ETSI SMG11 **Tdoc SMG11/S4 421/99**

Kyoto, Japan 6-10 December 1999

Title: **ETSI/AMR NS Selection Experiments 4&5 - COMSAT Results**

Source*: **COMSAT Laboratories**

Summary

This document presents a summary for ETSI/AMR Noise Suppression (NS) Selection Phase Experiments 4&5 conducted by COMSAT in the Spanish language.

1. Introduction

COMSAT Laboratories performed listening assessments in Spanish for the AMR Noise Selection Experiments 4 and 5. These experiments were designed to evaluate the performance of the six candidate noise suppression algorithms in the presence of three different noise types within the same experiment (car, street, and babble noise) for different signal-to-noise ratios (SNR) and two fixed bit rates (5.9 and 12.2 kbit/s). The six noise suppression candidate algorithms are presented in this contribution using blinded designations NS1 through NS6. A total of four experiments were designed by ETSI, as follows:

- Experiment 4A: Background noise performance for AMR NS at 5.9 kbit/s with low SNR values
- Experiment 4B: Background noise performance for AMR NS at 5.9 kbit/s with high SNR values
- Experiment 5A: Background noise performance for AMR NS at 12.2 kbit/s with low SNR values
- Experiment 5B: Background noise performance for AMR NS at 12.2 kbit/s with high SNR values

As specified in the AMR test plan, the assessment method was a modified version of the Category Comparison Rating (CCR) method defined in [3], in which the reference signal, instead of being the unprocessed ("Direct") speech, was the speech processed through the AMR codec without noise suppression. This was done because the main parameter of interest in this experiment was the NS performance, not the coding quality per se. The only exceptions were the MNRU conditions, whose Reference Signal was the Direct speech. In this report, the reference signal is frequently denoted as stimulus "A", and the test signal is denoted as stimulus "B". In the data reported, a positive CMOS value indicates preference towards the test sample, and a negative value indicates preference towards the reference sample.

After the performance of the subjective sessions, COMSAT discovered that up to version 2.0 of the Test Plan, there was a flaw in the presentation sequences for Experiments 4 & 5. This flaw caused condition 7 (7 dB Ideal noise suppression for car noise) to be systematically paired with the incorrect reference (condition 38, instead of 39) when the B/A direction was played. The problem was discovered in time for the English-language listening laboratory (Nortel) to perform the experiments using corrected presentation sequences. However, half of the votes (96) had to be discarded for condition 7 from COMSAT's analysis in each of the four experiments in their first realization.

Due to time constraints, it was not possible for COMSAT to repeat all the experiments. However, the experiments that could be repeated used corrected presentation sequences, and contain the full 192 votes for condition 7. Additional measures were taken to compensate the CCR presentation direction bias for condition 7 in the experiments that could not be repeated.

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

2. Source Material

COMSAT performed Experiments 4&5 using a subset of the Spanish speech material available in the NTT Speech Database, which was pre-processed in compliance with the specifications in the AMR NS Selection Processing Test Plan [2]. Twenty-four distinct native speakers of the Spanish language performed as subjects for each of the four sub-experiments, which were nominally balanced for gender. In total, 96 subjects were used. The raw data collected was used to derive gender-wise and combined-gender Comparison Mean Opinion Scores (CMOS) and standard deviation statistics for each sub-experiment. Additionally, rank-order analyses were performed for the different impairments in each sub-experiment.

Six sentence-pairs were selected for two male and two female Spanish-speaking talkers from the NTT Speech Database for each of the four experiments. This database contains quiet background speech sampled at 16 kHz, for a bandwidth of 8 kHz. The speech material was selected such that each experiment used a totally distinct set of speech material in the main part of the test.

The quiet background sentence pairs were provided to the Host Laboratory, also performed by COMSAT. The Host Laboratory was responsible for the pre- and post-processing of the speech material, including addition of the background noise.

3. Experiment Design

The experiment designs for Experiments 4 and 5 are defined in Section 10 of the AMR NS Selection Subjective Test Plan [1], and are summarized in Tables 1 and 2. From the tables, it can be seen that the test designs of Experiments 4A, 4B, 5A, and 5B were identical, with the exception of the specific SNR and bit rate values.

4. Processed Material

The host laboratory provided a CDROM for each of the four experiments. For each experiment, a total of 1568 files were produced (382 MB), which corresponds to the blocked processing of the source sentence-pairs through the 32 test conditions and the 32 reference conditions, plus 32 practice session files. See [4] 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 Experiments 4&5. All four experiments used the same sequences. After the performance of the subjective sessions, COMSAT discovered that up to version 2.0 of the Test Plan, there was a flaw in the presentation sequences. This flaw caused condition 7 to be systematically paired with the incorrect reference (condition 38, instead of 39) when the B/A direction was played. The problem was corrected in version 2.1 of the test plan.

5.2 Listeners

Each of the four subjective assessments was performed using 24 listeners (nominally balanced between male and female subjects), divided into six groups of four listeners each. In total, 96 different fluent speakers of the Spanish language performed as test subjects.

The listener selection criterion was 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 Spanish 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.

Table 1:
Factors and Conditions for Experiments 4&5.

Main Codec Conditions	**#**	**Notes**
Noise Suppresser Candidates	6	
Codec	1	AMR
Codec Modes (HR/FR)	HR FR	5.9 kbit/s rate for Experiment 4 12.2 kbit/s rate for Experiment 5
BERs	0	Clear channel, no transmission errors
Input level	1	Nominal (-26dBov)
Acoustic Background Noise	3	Car, street, and babble noise
Background noise SNR	2	Low in 4A and 5A and high in 4B and 5B (see Table 2)
Input Characteristic	1	GSM transmit filtered
Codec references	**#**	**Notes**
All Experiments	1	the same AMR rate without NS
Other references	**#**	**Notes**
Direct		nominal level, GSM transmit filtered
MNRU		nominal level, GSM transmit filtered, Q= 12, ΔQ= 4
Ideal noise suppression simulation		4, 7, and 10 dB
Common Conditions	**#**	**Notes**
GSM Channel	0	No channel model
Number of talkers	4	2 male + 2 female primary talkers
Number of speech samples	28	7 Sentence-pairs/primary talker (6 for Test, 1 for Practice)
Listening Level	1	-15dBPa (79dB SPL) at Ear Reference Point
Listeners	24	Naive Listeners
Randomizations	6	6 groups of 4 listeners
Rating Scale	1	Comparison Category Rating
Replications	1	Original Presentation Only

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. The handset was 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). The calibration was made 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 converter. The D/A converter was coupled to the input of the distribution amplifier through a Frequency Devices 9002 Eight-pole Elliptic Filter, which was 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, two sentence-pairs). 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 four seconds for each presented stimulus.

Table 2:
Allocation of conditions for Experiments 4A, 4B, 5A, and 5B

Cond.	Noise	4A&5A SNR (dB)	4B&5B SNR (dB)	Exp.4 Bit rate (kbit/s)	Exp.5 Bit rate (kbit/s)	Reference	Processed Codec	Processed Ideal NS
1	Car	6	12	5.9	12.2	AMR	AMR	-
2	Street	9	15	5.9	12.2	AMR	AMR	-
3	Babble	9	15	5.9	12.2	AMR	AMR	-
4	Car	6	12	-	-	MNRU-16	MNRU-12	-
5	Car	6	12	-	-	Direct	MNRU-12	-
4’	Street	9	15	-	-	MNRU-16	MNRU-12	-
5’	Street	9	15	-	-	Direct	MNRU-12	-
4”	Babble	9	15	-	-	MNRU-16	MNRU-12	-
5”	Babble	9	15	-	-	Direct	MNRU-12	-
6	Car	6	12	5.9	12.2	AMR	AMR	4
7	Car	6	12	5.9	12.2	AMR	AMR	7
8	Car	6	12	5.9	12.2	AMR	AMR	10
9	Street	9	15	5.9	12.2	AMR	AMR	4
10	Street	9	15	5.9	12.2	AMR	AMR	7
11	Street	9	15	5.9	12.2	AMR	AMR	10
12	Babble	9	15	5.9	12.2	AMR	AMR	4
13	Babble	9	15	5.9	12.2	AMR	AMR	7
14	Babble	9	15	5.9	12.2	AMR	AMR	10
15	Car	6	12	5.9	12.2	AMR	AMR/NS1	-
16	Car	6	12	5.9	12.2	AMR	AMR/NS2	-
17	Car	6	12	5.9	12.2	AMR	AMR/NS3	-
18	Car	6	12	5.9	12.2	AMR	AMR/NS4	-
19	Car	6	12	5.9	12.2	AMR	AMR/NS5	-
20	Car	6	12	5.9	12.2	AMR	AMR/NS6	-
21	Street	9	15	5.9	12.2	AMR	AMR/NS1	-
22	Street	9	15	5.9	12.2	AMR	AMR/NS2	-
23	Street	9	15	5.9	12.2	AMR	AMR/NS3	-
24	Street	9	15	5.9	12.2	AMR	AMR/NS4	-
25	Street	9	15	5.9	12.2	AMR	AMR/NS5	-
26	Street	9	15	5.9	12.2	AMR	AMR/NS6	-
27	Babble	9	15	5.9	12.2	AMR	AMR/NS1	-
28	Babble	9	15	5.9	12.2	AMR	AMR/NS2	-
29	Babble	9	15	5.9	12.2	AMR	AMR/NS3	-
30	Babble	9	15	5.9	12.2	AMR	AMR/NS4	-
31	Babble	9	15	5.9	12.2	AMR	AMR/NS5	-
32	Babble	9	15	5.9	12.2	AMR	AMR/NS6	-

5.4 Scoring

Within experiments using the Comparison Mean-Opinion-Score (CMOS) method of assessment, the presented (referenced) sentence-pairs were scored by the listeners using a seven-point perceived quality comparison scale as: “Much better, Better, Slightly Better, About the same, Slightly Worse, Worse, and Much Worse”. The quality comparison 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

In Experiments 4&5, CMOS expresses the quality degradation of stimulus “B” relative to stimulus “A”. In this test, stimulus A was in general the codec-processed speech without noise suppression, and stimulus B was in general the codec-processed speech with noise suppression (see Table 2 for precise descriptions). Positive CMOS values indicate a preference of stimulus B over stimulus A; negative values indicate a preference of stimulus A over stimulus B. Hence, positive CMOS values indicates a preference towards the NS-enabled vocoders. It should be noted that conditions 1, 2 and 3 were null conditions in the four experiments, i.e. stimuli A and B were identical. A total of 192 votes were cast per test condition for male and female talkers combined, or 96 for the gender-wise statistics.

As mentioned before, after the realization of the subjective sessions, a flaw was discovered in the presentation sequences given in the selection test plan. As a result, 96 votes had to be discarded for condition 7, which ended up only having valid data in the A/B presentation direction. Paired presentation experiment designs such as CCR experiments have a presentation direction effect, which is normally compensated by having data presented in both the A/B and B/A direction. Since this was not available for condition 7, an estimate of the direction bias was derived from the subject scores involving the AMR codec (conditions 1..3 and 6..32), and this bias was subtracted from the CMOS value for condition 7 in each experiment. Table 3 presents the CMOS values for the experimental data including this correction for condition 7.

In an effort to provide SMG 11 and SA4 with more reliable data, although limited by the Kyoto meeting date, COMSAT rerun two of the four experiments, Experiments 5A and 5B. The choice of experiments to rerun was based on the overall trends observed for the ideal noise condition scores as well as the magnitude of bias correction necessary. The experiments with larger bias were rerun first. The supplemental data for Experiments 5A and 5B are presented in Table 4.

Tables 3 and 4 contain the basic statistical analysis data produced by COMSAT for AMR NS Selection Experiment 7. This is the data as provided to the Global Analysis Laboratory, except for the confidence interval columns. Each test condition received a total of 192 votes for combined talkers, or 96 votes per talker for gender-wise statistics. In the tables, *Cnd* represents the test condition number, *Noise* is the background noise type, *Factor* is the circuit impairment, *CMOS* is the Comparison Mean Opinion Score, and *SD* is the standard deviation. The upper and lower 95% confidence intervals are denoted as *+95%* and *-95%*, respectively, in the table.

Complementarily, Table 5 contains a rank-ordered presentation of the combined talker data in Table 3(a), 3(b), 4(a), and 4(b) for Experiments 4A, 4B, 5A, and 5B, respectively, grouped by impairment type. Upper and lower 95% 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 multiple pair comparison against a control (D). 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 the CMOS for a test condition is significantly higher than (“>”), equivalent to (“=”), or significantly lower than (“<”) the CMOS values for the control condition (indicated by “-”).

Table 3(a)

CMOS and standard deviation for COMSAT's Experiment 4A (first realization)

Cnd.	Noise/SNR/Rate	Reference	Processed Codec/Ideal NS	Votes	Y(all)	SD(all)	+95%	-95%
1	Car/6dB/5.9	AMR	AMR	192	0.016	0.727	0.118	-0.087
2	Street/9dB/5.9	AMR	AMR	192	0.057	0.657	0.150	-0.036
3	Babble/9dB/5.9	AMR	AMR	192	-0.016	0.802	0.098	-0.129
4	Car/6dB	MNRU-16	MNRU-12	64	-0.453	0.795	-0.258	-0.648
4'	Street/9dB	MNRU-16	MNRU-12	64	-0.500	1.321	-0.176	-0.824
4"	Babble/9dB	MNRU-16	MNRU-12	64	-0.766	1.192	-0.474	-1.058
5	Car/6dB	Direct	MNRU-12	64	-1.469	1.181	-1.179	-1.758
5'	Street/9dB	Direct	MNRU-12	64	-1.297	1.191	-1.005	-1.589
5"	Babble/9dB	Direct	MNRU-12	64	-1.438	1.651	-1.033	-1.842
6	Car/6dB/5.9	AMR	AMR/4 dB	192	0.240	0.755	0.346	0.133
7	Car/6dB/5.9	AMR	AMR/7 dB	**96**	**0.389**	0.829	0.554	0.223
8	Car/6dB/5.9	AMR	AMR/10 dB	192	0.870	1.033	1.016	0.724
9	Street/9dB/5.9	AMR	AMR/4 dB	192	0.203	0.841	0.322	0.084
10	Street/9dB/5.9	AMR	AMR/7 dB	192	0.422	1.056	0.571	0.273
11	Street/9dB/5.9	AMR	AMR/10 dB	192	0.776	1.022	0.921	0.632
12	Babble/9dB/5.9	AMR	AMR/4 dB	192	0.224	1.027	0.369	0.079
13	Babble/9dB/5.9	AMR	AMR/7 dB	192	0.344	1.119	0.502	0.185
14	Babble/9dB/5.9	AMR	AMR/10 dB	192	0.557	1.143	0.719	0.396
15	Car/6dB/5.9	AMR	AMR/NS1	192	0.500	1.068	0.651	0.349
16	Car/6dB/5.9	AMR	AMR/NS2	192	0.682	1.157	0.846	0.519
17	Car/6dB/5.9	AMR	AMR/NS3	192	0.401	0.944	0.535	0.268
18	Car/6dB/5.9	AMR	AMR/NS4	192	0.620	1.037	0.766	0.473
19	Car/6dB/5.9	AMR	AMR/NS5	192	0.552	0.953	0.687	0.417
20	Car/6dB/5.9	AMR	AMR/NS6	192	0.734	1.057	0.884	0.585
21	Street/9dB/5.9	AMR	AMR/NS1	192	0.583	1.085	0.737	0.430
22	Street/9dB/5.9	AMR	AMR/NS2	192	0.542	1.028	0.687	0.396
23	Street/9dB/5.9	AMR	AMR/NS3	192	0.542	1.053	0.691	0.393
24	Street/9dB/5.9	AMR	AMR/NS4	192	0.599	1.107	0.756	0.442
25	Street/9dB/5.9	AMR	AMR/NS5	192	0.354	1.033	0.500	0.208
26	Street/9dB/5.9	AMR	AMR/NS6	192	0.807	1.240	0.983	0.632
27	Babble/9dB/5.9	AMR	AMR/NS1	192	0.302	1.045	0.450	0.154
28	Babble/9dB/5.9	AMR	AMR/NS2	192	0.318	1.087	0.471	0.164
29	Babble/9dB/5.9	AMR	AMR/NS3	192	0.208	1.101	0.364	0.053
30	Babble/9dB/5.9	AMR	AMR/NS4	192	0.344	1.042	0.491	0.196
31	Babble/9dB/5.9	AMR	AMR/NS5	192	0.266	0.942	0.399	0.132
32	Babble/9dB/5.9	AMR	AMR/NS6	192	0.500	1.176	0.666	0.334

Notes:
- Cnd: Condition number; Noise: Background noise type; SNR is signal-to-noise ratio in dB; Rate is bit rate in kbit/s; NSx is the noise suppressor candidate x. Default parameters are nominal level, 1 transcoding.
- CMOS is Comparison Mean Opinion Score, SD is standard deviation, Se is Standard Error, +95% and -95% are respectively the upper and lower 95% confidence interval.
- The bias compensation for Condition 7 was 0.278 in Experiment 4A.

Table 3(b)

CMOS and standard deviation for COMSAT's Experiment 4B (first realization)

Cnd.	Noise/SNR/Rate	Reference	Processed Codec/Ideal NS	Votes	Y(all)	SD(all)	+95%	-95%
1	Car/12dB/5.9	AMR	AMR	192	0.005	0.506	0.077	-0.066
2	Street/15dB/5.9	AMR	AMR	192	0.068	0.623	0.156	-0.020
3	Babble/15dB/5.9	AMR	AMR	192	0.021	0.639	0.111	-0.070
4	Car/12dB	MNRU-16	MNRU-12	64	-0.359	1.029	-0.107	-0.612
4'	Street/15dB	MNRU-16	MNRU-12	64	-0.328	0.909	-0.105	-0.551
4"	Babble/15dB	MNRU-16	MNRU-12	64	-0.922	1.251	-0.615	-1.228
5	Car/12dB	Direct	MNRU-12	64	-1.703	1.353	-1.372	-2.035
5'	Street/15dB	Direct	MNRU-12	64	-1.266	1.043	-1.010	-1.521
5"	Babble/15dB	Direct	MNRU-12	64	-2.359	1.045	-2.103	-2.615
6	Car/12dB/5.9	AMR	AMR/4 dB	192	0.214	0.753	0.320	0.107
7	Car/12dB/5.9	AMR	AMR/7 dB	**96**	**0.517**	0.790	0.676	0.359
8	Car/12dB/5.9	AMR	AMR/10 dB	192	0.760	0.968	0.897	0.623
9	Street/15dB/5.9	AMR	AMR/4 dB	192	0.245	0.692	0.343	0.147
10	Street/15dB/5.9	AMR	AMR/7 dB	192	0.432	0.728	0.535	0.329
11	Street/15dB/5.9	AMR	AMR/10 dB	192	0.599	0.875	0.723	0.475
12	Babble/15dB/5.9	AMR	AMR/4 dB	192	0.146	0.799	0.259	0.033
13	Babble/15dB/5.9	AMR	AMR/7 dB	192	0.359	0.773	0.469	0.250
14	Babble/15dB/5.9	AMR	AMR/10 dB	192	0.406	0.820	0.522	0.290
15	Car/12dB/5.9	AMR	AMR/NS1	192	0.609	0.965	0.746	0.473
16	Car/12dB/5.9	AMR	AMR/NS2	192	0.667	0.962	0.803	0.531
17	Car/12dB/5.9	AMR	AMR/NS3	192	0.495	0.949	0.629	0.361
18	Car/12dB/5.9	AMR	AMR/NS4	192	0.635	1.094	0.790	0.481
19	Car/12dB/5.9	AMR	AMR/NS5	192	0.583	0.999	0.725	0.442
20	Car/12dB/5.9	AMR	AMR/NS6	192	0.813	1.062	0.963	0.662
21	Street/15dB/5.9	AMR	AMR/NS1	192	0.448	0.861	0.570	0.326
22	Street/15dB/5.9	AMR	AMR/NS2	192	0.448	0.791	0.560	0.336
23	Street/15dB/5.9	AMR	AMR/NS3	192	0.370	0.871	0.493	0.247
24	Street/15dB/5.9	AMR	AMR/NS4	192	0.646	0.904	0.774	0.518
25	Street/15dB/5.9	AMR	AMR/NS5	192	0.438	0.860	0.559	0.316
26	Street/15dB/5.9	AMR	AMR/NS6	192	0.677	1.049	0.825	0.529
27	Babble/15dB/5.9	AMR	AMR/NS1	192	0.318	0.758	0.425	0.211
28	Babble/15dB/5.9	AMR	AMR/NS2	192	0.193	0.837	0.311	0.074
29	Babble/15dB/5.9	AMR	AMR/NS3	192	0.099	0.783	0.210	-0.012
30	Babble/15dB/5.9	AMR	AMR/NS4	192	0.260	0.834	0.378	0.142
31	Babble/15dB/5.9	AMR	AMR/NS5	192	0.115	0.791	0.226	0.003
32	Babble/15dB/5.9	AMR	AMR/NS6	192	0.417	0.870	0.540	0.294

Legend: - Cnd: Condition number; Noise: Background noise type; SNR is signal-to-noise ratio in dB; Rate is bit rate in kbit/s; NSx is the noise suppressor candidate x. Default parameters are nominal level, 1 transcoding.

- CMOS is Comparison Mean Opinion Score, SD is standard deviation, Se is Standard Error, +95% and -95% are respectively the upper and lower 95% confidence interval.

- The bias compensation for Condition 7 was 0.149 in Experiment 4B.

Table 3(c)
CMOS and standard deviation for COMSAT's Experiment 5A (first realization)

Cnd.	Noise/SNR/Rate	Reference	Processed Codec/Ideal NS	Votes	Y(all)	SD(all)	+95%	-95%
1	Car/6dB/12.2	AMR	AMR	192	0.031	0.909	0.160	-0.097
2	Street/9dB/12.2	AMR	AMR	192	0.031	0.943	0.165	-0.102
3	Babble/9dB/12.2	AMR	AMR	192	-0.010	1.033	0.136	-0.157
4	Car/6dB	MNRU-16	MNRU-12	64	-0.688	1.562	-0.305	-1.070
4'	Street/9dB	MNRU-16	MNRU-12	64	-0.672	1.169	-0.385	-0.958
4"	Babble/9dB	MNRU-16	MNRU-12	64	-0.656	1.116	-0.383	-0.930
5	Car/6dB	Direct	MNRU-12	64	-2.109	1.010	-1.862	-2.357
5'	Street/9dB	Direct	MNRU-12	64	-1.453	1.332	-1.127	-1.780
5"	Babble/9dB	Direct	MNRU-12	64	-1.688	1.531	-1.312	-2.063
6	Car/6dB/12.2	AMR	AMR/4 dB	192	0.313	1.124	0.471	0.154
7	Car/6dB/12.2	AMR	AMR/7 dB	**96**	**0.612**	1.088	0.829	0.394
8	Car/6dB/12.2	AMR	AMR/10 dB	192	1.031	1.058	1.181	0.882
9	Street/9dB/12.2	AMR	AMR/4 dB	192	0.333	1.035	0.480	0.187
10	Street/9dB/12.2	AMR	AMR/7 dB	192	0.438	1.138	0.598	0.277
11	Street/9dB/12.2	AMR	AMR/10 dB	192	0.885	1.166	1.050	0.721
12	Babble/9dB/12.2	AMR	AMR/4 dB	192	0.120	1.140	0.281	-0.041
13	Babble/9dB/12.2	AMR	AMR/7 dB	192	0.438	1.110	0.594	0.281
14	Babble/9dB/12.2	AMR	AMR/10 dB	192	0.510	1.193	0.679	0.342
15	Car/6dB/12.2	AMR	AMR/NS1	192	0.651	1.223	0.824	0.478
16	Car/6dB/12.2	AMR	AMR/NS2	192	0.828	1.110	0.985	0.671
17	Car/6dB/12.2	AMR	AMR/NS3	192	0.646	1.171	0.811	0.480
18	Car/6dB/12.2	AMR	AMR/NS4	192	0.854	1.167	1.019	0.689
19	Car/6dB/12.2	AMR	AMR/NS5	192	0.510	1.219	0.683	0.338
20	Car/6dB/12.2	AMR	AMR/NS6	192	0.859	1.252	1.036	0.682
21	Street/9dB/12.2	AMR	AMR/NS1	192	0.734	1.201	0.904	0.565
22	Street/9dB/12.2	AMR	AMR/NS2	192	0.745	1.208	0.916	0.574
23	Street/9dB/12.2	AMR	AMR/NS3	192	0.521	1.198	0.690	0.351
24	Street/9dB/12.2	AMR	AMR/NS4	192	0.495	1.290	0.677	0.312
25	Street/9dB/12.2	AMR	AMR/NS5	192	0.458	1.161	0.623	0.294
26	Street/9dB/12.2	AMR	AMR/NS6	192	0.870	1.236	1.045	0.695
27	Babble/9dB/12.2	AMR	AMR/NS1	192	0.396	1.002	0.538	0.254
28	Babble/9dB/12.2	AMR	AMR/NS2	192	0.396	1.053	0.545	0.247
29	Babble/9dB/12.2	AMR	AMR/NS3	192	0.172	1.022	0.316	0.027
30	Babble/9dB/12.2	AMR	AMR/NS4	192	0.406	1.250	0.583	0.229
31	Babble/9dB/12.2	AMR	AMR/NS5	192	0.135	1.079	0.288	-0.017
32	Babble/9dB/12.2	AMR	AMR/NS6	192	0.536	1.166	0.701	0.372

Legend: - Cnd: Condition number; Noise: Background noise type; SNR is signal-to-noise ratio in dB; Rate is bit rate in kbit/s; NSx is the noise suppressor candidate x. Default parameters are nominal level, 1 transcoding.
- CMOS is Comparison Mean Opinion Score, SD is standard deviation, Se is Standard Error, +95% and -95% are respectively the upper and lower 95% confidence interval.
- The bias compensation for Condition 7 was 0.315 in Experiment 5A.

Table 3(d)

CMOS and standard deviation for COMSAT's Experiment 5B (first realization)

Cnd.	Noise/SNR/Rate	Reference	Processed Codec/Ideal NS	Votes	Y(all)	SD(all)	+95%	-95%
1	Car/12dB/12.2	AMR	AMR	192	0.047	0.820	0.163	-0.069
2	Street/15dB/12.2	AMR	AMR	192	-0.031	0.744	0.074	-0.137
3	Babble/15dB/12.2	AMR	AMR	192	-0.042	0.792	0.070	-0.154
4	Car/12dB	MNRU-16	MNRU-12	64	-0.422	1.020	-0.172	-0.672
4'	Street/15dB	MNRU-16	MNRU-12	64	-0.422	1.467	-0.062	-0.781
4"	Babble/15dB	MNRU-16	MNRU-12	64	-0.469	1.333	-0.142	-0.795
5	Car/12dB	Direct	MNRU-12	64	-1.703	1.305	-1.383	-2.023
5'	Street/15dB	Direct	MNRU-12	64	-1.594	1.561	-1.211	-1.976
5"	Babble/15dB	Direct	MNRU-12	64	-2.063	0.974	-1.824	-2.301
6	Car/12dB/12.2	AMR	AMR/4 dB	192	0.307	0.834	0.425	0.189
7	Car/12dB/12.2	AMR	AMR/7 dB	**96**	**0.354**	0.826	0.519	0.189
8	Car/12dB/12.2	AMR	AMR/10 dB	192	0.771	1.002	0.913	0.629
9	Street/15dB/12.2	AMR	AMR/4 dB	192	0.193	0.745	0.298	0.087
10	Street/15dB/12.2	AMR	AMR/7 dB	192	0.385	0.879	0.510	0.261
11	Street/15dB/12.2	AMR	AMR/10 dB	192	0.776	0.964	0.912	0.640
12	Babble/15dB/12.2	AMR	AMR/4 dB	192	0.104	0.850	0.224	-0.016
13	Babble/15dB/12.2	AMR	AMR/7 dB	192	0.313	1.001	0.454	0.171
14	Babble/15dB/12.2	AMR	AMR/10 dB	192	0.391	1.236	0.565	0.216
15	Car/12dB/12.2	AMR	AMR/NS1	192	0.802	1.079	0.955	0.649
16	Car/12dB/12.2	AMR	AMR/NS2	192	0.708	1.053	0.857	0.559
17	Car/12dB/12.2	AMR	AMR/NS3	192	0.625	1.005	0.767	0.483
18	Car/12dB/12.2	AMR	AMR/NS4	192	0.807	1.038	0.954	0.660
19	Car/12dB/12.2	AMR	AMR/NS5	192	0.641	1.093	0.795	0.486
20	Car/12dB/12.2	AMR	AMR/NS6	192	1.021	1.162	1.185	0.856
21	Street/15dB/12.2	AMR	AMR/NS1	192	0.635	0.905	0.763	0.507
22	Street/15dB/12.2	AMR	AMR/NS2	192	0.693	1.020	0.837	0.548
23	Street/15dB/12.2	AMR	AMR/NS3	192	0.479	0.965	0.616	0.343
24	Street/15dB/12.2	AMR	AMR/NS4	192	0.708	1.022	0.853	0.564
25	Street/15dB/12.2	AMR	AMR/NS5	192	0.464	0.891	0.590	0.337
26	Street/15dB/12.2	AMR	AMR/NS6	192	0.943	1.159	1.107	0.779
27	Babble/15dB/12.2	AMR	AMR/NS1	192	0.297	1.107	0.453	0.140
28	Babble/15dB/12.2	AMR	AMR/NS2	192	0.375	1.142	0.537	0.213
29	Babble/15dB/12.2	AMR	AMR/NS3	192	0.177	1.018	0.321	0.033
30	Babble/15dB/12.2	AMR	AMR/NS4	192	0.396	1.153	0.559	0.233
31	Babble/15dB/12.2	AMR	AMR/NS5	192	0.089	0.953	0.223	-0.046
32	Babble/15dB/12.2	AMR	AMR/NS6	192	0.557	1.143	0.719	0.396

Legend: - Cnd: Condition number; Noise: Background noise type; SNR is signal-to-noise ratio in dB; Rate is bit rate in kbit/s; NSx is the noise suppressor candidate x. Default parameters are nominal level, 1 transcoding.

- CMOS is Comparison Mean Opinion Score, SD is standard deviation, Se is Standard Error, +95% and -95% are respectively the upper and lower 95% confidence interval.

- The bias compensation for Condition 7 was 0.261 in Experiment 5B.

Table 4(a)

CMOS and standard deviation for COMSAT's Experiment 5A (second realization)

Cnd.	Noise/SNR/Rate	Reference	Processed Codec/Ideal NS	Votes	Y(all)	SD(all)	+95%	-95%
1	Car/6dB/12.2	AMR	AMR	192	0.021	0.898	0.148	-0.106
2	Street/9dB/12.2	AMR	AMR	192	0.031	0.805	0.145	-0.083
3	Babble/9dB/12.2	AMR	AMR	192	0.109	0.870	0.232	-0.014
4	Car/6dB	MNRU-16	MNRU-12	64	-0.250	1.512	0.120	-0.620
4'	Street/9dB	MNRU-16	MNRU-12	64	-0.484	1.357	-0.152	-0.817
4"	Babble/9dB	MNRU-16	MNRU-12	64	-0.453	1.068	-0.191	-0.715
5	Car/6dB	Direct	MNRU-12	64	-1.313	1.332	-0.986	-1.639
5'	Street/9dB	Direct	MNRU-12	64	-0.891	1.323	-0.567	-1.215
5"	Babble/9dB	Direct	MNRU-12	64	-1.453	1.321	-1.130	-1.777
6	Car/6dB/12.2	AMR	AMR/4 dB	192	0.328	0.916	0.458	0.199
7	Car/6dB/12.2	AMR	AMR/7 dB	192	0.422	1.010	0.565	0.279
8	Car/6dB/12.2	AMR	AMR/10 dB	192	0.807	1.063	0.958	0.657
9	Street/9dB/12.2	AMR	AMR/4 dB	192	0.203	0.878	0.327	0.079
10	Street/9dB/12.2	AMR	AMR/7 dB	192	0.427	0.968	0.564	0.290
11	Street/9dB/12.2	AMR	AMR/10 dB	192	0.703	1.098	0.858	0.548
12	Babble/9dB/12.2	AMR	AMR/4 dB	192	0.260	0.877	0.385	0.136
13	Babble/9dB/12.2	AMR	AMR/7 dB	192	0.380	0.952	0.515	0.245
14	Babble/9dB/12.2	AMR	AMR/10 dB	192	0.458	1.092	0.613	0.304
15	Car/6dB/12.2	AMR	AMR/NS1	192	0.594	1.019	0.738	0.450
16	Car/6dB/12.2	AMR	AMR/NS2	192	0.729	1.180	0.896	0.562
17	Car/6dB/12.2	AMR	AMR/NS3	192	0.438	1.042	0.585	0.290
18	Car/6dB/12.2	AMR	AMR/NS4	192	0.776	1.096	0.931	0.621
19	Car/6dB/12.2	AMR	AMR/NS5	192	0.604	1.102	0.760	0.448
20	Car/6dB/12.2	AMR	AMR/NS6	192	0.729	1.144	0.891	0.567
21	Street/9dB/12.2	AMR	AMR/NS1	192	0.625	1.056	0.774	0.476
22	Street/9dB/12.2	AMR	AMR/NS2	192	0.536	1.058	0.686	0.387
23	Street/9dB/12.2	AMR	AMR/NS3	192	0.411	1.025	0.556	0.267
24	Street/9dB/12.2	AMR	AMR/NS4	192	0.583	1.080	0.736	0.431
25	Street/9dB/12.2	AMR	AMR/NS5	192	0.281	0.984	0.420	0.142
26	Street/9dB/12.2	AMR	AMR/NS6	192	0.719	1.213	0.890	0.547
27	Babble/9dB/12.2	AMR	AMR/NS1	192	0.313	1.071	0.464	0.161
28	Babble/9dB/12.2	AMR	AMR/NS2	192	0.240	0.995	0.380	0.099
29	Babble/9dB/12.2	AMR	AMR/NS3	192	0.255	0.894	0.382	0.129
30	Babble/9dB/12.2	AMR	AMR/NS4	192	0.401	0.987	0.541	0.261
31	Babble/9dB/12.2	AMR	AMR/NS5	192	0.161	0.868	0.284	0.039
32	Babble/9dB/12.2	AMR	AMR/NS6	192	0.500	1.102	0.656	0.344

Legend: - Cnd: Condition number; Noise: Background noise type; SNR is signal-to-noise ratio in dB; Rate is bit rate in kbit/s; NSx is the noise suppressor candidate x. Default parameters are nominal level, 1 transcoding.

- CMOS is Comparison Mean Opinion Score, SD is standard deviation, Se is Standard Error, +95% and -95% are respectively the upper and lower 95% confidence interval.

Table 4(b)

CMOS and standard deviation for COMSAT's Experiment 5B (second realization)

Cnd.	Noise/SNR/Rate	Reference	Processed Codec/Ideal NS	Votes	Y(all)	SD(all)	+95%	-95%
1	Car/12dB/12.2	AMR	AMR	192	-0.026	0.618	0.061	-0.113
2	Street/15dB/12.2	AMR	AMR	192	0.073	0.720	0.175	-0.029
3	Babble/15dB/12.2	AMR	AMR	192	-0.026	0.802	0.087	-0.139
4	Car/12dB	MNRU-16	MNRU-12	64	-0.344	1.130	-0.067	-0.621
4'	Street/15dB	MNRU-16	MNRU-12	64	-0.656	1.011	-0.408	-0.904
4"	Babble/15dB	MNRU-16	MNRU-12	64	-0.891	1.236	-0.588	-1.193
5	Car/12dB	Direct	MNRU-12	64	-1.844	0.877	-1.629	-2.059
5'	Street/15dB	Direct	MNRU-12	64	-2.016	1.134	-1.738	-2.293
5"	Babble/15dB	Direct	MNRU-12	64	-2.203	0.962	-1.967	-2.439
6	Car/12dB/12.2	AMR	AMR/4 dB	192	0.359	0.800	0.473	0.246
7	Car/12dB/12.2	AMR	AMR/7 dB	192	0.682	0.843	0.801	0.563
8	Car/12dB/12.2	AMR	AMR/10 dB	192	0.776	0.964	0.912	0.640
9	Street/15dB/12.2	AMR	AMR/4 dB	192	0.224	0.784	0.335	0.113
10	Street/15dB/12.2	AMR	AMR/7 dB	192	0.510	0.880	0.635	0.386
11	Street/15dB/12.2	AMR	AMR/10 dB	192	0.646	0.965	0.782	0.509
12	Babble/15dB/12.2	AMR	AMR/4 dB	192	0.281	0.894	0.408	0.155
13	Babble/15dB/12.2	AMR	AMR/7 dB	192	0.443	0.925	0.574	0.312
14	Babble/15dB/12.2	AMR	AMR/10 dB	192	0.563	1.062	0.713	0.412
15	Car/12dB/12.2	AMR	AMR/NS1	192	0.854	0.943	0.988	0.721
16	Car/12dB/12.2	AMR	AMR/NS2	192	0.766	0.905	0.894	0.638
17	Car/12dB/12.2	AMR	AMR/NS3	192	0.656	0.835	0.774	0.538
18	Car/12dB/12.2	AMR	AMR/NS4	192	0.807	0.970	0.945	0.670
19	Car/12dB/12.2	AMR	AMR/NS5	192	0.703	0.921	0.833	0.573
20	Car/12dB/12.2	AMR	AMR/NS6	192	1.068	1.093	1.222	0.913
21	Street/15dB/12.2	AMR	AMR/NS1	192	0.682	0.948	0.816	0.548
22	Street/15dB/12.2	AMR	AMR/NS2	192	0.505	0.812	0.620	0.390
23	Street/15dB/12.2	AMR	AMR/NS3	192	0.417	0.852	0.537	0.296
24	Street/15dB/12.2	AMR	AMR/NS4	192	0.646	1.043	0.793	0.498
25	Street/15dB/12.2	AMR	AMR/NS5	192	0.422	0.782	0.533	0.311
26	Street/15dB/12.2	AMR	AMR/NS6	192	0.948	1.001	1.090	0.806
27	Babble/15dB/12.2	AMR	AMR/NS1	192	0.401	1.044	0.549	0.253
28	Babble/15dB/12.2	AMR	AMR/NS2	192	0.406	0.881	0.531	0.282
29	Babble/15dB/12.2	AMR	AMR/NS3	192	0.281	0.918	0.411	0.151
30	Babble/15dB/12.2	AMR	AMR/NS4	192	0.443	0.990	0.583	0.303
31	Babble/15dB/12.2	AMR	AMR/NS5	192	0.302	0.876	0.426	0.178
32	Babble/15dB/12.2	AMR	AMR/NS6	192	0.578	1.010	0.721	0.435

Legend: - Cnd: Condition number; Noise: Background noise type; SNR is signal-to-noise ratio in dB; Rate is bit rate in kbit/s; NSx is the noise suppressor candidate x. Default parameters are nominal level, 1 transcoding.

- CMOS is Comparison Mean Opinion Score, SD is standard deviation, Se is Standard Error, +95% and -95% are respectively the upper and lower 95% confidence interval.

Table 5(a)

Sorted CMOS presentation organized by impairment for COMSAT Experiment 4A (*first* realization)

Cnd.	Noise/SNR/Rate	A x B	Votes	Y(all)	+95%	-95%	t	HSD	D
4	Car/6dB	MNRU16 x MNRU12	64	-0.453	-0.258	-0.648			
4'	Street/9dB	MNRU16 x MNRU12	64	-0.500	-0.176	-0.824			
4"	Babble/9dB	MNRU16 x MNRU12	64	-0.766	-0.474	-1.058			N/A
5'	Street/9dB	Direct x MNRU12	64	-1.297	-1.005	-1.589			
5"	Babble/9dB	Direct x MNRU12	64	-1.438	-1.033	-1.842			
5	Car/6dB	Direct x MNRU12	64	-1.469	-1.179	-1.758			
8	Car/6dB/12.2	*AMR x AMR/10 dB*	192	0.870	1.016	0.724			>
20	Car/6dB/12.2	AMR x AMR/NS6	192	0.734	0.884	0.585			>
16	Car/6dB/12.2	AMR x AMR/NS2	192	0.682	0.846	0.519			>
18	Car/6dB/12.2	AMR x AMR/NS4	192	0.620	0.766	0.473			>
19	Car/6dB/12.2	AMR x AMR/NS5	192	0.552	0.687	0.417			>
15	Car/6dB/12.2	AMR x AMR/NS1	192	0.500	0.651	0.349			>
7	Car/6dB/12.2	*AMR x AMR/7 dB*	**96**	**0.389**	**0.554**	**0.223**			>
17	Car/6dB/12.2	AMR x AMR/NS3	192	0.401	0.535	0.268			>
6	Car/6dB/12.2	*AMR x AMR/4 dB*	192	0.240	0.346	0.133			=
1	Car/6dB/12.2	AMR x AMR	192	0.016	0.118	-0.087			-
26	Street/9dB/12.2	AMR x AMR/NS6	192	0.807	0.983	0.632			>
11	Street/9dB/12.2	*AMR x AMR/10 dB*	192	0.776	0.921	0.632			>
24	Street/9dB/12.2	AMR x AMR/NS4	192	0.599	0.756	0.442			>
21	Street/9dB/12.2	AMR x AMR/NS1	192	0.583	0.737	0.430			>
22	Street/9dB/12.2	AMR x AMR/NS2	192	0.542	0.687	0.396			>
23	Street/9dB/12.2	AMR x AMR/NS3	192	0.542	0.691	0.393			>
10	Street/9dB/12.2	*AMR x AMR/7 dB*	192	0.422	0.571	0.273			>
25	Street/9dB/12.2	AMR x AMR/NS5	192	0.354	0.500	0.208			>
9	Street/9dB/12.2	*AMR x AMR/4 dB*	192	0.203	0.322	0.084			=
2	Street/9dB/12.2	AMR x AMR	192	0.057	0.150	-0.036			-
14	Babble/9dB/12.2	*AMR x AMR/10 dB*	192	0.557	0.719	0.396			>
32	Babble/9dB/12.2	AMR x AMR/NS6	192	0.500	0.666	0.334			>
13	Babble/9dB/12.2	*AMR x AMR/7 dB*	192	0.344	0.502	0.185			>
30	Babble/9dB/12.2	AMR x AMR/NS4	192	0.344	0.491	0.196			>
28	Babble/9dB/12.2	AMR x AMR/NS2	192	0.318	0.471	0.164			>
27	Babble/9dB/12.2	AMR x AMR/NS1	192	0.302	0.450	0.154			>
31	Babble/9dB/12.2	AMR x AMR/NS5	192	0.266	0.399	0.132			=
12	Babble/9dB/12.2	*AMR x AMR/4 dB*	192	0.224	0.369	0.079			=
29	Babble/9dB/12.2	AMR x AMR/NS3	192	0.208	0.364	0.053			=
3	Babble/9dB/12.2	AMR x AMR	192	-0.016	0.098	-0.129			-

Note: Condition 7 is purposely out of rank-order in order to reflect its larger confidence interval due to the reduced number of votes. Also, under the HSD criterion, conditions 7, 6, and 1 have equivalent scores.

Table 5(b)

Sorted CMOS presentation organized by impairment for COMSAT Experiment 4B (*first* realization)

Cnd.	Noise/SNR/Rate	A x B	Votes	Y(all)	+95%	-95%	t	HSD	D
4’	Street/15dB	MNRU16 x MNRU12	64	-0.328	-0.105	-0.551			
4	Car/12dB	MNRU16 x MNRU12	64	-0.359	-0.107	-0.612			
4”	Babble/15dB	MNRU16 x MNRU12	64	-0.922	-0.615	-1.228			N/A
5’	Street/15dB	Direct x MNRU12	64	-1.266	-1.010	-1.521			
5	Car/12dB	Direct x MNRU12	64	-1.703	-1.372	-2.035			
5”	Babble/15dB	Direct x MNRU12	64	-2.359	-2.103	-2.615			
20	Car/12dB/12.2	AMR x AMR/NS6	192	0.813	0.963	0.662			>
8	Car/12dB/12.2	*AMR x AMR/10 dB*	192	0.760	0.897	0.623			>
16	Car/12dB/12.2	AMR x AMR/NS2	192	0.667	0.803	0.531			>
18	Car/12dB/12.2	AMR x AMR/NS4	192	0.635	0.790	0.481			>
15	Car/12dB/12.2	AMR x AMR/NS1	192	0.609	0.746	0.473			>
19	Car/12dB/12.2	AMR x AMR/NS5	192	0.583	0.725	0.442			>
7	Car/12dB/12.2	*AMR x AMR/7 dB*	**96**	**0.517**	**0.676**	**0.359**			>
17	Car/12dB/12.2	AMR x AMR/NS3	192	0.495	0.629	0.361			>
6	Car/12dB/12.2	*AMR x AMR/4 dB*	192	0.214	0.320	0.107			=
1	Car/12dB/12.2	AMR x AMR	192	0.005	0.077	-0.066			-
26	Street/15dB/12.2	AMR x AMR/NS6	192	0.677	0.825	0.529			>
24	Street/15dB/12.2	AMR x AMR/NS4	192	0.646	0.774	0.518			>
11	Street/15dB/12.2	*AMR x AMR/10 dB*	192	0.599	0.723	0.475			>
21	Street/15dB/12.2	AMR x AMR/NS1	192	0.448	0.570	0.326			>
22	Street/15dB/12.2	AMR x AMR/NS2	192	0.448	0.560	0.336			>
25	Street/15dB/12.2	AMR x AMR/NS5	192	0.438	0.559	0.316			>
10	Street/15dB/12.2	*AMR x AMR/7 dB*	192	0.432	0.535	0.329			>
23	Street/15dB/12.2	AMR x AMR/NS3	192	0.370	0.493	0.247			>
9	Street/15dB/12.2	*AMR x AMR/4 dB*	192	0.245	0.343	0.147			=
2	Street/15dB/12.2	AMR x AMR	192	0.068	0.156	-0.020			-
32	Babble/15dB/12.2	AMR x AMR/NS6	192	0.417	0.540	0.294			>
14	Babble/15dB/12.2	*AMR x AMR/10 dB*	192	0.406	0.522	0.290			>
13	Babble/15dB/12.2	*AMR x AMR/7 dB*	192	0.359	0.469	0.250			>
27	Babble/15dB/12.2	AMR x AMR/NS1	192	0.318	0.425	0.211			>
30	Babble/15dB/12.2	AMR x AMR/NS4	192	0.260	0.378	0.142			>
28	Babble/15dB/12.2	AMR x AMR/NS2	192	0.193	0.311	0.074			=
12	Babble/15dB/12.2	*AMR x AMR/4 dB*	192	0.146	0.259	0.033			=
31	Babble/15dB/12.2	AMR x AMR/NS5	192	0.115	0.226	0.003			=
29	Babble/15dB/12.2	AMR x AMR/NS3	192	0.099	0.210	-0.012			=
3	Babble/15dB/12.2	AMR x AMR	192	0.021	0.111	-0.070			-

Table 5(c)

Sorted CMOS presentation organized by impairment for COMSAT Experiment 5A (second realization)

Cnd.	Noise/SNR/Rate	A x B	Votes	Y(all)	+95%	-95%	t	HSD	D
4	Car/6dB	MNRU16 x MNRU12	64	-0.250	0.120	-0.620			
4”	Babble/9dB	MNRU16 x MNRU12	64	-0.453	-0.191	-0.715			
4’	Street/9dB	MNRU16 x MNRU12	64	-0.484	-0.152	-0.817			N/A
5’	Street/9dB	Direct x MNRU12	64	-0.891	-0.567	-1.215			
5	Car/6dB	Direct x MNRU12	64	-1.313	-0.986	-1.639			
5”	Babble/9dB	Direct x MNRU12	64	-1.453	-1.130	-1.777			
8	Car/6dB/12.2	*AMR x AMR/10 dB*	192	0.807	0.958	0.657			>
18	Car/6dB/12.2	AMR x AMR/NS4	192	0.776	0.931	0.621			>
16	Car/6dB/12.2	AMR x AMR/NS2	192	0.729	0.896	0.562			>
20	Car/6dB/12.2	AMR x AMR/NS6	192	0.729	0.891	0.567			>
19	Car/6dB/12.2	AMR x AMR/NS5	192	0.604	0.760	0.448			>
15	Car/6dB/12.2	AMR x AMR/NS1	192	0.594	0.738	0.450			>
17	Car/6dB/12.2	AMR x AMR/NS3	192	0.438	0.585	0.290			>
7	Car/6dB/12.2	*AMR x AMR/7 dB*	192	0.422	0.565	0.279			>
6	Car/6dB/12.2	*AMR x AMR/4 dB*	192	0.328	0.458	0.199			>
1	Car/6dB/12.2	AMR x AMR	192	0.021	0.148	-0.106			-
26	Street/9dB/12.2	AMR x AMR/NS6	192	0.719	0.890	0.547			>
11	Street/9dB/12.2	*AMR x AMR/10 dB*	192	0.703	0.858	0.548			>
21	Street/9dB/12.2	AMR x AMR/NS1	192	0.625	0.774	0.476			>
24	Street/9dB/12.2	AMR x AMR/NS4	192	0.583	0.736	0.431			>
22	Street/9dB/12.2	AMR x AMR/NS2	192	0.536	0.686	0.387			>
10	Street/9dB/12.2	*AMR x AMR/7 dB*	192	0.427	0.564	0.290			>
23	Street/9dB/12.2	AMR x AMR/NS3	192	0.411	0.556	0.267			>
25	Street/9dB/12.2	AMR x AMR/NS5	192	0.281	0.420	0.142			=
9	Street/9dB/12.2	*AMR x AMR/4 dB*	192	0.203	0.327	0.079			=
2	Street/9dB/12.2	AMR x AMR	192	0.031	0.145	-0.083			-
32	Babble/9dB/12.2	AMR x AMR/NS6	192	0.500	0.656	0.344			>
14	Babble/9dB/12.2	*AMR x AMR/10 dB*	192	0.458	0.613	0.304			>
30	Babble/9dB/12.2	AMR x AMR/NS4	192	0.401	0.541	0.261			>
13	Babble/9dB/12.2	*AMR x AMR/7 dB*	192	0.380	0.515	0.245			>
27	Babble/9dB/12.2	AMR x AMR/NS1	192	0.313	0.464	0.161			=
12	Babble/9dB/12.2	*AMR x AMR/4 dB*	192	0.260	0.385	0.136			=
29	Babble/9dB/12.2	AMR x AMR/NS3	192	0.255	0.382	0.129			=
28	Babble/9dB/12.2	AMR x AMR/NS2	192	0.240	0.380	0.099			=
31	Babble/9dB/12.2	AMR x AMR/NS5	192	0.161	0.284	0.039			=
3	Babble/9dB/12.2	AMR x AMR	192	0.109	0.232	-0.014			-

Table 5(d)

Sorted CMOS presentation organized by impairment for COMSAT Experiment 5B (second realization)

Cnd.	Noise/SNR/Rate	A x B	Votes	Y(all)	+95%	-95%	t	HSD	D
4	Car/12dB	MNRU16 x MNRU12	64	-0.344	-0.067	-0.621			
4’	Street/15dB	MNRU16 x MNRU12	64	-0.656	-0.408	-0.904			
4”	Babble/15dB	MNRU16 x MNRU12	64	-0.891	-0.588	-1.193			N/A
5	Car/12dB	Direct x MNRU12	64	-1.844	-1.629	-2.059			
5’	Street/15dB	Direct x MNRU12	64	-2.016	-1.738	-2.293			
5”	Babble/15dB	Direct x MNRU12	64	-2.203	-1.967	-2.439			
20	Car/12dB/12.2	AMR x AMR/NS6	192	1.068	1.222	0.913			>
15	Car/12dB/12.2	AMR x AMR/NS1	192	0.854	0.988	0.721			>
18	Car/12dB/12.2	AMR x AMR/NS4	192	0.807	0.945	0.670			>
8	Car/12dB/12.2	*AMR x AMR/10 dB*	192	0.776	0.912	0.640			>
16	Car/12dB/12.2	AMR x AMR/NS2	192	0.766	0.894	0.638			>
19	Car/12dB/12.2	AMR x AMR/NS5	192	0.703	0.833	0.573			>
7	Car/12dB/12.2	*AMR x AMR/7 dB*	192	0.682	0.801	0.563			>
17	Car/12dB/12.2	AMR x AMR/NS3	192	0.656	0.774	0.538			>
6	Car/12dB/12.2	*AMR x AMR/4 dB*	192	0.359	0.473	0.246			>
1	Car/12dB/12.2	AMR x AMR	192	-0.026	0.061	-0.113			-
26	Street/15dB/12.2	AMR x AMR/NS6	192	0.948	1.090	0.806			>
21	Street/15dB/12.2	AMR x AMR/NS1	192	0.682	0.816	0.548			>
11	Street/15dB/12.2	*AMR x AMR/10 dB*	192	0.646	0.782	0.509			>
24	Street/15dB/12.2	AMR x AMR/NS4	192	0.646	0.793	0.498			>
10	Street/15dB/12.2	*AMR x AMR/7 dB*	192	0.510	0.635	0.386			>
22	Street/15dB/12.2	AMR x AMR/NS2	192	0.505	0.620	0.390			>
25	Street/15dB/12.2	AMR x AMR/NS5	192	0.422	0.533	0.311			>
23	Street/15dB/12.2	AMR x AMR/NS3	192	0.417	0.537	0.296			>
9	Street/15dB/12.2	*AMR x AMR/4 dB*	192	0.224	0.335	0.113			>
2	Street/15dB/12.2	AMR x AMR	192	0.073	0.175	-0.029			-
32	Babble/15dB/12.2	AMR x AMR/NS6	192	0.578	0.721	0.435			>
14	Babble/15dB/12.2	*AMR x AMR/10 dB*	192	0.563	0.713	0.412			>
13	Babble/15dB/12.2	*AMR x AMR/7 dB*	192	0.443	0.574	0.312			>
30	Babble/15dB/12.2	AMR x AMR/NS4	192	0.443	0.583	0.303			>
28	Babble/15dB/12.2	AMR x AMR/NS2	192	0.406	0.531	0.282			>
27	Babble/15dB/12.2	AMR x AMR/NS1	192	0.401	0.549	0.253			>
31	Babble/15dB/12.2	AMR x AMR/NS5	192	0.302	0.426	0.178			>
12	Babble/15dB/12.2	*AMR x AMR/4 dB*	192	0.281	0.408	0.155			>
29	Babble/15dB/12.2	AMR x AMR/NS3	192	0.281	0.411	0.151			>
3	Babble/15dB/12.2	AMR x AMR	192	-0.026	0.087	-0.139			-

Table 5 provides the most convenient presentation of the test data for statistical comparisons, in particular using the HSD and Dunnet criteria. Overall across the 4 experiments, it can be seen that the rank-order of the NS-enabled solutions was in all cases higher than AMR without NS. However, for car and street noise, most of the NS-enabled AMR solutions were significantly preferred than AMR without NS, while for babble noise a smaller number of NS solutions were significantly preferred to AMR without NS. The cases when no significantly observed were as follows:

- NS5 for street noise and NS 1,2, 3, and 5 in Experiment 4A (HSD). Under the Dunnet criterion, only NS 3 and 5 are equivalent to AMR without NS when babble noise is present, and all NS solutions performed better that AMR without NS for street noise.
- In Experiment 4B, NS 2, 3, 4 and 5 were equivalent under the HSD criterion to AMR without NS when babble noise is present. Under the Dunnet criterion, only NS 2, 3, and 5 were equivalent to AMR without NS when babble noise is present..
- In Experiment 5A, NS 5 under street noise and NS 1, 2, 3, 4, and 5 under babble noise were equivalent to AMR without NS using the HSD criterion. With the Dunnet criterion, NS 5 under street noise and NS 1, 2, 3, and 5 under babble noise were again equivalent to AMR without NS.
- In Experiment 5B, all NS solutions were significantly preferred to AMR without NS for any of the three noises (using either HSD or Dunnet criteria).

As regards the ideal noise suppression scores, Table 6 summarizes the HSD equivalencies of the ideal NS cases from Table 5. It can be seen that there is a trend that indicates that for the car noise cases, 4 and 7 dB ideal NS were equally preferred for low SNR cases and 10 and 7 dB ideal NS were equally preferred for high SNR cases. In the presence of street noise, 4 and 7 dB ideal NS were equally preferred for low SNR; however, in this case, 7 and 10 dB ideal NS were also equivalently preferred. No distinction in preference was observed for the ideal NS cases with AMR at 5.9 kbit/s in high SNR street noise, while for the higher bit rate (12.2 kbit/s), 7 and 10 dB ideal NS were also equally preferred, but significantly more preferred than 4 dB ideal NS. For all babble noise cases but the high SNR at 5.9 kbit/s case, all three ideal SNR cases were equally preferred.

Table 7 contains a summary of the rank-orders (irrespective of statistical significance) observed for the six NS solutions studied in Experiments 4 and 5. Overall, it can be seen that NS 6 held the highest-ranking position in 8 out of 9 cases. NS 1, 2, and 4 also showed up in the upper ranking positions, with a small advantage towards NS 4. Alternatively, NS 3 and 5 most of the time occupied the lower ranking positions.

Table 6:
Summary of HSD criterion for Ideal Noise Suppression in Experiments 4&5

Noise Type	Ideal NS	Low SNR (6/9 dB)		High SNR (12/15 dB)	
		4A	5A	4B	5B
Car	10				
	7				
	4				
Street	10				
	7				
	4				
Babble	10				
	7				
	4				

Table 7:
Summary presentation of rank-order for the NS candidates in Experiments 4&5

4A			4B			5A			5B		
Car	Street	Babble	Car	Street	Babble	Car	Street	Babble	Car	Street	Babble
6	6	6	6	6	6	4	6	6	6	6	6
2	4	4	2	4	1	2	1	4	1	1	4
4	1	2	4	1	4	6	4	1	4	4	2
5	2	1	1	2	2	5	2	3	2	2	1
1	3	5	5	5	5	1	3	2	5	5	5
3	5	3	3	3	3	3	5	5	3	3	3

7. Conclusion

COMSAT performed Experiments 4&5 of the AMR NS Selection Phase in the Spanish language in compliance with the test plans [1,2]. These experiments were designed to evaluate the performance of the six candidate noise suppression algorithms in the presence of three different noise types within the same experiment (Car, street, and babble noise) for different signal-to-noise ratios (SNR) and two fixed bit rates (5.9 and 12.2 kbit/s), for a total of four subjective experiments. The six noise suppression candidate algorithms are presented in this contribution using blinded designations NS1 through NS6.

The assessment method specified in the Test Plan was a modified version of the Category Comparison Rating (CCR) method defined in [3]. Here, the reference signal, instead of being the unprocessed ("Direct") speech, was the speech processed through the AMR codec without noise suppression for all conditions involving the NS candidates. In the data reported, a positive CMOS value indicates preference towards the test sample, and a negative value indicates preference towards the reference sample.

After the performance of the subjective sessions, COMSAT discovered that up to version 2.0 of the Test Plan, there was a flaw in the presentation sequences for Experiments 4 & 5. This flaw caused condition 7 (7 dB Ideal noise suppression for car noise) to be systematically paired with the incorrect reference (condition 38, instead of 39) when the B/A direction was played. The problem was discovered in time for the English-language listening laboratory (Nortel) to perform the experiments using corrected presentation sequences. However, half of the votes (96) had to be discarded for condition 7 from COMSAT's analysis in each of the four experiments in their first realization. Due to time constraints, it was possible for COMSAT to repeat only Experiments 5A and 5B. The second realization of Experiments 5A and 5B used the corrected presentation sequences, and contain the full 192 votes for condition 7. Additional measures were taken to compensate the CCR presentation direction bias for condition 7 in Experiments 4A and 4B, which could not be repeated.

As regards the ideal noise suppression scores, a trend indicates that for the car noise cases, 4 and 7 dB ideal NS were equally preferred for low SNR cases and 10 and 7 dB ideal NS were equally preferred for high SNR cases. In the presence of street noise, 4 and 7 dB ideal NS were equally preferred for low SNR; however, in this case, 7 and 10 dB ideal NS were also equivalently preferred. No distinction in preference was observed for the ideal NS cases with AMR at 5.9 kbit/s in high SNR street noise, while for the higher bit rate (12.2 kbit/s), 7 and 10 dB ideal NS were also equally preferred, but significantly more preferred than 4 dB ideal NS. For all babble noise cases but the high SNR at 5.9 kbit/s case, all three ideal SNR cases were equally preferred.

The rank-orders observed for the six NS solutions studied in Experiments 4 and 5 were summarized, irrespective of statistical significance in differences. Overall, NS 6 held the highest-ranking position in 8 out of 9 cases. NS 1, 2, and 4 also showed up in the upper ranking positions, with a small advantage towards NS 4. Alternatively, NS 3 and 5 most of the time occupied the lower ranking positions.

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] ITU Rec.P.800, *"Methods for Subjective Determination of Transmission Quality"*; Geneva, August 1996.

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