2	IT/Nom/15dB	96	3.198	3.440	2.956			-
43	AMR/NS1@12.2	IT/Nom/15dB	96	3.188	3.443	2.932			=
44	AMR/NS2@12.2	IT/Nom/15dB	96	3.167	3.427	2.906			=
46	AMR/NS4@12.2	IT/Nom/15dB	96	3.115	3.376	2.854			=
12	AMR@12.2	IT/Nom/15dB	96	3.073	3.318	2.828			=
45	AMR/NS3@12.2	IT/Nom/15dB	96	3.063	3.314	2.811			=
47	AMR/NS5@12.2	IT/Nom/15dB	96	3.010	3.264	2.757			=

Complementarily, Table 4 contains a rank-ordered presentation of the combined talker data in Table 3, grouped by impairment type. Upper and lower confidence intervals are also reported. Statistically equivalent test conditions are indicated using Student's t-test Least Significant Difference (LSD) criterion, Tukey-Kramer's Honestly Significant Difference (HSD) Criterion, and Dunnet's multiple-pair comparison to a control. In the table, *Condition* represents the test condition number, *Factor* is the circuit impairment, *MOS* is the Mean Opinion Score, *Se* is the standard error, and the *+95%* and *–95%* columns represent the upper and lower 95% confidence interval, respectively. The *LSD* column shows which test conditions can be considered equivalent under the LSD criterion (indicated by contiguous vertical lines within each test factor) for a given impairment. The *HSD* column indicates which test conditions can be considered equivalent by the HSD criterion (indicated by contiguous vertical lines within each test factor). The LSD criterion is used to compare a pair of conditions, while the HSD criterion is used to determine groups of equivalency within a set of samples. The last column, *D*, indicates whether, under the Dunnet test, the MOS for a test condition is significantly higher than (">"), equivalent to ("="), or significantly lower than ("<") the MOS values for the control condition (indicated by ""). For the analysis that follows, the HSD and Dunnet criteria were used.

It can be seen from Table 4 that for high input level and in the presence of static car noise, NS 2 through 6 had an equivalent performance (HSD), which was better that that for the AMR coder without noise suppression. Candidate NS 1 performed equivalently to AMR without NS for high level input. For nominal input level and in the presence of static car noise, all NS algorithms scored equivalently, but in general significantly better than AMR without noise suppression. For the HSD criterion, NS3 was equivalent to AMR without NS with nominal input level. For low level input and in the presence of static car noise, as well as in the presence of dynamic car noise, all NS algorithms again scored equivalently, but significantly better than AMR without noise suppression in all cases. In the presence of background music and of background interfering talker, all NS algorithms performed equivalently to the AMR codec without noise suppression.

A general observation of the MOS scores indicates a low overall average mean for the experiment, in particular of the MNRU conditions. This can be explained by cultural factors, in which lower subjective scores are generally observed for oriental languages, as well as by the fact that this was a noisy ACR experiment where some of the processed speech was subjected to noise suppression.

7. Conclusion

COMSAT performed AMR NS Selection Experiment 10 for the Japanese language in compliance with the test plan. This experiment verified the performance of the noise suppression algorithms for the AMR codec in the presence of a variety of background noise types and in the presence of speech input level variation and tandem. It was observed that there was a statistically significant improvement in the perceived quality for (static and dynamic at 6 dB SNR) car noise cases for most of the NS algorithms studied when compared to the performance of the AMR codec without noise suppression. In general, however, the NS algorithms performed equivalently, with the exception of NS1 for high input level and static car noise. When background music or interfering talkers (15 dB SNR) were present, no improvement in performance was observed, if compared to the AMR codec without NS.

References

[1] SMG11SQ, *"Test Plan for the AMR Specification for the AMR-NS Selection Phase"*; Tdoc SMG11 288/99.

[2] SMG11SQ, *"Processing Functions for the GSM AMR Noise Suppression Selection Tests"*; Tdoc SMG11 281R/99

[3] COMSAT Laboratories, "*Host Laboratory Processing for ETSI/AMR Noise Suppression Selection Tests"*, Tdoc SMG11 417/99