Joint TSG-S4#10 - SMG11#15 Meeting
Helsinki, Finland
February 28th - March 3rd 2000

S4/SMG11 Tdoc 116/00

ETSI TS xxx xxx V0.1.0 (2000-02)

European Standard (Telecommunications series)

Digital cellular telecommunication system (Phase 2+); Results of the AMR Noise Suppression Selection Phase (GSM xx.xx version 0.1.0)

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

Reference

DTS/SMG-110675Q7 (?.PDF)

Keywords

AMR, CODEC

ETSI

Postal address

F-06921 Sophia Antipolis Cedex - FRANCE

Office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16
Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Internet

secretariat@etsi.fr
Individual copies of this ETSI deliverable
can be downloaded from
http://www.etsi.org

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available **free of charge** from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://www.etsi.org/ipr).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This ETSI Technical Report has been produced by ETSI Technical Committee Special Mobile Group (SMG).

The technical report provides the performance results of the Selection phase of testing of the GSM Adaptive Multi-Rate Noise Suppression (AMR/NS) Work Item.

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

Version 8.x.y, where:

8 indicates Release 1999 of GSM Phase 2+

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

y the third digit is incremented when editorial only changes have been incorporated in the specification.

1 Scope

This technical report provides background information on the performance of the six candiates which were proposed as solutions for publication of an example noise suppression solution for application to the GSM Adaptive Multi-Rate (AMR) speech codec. Experimental test results from the seech quality related testing are reported to illustrate the behaviour of the candidate algorithms in multiple operational conditions. Additional information is also provided covering data not necessarily directly associated with speech quality (such as complexity, delay, effect on voice activity factor).

2 References

The following documents contain provisions that, 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.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.

[1] GSM 01.04: "Digital cellular telecommunications system (Phase 2+); Abbreviations and acronyms".

[2] GSM 02.76: "Noise Suppression for the AMR Codec; Service Description; Stage 1"

[3] GSM 03.50: "Transmission planning aspects of the speech service in the GSM Public Land Mobile Network (PLMN) system".

[4] GSM 06.08: "Digital cellular telecommunications system; Half rate speech; Performance of the GSM half rate speech codec".

[5] GSM 06.55: "Digital cellular telecommunications system; Performance Characterisation of the GSM Enhanced Full Rate (EFR) speech codec".

[6] GSM 06.75: "Digital cellular telecommunications system; ; Performance Characterisation of the GSM Adaptive Multi-Rate (AMR) speech codec".

3 Definitions and abbreviations

3.1 Definitions

The following terminology is used throughout this report.

Adaptive Multi-Rate (AMR) codec; Speech and channel codec capable of operating at gross bit-rates of 11.4 kbit/s ("*half-rate*") and 22.8 kbit/s ("*full-rate*"). In addition, the codec may operate at various combinations of speech and channel coding (*codec mode*) bit-rates for each *channel mode*.

Channel mode; Half-rate or full-rate operation

Codec mode; For a given *channel mode*, the bit partitioning between the speech and channel codecs.

Error Patterns

Error Insertion Device; Result of offline simulations stored on files. To be used by the "Error Insertion Device" to model the radio transmission from the output of the channel decoder and interleaver to the input of the deinterleaver and channel decoder.

Full-rate (FR); Full-rate channel or *channel mode*

Half-rate (HR); Half-rate channel or *channel mode*

Toll Quality; Speech quality normally achieved on modern wireline telephones. Synonymous with "ISDN quality" in most western countries.

Wireline quality; Speech quality provided by modern wireline networks. Normally taken to imply quality at least as good as that of 32kbit/s G.726 or G.728 16 kbit/s codecs.

3.2 Abbreviations

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

A/D	Analogue to Digital
ACR	Absolute Category Rating
ADPCM	Adaptive Differential Pulse Code Modulation
AMR	Adaptive Multi-Rate
AMR-NS	AMR Noise Suppression
BSC	Base Station Controller
BTS	Base Transceiver Station
CCR	Comparison Category Rating
C/I	Carrier-to-Interfere ratio
CI	Confidence Interval
CNI	Comfort Noise Insertion
CRC	Cyclic Redundancy Check
D/A	Digital to Analogue
DAT	Digital Audio Tape
DCR	Degradation Category Rating
DSP	Digital Signal Processor
DTMF	Dual Tone Multi Frequency
DTX	Discontinuous Transmission for power consumption and interference reduction
EFR	Enhanced Full Rate
ESP	Product of E (Efficiency), S (Speed) and P (Percentage of Power) of the DSP
FR	Full Rate (also GSM FR)
FH	Frequency Hopping
G.726	ITU 16/24/32kbit/s ADPCM codec
G.728	ITU 16kbit/s LD-CELP codec
G.729	ITU 8/6.4/11.8 kbit/s speech codec
GBER	Average gross bit error rate
GSM	Global System for Mobile communications
HR	Half Rate (also GSM HR)
IRS	Intermediate Reference System
ITU-T	International Telecommunication Union - Telecommunications Standardisation Sector
MNRU	Modulated Noise Reference Unit
Mod. IRS	Modified IRS
MOPS	Million of Operation per Seconds
MOS	Mean Opinion Score
MS	Mobile Station
MSC	Mobile Switching Center
PCM	Pulse Code Modulation
PSTN	Public Switched Telecommunications Network
Q	Speech-to-speech correlated noise power ratio in dB

SD	Standard Deviation
SID	Silence Descriptor
SMG	Special Mobile Group
SNR	Signal To Noise Ratio
TCH-AFS	Traffic CHannel Adaptive Full rate Speech
TCH-AHS	Traffic CHannel Adaptive Half rate Speech
TDMA	Time Division Multiple Access
TFO	Tandem Free Operation
tMOPS	true Million of Operations per Seconds
TUx	Typical Urban at multipath propagation profile at x km/s
VAD	Voice Activity Detector
VAF	Voice Activity Factor
wMOPS	weighted Million of Operations per Seconds

Multiple Error Patterns were used during the Characterisation tests. They are identified by the propagation Error Conditions from which they are derived. The following conventions are used:

ECx	Error Conditions at x dB C/I simulating a radio channel under static C/I using ideal Frequency Hopping in a TU3 multipath propagation profile

For abbreviations not given in this sub-clause, see GSM 01.04 [1].

4 General

4.1 Project History

In June 1998 during SMG#26, SMG approved a Work Item to develop and standardise a noise suppression solution for the Adaptive Multi-rate (AMR) speech codec. SMG11 have carried out this work since September 1998 (SMG11#7).

The work in SMG11 focussed on the preparations for a Selection Phase with the intention of choosing an example optional noise suppression solution. In the course of this work, documentation covering Requirements [2], Design Constraints, Selection Phase Deliverables, Selection Phase Rules, and a Selection Phase Test Plan were drafted.

In August 1999 the Selection Phase commenced. Six Noise Suppression algorithms were submitted as candidates. The algorithm proposals came from Ericsson (NS5), Matra Nortel Communications (MNC) (NS4), Mitsubishi Electric Corporation (NS1), Motorola (NS6), Nokia (NS3) and Siemens AG (NS2). Testing of candidate solutions was carried out during September-November 1999, and the listening test results were analysed at two meetings: SMG11#13 (December 1999) and SMG11#14 (January 2000). Listening test results and deliverables from proponents (technical descriptions of the algorithms, analysis of compliance to design constraints, additional information such as objective measurements) were reviewed within SMG11.

SMG11 were not able to reach a consensus on selecting an example solution, and as a consequence the deliverables from the Work Item were amended during SMG#31 to comprise of a specification defining Recommended Minimum Performance Requirements [TBA], an associated Subjective Listening Test Plan, and a Technical Report recording all pertinent information arising from the Selection Phase. This document forms the latter deliverable.

4.2 Overview of the AMR-NS Work Item

The Work Item covered the development of a noise suppression algorithm as an example optional feature designed to enhance speech quality in a range of environments where there is significant (acoustic) background noise. The noise suppression function is a preprocessing module that is used to improve the signal to noise ratio of a speech signal prior to voice coding. Solutions implementing noise suppression as a separate preprocessing module prior to the AMR speech encoder or as an embedded module operating on the input speech buffer were considered. AMR Noise Suppression (AMR-NS) is intended to be used in the mobile station (operating on the uplink speech signal). The possibility to implement AMR Noise Suppression in the network (operating on the downlink speech signal) was considered for feasibility purposes only. As part of this study, tests with noise suppression in both uplink and downlink (tandem noise suppression) were included and the results are included in this report. It should be noted that the Recommended

Minimum Performance Requirements Specification [TBA] covers only the uplink case where the algorithm is implemented in the mobile station.

4.3 Presentation of the following sections

The following sections provide a summary of the Selection Phase test results, including the results of ojective performance measurements, and a record of relevant other information for each of the candidate algorithms.

- Section 5 defines the minimum performance requirements defined for the Selection Phase.
- Section 6 defines the means used to compare candidate algorithms directly in terms of speech quality performance.
- Section 7 describes the subjective listening tests undertaken and summarises the results achieved (covering the requirements of Section 5 and the means of comparison of Section 6).
- Section 8 summarises the design constraints defined for the Selection Phase.
- Section 9 summarises the effect on the existing AMR Voice Activity Detector (VAD) function, in the form of voice activity factor (VAF) measurements.
- Section 10 summarises the results of an Objective Performance Measure used to characterise the noise suppression algorithms.
- Section 11 summarises the results of the Feasibility study into Implementing Noise Suppression in the downlink.

Annex A contains the final versions of the Design Constraints, Selection Rules, and Selection Phase Deliverables defined for the Selection Phase

Annex B (a separate component of the archive file comprising this report) is the final version of the Selection Phase Test Plan.

Annex C (a separate component of the archive file comprising this report) is the final version of the Selection Phase Global Analysis Spreadsheet, and is the full record of the results achieved from the subjective listening tests.

Annex D contains the methodologies used to derive signal to noise ratio improvement values from the subjective listening tests.

Annex E defines the methodology used to generate the objective measures of performance reported in section 10.

Annex F defines the methodology used to determine impact on Voice Activity Factor.

Annex G provides a reference list of SMG11 temporary documents which contain relevant information used during the Selection Phase.

5 Minimum Performance Requirements

Performance requirements were established during the AMR-NS development phase which reflected the understanding that there was no clear risk-free means of identifying minimum performance, because this was the first time such a standardisation effort for noise suppression functionality had been undertaken in ETSI. As a result, failure to meet some minimum performance requirements was not considered to be a reason for disqualification, particularly if such failures were not consistent across all listening laboratories undertaking a particular test (where the term systematic failure is used to describe failures consistent across laboratories).

Table 5.1 lists the minimum requirements as stated in the Stage 1 specification [2] and, for each requirement, defines the associated experiment or experiments defined to check compliance to the requirement. In each case a criterion is defined to determine failure to meet the requirement. The reference condition is AMR without noise suppression in all cases, except for the evaluation of speech quality during the Initial Convergence Time (Experiment 1).

The possibility to implement AMR Noise Suppression in the network (operating on the downlink speech signal) was defined to be part of a feasibility study. This is considered further in Section 11.

Associated Section in Stage 1 Description [3]	Requirement (Title)	Relevant Tests
4.6.1.1	Initial Convergence	Experiment 1: Expert/Informal listening test Any candidate for which the Listening Experts determine that the quality degradation in the initial convergence time is unacceptable will be ragarded as failing the requirement.
4.6.1.2	No degradation in clean speech	Experiment 2: Degradation in Clean (Pair comparison) Any candidate failing to be preferred with a 50% probability in any test condition will be regarded as failing the requirement
4.6.1.3/4.6.1.4	No artifacts in residual noise & No speech clipping or reduction in intelligibility	Experiment 3: Performance in Background Noise (ACR) A candidate failing to be at least as good as AMR w/o NS at the same noise level will be regarded as failing the requirement
4.6.1.5	AMR+NS preferred to AMR w/o NS	Experiments 4-10: Performance under background noise. Any candidate failing to be preferred to the reference (AMR w/o NS) with a 95% probability for any condition will be regarded as failing the requirement.
4.8	Voice Activity Factor	Test defined in [3] section 4.8:Any candidate failing to meet the requirement stated in section 4.8 of [2] will be regarded as failing the requirement. (This requirement states that the use of noise suppression should not significantly increase channel activity when used in conjunction with DTX.)

Table 5.1: *Minimum performance requirements*

The total number of simple failures and number of systematic failures (failure of the same test condition in all tests performed for the same experiment) were recorded and the candidates were ranked accordingly.

In order to generate additional information, the candidates were also ranked according to the number of simple and systematic failures, assuming that a candidate only fails as a result of Experiments 4-10 if it is not found at least as good as the reference at the 95% confidence interval.

6 Comparison of Candidates by Subjective Means

A number of means of ranking candidates were developed based on figures of merit (FOMs). These FOMs were derived from the listening test results and are defined below. The FOMs covering downlink operation are not included (see Section 11), and FOMs defined in the Selection Rules which are not distinct are also not included. Additionally, FOMs not associated with subjective listening test results are not included.

Two sets of Figure of Merits are described here. The FOM numbering definitions used during the Selection Phase are retained for ease of cross-referencing. The first set (FoMs#1, 3, 4, 6) is based on the CCR test results (AMR/NS Selection Experiments 4 to 9). The second set (FoMs#7, 10) is derived from the ACR Test results.

FOM#1	Summation of CMOS scores for all conditions within an experiment, summed (unweighted) across all experiments, excluding all conditions including NS in the downlink direction.
	Repeat measures as per FOM#1above but restricted to:
FOM#3a	All conditions where the SNR >= 10dB (but not including conditions where the SNR >= 30dB)
FOM#3c	All conditions where the SNR < 10dB
	Repeat measures as per FOM#1 above but restricted to:
FOM#4a	All conditions with DTX on
FOM#4b	All conditions with DTX off
	Subjective SNR improvement per Noise Type based on the CCR test results evaluated using the methodology defined in Annex?:
FOM#6a	For car noise
FOM#6b	For street noise
FOM#6c	For babble noise
FOM#7a	Summation of the delta MOS scores per Experiment summed across all ACR Experiments (Experiment 3 and 10), excluding all test conditions using noise suppression in the downlink direction. The delta MOS score is identified as the difference between the MOS score obtained by the candidate for a specific test condition and the MOS score for the reference (AMR without noise suppression) in the same test condition.
FOM#7b	Unweighted summation of the delta dBq scores per Experiment summed across all ACR Experiments (Experiment 3 and 10), excluding all test conditions using noise supprssion in the downlink direction. The delta dBq score is identified as the difference between the dBq score obtained by the candidate for a specific test condition and the dBq score for the reference (AMR without noise suppression) in the same test condition. When a dBq score is outside the linear part of the MNRU curve, it should be replaced by the dBq value obtained by replacing the non-linear part of the MNRU curve with a linear extrapolation with slope 0.05. The linear part of the MNRU curve is identified as the area of the *MOS=f(dBq)* curve where the slope is higher than 0.05.
FOM#10	Subjective SNR improvement per Noise Type based on the ACR test results evaluated using the methodology defined in Annex ? (Care must be taken in analysing the results according to this FOM. Experiment 3 is designed to look for unnatural effects in the suppressed speech, and is therefore not attempting to measure SNR improvement at all. Experiment 3 has a large influence; hence the low and often negative values for this FOM. FOM#6 is a more representative measure for SNR improvement.)

7 Selection Phase Listening Tests and Results

The candidates were referred to as NS1, NS2,..., NS6 during the analysis. The mapping to particular candidates is defined below.

NS1 = Mitsubishi Electric Corporation

NS2 = Siemens AG

NS3 = Nokia

NS4 = Matra Nortel Communications

NS5 = Ericsson

NS6 = Motorola

7.1 Summary of Selection Tests undertaken

The six candidates were tested in a variety of test conditions in 5 independent test laboratories. Testing was carried out using 7 languages. The tests took place during a period from September to November 1999.

Candidate performances were evaluated across many test conditions consisting of 10 experiments and 14 sub-experiments [Annex B]:

Experiment 1:	*Quality During the Initial Convergence Time (informal test)*
Experiment 2:	*Degradation in Clean Speech (pair comparison test)*
Experiment 3:	*Artifacts and Clipping Effects in Background Noise Conditions (ACR test)* *Experiment 3a: car noise*

Experiment 3b: street noise
Experiment 3c: babble noise

Experiments 4 and 5: *Performances in Background Noise Conditions (CCR test)*
Experiment 4a: low SNR, with AMR 5.9 kbit/s
Experiment 4b: high SNR, with AMR 5.9 kbit/s
Experiment 5a: low SNR, with AMR 12.2 kbit/s
Experiment 5b: high SNR, with AMR 12.2 kbit/s

Experiments 6 and 7: *Performance in Background Noise: Influence of Propagation Errors (CCR test)*
Experiment 6: car noise at SNR of 6 dB with C/I=10 dB in uplink and error-free in downlink
Experiment 7: street noise at SNR of 9 dB with C/I=10 dB in uplink and error-free in downlink

Experiments 8 and 9: *Performances in Background Noise: Influence of VAD/DTX (CCR test)*

Experiment 10: *Influence of the Input Signal + Noise Level and Performances with Special Noises (ACR test)*

Experiment 1 is an informal test with expert listeners analysing any negative impact the noise suppressors may have during convergence time. *Experiment 2* is based on pair comparison to test if there is any degradation when using noise suppression compared to the coder without noise suppression. *Experiment 3* is an Absolute Category Rating (ACR) test analysing any artefacts and clipping effects in background noise. *Experiments 4 to 9* are Comparison Category Rating (CCR) tests analysing performances in background noise conditions with and without propagation errors, and also the influence of VAD/DTX. *Experiment 10* is an ACR test investigating the influence of the level of input signal and noise, and also assessing the performance for special noise types.

Most of the testing was carried out either as ACR or CCR tests. These two differ from each other in the methodology. ACR tests ask the listeners to assess the quality of each speech sample under test while CCR tests are based on asking the listeners to assess the quality differences between two samples. ACR and CCR tests are both well established and recognised speech quality testing methodologies.

The listening test laboratories performing the selection tests were: Arcon (English language), AT&T (Mandarin, Spanish and English), Nortel Networks (English), FUB (Italian), and COMSAT (French, Spanish and Japanese). All experiments and sub-experiments were carried out with 2 languages. The allocation of experiments to listening laboratories, and the languages used for each experiment, are shown in Table 7.1.

	Arcon	**AT&T**	**Nortel Networks**	**FUB**	**COMSAT**
	English	Mandarin Spanish, or English	English	Italian	French, Spanish, or Japanese
1			*X*		*Spanish*
2	*X*				*French*
3a	*X*			*X*	
3b	*X*			*X*	
3c	*X*			*X*	
4a			*X*		*Spanish*
4b			*X*		*Spanish*
5a			*X*		*Spanish*
5b			*X*		*Spanish*
6		*Spanish*	*X*		
7		*Spanish*	*X*		
8		*Mandarin, English*			
9		*Mandarin, English*			
10	*X*				*Japanese*
Host lab	ARCON	*COMSAT*	ARCON	*COMSAT*	*COMSAT*

Table 7.1: *Allocation of Experiments to Listening Laboratories*

The reference conditions were processed by Arcon and COMSAT, while the test samples were processed through the candidate algorithms by the candidate organisations themselves and were cross checked by other candidates. A simple blind procedure was followed to ensure that the test laboratories and the test subjects had no knowledge of the test conditions.

7.2 Summary of Listening Test Results Covering Minimum Performance Requirements

The candidates were ranked according to the number of simple and systematic failures (with the latter meaning failure of the same test condition in all tests performed for the same experiment).

All candidate algorithms failed to fulfil some of the minimum performance requirements. Table 7.2 records the number of failures for each candidate according to the minimum performance requirements as stated in Table 5.1 (excluding those not associated with listening tests).

Simple Failures (excluding noise suppression in the downlink)	5	6	9	9	9	13
	1. NS5	**2. NS2**	**3. NS3**	**3. NS4**	**3. NS6**	**6. NS1**
Systematic Failures (excluding noise suppression in the downlink)	2	2	2	2	2	4
	1. NS2	**1. NS3**	**1. NS4**	**1. NS5**	**1. NS6**	**6. NS1**

Table 7.2: *Failures per candidate using the Minimum Performance Requirements*

Additionally results for the number of failures are presented in Table 7.3 where the requirements of Table 5.1 are relaxed for Experiments 4-10 such that a failure is noted if a candidate is not found at least as good as the reference at the 95% confidence interval ("equal or better than" criterion).

Simple Failures (excluding noise suppression in the downlink)	0	2	3	5	5	7
	1. NS5	**2. NS2**	**3. NS3**	**4. NS4**	**4. NS6**	**6. NS1**
Systematic Failures (excluding noise suppression in the downlink)	0	0	0	0	0	1
	1. NS2	**1. NS3**	**1. NS4**	**1. NS5**	**1. NS6**	**6. NS1**

Table 7.3: *Failures per candidate using Relaxed Performance Requirements*

It can be readily seen (Annex C) that all candidates have systematic failures in Experiment 10 for two special noise types: music noise and multiple interfering talkers. For additional information, the calculation of the number of failures was carried out also for the case when music noise and multiple interfering talker noise (in Experiment 10) are excluded in the analysis, but where otherwise the Minimum Performance criteria of Table 5.1 are applied. This is justified by noting that it is not at all clear whether noise suppression functionality should attempt to suppress such background signals.

Simple Failures (excluding noise suppression in the downlink)	1	2	5	5	5	9
	1. NS5	**2. NS2**	**3. NS3**	**3. NS4**	**3. NS6**	**6. NS1**
Systematic Failures (excluding noise suppression in the downlink)	0	0	0	0	0	2
	1. NS2	**1. NS3**	**1. NS4**	**1. NS5**	**1. NS6**	**6. NS1**

Table 7.3: *Failures per candidate excluding Conditions with Music and Interfering Talkers*

7.3 Summary of Listening Test Results Covering Comparison of Candidates

This summary is presented in the form of tables of the FOMs defined in Section 6. It should be noted that SMG11 had stated that FOM#1 is the preferred Figure of Merit. This measures the ability of the solutions to suppress noise in terms of resulting speech quality compared to the non-suppressed speech.

Having said this, it is clear that other Figures of Merit are significant. In particular note should be taken of FOM#7, derived from Experiment 3 which is used to detect unnatural effects in the noise-suppressed signal.

FOM#1	21.0962	17.0745	15.9298	15.5055	12.0572	12.0193
	NS6	**NS4**	**NS2**	**NS1**	**NS3**	**NS5**
FOM#3a	8.2708	6.8802	5.8854	5.7708	4.8906	4.7969
	NS6	**NS4**	**NS1**	**NS2**	**NS5**	**NS3**
FOM#3c	12.8253	10.1943	10.1590	9.6201	7.2603	7.1287
	NS6	**NS4**	**NS2**	**NS1**	**NS3**	**NS5**
FOM#4a	2.4167	1.9531	1.9167	1.6510	1.6302	1.5156
	NS6	**NS2**	**NS4**	**NS1**	**NS5**	**NS3**
FOM#4b	18.6795	15.1578	13.9767	13.8545	10.5416	10.3801
	NS6	**NS4**	**NS2**	**NS1**	**NS3**	**NS5**
FOM#6a	53.0826	43.9323	40.3828	39.0471	31.3767	27.6231
	NS6	**NS4**	**NS1**	**NS2**	**NS3**	**NS5**
FOM#6b	49.9855	41.8609	40.0272	38.9033	30.9570	24.0992
	NS6	**NS4**	**NS1**	**NS2**	**NS3**	**NS5**
FOM#6c	57.7283	47.0396	40.9161	39.2628	32.9089	32.0062
	NS6	**NS4**	**NS1**	**NS2**	**NS5**	**NS3**
FOM#7a	6.5104	5.3229	4.8125	3.8021	3.4688	3.2917
	NS5	**NS2**	**NS4**	**NS3**	**NS6**	**NS1**
FOM#7b	4.2432	-2.3479	-9.1099	-20.4534	-31.8353	-42.9414
	NS5	**NS2**	**NS3**	**NS4**	**NS1**	**NS6**
FOM#10	3.6177	-4.0464	-4.1682	-8.8849	-12.2269	-16.1904
	NS5	**NS2**	**NS3**	**NS4**	**NS1**	**NS6**

7.4 Graphical Representation of Results from all Formal Listening Tests

This section provides the results of all 14 sub-experiments from all labs in graphical form. For more detailed information see Annex C. Note these graphs have been imported directly from the Global Analysis spreadsheet, and therefore also contain data for noise suppression in tandem in the uplink and downlink (which formed part of the feasibility study).

The following abreviations are used in conjunction with these graphs:

@x	Defined bit rate of AMR speech codec
AMR/NS	AMR with noise suppression active
DL	Downlink
T1	Single Connection, i.e. noise suppression present in the uplink only
T2	Tandem Connection, i.e. noise suppression present in the uplink and the downlink
UL	Uplink
w/DTX	with VAD/DTX active

7.4.1 Experiment 2: Degradation in Clean Speech (pair comparison test)

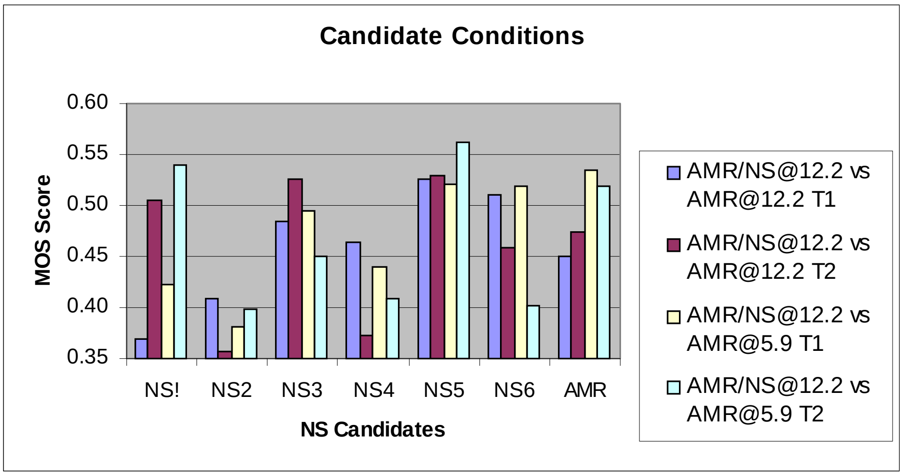

Figure 7.1: *Experiment 2 Results: English Language*

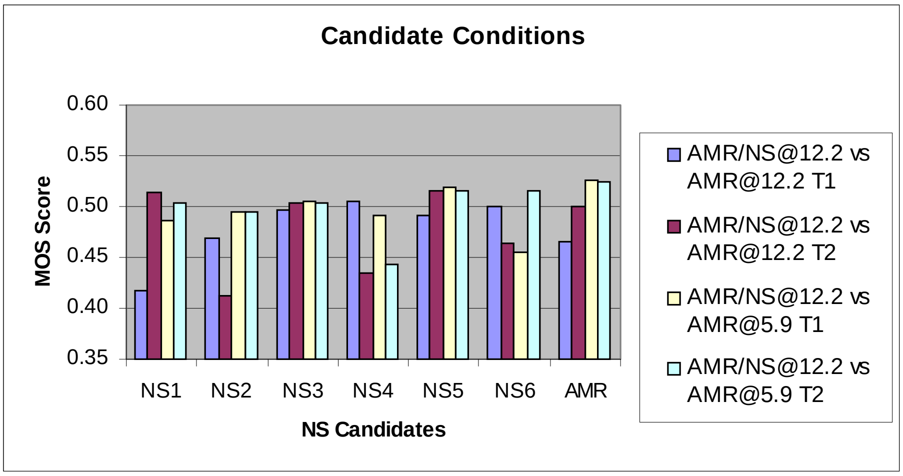

Figure 7.2: *Experiment 2 Results: French Language*

7.4.2 Experiment 3: Artifacts and Clipping in Background Noise

7.4.2.1 Car Noise

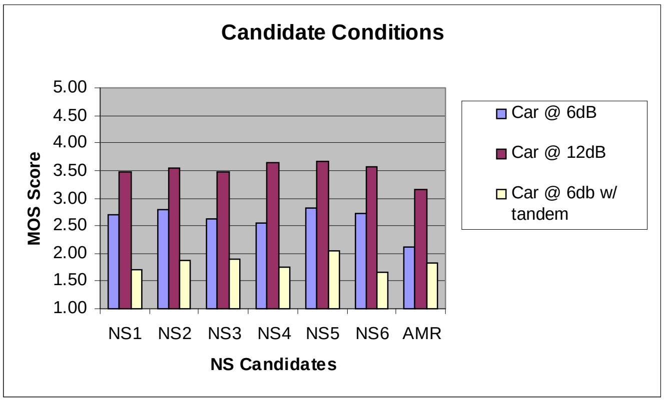

Figure 7.3: *Experiment 3 Results: Car Noise, English Language*

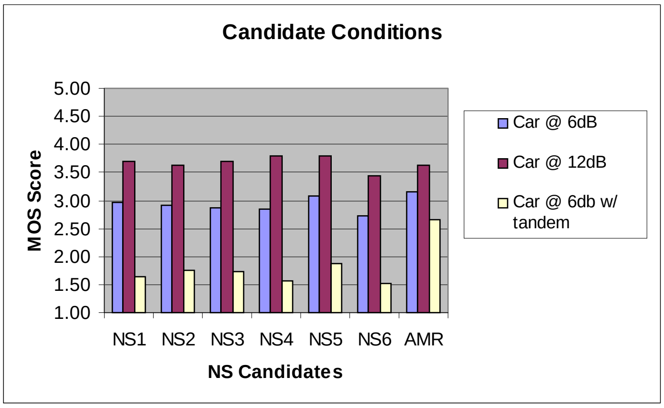

Figure 7.4: *Experiment 3 Results: Car Noise,* Italian *Language*

7.4.2.2 Street Noise

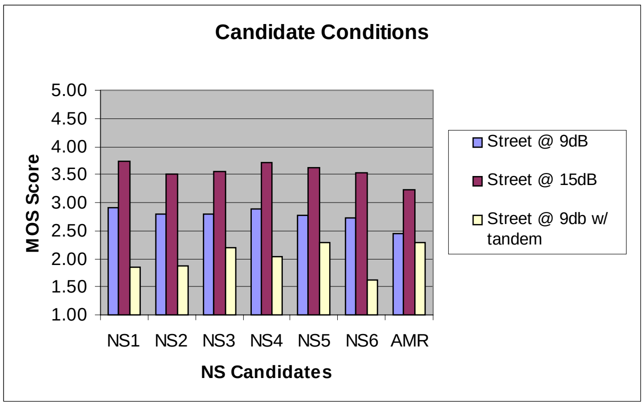

Figure 7.5: *Experiment 3 Results:Street Noise, English Language*

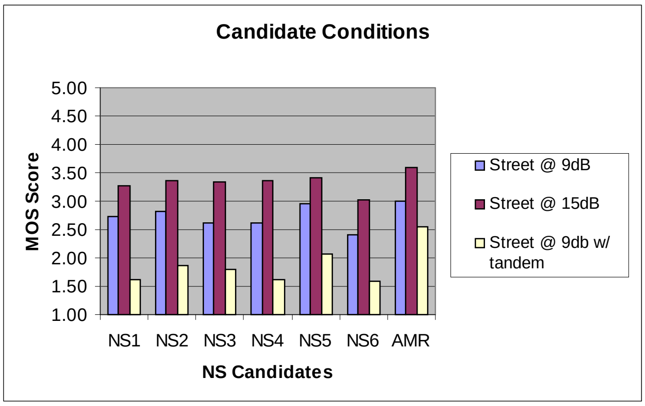

***Figure 7.6:** Experiment 3 Results: Street Noise,* Italian *Language*

7.4.2.3 Babble Noise

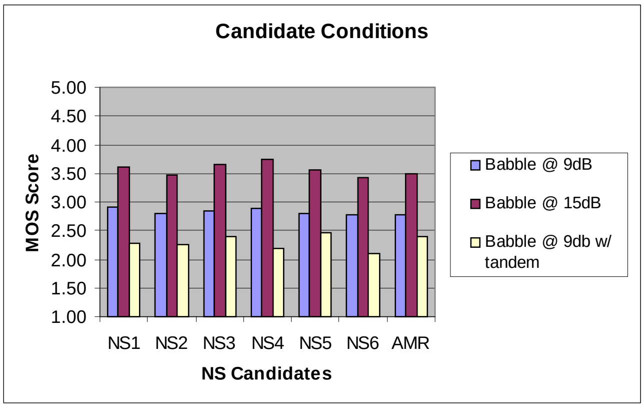

***Figure 7.7:** Experiment 3 Results:Babble Noise, English Language*

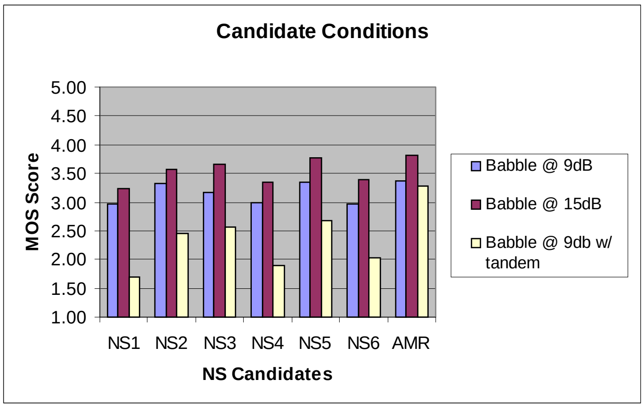

***Figure 7.8:** Experiment 3 Results: Babble Noise,* Italian *Language*

7.4.3 Experiment 4: Performance in Background Noise (5.9kbps AMR Speech Codec)

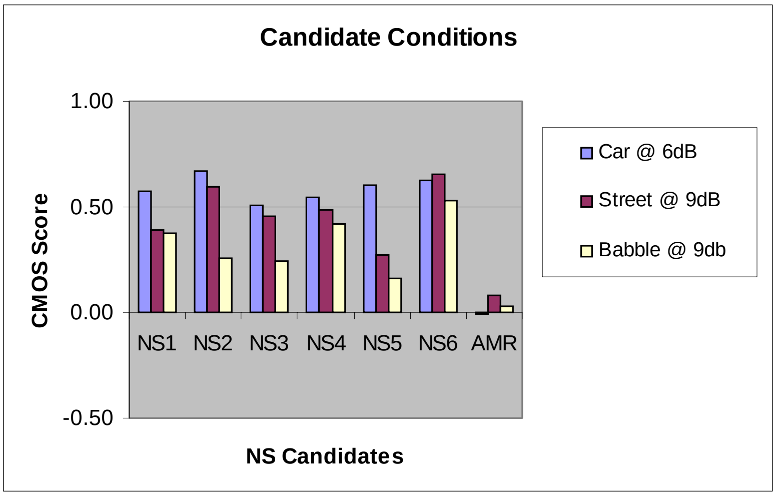

Figure 7.9: *Experiment 4 Results: Low SNR, English Language*

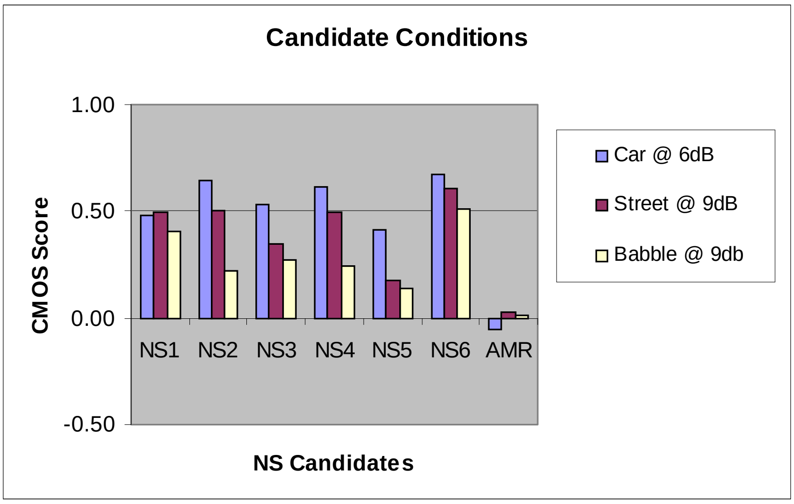

Figure 7.10: *Experiment 4 Results: Low SNR, Spanish Language*

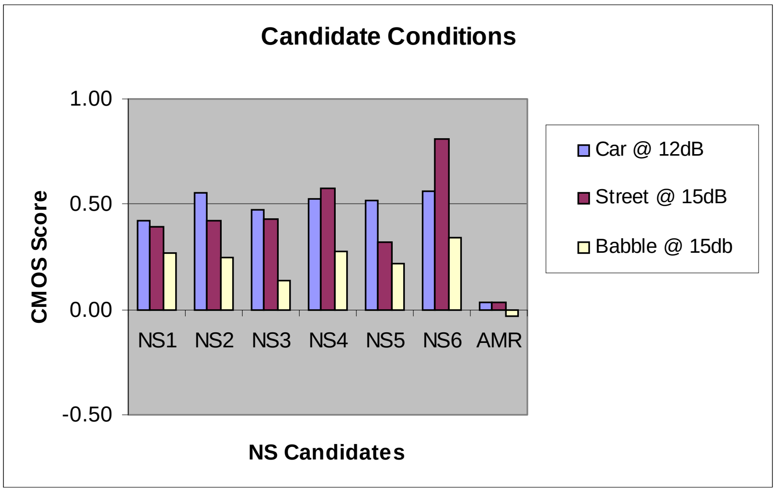

Figure 7.11: *Experiment 4 Results: High SNR, English Language*

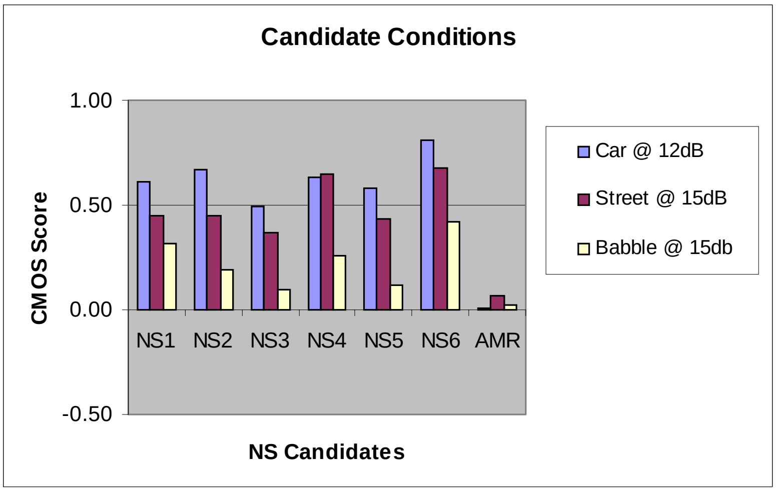

Figure 7.12: *Experiment 4 Results: High SNR, Spanish Language*

7.4.4 Experiment 5: Performance in Background Noise (12.2kbps AMR Speech Codec)

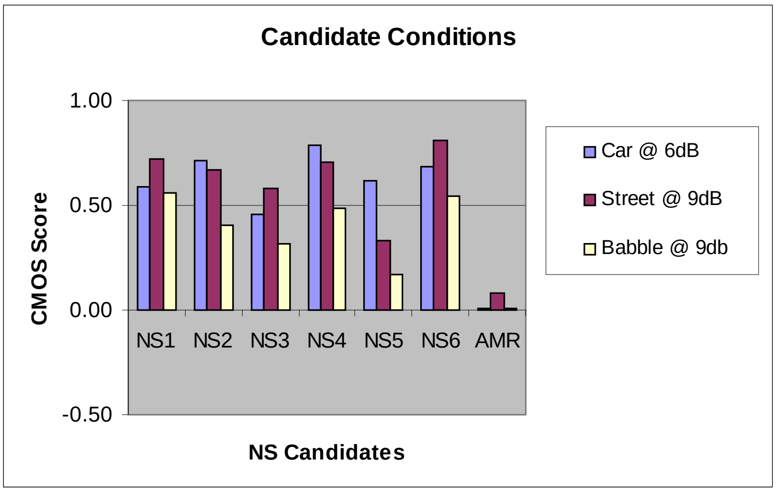

Figure 7.13: *Experiment 5 Results: Low SNR, English Language*

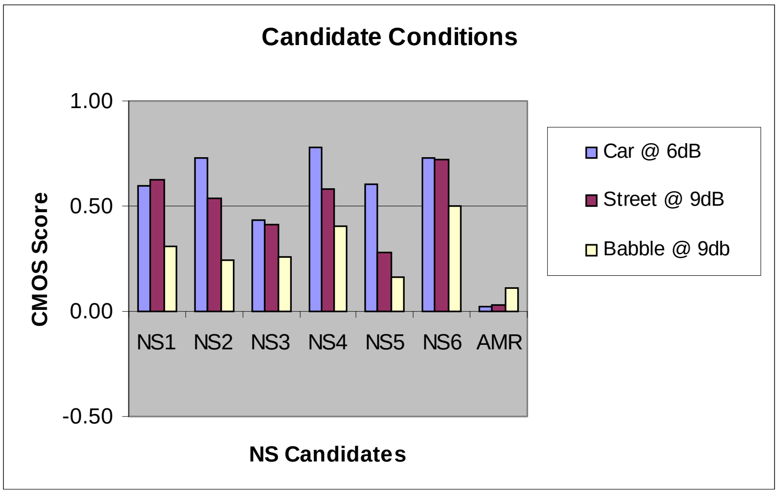

Figure 7.14: *Experiment 5 Results: Low SNR, Spanish Language*

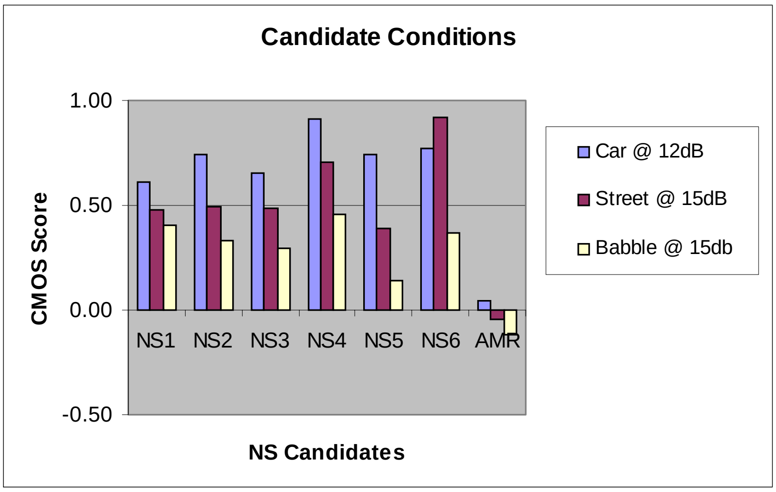

Figure 7.15: *Experiment 5 Results: High SNR, English Language*

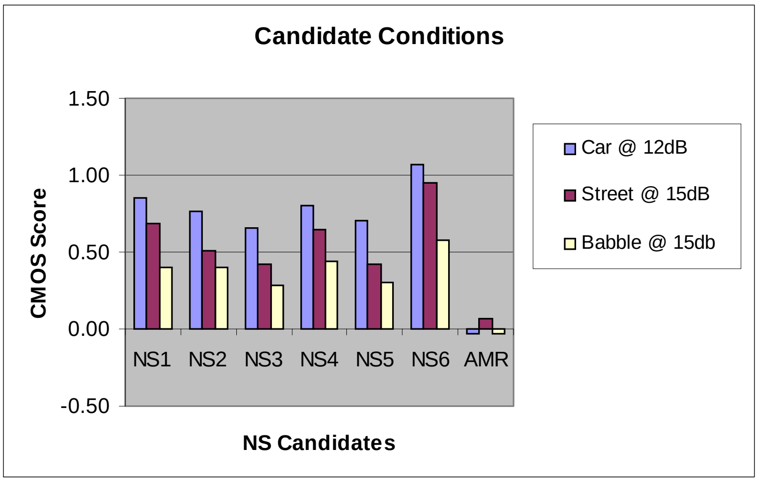

Figure 7.16: *Experiment 5 Results: High SNR, Spanish Language*

7.4.5 Experiment 6: Performance in Background Noise with Channel Errors (Car Noise with 6dB SNR)

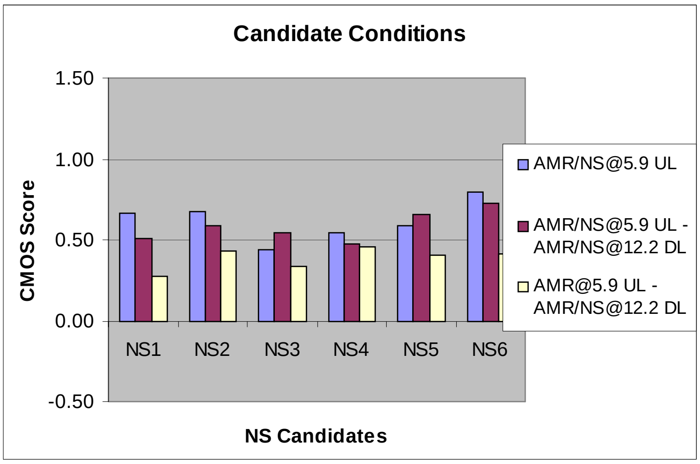

Figure 7.17: *Experiment 6 Results: English Language*

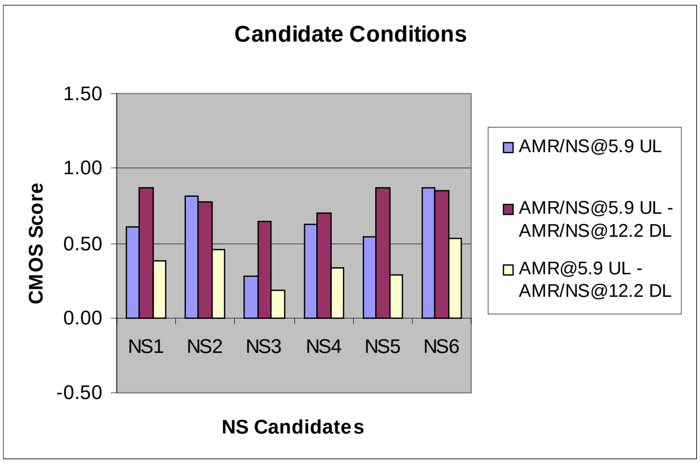

Figure 7.18: *Experiment 6 Results: Spanish Language*

7.4.6 Experiment 7: Performance in Background Noise with Channel Errors (Street Noise with 9dB SNR)

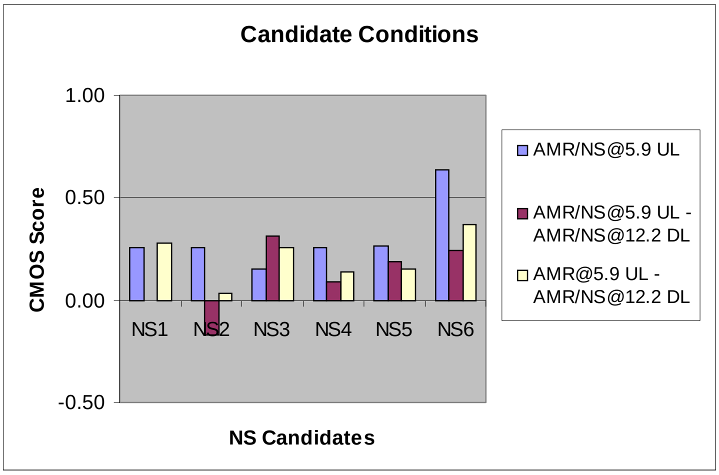

Figure 7.19: *Experiment 7 Results: English Language*

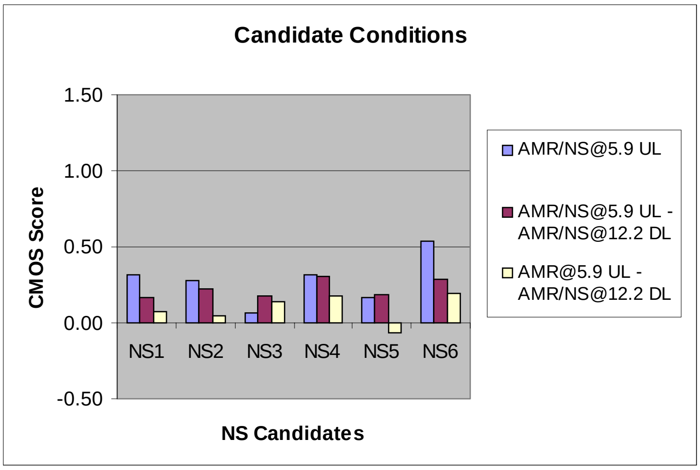

Figure 7.20: *Experiment 7 Results: Spanish Language*

7.4.7 Experiment 8: Performance in Car Noise with VAD/DTX active (VAD Option 1)

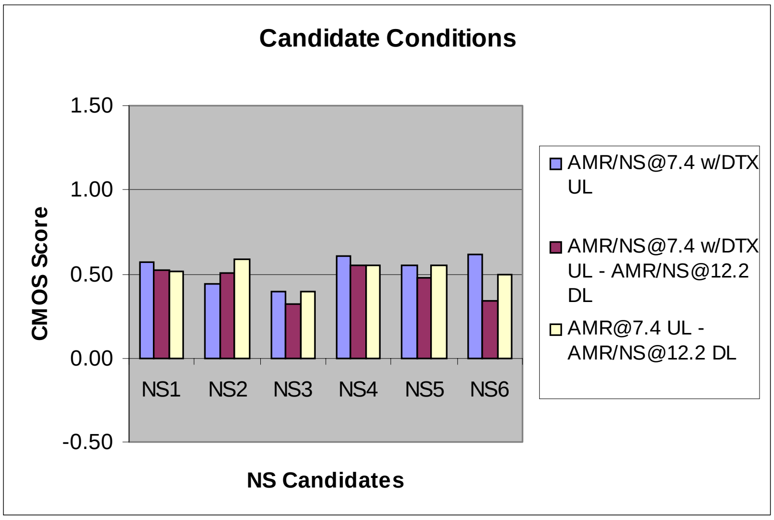

Figure 7.21: *Experiment 8 Results: English Language*

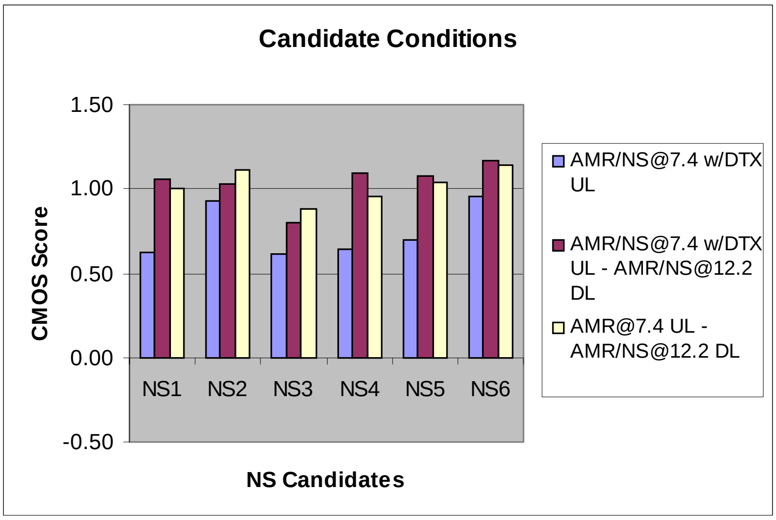

Figure 7.22: *Experiment 8 Results: Mandarin Language*

7.4.8 Experiment 9: Performance in Street Noise with VAD/DTX active (VAD Option 2)

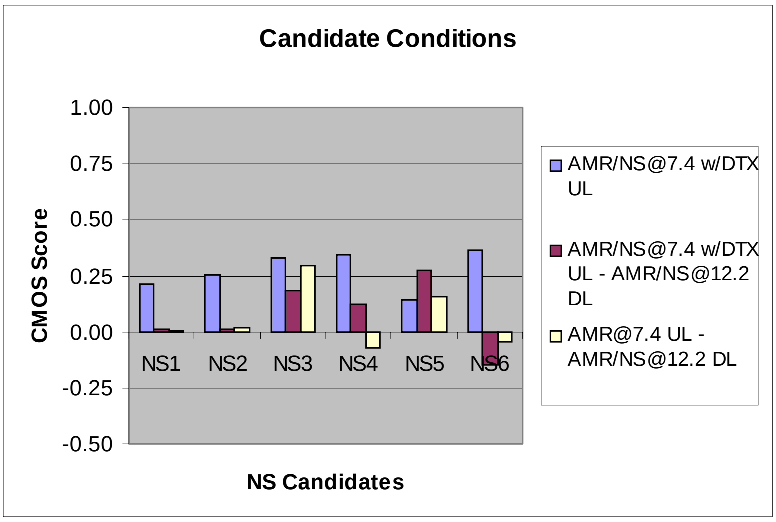

Figure 7.23: *Experiment 9 Results: English Language*

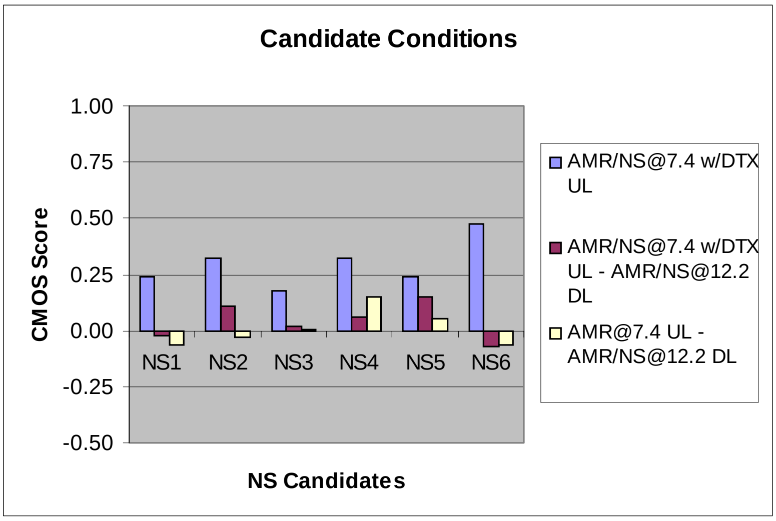

Figure 7.24: *Experiment 9 Results: Mandarin Language*

7.4.9 Experiment 10: Influence of Input Signal Level and Special Noise Types.

7.4.9.1. Influence of Input Level

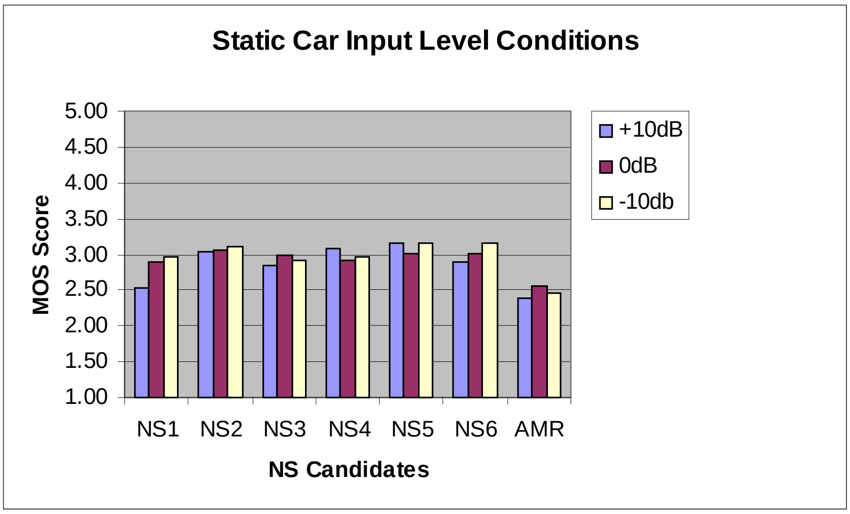

Figure 7.25: *Experiment 10 Results: Effect of Input Level, Car Noise, English Language*

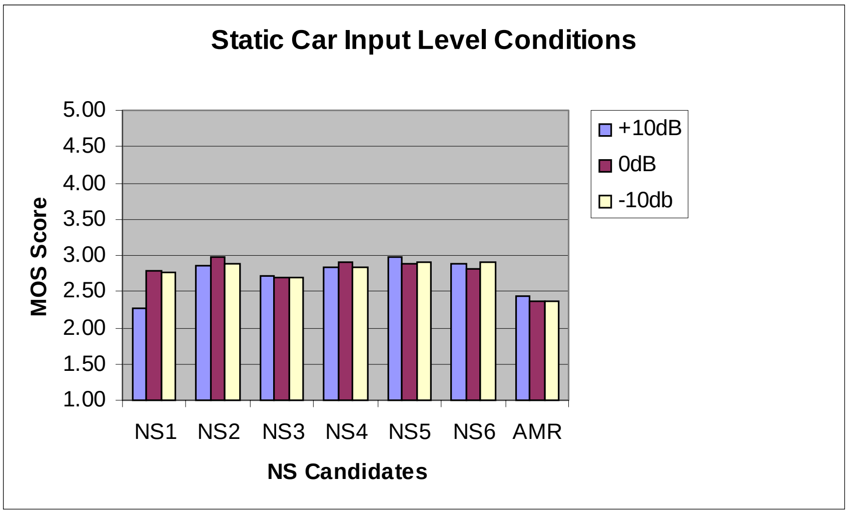

Figure 7.25: Experiment 10 Results: Effect of Input Level, Car Noise, Japanese Language

7.4.9.2 Performance wirh Special Noise Types

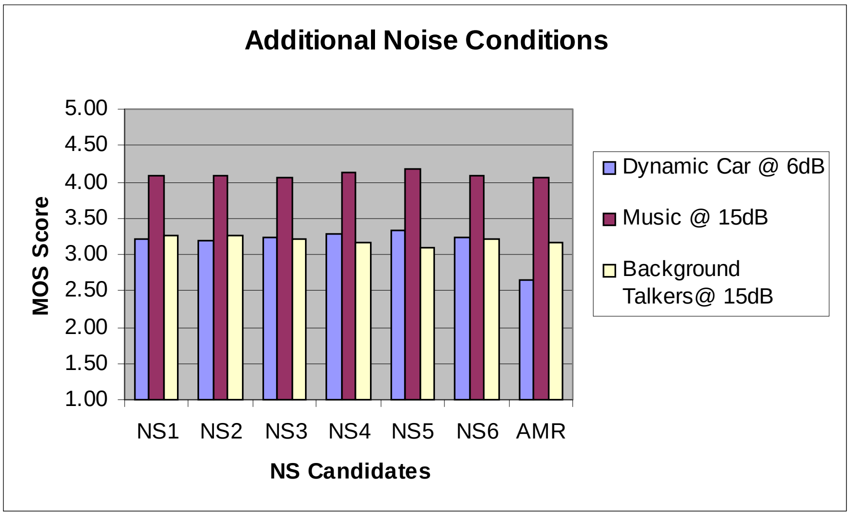

Figure 7.27: *Experiment 10 Results:Special Noises, English Language*

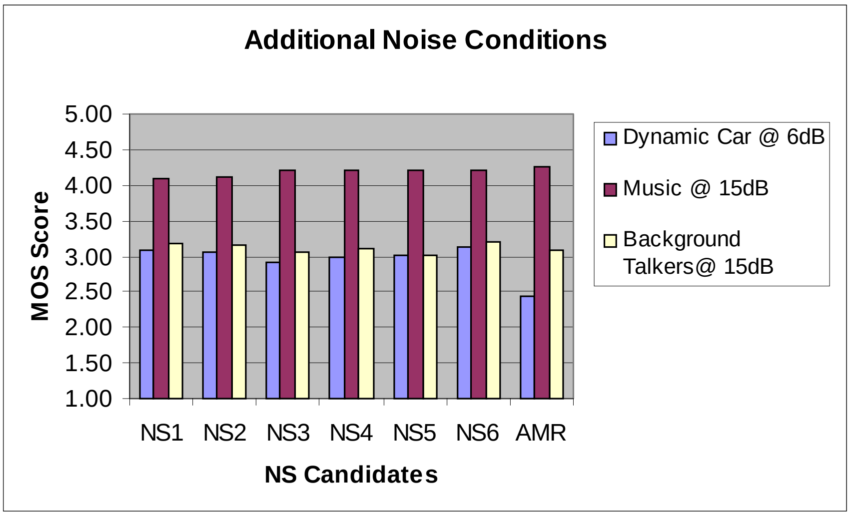

Figure 7.28: *Experiment 10 Results:Special Noises, Japanese Language*

8 Design Constraints

This section summarises the design constraints (limits on complexity, delay) and details the related values for all the candidates who took part in the Selection Phase.

Both the requirements (limits) and values for each candidate are provided in the Table 8.1.

In the context of this table, the following definitions are made. The DSP that runs the algorithm has been modelled through three parameters E, S and P. **E** stands for the Efficiency of the DSP. This corresponds to the ratio TMOPS/WMOPS of the implementation of the codec on the DSP. **S** stands for the Speed of the DSP: Maximum Number of Operations that the DSP can run in 1 second. This number is expressed in MOPS. **P** stands for the percentage of DSP processing power assigned to the codec. The processing delay of a task whose complexity is X can then be computed using the formula: D = X*20/ESP, the time unit being ms.

	NS1	NS2	NS3	NS4	NS5	NS6	Requirement
WMOPS	2,910	3,386	2,432	3,623	4,472	3,934	5,000
Dynamic RAM (words)	*770*	*2234*	*781*	*768*	*1529*	*1073*	*3039*
Static RAM (words)	262	718	168	577	850	239	1500
Data ROM (words)	312	863	302	731	877	537	1000
Program ROM (basic ETSI operations)	754	772	1018	907	884	581	2000
Delay (ms)	5,00	5,00	0,00	2,00	0,00	5,00	7,00
Delay-stand alone (ms)		5,00	1,50	10,75	0,00	5,00	
Implementation	*embedded*	*stand alone*	*embedded*	*embedded*	*stand alone*	*embedded*	
FOM(1)	5,72	8,49	5,38	8,33	10,34	6,48	15,80
FOM(2)@ESP25	7,33	7,71	1,95	4,90	3,58	8,15	11,00
FOM(2)@ESP50	6,16	6,35	0,97	3,45	1,79	6,57	9,00
FOM(2)@ESP100	5,58	5,68	0,49	2,72	0,89	5,79	8,00

FOM(1) = WMOPS + 2*sRAM + (2/5)*dROM + 2*pROM **sRam, dROM in kbytes, pROM in kbasic ETSI ops**

FOM(2) = delay(proc) + delay(algor) **delay(proc) = WMOPS * 20 /(E*S*P))** [i]**; in ms**

Table 8.1: *Summary of Design Constraints Information.*

9 Impact on Voice Activity Factor VAF (with VAD/DTX active)

The Selection Phase Requirement concerning impact on VAD/DTX stated that the AMR speech codec with noise suppression activated should not significantly increase channel activity when used in conjunction with DTX.

Table 9.1 details the VAF increase for each candidate for each VAD option, as an average across all tested speech plus noise samples. In this table a positive value denotes an increase in VAF, whereas a negative value denotes a decrease in VAF.

Candidate	NS1	NS2	NS3	NS4	NS5	NS6
VAF increase for VAD Option 1 (%)	+1.77	+2.72	+0.11	-0.79	+0.20	+0.68
VAF increase for VAD Option 2 (%)	+0.09	+0.30	+0.00	-2.22	-0.42	+0.03

Table 9.1: *Summary of Impact on VAF*

10 Objective Performance Measurements

A tool was used to generate objective measures of performance (in terms of speech quality). This information is regarded as additional, and is in all cases secondary to the results obtained by subjective listening (as reported in Section 7). Two measures were undertaken on a subset of the material utilised in the listening tests. These were Noise Power Level Reduction (NPLR) and Signal to Noise Improvement (SNRI). Further details can be found in Annex E. The following tables provide the results of the analysis for each candidate.

[Results tables to be added]

11 Feasibility Study: Downlink Noise Suppression for AMR

During the selection testing of the NS candidates, conditions including the noise suppression algorithm in the downlink path were tested. The aim was to assess the feasibility of putting the noise suppression algorithm in the network on the downlink path. Because the selection process was focused on the uplink, those conditions were not taken into account in the selection results. However, results are available and are noted here.

It was decided not to test the downlink path in isolation to avoid doubling the amount of testing required. Moreover, to perform a fair comparison, no different tuning of algorithm behaviour was allowed between the downlink and the uplink noise suppression algorithms.

[Detail to be added]

Annex A: Key Selection Phase Documents

Annex B: Selection Phase Test Plan

See associated file [TBA]

Annex C: Global Analysis Spreadsheet

See associated file [TBA]

Annex D: Methodologies for Measuring Subjective SNR Improvement

Annex E: Methodology for Speech Quality by Objective Means

Annex F: Methodology for Measuring Impact on Voice Activity Factor (VAF)

Annex G: List of Key SMG11 Documents

Annex ? (informative): Change Request History

SMG#	Tdoc SMG	Spec	CR	Cat	PH	Vers	New Vers	Subject
sa6	570/99	06.75	A001	F	R98	7.0.0	7.1.0	Update of AMR Transmission Delay Figures

History

Document history		
v0.1	February 2000	First Draft