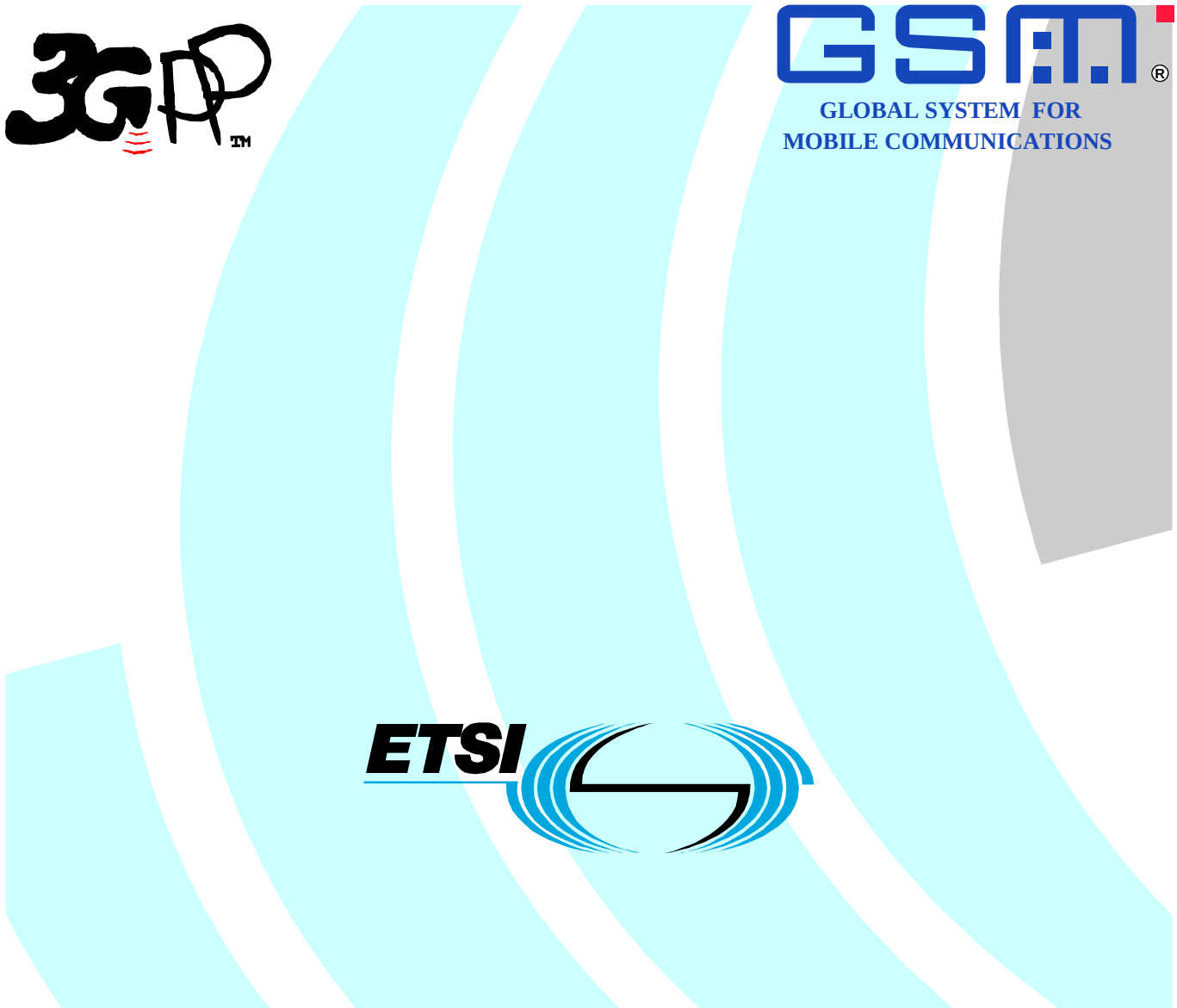
ETSI TR 126 976 V5.1.0 (2003-09)

Technical Report

**Digital cellular telecommunications system (Phase 2+);
Universal Mobile Telecommunications System (UMTS);
Performance characterization of the Adaptive
Multi-Rate Wideband (AMR-WB) speech codec
(3GPP TR 26.976 version 5.1.0 Release 5)**



Reference
RTR/TSGS-0426976v510

Keywords
GSM, UMTS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - 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

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 ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

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 ETSI 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 Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp .

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

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

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

The present document provides information of the AMR Wideband (AMR-WB) Characterisation, Verification and Selection Phases. Experimental test results from the speech quality related testing are reported to illustrate the behaviour of the AMR-WB codec. Additional information is provided, e.g., on implementation complexity of the AMR-WB codec. Also the verification results for the floating-point version of the AMR-WB codec (3GPP TS 26.204) are presented.

2 References

The following documents contain provisions, which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document in the same Release as the present document.

[1] 3GPP TR 26.901: "AMR wideband speech codec; Feasibility study report (Release 4)".

[2] Tdoc SP-99060: "Proposed TSG-S4 Work Items for approval", 3GPP TSG-SA meeting #2, 2-4 March, 1999 (Fort Lauderdale, USA).

[3] Tdoc SP-99354: "Common WI description for the Wideband Codec", 3GPP TSG-SA meeting #5, 11-13 October, 1999 (Kjongju, South Korea).

[4] Tdoc SP-000259: "AMR Wideband Speech Codec Qualification Phase Report", 3GPP TSG-SA#8, 26-28 June, 2000 (Dusseldorf, Germany)

[5] Tdoc SP-000555: "Results of AMR Wideband (AMR-WB) Codec Selection Phase", 3GPP TSG-SA, Bangkok, Thailand, December 2000.

[6] Tdoc S4-000321: "Permanent Project Document: AMR Wideband Performance Requirements (WB-3, version 2.2)", 3GPP TSG-S4.

[7] Tdoc S4-000508: "Permanent Project Document: Selection Rules for AMR-WB (WB-5b, version 1.1)", 3GPP TSG-S4.

[8] Tdoc S4-000340: "Permanent Project Document: Design Constraints (WB-4, version 1.3)", 3GPP TSG-S4.

[9] Tdoc S4-000427: "Permanent Project Document: AMR Wideband Codec Development Project Deliverables for the Selection Test (WB-6b, version 2.0)", 3GPP TSG-S4.

[10] Tdoc S4-000382: "Permanent Project Document: AMR-WB Selection Test Plan (WB-8b, version 1.0)", 3GPP TSG-S4.

[11] Tdoc S4-000389: "Permanent Project Document: Processing Functions for WB-AMR Subjective Experiments (WB-7, v.1.0)", 3GPP TSG-S4.

[12] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[13] Tdoc S4-010463: "Test Plan for the AMR Wideband Characterisation Phase 1 v.1.2", 3GPP TSG-S4.

[14] Tdoc S4-010008: "Complexity verification report of the AMR-WB codec", 3GPP TSG-S4.

[15] Tdoc S4-010393: "Results of cross-language comparisons for Experiments 1, 2 and 5 of the AMR-WB Characterisation Phase 1A", 3GPP TSG-S4.

[16] Tdoc S4-010021: "DTMF transparency of the AMR-WB speech codec", 3GPP TSG-S4.

[17] Tdoc S4-010050: "AMR WB Verification: Switching Performance Between AMR WB and AMR", 3GPP TSG-S4.

[18] Tdoc S4-010052: "Verification of the delays for the Wideband AMR codec", 3GPP TSG-S4.

[19] Tdoc S4-010158: "WB-AMR Verification results: Performance with music signals (expert Listening tests)", 3GPP TSG-S4.

[20] Tdoc S4-010228: "AMR-WB Verification: Special input signals", 3GPP TSG-S4.

[21] Tdoc S4-010230: "AMR-WB verification: Testing of Comfort Noise Generation System", 3GPP TSG-S4.

[22] Tdoc S4-010379: "AMR-WB verification : frequency response", 3GPP TSG-S4.

[23] Tdoc S4-010608: "AMR-WB verification: Signaling Tones", 3GPP TSG-S4.

[24] Tdoc S4-010040: "AMR Wideband Verification Phase - Muting Behaviour", 3GPP TSG-S4.

[25] Tdoc S4-010330: "AMR WB Verification: Overload Performance", 3GPP TSG-S4.

[26] Tdoc S4-020049r1: "Verification of floating-point implementation of AMR-WB using Wideband-PESQ", 3GPP TSG-S4.

[27] Tdoc S4-020124: "Addendum to Verification of floating-point implementation of AMR-WB using Wideband-PESQ", 3GPP TSG-S4.

[28] Tdoc S4-020270: "AMR-WB Floating-Point Verification: VAD and Comfort Noise Performance", 3GPP TSG-S4.

[29] Tdoc S4-020113: "AMR WB Floating-point C-Code Verification: Overload Performance", 3GPP TSG-S4.

[30] Tdoc S4-020080: "AMR-WB Floating-Point Verification: Music Performance (Expert Listening Tests)", 3GPP TSG-S4.

[31] Tdoc S4-020079: "AMR-WB Floating-Point Verification: Bit-Exactness, Idle-Channel Behavior and Long-Term Stability Performance", 3GPP TSG-S4.

[32] Tdoc S4-020114: "Transparency of AMR-WB (Floating-Point) Codec for DTMF signals", 3GPP TSG-S4.

[33] Tdoc S4-020077: "Verification of AMR-WB floating point", 3GPP TSG-S4.

[34] Tdoc S4-020062: "Verification results of the AMR-WB floating-point codec", 3GPP TSG-S4.

[35] Tdoc S4-020064: "Subjective test results of the AMR-WB floating-point codec", 3GPP TSG-S4.

[36] Tdoc S4-010230: "AMR-WB verification: Testing of Comfort Noise Generation System", 3GPP TSG-S4.

[37] TSG S4#12(00): "Processing Functions for WB-AMR Subjective Experiments", Annex A, 3GPP TSG-S4#12(00).

[38] Tdoc 304/98: "On the Performance of proposed AMR VAD", ETSI SMG11.

[39] Q.7/16: "Subjective Characterization Test Plan for the ITU-T Wideband (7 kHz) Speech Coding Algorithm around 16 kbit/s"; Version 0.7, March 29, 2002.

[40] Q.7/12: "Report of the Global Analysis Laboratory for the ITU-T Q.7/16 Wideband Characterization Test", Geneva, 27 – 31 May 2002.

[41] Tdoc GP-020152: "Channel coding for O-TCH/WFS and O-TCH/WHS: High Level Description", 3GPP TSG-GERAN.

[42] Tdoc GP-020153 "Channel coding for O-TCH/WFS and O-TCH/WHS: Listening Test Plan", 3GPP TSG-GERAN.

[43] Tdoc GP-031432 "Listening Test Results for AMR-WB", 3GPP TSG-GERAN.

[44] Tdoc GP-020155 "Channel coding for O-TCH/WFS and O-TCH/WHS: Objective Measurements", 3GPP TSG-GERAN.

[45] Tdoc GP-020156: "CR 45.003-016 Channel coding for O-TCH/WFS and O-TCH/WHS", 3GPP TSG-GERAN.

[46] ITU-T Recommendation Q.23: "Technical features of push-button telephone sets".

[47] ITU-T Recommendation Q.24: "Multifrequency push-button signal reception".

[48] ITU-T Recommendation G.191: "Software tools for speech and audio coding standardization".

[49] ITU-T Recommendation G.711: "Pulse code modulation (PCM) of voice frequencies".

[50] ITU-T Recommendation G.722: "7 kHz audio-coding within 64 kbit/s".

[50a] ITU-T Recommendation G.722.1: "Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss".

[50b] ITU-T Recommendation G.722.2: "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-rate Wideband (AMR-WB)".

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

[52] ITU-T Recommendation E.180: "Technical characteristics of tones for the telephone service".

[53] ITU-T Recommendation P.862: "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs".

[54] 3GPP TS 26.173: "ANSI-C code for the Adaptive Multi Rate (AMR) Wideband speech codec".

[55] 3GPP TS 26.174: "AMR speech codec, wideband; Test sequences".

[56] 3GPP TS 26.204: "ANSI-C code for the floating-point Adaptive Multi-Rate (AMR) wideband speech codec".

[57] 3GPP TS 26.190: "Mandatory Speech Codec speech processing functions AMR Wideband speech codec; Transcoding functions".

[58] 3GPP TS 26.191: "AMR speech codec, wideband; Error concealment of lost frames".

[59] 3GPP TS 26.192: "Mandatory Speech Codec speech processing functions AMR Wideband Speech Codec; Comfort noise aspects".

[60] 3GPP TS 26.193: "AMR speech codec, wideband; Source Controlled Rate operation".

[61] 3GPP TS 26.194: "Mandatory Speech Codec speech processing functions AMR Wideband speech codec; Voice Activity Detector (VAD)".

3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [12] and the following apply:

ACR	Absolute Category Rating
AMR	Adaptive Multi-Rate
AMR-WB	Adaptive Multi-Rate Wideband
C/I	Carrier-to-Interfere ratio
CCR	Comparison Category Rating
CI	Confidence Interval
CMOS	Comparison MOS
DCR	Degradation Category Rating
DMOS	Differential MOS
DTMF	Dual Tone Multi Frequency
DTX	Discontinuous Transmission for power consumption and interference reduction
EDGE	Enhanced Data rates for GSM Evolution
EFR	Enhanced Full-Rate
ETSI	European Telecommunication Standards Institute
FoM	Figure of Merit
FR	Full-Rate
G.722	ITU 48/56/64kbit/s wideband codec
G.722-48k	ITU 48 kbit/s wideband codec
G.722-56k	ITU 56 kbit/s wideband codec
G.722-64k	ITU 64kbit/s wideband codec
GBER	Average gross bit error rate
GERAN	GSM/EDGE Radio Access Network
GSM	Global System for Mobile communications
HR	Half-Rate
ITU-T	International Telecommunication Union – Telecommunications Standardisation Sector
MNRU	Modulated Noise Reference Unit
MOPS	Million of Operation per Seconds
MOS	Mean Opinion Score
PoW	Poor or Worse
PSK	Phase Shift Key
SMG	Special Mobile Group
TSG-SA	Technical Specification Group - Service and System Aspects
SA4	Service and System Aspects Working Group 4 (TSG-SA WG4)
SNR	Signal To Noise Ratio
TFO	Tandem Free Operation
TSG	Technical Specification Group
UMTS	Universal Mobile Telecommunication System
UTRAN	Universal Terrestrial Radio Access network
VAD	Voice Activity Detection
wMOPS	weighted Million of Operations per Seconds

4 General

4.1 Project history

The possibility to develop a wideband speech codec for GSM, with audio bandwidth up to 7 kHz instead of 3.4 kHz, was noted already during the feasibility study of the (narrowband) Adaptive Multi-Rate (AMR) codec. When the AMR codec standardisation was launched at ETSI SMG#23 in October 1997, the work was focused on developing narrowband coding. Wideband coding was set as a possible longer-term target.

ETSI SMG11 then carried out a feasibility study on wideband coding by June 1999. The results showed that wideband coding is feasible for mobile communication for the applicable bit-rates and error conditions. The feasibility study considered development of wideband coding not only for GSM Full-Rate channel, but also for GSM EDGE channels, and for UMTS [1].

3GPP TSG-SA approved a work item on UMTS wideband coding at TSG-SA#2 in March 1999 [2]. This took place couple of months before the end of the wideband feasibility study in ETSI SMG11. However, the effective start of the work was pending on the results of SMG11 feasibility study. Upon finalisation of the feasibility study, the wideband codec development and standardisation work was started. The work was carried out jointly by SA4 and SMG11 under a common SA4/SMG11 work item. The common harmonised WI description was approved in ETSI SMG#29 (June 1999) and in TSG-SA#5 (October 1999) [3].

The codec selection was carried out as a competitive selection process consisting of two phases: a Qualification (Pre-Selection) Phase and a Selection Phase. The Qualification Phase was carried out by June 2000 and the Selection Phase from July to October 2000. From altogether nine codec candidates, seven codecs were submitted for the Qualification Phase. One candidate was later withdrawn and the remaining six codecs were accepted at TSG-SA#8 in June 2000 to proceed into the Selection Phase [4]. After that two codec proponents joined their codec development effort reducing the number of codec candidates to five for the Selection Phase. The codecs that participated into the Selection Phase came from Ericsson, FDNS consortium (consisting of France Télécom, Deutsche Telekom, Nortel Networks and Siemens), Motorola, Nokia and Texas Instruments.

The Selection Phase results were reviewed, analysed and debated during SA4#13 in October 2000. A recommendation for the Nokia codec candidate to be selected was made [5]. The selection phase results and the codec selection were approved at TSG-SA#10 in December 2000 completing the development and selection of the wideband codec.

The completion of the codec standardisation development included also Verification Phase whose results are reported in this technical report. The phase was conducted in order to check the correctness of the code and behaviour in special conditions. Also, detailed analysis of the implementation complexity and transmission delay was performed during this phase. Verification was carried out, for most parts, by TSG-SA#11 in March 2001.

The Characterisation Phase is the latest phase. During this phase the codec was tested in a more complete manner than in the selection phase. Characterisation will be completed by the end of the year 2002.

The selected codec fulfils the project targets. It met all speech quality requirements covered in the selection tests. No failures were found in any of the participated listening test laboratories in any of the tested conditions. The codec fulfils all the design constraints.

3GPP has also specified a floating-point version of the AMR-WB codec (3GPP TS 26.204). This work started in the end of 2001 and was completed by TSG-SA#15 in March 2002.

4.2 Overview of the wideband codec work item

Wideband coding brings quality improvement over the existing narrowband telephony through the use of extended audio bandwidth. The AMR codec, standardised for GSM Release 98 and 3GPP Release 99, provides good performance for telephone bandwidth speech (audio bandwidth limited to 3.4 kHz). However, the introduction of a wideband speech service (audio bandwidth extended to 7 kHz) brings improved voice quality especially in terms of increased voice naturalness. Wideband coding brings speech quality exceeding that of (narrowband) wireline quality to 3G and GSM/GERAN systems.

The wideband codec was developed as a multi-rate codec consisting of several codec modes like the AMR codec. Consequently, the wideband codec is referred to as AMR Wideband (AMR-WB) codec. Like in AMR, the codec mode is chosen based on the operating conditions on the radio channel. Adapting coding depending on the channel quality provides high robustness against transmission errors. The codec also includes a source controlled rate operation mechanism, which allows it to encode speech at a lower average rate by taking speech inactivity into account.

The AMR-WB codec was developed to operate in the following multiple applications (see note):

- Application A: GSM full-rate traffic channel with an additional constraint of 16 kbit/s A-ter sub-multiplexing.
- Application B: GSM full-rate traffic channel.
- Application C: Circuit Switched EDGE/GERAN 8-PSK Phase II radio channels.

- Application E: 3G UTRAN WCDMA radio channel.

NOTE: Letter "D" was reserved for an intended GSM multi-slot application. However, this was not found needed and was withdrawn later during standardisation.

The codec mode can be changed every 20 ms in 3G WCDMA channels and every 40 ms in GSM/GERAN channels. (For Tandem Free Operation interoperability with GSM/GERAN, mode change rate is restricted in 3G to 40 ms in AMR-WB encoder.)

4.3 Presentation of the following clauses

The following clausess provide a summary of the Selection, Verification and Characterisation Phase test results, including a review of the performance requirements and selection criteria. Clause 5 defines the minimum performance requirements for speech quality. Clause 6 will give short summary of the experiments performed (and to be performed) during the characterisation and verification phases of testing. Clause 7 gives some guidance about interpretation of the subjective test results. Clauses 8-19 describe the results of the subjective listening tests undertaken during the characterisation phase. Clauses 19-30 contain results from the Verification Phase.

Annex A contains detailed information about the AMR-WB selection phase. In addition, Annex B contains results from the AMR-WB floating-point Verification Phase.

5 Performance requirements

The speech quality performance requirements are specified separately for each application.

In Application A, the general quality requirement is to be better than ITU-T Recommendation G.722 wideband codec at 48 kbit/s (G.722-48k). In Application B, quality equal to G.722-56k is required. For applications C and E a higher quality requirement is set requiring quality to be equal to G.722-64k. These are general requirements for clean channel performance (no transmission errors). Under the impact of background noise, relaxation is allowed in some cases (e.g. in Application A quality equal to G.722-48k is required in tandem conditions under background noise). In erroneous transmission, the codec should be robust against transmission errors. An illustrative diagram of the setting of quality requirements is given in figure 5.1 [4].

In Application A, the speech coding rate is restricted below 14.4 kbit/s, while in Application B rates up to the GSM FR transmission channel bit-rate of 22.8 kbit/s are possible. Due to this restriction, Application B can provide better maximum quality (at low error-rate conditions) than Application A.

The requirements are explained in more detail in Annex A. A full description of the performance requirements can be fo und in Permanent AMR-WB Project Document: Performance Requirements [6].

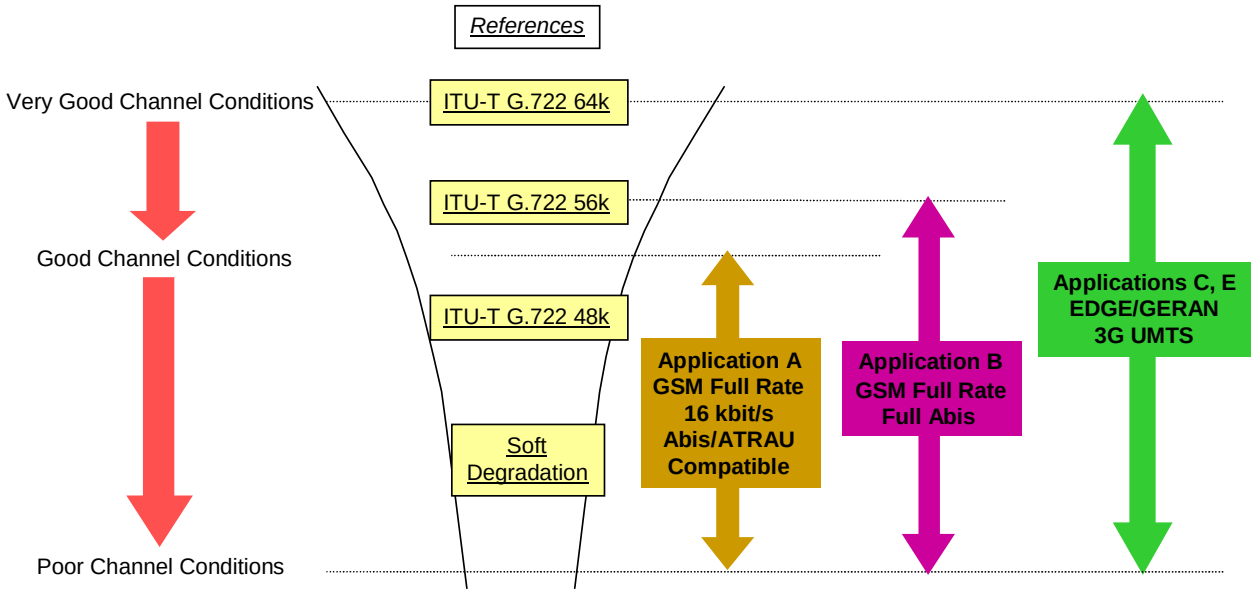


Figure 5.1: Quality requirements for the AMR-WB codec for the various applications [4].

6 Introduction to the testing of AMR-WB speech codec

6.1 AMR-WB Characterisation Phase

AMR-WB speech codec was characterised first by 3GPP and later by ITU, after it adopted AMR-WB speech codec as ITU standard G.722.2. Results from both tests are reported in this technical report.

The Characterisation Tests in 3GPP, consist of 8 main experiments, some of which contain a number of sub-experiments. Some experiments were tested twice with two different languages. For practical reasons some of the experiments were performed with one language. For example, experiments with different background noise types use only one language per noise type. The summary of the experiments is presented in table 6.1.

Table 6.1: Summary of 3GPP characterisation phase experiments

Exp.	Characterise:	Test	Title	Cond.	Languages
1	All systems	ACR	Input levels and self-tandeming	56	2
2	All systems	ACR	Interoperability Performance in Real World Wideband Scenarios	56	2
3	All systems	ACR	Interoperability Performance in Real World Narrowband Scenarios	56	1
4	All systems (GSM GMSK)	DCR	Performance of VAD/DTX/CNG Algorithm	40	1
5	GSM GMSK	ACR	The Effect of Static Errors under Clean Speech Conditions	48	2
6a	GSM GMSK	DCR	The Effect of Background Noise 1 in Static C/I Conditions	40	1
6b	GSM GMSK	DCR	The Effect of Background Noise 2 in Static C/I Conditions	40	1
7a	3G	ACR	The Effect of Static Errors under Clean Speech Conditions	56	1
7b	3G	ACR	The Effect of Static Errors under Clean Speech Conditions	56	1
8a	3G	DCR	The Effect of Background Noise 3 in Static C/I Conditions	48	1
8b	3G	DCR	The Effect of Background Noise 4 in Static C/I Conditions	48	1
8c	3G	DCR	The Effect of Background Noise 5 in Static C/I Conditions	48	1
			Total		**15**

3GPP Characterisation was carried out as a collaborative activity of several test laboratories. It was carried out based on a common test plan [13]. The testing was divided between several laboratories using different speech databases and languages. Special laboratories were allocated for host lab and cross-checking functions. The work division is described in table 6.2. Clauses 7-15 contain the complete set of test results for the AMR-WB speech codec Characterisation Phase, i.e. all systems (no channel errors) and GSM GMSK and 3G WCDMA channels.

Table 6.2: Allocation of listening and host laboratories to experiments

			Host Lab		Cross-check Lab	
Exp.	**Noise**	**Language**	**LMGT**	**ARCON**	**LMGT**	**ARCON**
1	Quiet	En/Fi	BT	NO	NO	BT
2	Quiet	En/Fr	LM	FT	FT	LM
3	Quiet	En	DY	-	-	DY
4	Ofc, Str, Car(15), Caf	En	NN	-	-	NN
5	Quiet	Fr/Ge	FT	DT	DT	FT
6a	Car(15)	En	LM	-	-	LM
6b	Ofc	Fi	-	NO	NO	-
7a	Quiet	Ge	-	DT	DT	-
7b	Quiet	En	BT	-	-	BT
8a	Car(10)	Ja	NA	-	-	NA
8b	Str	Sp	-	DY	DY	-
8c	Caf	En	-	AR	AR	-
Legend:	- Ofc: Office noise at 20 dB SNR; Str: Street noise at 15 dB SNR; Car(15): Static car noise at 15 dB SNR; - Car(10): Static car noise at 10 dB SNR; Caf: cafeteria noise at 15 dB SNR; - En: English; Fi: Finnish; Fr: French; Ge: German; Ja: Japanese; Sp: Spanish; - AR: ARCON; BT; DT ;DY: Dynastat; FT; LM: LMGT; NA: NTT-AT; NN: Nortel Networks; NO: Nokia.					
NOTE:	In the characterisation testing, experiments 1, 2 and 5 were conducted twice using different listening laboratories and languages. Tdoc S4-010393 from Dynastat presents the results of statistical analyses designed to determine if the subjective data from separate Listening Labs (i.e., different languages) could be combined to summarise the results of Experiments 1, 2 and 5. The results from these analyses indicate that the subjective data can not be combined in a statistically meaningful way across Listening Labs for any of the experiments.					

6.1.1 Characterisation testing in ITU

Additional characterisation testing was performed in ITU after AMR-WB codec was selected as an ITU standard G.722.2. The summary of the experiments is presented in table 6.3. Testing consisted of additional experiments not conducted during the 3GPP characterisation.

Table 6.3: Summary of different characterisation phase experiments

Exp.	Test	Title	Cond.	Languages
1	DCR	Effects of Bit Rate, Input Level, and VAD/DTX	30	2
2	DCR	Effects of Bit Rate, Tandeming, and Background Noise	40	2
3a	ACR	Effects of Wideband Coding and Test Method on Music Quality	6 music classes	1
3b	DCR	Effects of Wideband Coding and Test Method on Music Quality	6 music classes	1

Characterisation was carried out based on a common test plan [39]. The testing was divided between two laboratories using different speech databases and languages. The work division is described in table 6.4. Clauses 16 to 18 contain the complete set of test results for the AMR-WB speech codec Characterisation in ITU.

Table 6.4: Allocation of listening laboratories and host laboratories to experiments. The cross-checking were performed between the two host labs Arcon and Nokia.

		Host Lab	
Exp.	**Noise**	**ARCON**	**Nokia**
1	Quiet	Dynastat/English	Nokia/Finnish
2	Bable and interfering talker	Dynastat/English	Nokia/Finnish
3a	Quiet	-	Nokia/Finnish
3b	Quiet	-	Nokia/Finnish

6.1.2 Characterisation testing in TSG-GERAN

After selection of the AMR-WB codec, the channel coding for AMR-WB in 8-PSK channels was modified in order to harmonise it with the channel coding already specified for AMR-NB codec in 8-PSK channels. Additional characterisation test results were presented in TSG-GERAN to verify the performance of the new channel coding. Testing consisted of two experiments: Experiment 1 in clean speech with channel errors in 8-PSK FR channel and experiment 2 in clean speech with channel errors in 8-PSK HR channel. The detailed description of the test conditions and procedures can be found from [41] to [45].

6.2 AMR-WB Verification Phase

Table 6.5 lists the verification items relevant for performance characterisation and corresponding contributing organisations. The verification results are contained in clauses 19 to 30.

Table 6.5: Verification tasks and their allocation to the volunteering laboratories

	Description	Contributing Organisation(s)	Tdoc
1	Performances with DTMF Tones	BT	S4-010021
2	Performances with Special Input Signals	Nokia	S4-010228
3	Overload Performance (objective tests and informal listening)	Matsushita	S4-010330
4	Muting Behaviour	Nortel Networks	S4-010040
5	Transmission Delay (Round Trip) (TFO guidance)	Nortel Networks	S4-010052
6	Frequency Response	France Telecom	S4-010379
7	Signalling Tones	France Telecom	S4-010608
8	Complexity Analysis	Alcatel, STMicroelectronics, Philips Semiconductor	S4-010008
9	Comfort Noise Generation	Ericsson	S4-010230
10	Performance with music signals (informal expert listening)	Deutsche Telekom	S4-010158
11	Switching Performance between AMR and AMR-WB modes (note AMR-WB code does not include this switching capability)	Siemens	S4-010050

6.3 AMR-WB floating-point verification phase

Table 6.6 lists the verification items relevant for performance characterisation and corresponding contributing organisations for specifying the AMR-WB floating-point standard 3GPP TS 26.204. The verification results are contained in annex B.

Table 6.6: Verification tasks and their allocation to the volunteering laboratories

	Description	Contributing Organisation(s)	Tdoc
1	Verification of subjective speech quality with respect to the existing AMR-WB fixed-point codec (subjective testing): clean speech, input levels, tandeming, background noise	Nokia, Ericsson	S4-020064
2	Verification of speech quality using objective measurements (wideband extension of P.862, Annex B)	BT (Psytechnics)	S4-020124 S4-020049r1
3	DTMF- and signalling tones	Hughes Software Systems	S4-020114
4	Performance with music signals	Siemens	S4-020080
5	Special signals (in particular, non-speech signals)	FT	S4-020077
6	Check of overload performance	NEC	S4-020113
7	Idle channel behaviour (output signal when low noise input signal)	Siemens	S4-020079
8	Operation of the VAD and comfort noise	Ericsson	S4-020270
9	Stability of the codec over time	Nokia, Siemens	S4-020079 S4-020062
10	Bit-exactness of the decoder	Nokia, Siemens, FT	S4-020079 S4-020062

7 Important notes about the interpretation of test results

Mean Opinion Scores can only be representative of the test conditions in which they were recorded (speech material, speech processing, listening conditions, language, and cultural background of the listening subject). Listening tests performed with other conditions than those used in the AMR-WB Characterisation phase of testing could lead to a different set of MOS results. On the other hand, the relative performances of different codec under tests is considered more reliable and less impacted by cultural difference between listening subjects than absolute MOS values. When looking at the relative differences of the codecs in the same test, it should be noted that a difference of 0.2 MOS between two test results was usually found not statistically significant.

The subjective testing is conducted using limited amount of speech material in order to keep the size of the experiment within reasonable limits. Sometimes this can cause some irregularities to the test results. Also the performance of the tested codecs is not always known when designing the test, thus balancing the test conditions may not always be perfect. This may result imperfect utilisation of the ranking scale and difficulties to discriminate the codecs with quality very close to each other.

For example, higher error-rate condition may sometimes get better MOS values than the lower error-rate condition. In the lower error-rate condition those few errors can hit for the onset parts of the speech sentences, thus dramatically increasing the effect of errors. If two conditions have error-rate close to each other, this "random" effect can change the ordering of these conditions because we do not have enough test material to get statistically enough occurrences of errors.

The resolution of the testing is limited. The listeners are usually using scale from 1 to 5 to rank the different codecs. However, during the tests presented in the present document, we are characterising nine different AMR-WB modes, most of which are very high quality codecs and this causes sometimes a "saturation" effect in the test, i.e. the listeners can not discriminate the different codecs because of the limited dynamics in the ranking scale.

Also the listening environment will affect the scale of the results. For example, the results can be very different if the same stimulus is presented to the listener through monaural or binaural headphones.

Taking account the comments presented above, the reader is advised to exercise some precautions when looking and comparing the individual scores of the tests. Usually, looking at the whole picture and overall trends in the test in question may give better interpretation of the performance of the codecs. This precaution should be especially taken account when looking at the experiments conducted using erroneous channels which may present rather big variability of results over the limited amount of tested conditions.

8 Performance in self-tandeming and with variation of the input speech level

Experiment 1 was designed to evaluate the error-free clean-speech performance of all the AMR-WB codec modes in tandeming conditions and with a variety of input levels. Tests were conducted using two languages: Finnish and English.

Looking at the results in figure 8.1 and figure 8.2, both tests show very good results for the AMR-WB modes with bit-rates 12.65 kbit/s and upwards. For these the quality is equal or better than for G.722 at 64 kbit/s. Results are consistent over all the tested input levels and tandeming. The 8.85 kbit/s mode gives quality equal to G.722 at 48 kbit/s. The lowest mode 6.6 kbit/s provides quality, which is lower than quality of G.722-48. This is clear especially in tandeming and with high input level. However, the two lowest modes are designed to be used only temporarily in poor radio channel conditions. The error bars in figures 8.1 and 8.2 represent the 95 % confidence intervals.

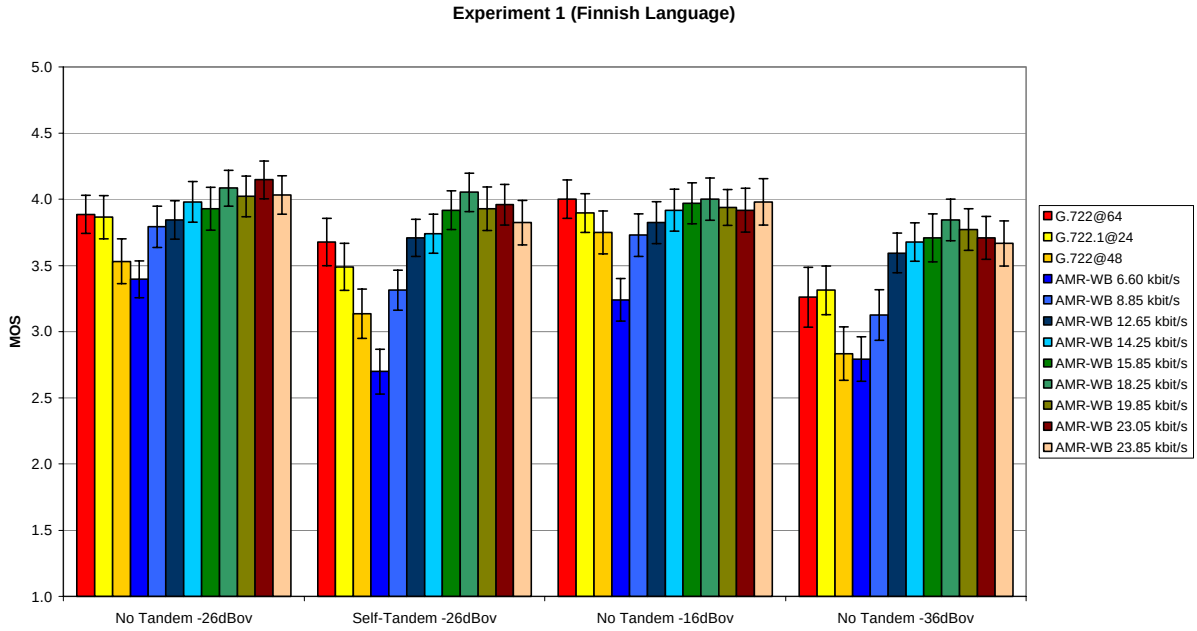


Figure 8.1: Experiment 1, testing Tandeming and input levels with Finnish language

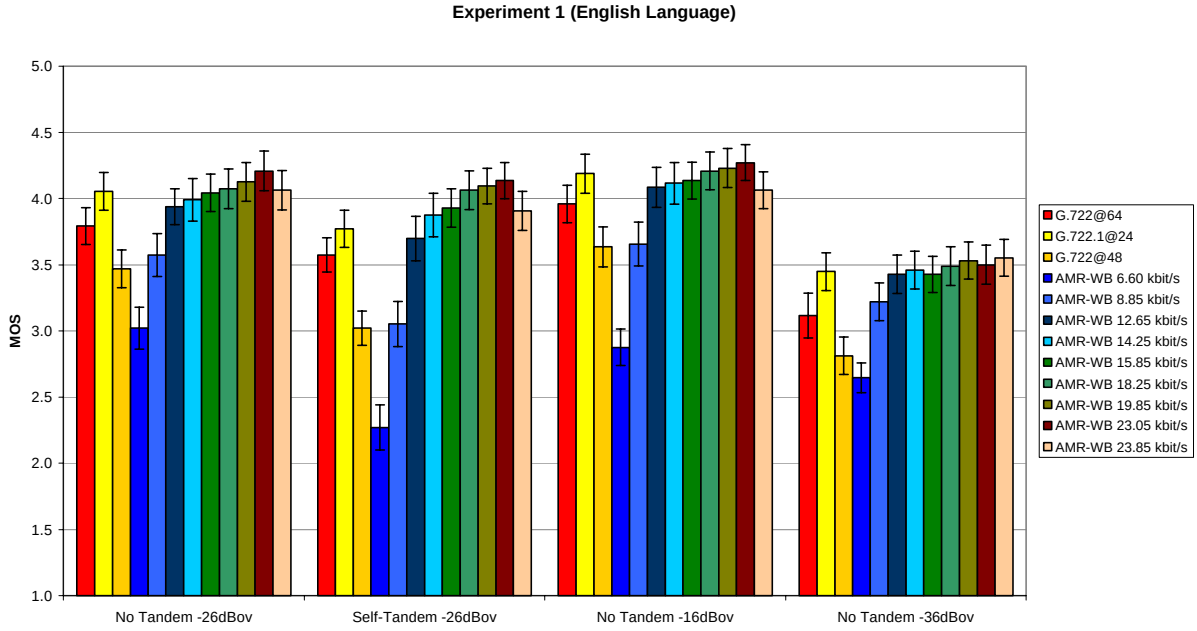


Figure 8.2: Experiment 1, testing tandeming and input levels with English language

9 Interoperability Performance in Real World Wideband Scenarios

The purpose of Experiment 2 was to characterise the error-free, clean-speech performance of all the AMR-WB codec modes in tandem with other wideband standards, e.g. with G.722/G.722.1. Two different languages were used, English and French. All nine AMR-WB modes were tested with the following tandeming scenarios shown in table 9.1.

Table 9.1: Naming in Figure 9.1

	Naming in Figure 9.1
No Tandem	No Tandem
AMR-WB mode [0…8] -> G.722@64	G.722@64 Tandem 2nd
AMR-WB mode [0…8] -> G.722@48	G.722@48 Tandem 2nd
G.722@48 -> AMR-WB mode [0…8]	G.722@48 Tandem 1st
AMR-WB mode [0…8] -> G.722.1@24	G.722.1@24 Tandem 2nd

The results show that in Experiment 2 the overall tandem performance of the AMR-WB codec is independent of the combination of AMR-WB with G.722 at 64 kbit/s or G.722.1 at 24 kbit/s, or for the AMR-WB codec preceded by the G.722 codec at 48 kbit/s. However, the connections with the AMR-WB codec followed by G.722 at 48 kbit/s in general resulted in a significantly poorer connection than the other tandem connections studied. This probably happens because of the multiplicative noise distortion that the G.722 ADPCM algorithm introduces in the second stage of processing (as opposed to the relatively smooth output of coders like AMR-WB and G.722.1, which introduce a different type of distortion). The error bars in figures 9.1 and 9.2 represent the 95 % confidence intervals.

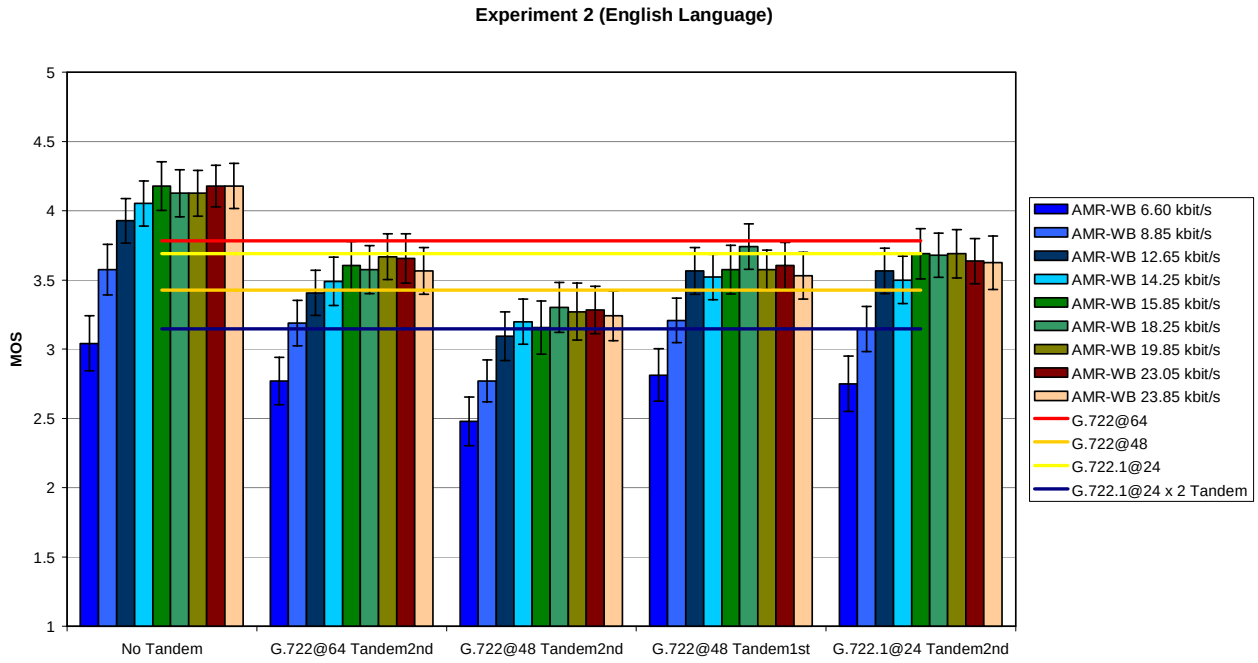


Figure 9.1: Experiment 2, testing tandeming with other standards with English language

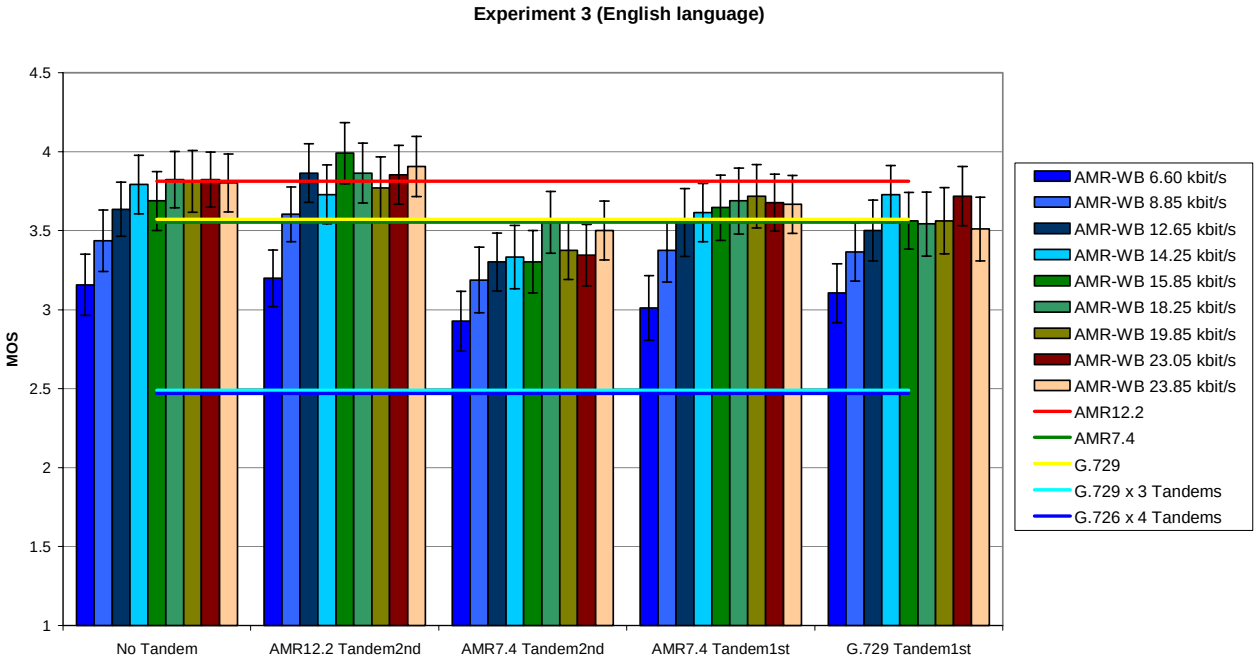


Figure 9.2: Experiment 2, testing tandeming with other standards with French language

10 Interoperability Performance in Real World Narrowband Scenarios

The purpose of Experiment 3 was to characterise the performances of the different AMR-WB codec modes in tandem with narrowband standards, e.g. with AMR-NB 12.2 and 7.4 kbit/s modes and with ITU-T Recommendation G.729. English language was used in testing. All nine AMR-WB modes were tested with the following tandeming scenarios shown in table 10.1.

Table 10.1: Naming in Figure 10.1

	Naming in Figure 10.1
No Tandem	No Tandem
AMR-WB mode [0…8] -> AMR-NB 12.2 kbit/s	AMR12.2 Tandem 2nd
AMR-WB mode [0…8] -> AMR-NB 7.4 kbit/s	AMR7.4 Tandem 2nd
AMR-NB 7.4 kbit/s -> AMR-WB mode [0…8]	AMR7.4 Tandem 1st
G.729 -> AMR-WB mode [0…8]	G.729 Tandem 1st

It can be seen in figure 10.1, that for narrowband speech, AMR-WB offers similar performance as AMR 12.2 kbit/s mode, when the bit-rate of the AMR-WB is 12.65 kbit/s or higher. For the two lowest AMR-WB modes 8.85 kbit/s and 6.6 kbit/s, the quality is worse than the quality of AMR 7.4 kbit/s and 8 kbit/s G.729.

In general, tandeming AMR-WB with narrow band codecs does not degrade the quality very much when compared to the single coding of the same narrow band codec, except for cases when the two lowest bit-rates of the AMR-WB codec are used. Only in the condition where AMR-NB 7.4 kbit/s coding is after the AMR-WB coding, some quality degradation can be observed. The error bars in figure 10.1 represent the 95 % confidence intervals.

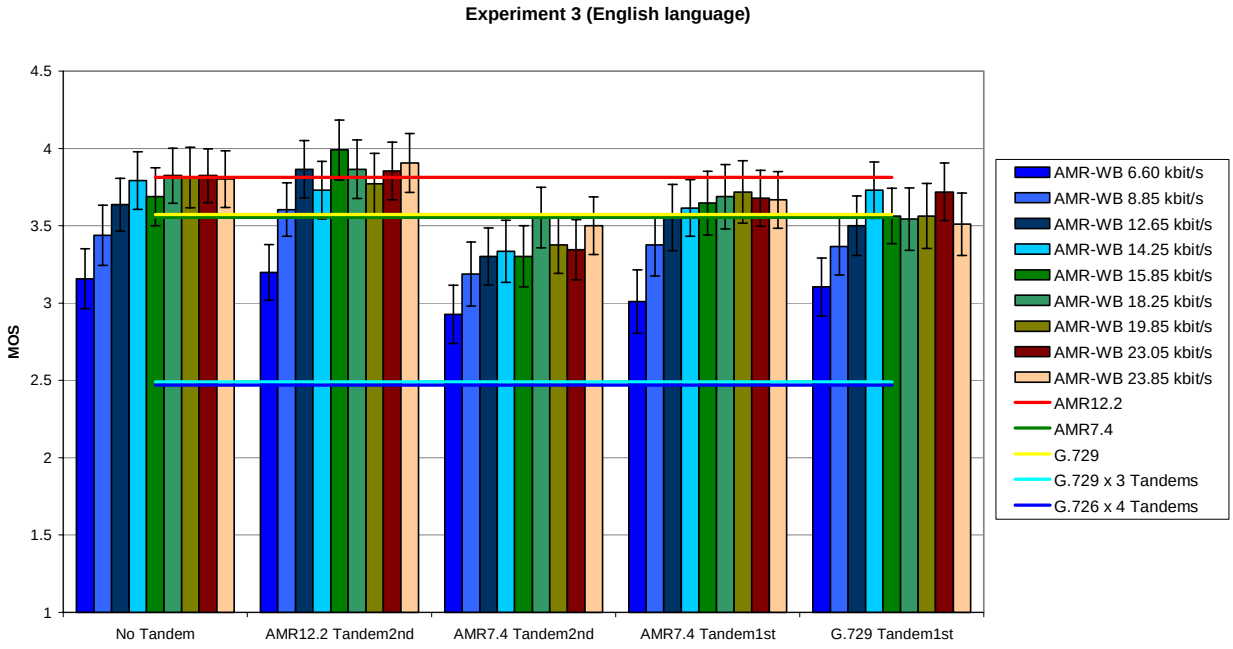


Figure 10.1: Experiment 3, testing tandeming with narrowband standards with English language

11 Performance of VAD/DTX/CNG Algorithm

The objective of Experiment 4 was to evaluate the degradation induced by the activation of the voice activity detection and discontinuous transmission on the link under test. The test used a 5-point Degradation Category Rating (DCR). English language was used in testing the experiment 4.

The tests were performed using modes 12.65 kbit/s and 18.25 kbit/s. Both modes were tested with and without errors. ETSI GSM FR error profiles were used. Table 11.1 describes the conditions in which the codec were tested with VAD=ON and VAD=OFF. Note, that after the characterisation, the support for bit-rates above 12.65 kbit/s was dropped from the GSM GMSK FR channel. This means, that the channel coding and the results for 18.25 kbit/s mode for GSM FR channel are obsolete.

Table 11.1: List of tested conditions with VAD=ON and VAD=OFF

Noise types	No errors		C/I=9 dB (FER ~ 1.0 %)	C/I=15 dB (FER ~ 0.6 %)
Office noise at 20 dB	12.65 kbit/s	18.25 kbit/s	12.65 kbit/s	18.25 kbit/s
Street noise at 15 dB	12.65 kbit/s	18.25 kbit/s	12.65 kbit/s	18.25 kbit/s
Car noise at 15 dB	12.65 kbit/s	18.25 kbit/s	12.65 kbit/s	18.25 kbit/s
Cafeteria noise at 15 dB	12.65 kbit/s	18.25 kbit/s	12.65 kbit/s	18.25 kbit/s

From the results in figure 11.1, it can be seen that, conditions using VAD/DTX/CNG in the processing were statistically rated at least no worse than samples without VAD/DTX/CNG. This result supports the conclusion that the VAD/DTX/CNG operation is transparent to the listener. The error bars in figure 11.1 represent the 95 % confidence intervals.

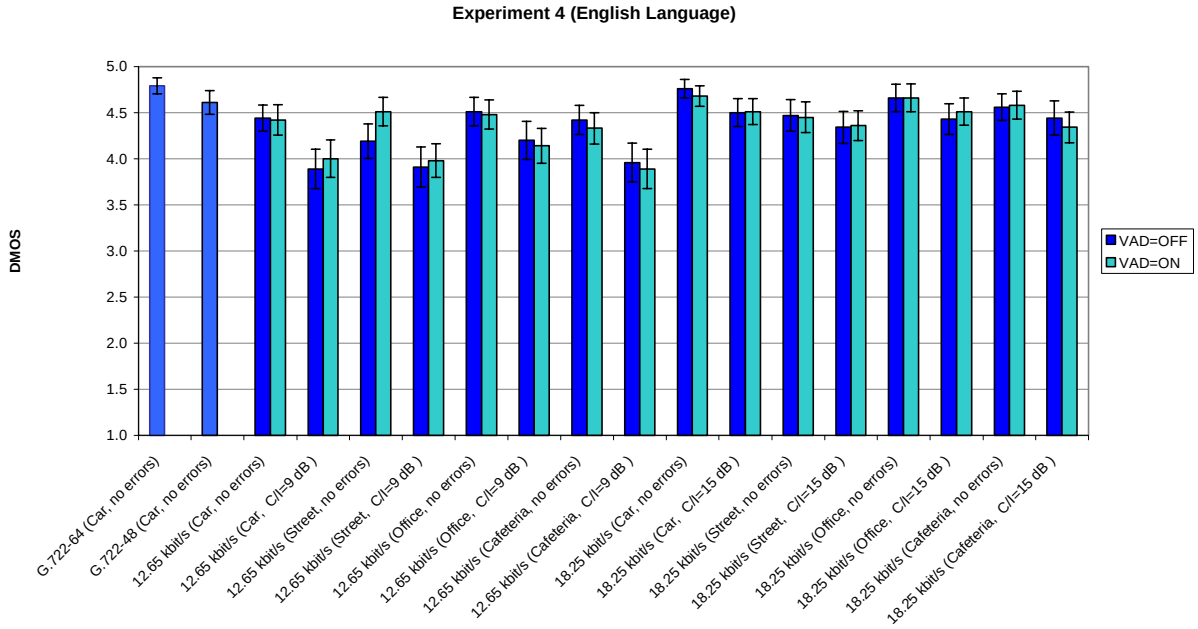


Figure 11.1: Experiment 4, testing VAD/DTX with English language

12 Performance in Static Errors under Clean Speech Conditions in GSM GMSK

The purpose of Experiment 5 was to characterise the performances of different AMR-WB codec modes in GSM GMSK FR channel. Experiment 5 was tested using two languages, German and French.

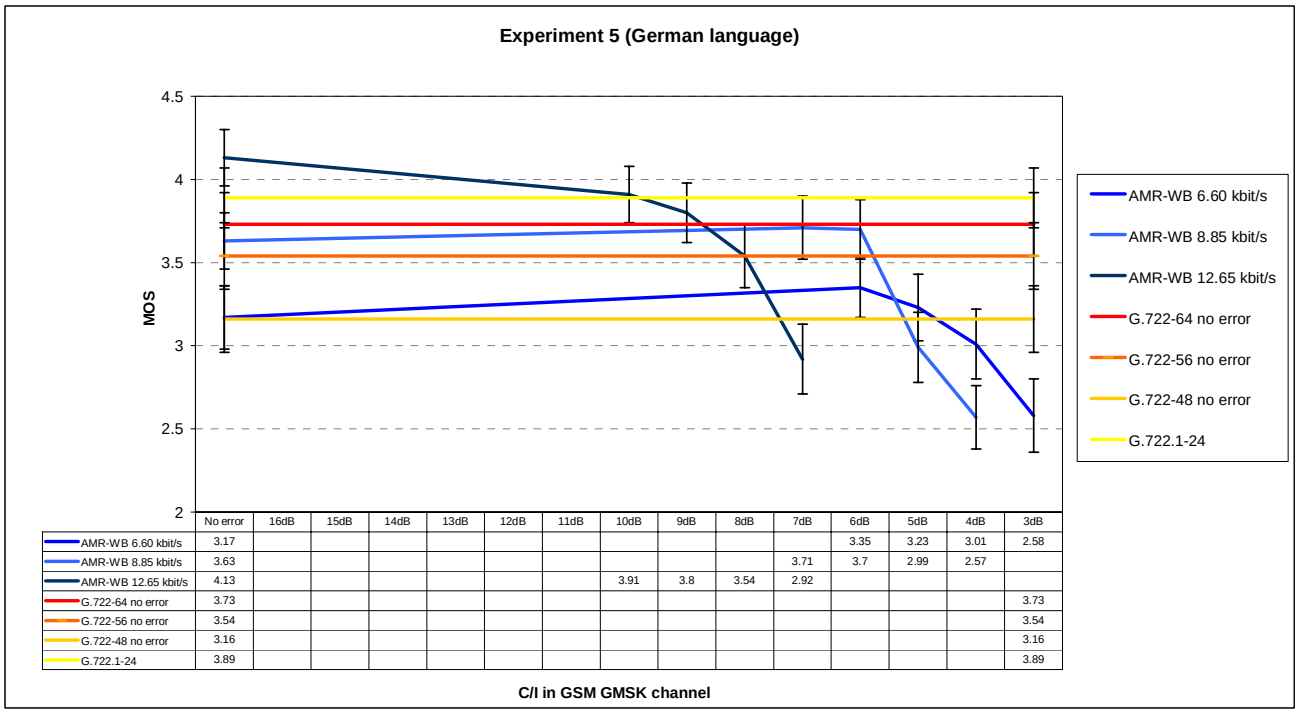
In Experiments 5, static C/I conditions are used. Their value is quoted in terms of Carrier to Interference Ratio (C/I), and the average C/I over the duration of the test condition is set to a fixed value. In these experiments, a selection of static C/I values varying from 3 dB to 16 dB are used, in addition to the error-free case.

The experiments are designed to characterise the performance of the codec in each of its modes over a range of channel conditions, producing what has been termed a family of curves. For each mode, error free and 4 different error conditions was tested. Two different languages were used.

From both figures it can be seen that the quality of at least G.722 at 56 kbit/s can be achieved at about 10 dB C/I and above. The quality better or equal of at least G.722 at 64 kbit/s can be achieved at about 11 dB C/I and above. The error bars in figures 12.1 and 12.2 represent the 95 % confidence intervals.

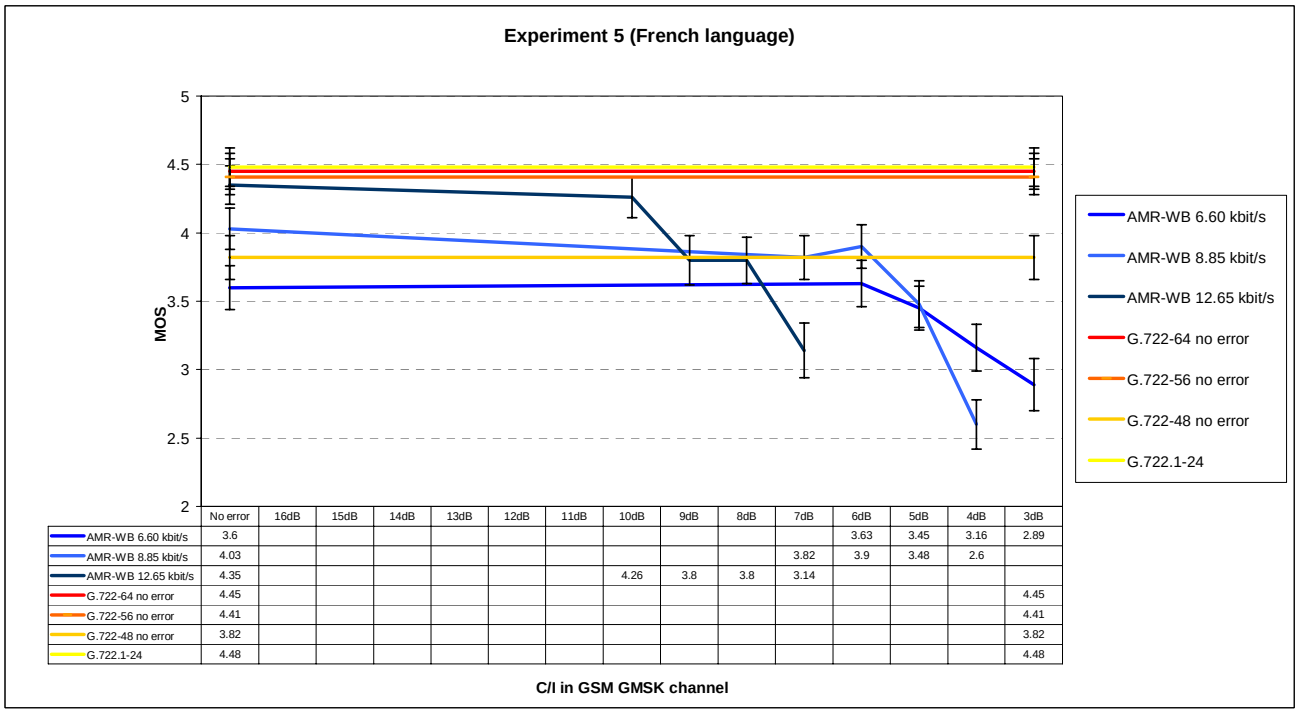
NOTE 1: After the characterisation, the support for bit-rates above 12.65 kbit/s was dropped from the GSM GMSK FR channel. This means, that the channel coding and the results for 14.25 kbit/s, 15.85 kbit/s, 18.25 kbit/s and 19.85 kbit/s modes for GSM FR channel are not shown in the figures 12.1 and 12.2, even they were originally tested during the characterisation.

NOTE 2: G.722 reference codecs, shown in figures 12.1 and 12.2, were tested in error-free conditions only.



	No error	16dB	15dB	14dB	13dB	12dB	11dB	10dB	9dB	8dB	7dB	6dB	5dB	4dB	3dB
AMR-WB 6.60 kbit/s	3.17											3.35	3.23	3.01	2.58
AMR-WB 8.85 kbit/s	3.63										3.71	3.7	2.99	2.57	
AMR-WB 12.65 kbit/s	4.13							3.91	3.8	3.54	2.92				
G.722-64 no error	3.73														3.73
G.722-56 no error	3.54														3.54
G.722-48 no error	3.16														3.16
G.722.1-24	3.89														3.89

Figure 12.1: Experiment 5, testing GSM FR channel with German language



	No error	16dB	15dB	14dB	13dB	12dB	11dB	10dB	9dB	8dB	7dB	6dB	5dB	4dB	3dB
AMR-WB 6.60 kbit/s	3.6											3.63	3.45	3.16	2.89
AMR-WB 8.85 kbit/s	4.03										3.82	3.9	3.48	2.6	
AMR-WB 12.65 kbit/s	4.35							4.26	3.8	3.8	3.14				
G.722-64 no error	4.45														4.45
G.722-56 no error	4.41														4.41
G.722-48 no error	3.82														3.82
G.722.1-24	4.48														4.48

Figure 12.2: Experiment 5, testing GSM FR channel with French language

13 Performance in Background Noise in Static C/I Conditions in GSM GMSK

The purpose of Experiments 6a and 6b were to characterise the performances of the different AMR-WB codec modes in static error conditions in the presence of background noise. For each mode, 3 different error conditions can be tested (in addition to error free case). Experiment 6a was conducted using English language and experiment 6b using Finnish language. The noise types and levels used are described in table 13.1.

Table 13.1: Noise types and levels for experiments 6a and 6b

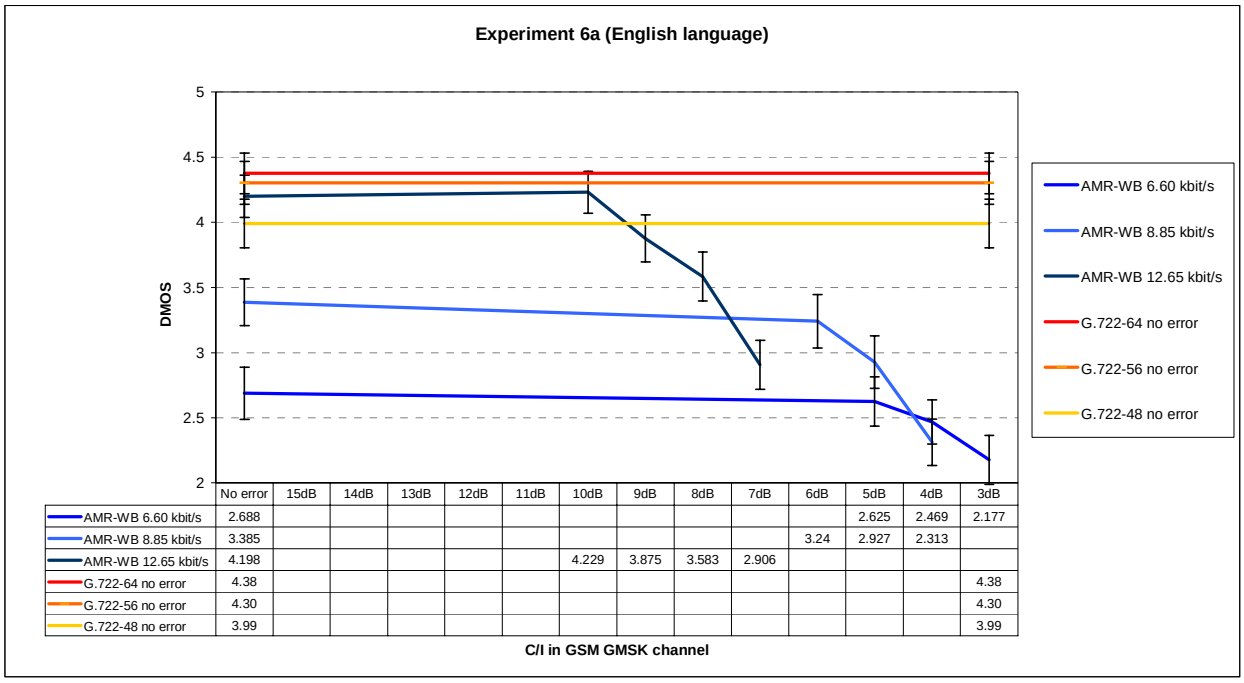
Experiment	Noise type	Level
Exp. 6a (GSM GMSK)	Car	15 dB
Exp. 6b (GSM GMSK)	Office	20 dB

In Experiments 6a and 6b, static C/I conditions are used. Their value is quoted in terms of Carrier to Interference Ratio (C/I), and the average C/I over the duration of the test condition is set to a fixed value. In these experiments, a selection of static C/I values varying from 3 dB to 15 dB are used, in addition to the error-free case.

It seems, that both experiments give very similar results about the performance of the different AMR-WB modes in the presence of background noise. From both figures it can be seen that the quality of G.722 at 56 kbit/s can be achieved in C/I-ratios 10 dB and above. The quality better or equal to G.722 at 64 kbit/s can be achieved in C/I-ratios 12 dB and above. The error bars in figures 13.1 and 13.2 represent the 95 % confidence intervals.

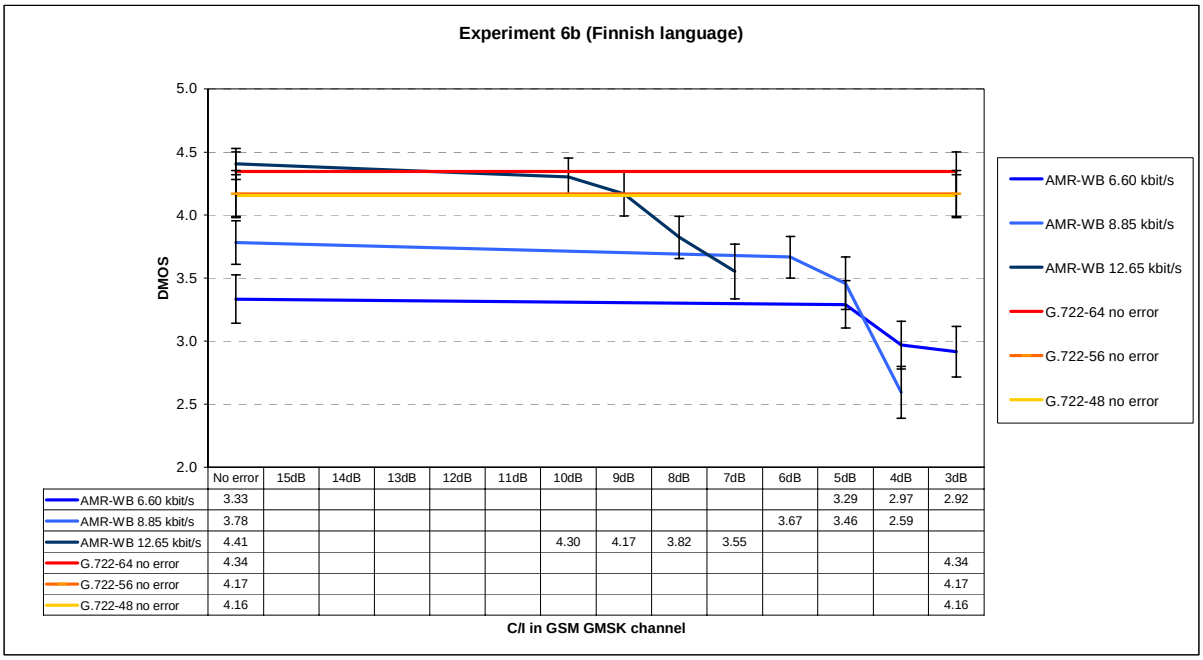
Note, that after the characterisation, the support for bit-rates above 12.65 kbit/s was dropped from the GSM GMSK FR channel. This means, that the channel coding and the results for 14.25 kbit/s, 15.85 kbit/s, 18.25 kbit/s and 19.85 kbit/s modes for GSM FR channel are not shown in the figures 12.1 and 12.2, even they were originally tested during the characterisation.

NOTE: G.722 reference codecs, shown in Figures 13.1 and 13.2, were tested in error-free conditions only.



	No error	15dB	14dB	13dB	12dB	11dB	10dB	9dB	8dB	7dB	6dB	5dB	4dB	3dB
AMR-WB 6.60 kbit/s	2.688											2.625	2.469	2.177
AMR-WB 8.85 kbit/s	3.385										3.24	2.927	2.313	
AMR-WB 12.65 kbit/s	4.198						4.229	3.875	3.583	2.906				
G.722-64 no error	4.38													4.38
G.722-56 no error	4.30													4.30
G.722-48 no error	3.99													3.99

Figure 13.1: Experiment 6a, testing GSM FR channel with English language



	No error	15dB	14dB	13dB	12dB	11dB	10dB	9dB	8dB	7dB	6dB	5dB	4dB	3dB
AMR-WB 6.60 kbit/s	3.33											3.29	2.97	2.92
AMR-WB 8.85 kbit/s	3.78										3.67	3.46	2.59	
AMR-WB 12.65 kbit/s	4.41						4.30	4.17	3.82	3.55				
G.722-64 no error	4.34													4.34
G.722-56 no error	4.17													4.17
G.722-48 no error	4.16													4.16

Figure 13.2: Experiment 6b, testing GSM FR channel with Finnish language

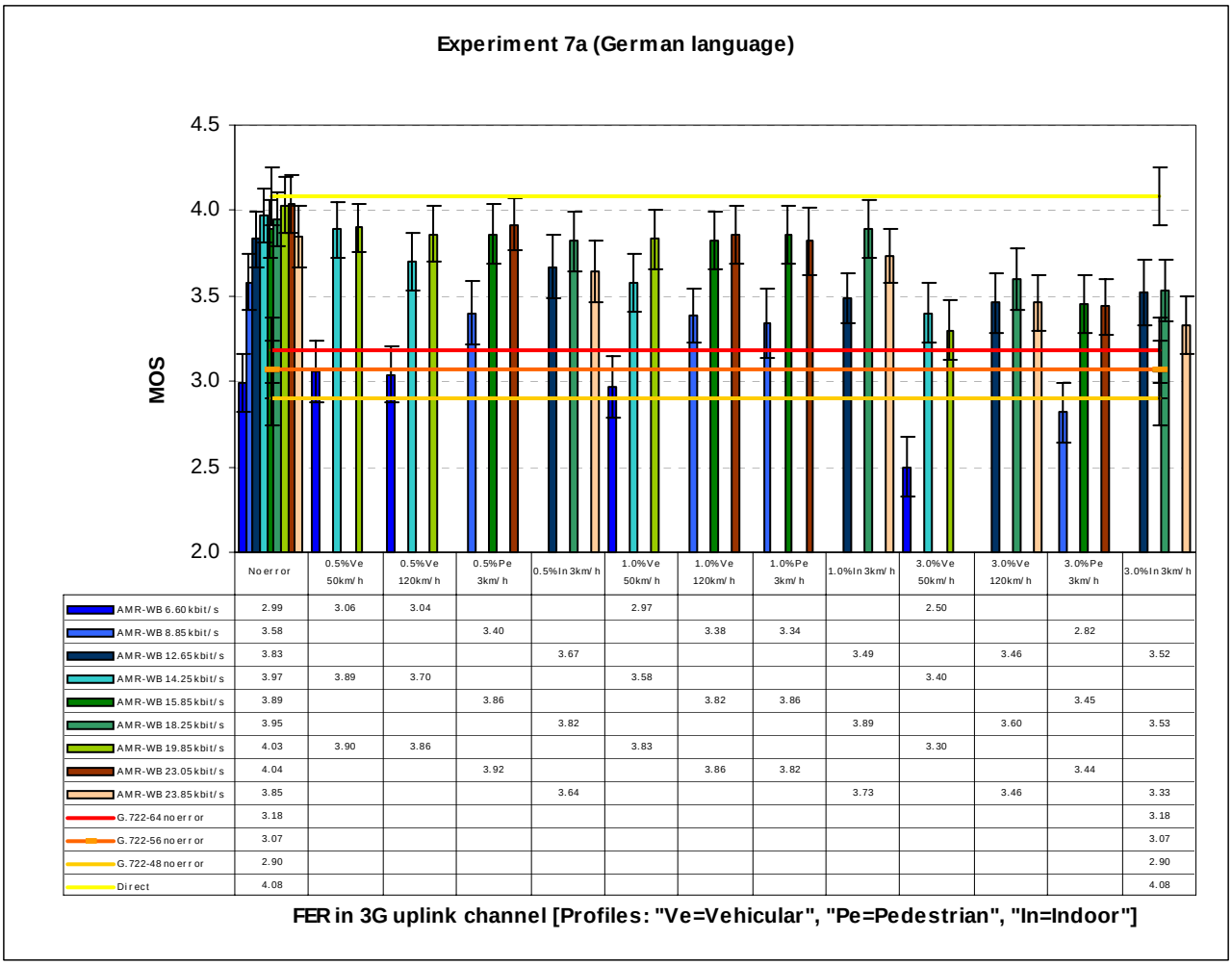
14 Performance in Static Errors under Clean Speech Conditions in 3G

The experiments 7a and 7b are designed to characterise the performance of the codec in each of its modes over a range of 3G channel conditions (for clean speech), producing what has been termed a family of curves.

Due to the number of modes available (9), and the range of C/I conditions over which each of these modes could be tested, it will not be possible to characterise all possible combinations. For each mode, 4 different error conditions were tested (in addition to error free). The test methodology was Absolute Category Rating (ACR).

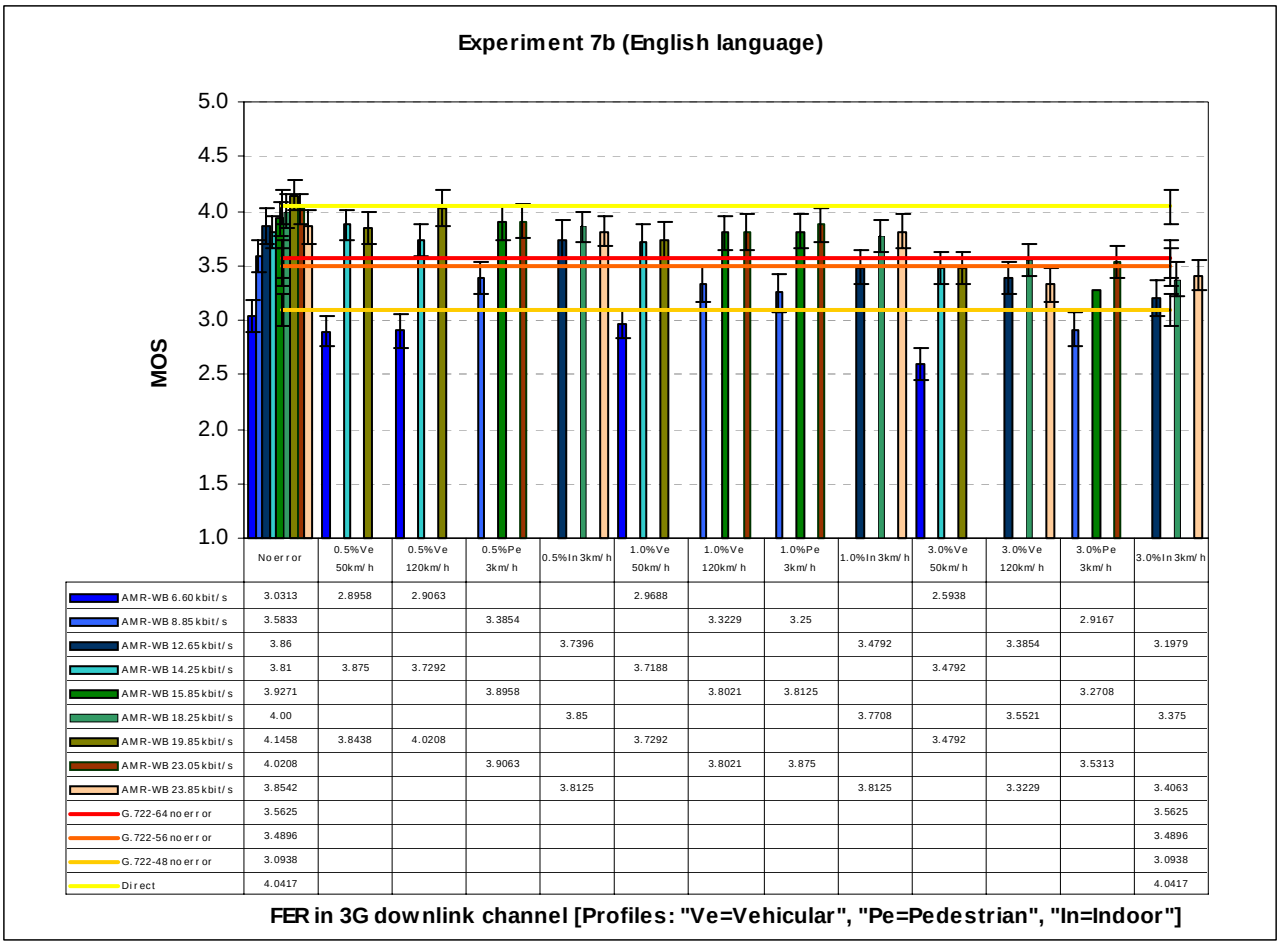
The sub-experiment 7a was performed in German language and 7b in English language. The sub-experiments are identical with an exception that experiment 7a uses uplink and experiment 7b downlink 3G channels. The error bars in figures 14.1 and 14.2 represent the 95 % confidence intervals.

NOTE: G.722 reference codecs, shown in figures 14.1 and 14.2, were tested in error-free conditions only.



	No error	0.5%Ve 50km/h	0.5%Ve 120km/h	0.5%Pe 3km/h	0.5%In 3km/h	1.0%Ve 50km/h	1.0%Ve 120km/h	1.0%Pe 3km/h	1.0%In 3km/h	3.0%Ve 50km/h	3.0%Ve 120km/h	3.0%Pe 3km/h	3.0%In 3km/h
AMR-WB 6.60 kbit/s	2.99	3.06	3.04			2.97				2.50			
AMR-WB 8.85 kbit/s	3.58			3.40			3.38	3.34				2.82	
AMR-WB 12.65 kbit/s	3.83				3.67				3.49		3.46		3.52
AMR-WB 14.25 kbit/s	3.97	3.89	3.70			3.58				3.40			
AMR-WB 15.85 kbit/s	3.89			3.86			3.82	3.86				3.45	
AMR-WB 18.25 kbit/s	3.95				3.82				3.89		3.60		3.53
AMR-WB 19.85 kbit/s	4.03	3.90	3.86			3.83				3.30			
AMR-WB 23.05 kbit/s	4.04			3.92			3.86	3.82				3.44	
AMR-WB 23.85 kbit/s	3.85				3.64				3.73		3.46		3.33
G.722-64 no error	3.18												3.18
G.722-56 no error	3.07												3.07
G.722-48 no error	2.90												2.90
Direct	4.08												4.08

Figure 14.1: Experiment 7a, testing 3G uplink channel with German language



	No error	0.5%Ve 50km/h	0.5%Ve 120km/h	0.5%Pe 3km/h	0.5%In 3km/h	1.0%Ve 50km/h	1.0%Ve 120km/h	1.0%Pe 3km/h	1.0%In 3km/h	3.0%Ve 50km/h	3.0%Ve 120km/h	3.0%Pe 3km/h	3.0%In 3km/h
AMR-WB 6.60 kbit/s	3.0313	2.8958	2.9063			2.9688				2.5938			
AMR-WB 8.85 kbit/s	3.5833			3.3854			3.3229	3.25				2.9167	
AMR-WB 12.65 kbit/s	3.86				3.7396				3.4792		3.3854		3.1979
AMR-WB 14.25 kbit/s	3.81	3.875	3.7292			3.7188				3.4792			
AMR-WB 15.85 kbit/s	3.9271			3.8958			3.8021	3.8125				3.2708	
AMR-WB 18.25 kbit/s	4.00				3.85				3.7708		3.5521		3.375
AMR-WB 19.85 kbit/s	4.1458	3.8438	4.0208			3.7292				3.4792			
AMR-WB 23.05 kbit/s	4.0208			3.9063			3.8021	3.875				3.5313	
AMR-WB 23.85 kbit/s	3.8542				3.8125				3.8125		3.3229		3.4063
G.722-64 no error	3.5625												3.5625
G.722-56 no error	3.4896												3.4896
G.722-48 no error	3.0938												3.0938
Direct	4.0417												4.0417

Figure 14.2: Experiment 7b, testing 3G downlink channel with English language

15 Performance in Background Noise in Static C/I Conditions in 3G

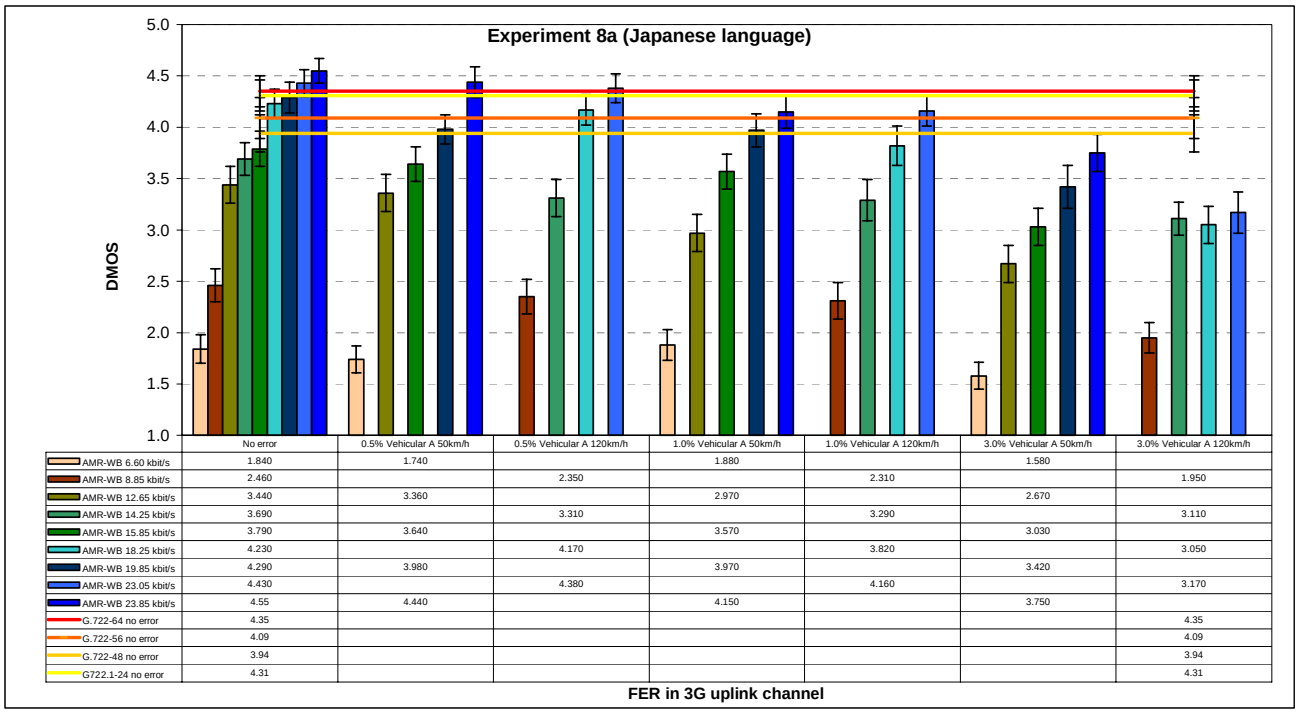
The purpose of Experiment 8 is to characterise the performances of the different AMR-WB codec modes in static error conditions in the presence of background noise. Experiment 8 will use different noise samples than those tested in experiments 6a and 6b. The noise types and levels used are described in table 15.1.

Table 15.1: Noise types and levels for experiments 8a, 8b and 8c

Experiment	Noise type	Level
Exp. 8a (3G)	Car	10 dB
Exp. 8b (3G)	Street	15 dB
Exp. 8c (3G)	Cafeteria	15 dB

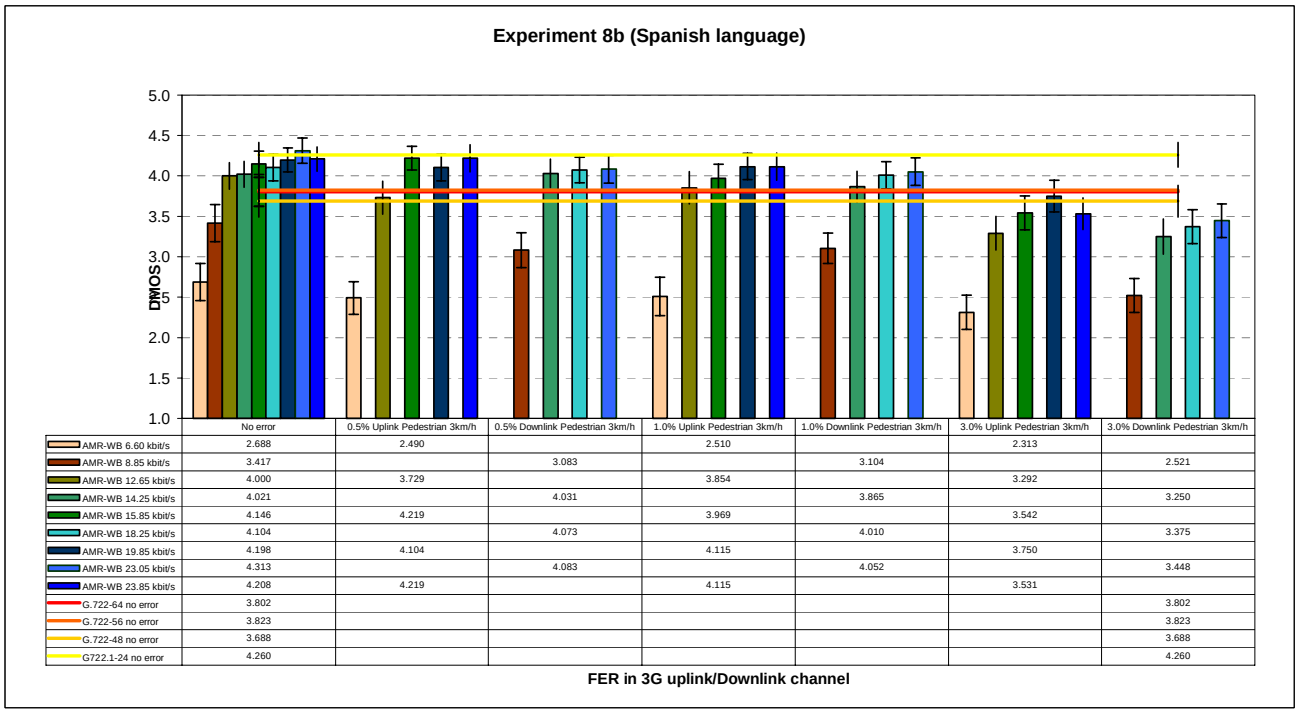
The test methodology was Degradation Category Rating (DCR). The sub-experiment 8a was performed in Japanese language, 8b in Spanish language and 8c in English language. The error bars in figures 15.1, 15.2 and 15.3 represent the 95 % confidence intervals

NOTE: G.722 reference codecs, shown in figures 15.1, 15.2 and 15.3, were tested in error-free conditions only.



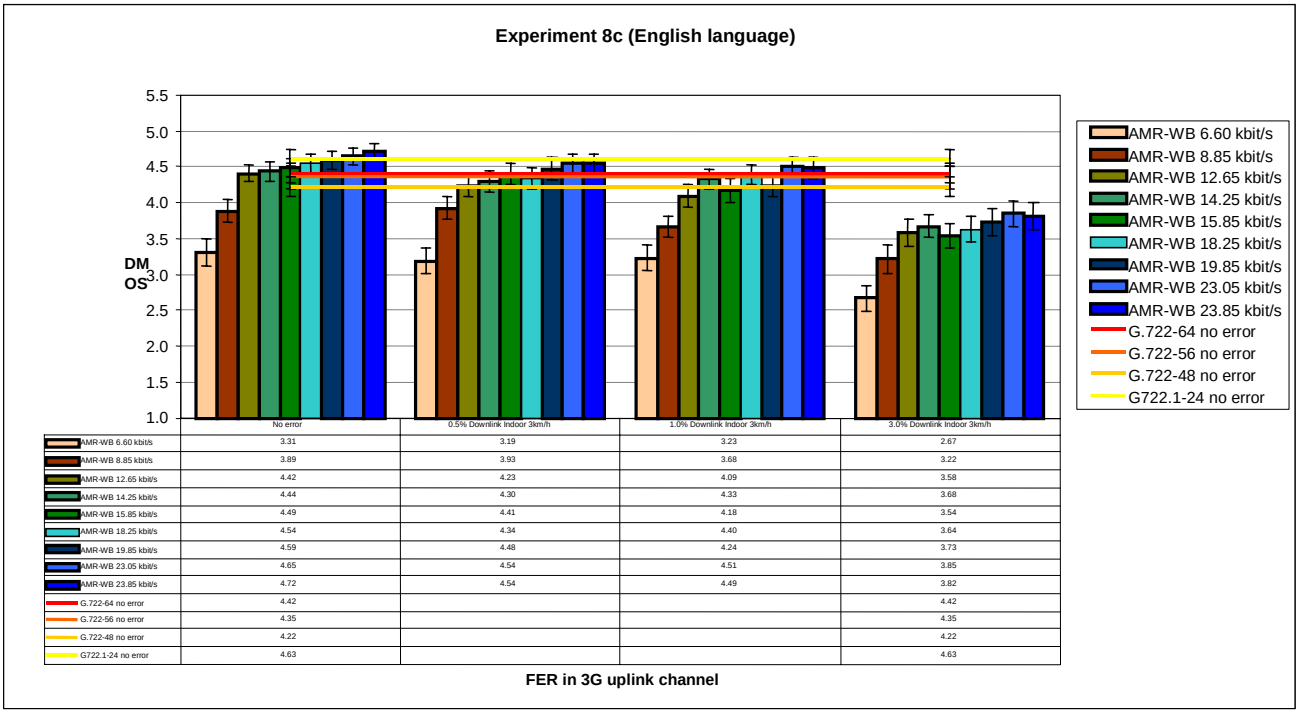
	No error	0.5% Vehicular A 50km/h	0.5% Vehicular A 120km/h	1.0% Vehicular A 50km/h	1.0% Vehicular A 120km/h	3.0% Vehicular A 50km/h	3.0% Vehicular A 120km/h
AMR-WB 6.60 kbit/s	1.840	1.740		1.880		1.580	
AMR-WB 8.85 kbit/s	2.460		2.350		2.310		1.950
AMR-WB 12.65 kbit/s	3.440	3.360		2.970		2.670	
AMR-WB 14.25 kbit/s	3.690		3.310		3.290		3.110
AMR-WB 15.85 kbit/s	3.790	3.640		3.570		3.030	
AMR-WB 18.25 kbit/s	4.230		4.170		3.820		3.050
AMR-WB 19.85 kbit/s	4.290	3.980		3.970		3.420	
AMR-WB 23.05 kbit/s	4.430		4.380		4.160		3.170
AMR-WB 23.85 kbit/s	4.55	4.440		4.150		3.750	
G.722-64 no error	4.35						4.35
G.722-56 no error	4.09						4.09
G.722-48 no error	3.94						3.94
G722.1-24 no error	4.31						4.31

Figure 15.1: Experiment 8a, testing 3G channel with Japanese language



	No error	0.5% Uplink Pedestrian 3km/h	0.5% Downlink Pedestrian 3km/h	1.0% Uplink Pedestrian 3km/h	1.0% Downlink Pedestrian 3km/h	3.0% Uplink Pedestrian 3km/h	3.0% Downlink Pedestrian 3km/h
AMR-WB 6.60 kbit/s	2.688	2.490		2.510		2.313	
AMR-WB 8.85 kbit/s	3.417		3.083		3.104		2.521
AMR-WB 12.65 kbit/s	4.000	3.729		3.854		3.292	
AMR-WB 14.25 kbit/s	4.021		4.031		3.865		3.250
AMR-WB 15.85 kbit/s	4.146	4.219		3.969		3.542	
AMR-WB 18.25 kbit/s	4.104		4.073		4.010		3.375
AMR-WB 19.85 kbit/s	4.198	4.104		4.115		3.750	
AMR-WB 23.05 kbit/s	4.313		4.083		4.052		3.448
AMR-WB 23.85 kbit/s	4.208	4.219		4.115		3.531	
G.722-64 no error	3.802						3.802
G.722-56 no error	3.823						3.823
G.722-48 no error	3.688						3.688
G722.1-24 no error	4.260						4.260

Figure 15.2: Experiment 8b, testing 3G channel with Spanish language



	No error	0.5% Downlink Indoor 3km/h	1.0% Downlink Indoor 3km/h	3.0% Downlink Indoor 3km/h
AMR-WB 6.60 kbit/s	3.31	3.19	3.23	2.67
AMR-WB 8.85 kbit/s	3.89	3.93	3.68	3.22
AMR-WB 12.65 kbit/s	4.42	4.23	4.09	3.58
AMR-WB 14.25 kbit/s	4.44	4.30	4.33	3.68
AMR-WB 15.85 kbit/s	4.49	4.41	4.18	3.54
AMR-WB 18.25 kbit/s	4.54	4.34	4.40	3.64
AMR-WB 19.85 kbit/s	4.59	4.48	4.24	3.73
AMR-WB 23.05 kbit/s	4.65	4.54	4.51	3.85
AMR-WB 23.85 kbit/s	4.72	4.54	4.49	3.82
G.722-64 no error	4.42			4.42
G.722-56 no error	4.35			4.35
G.722-48 no error	4.22			4.22
G722.1-24 no error	4.63			4.63

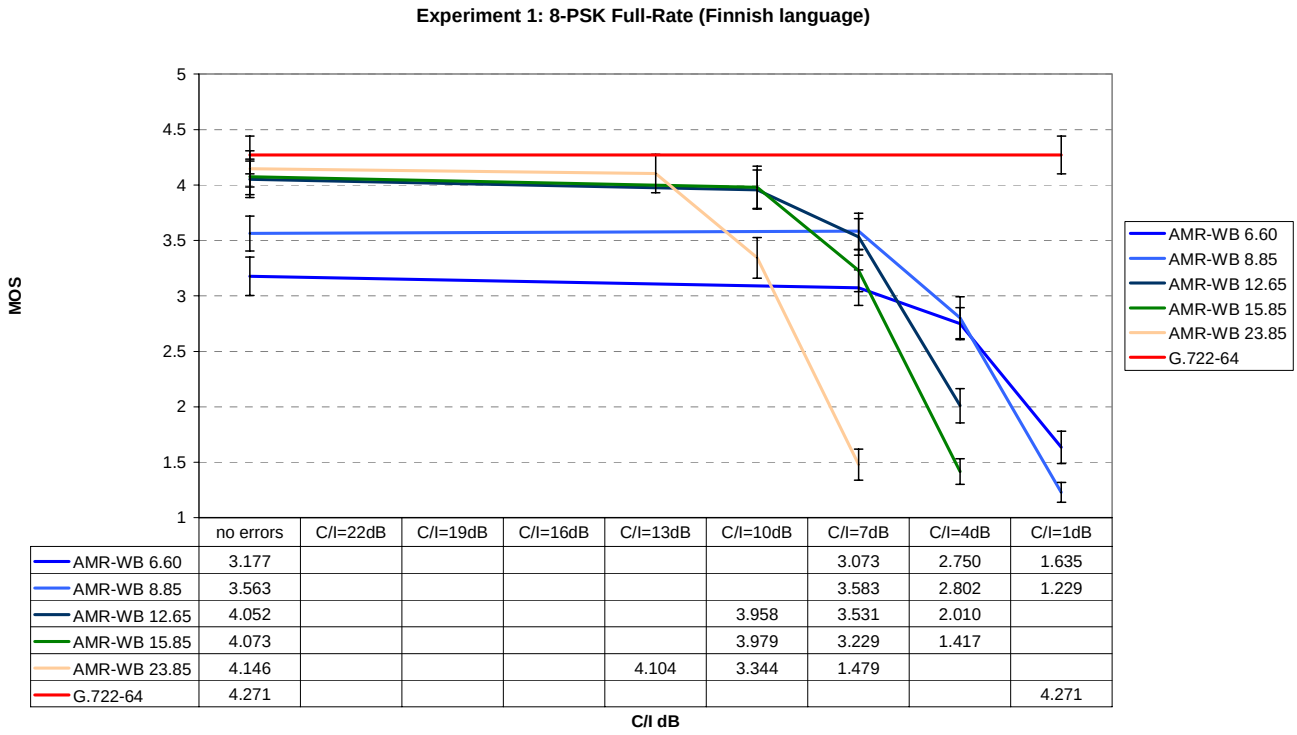
Figure 15.3: Experiment 8c, testing 3G channel with English language

16 Performance in Static Errors under Clean Speech Conditions in GERAN 8-PSK FR and HR channels

The experimental results contained in this clause were presented in TSG-GERAN. The purpose of the experiment was to verify the operation of AMR-WB channel coding in 8-PSK FR- and HR-channels after the channel coding was modified to harmonise it with already existing AMR-NB 8-PSK channel coding. The experiment was designed to test the degradation of quality as a function of channel errors for each tested AMR-WB mode, i.e. to verify the performance of the channel coding for each of the modes.

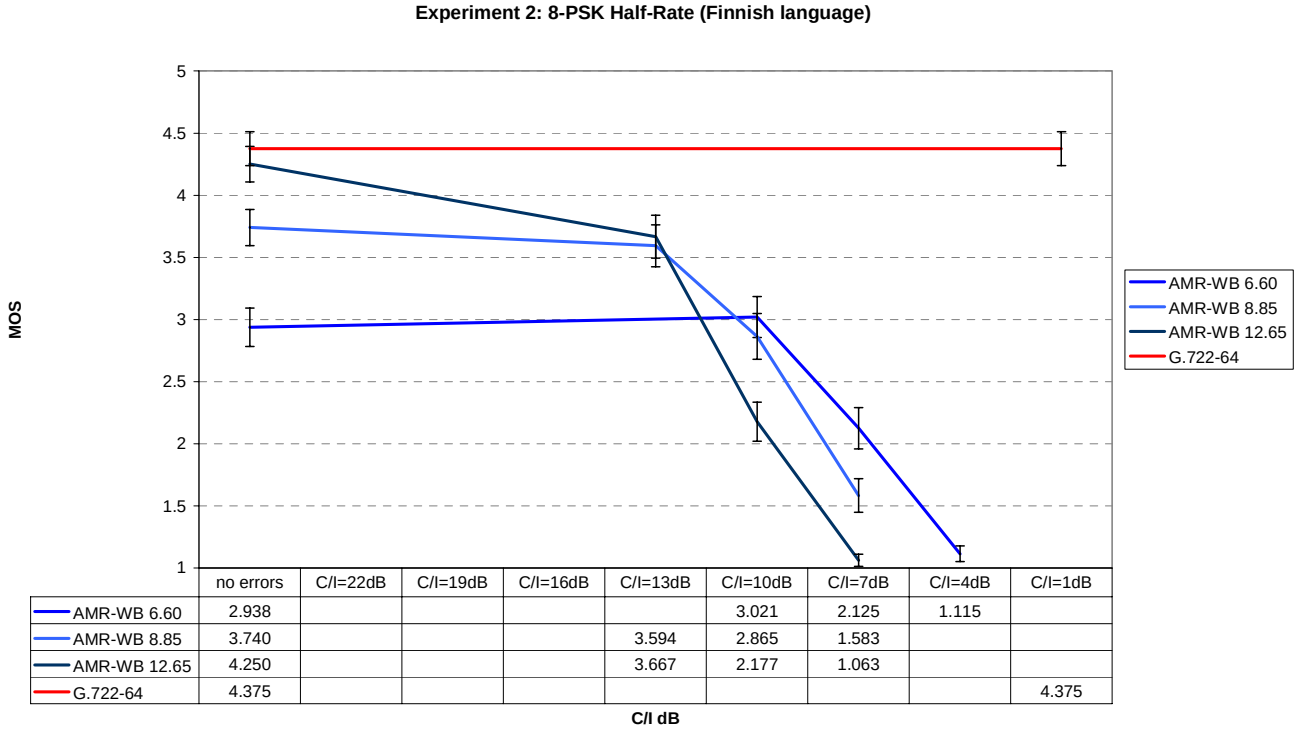
Experiment was performed in one language (Finnish). The presentation of the results in this clause are extract from the TSG-GERAN contribution [43]. A detailed test plan for this experiment is shown in [42]. The error bars in figures 16.1 and 16.2 represent the 95 % confidence intervals.

NOTE: G.722-64 reference codec, shown in figures 16.1 and 16.2, was tested in error-free conditions only.



	no errors	C/I=22dB	C/I=19dB	C/I=16dB	C/I=13dB	C/I=10dB	C/I=7dB	C/I=4dB	C/I=1dB
AMR-WB 6.60	3.177						3.073	2.750	1.635
AMR-WB 8.85	3.563						3.583	2.802	1.229
AMR-WB 12.65	4.052					3.958	3.531	2.010	
AMR-WB 15.85	4.073					3.979	3.229	1.417	
AMR-WB 23.85	4.146				4.104	3.344	1.479		
G.722-64	4.271								4.271

Figure 16.1: Experiment 1, testing GERAN 8-PSK FR channel with Finnish language



	no errors	C/I=22dB	C/I=19dB	C/I=16dB	C/I=13dB	C/I=10dB	C/I=7dB	C/I=4dB	C/I=1dB
AMR-WB 6.60	2.938					3.021	2.125	1.115	
AMR-WB 8.85	3.740				3.594	2.865	1.583		
AMR-WB 12.65	4.250				3.667	2.177	1.063		
G.722-64	4.375								4.375

Figure 16.2: Experiment 2, testing GERAN 8-PSK HR channel with Finnish language

17 Effects of Bit Rate, Input Level, and VAD/DTX (DCR)

The experiment in this clause, was conducted by ITU. The purpose of experiment was to test the operation of VAD/DTX with different input levels and without background noise.

The test methodology was Degradation Category Rating (DCR). Experiment was performed in two language: English and Finnish. The presentation of the results in this clause are extract from the ITU global analysis document [40].

Table 17.1 shows summary results for Exp.1 for the Dynastat and Nokia Listening Labs (LL's). Results are presented for each of the 30 conditions (Mean and Standard Deviation) computed over the six talkers and 32 listeners. The DMOS scores are strongly correlated across the two LL's (r = .930). The averages across conditions for the two LL's are equivalent ($Mean_{Dyn}$ = 3.804, $Mean_{Nok}$ = 3.830) but the Nokia scores have slightly more variation ($StdDev_{Dyn}$ = 0.898, $StdDev_{Nok}$ = 1.011). Figure 17.1 shows a scattergram of the Dynastat vs. Nokia DMOS scores for the conditions tested in Exp.1. Figure 17.2 compares the DMOS scores for the MNRU reference conditions for the two LL's. The slope of the functions is similar in the lower range of MNRU but begins to diverge around 40dB where the Dynastat (NAE) listeners appear to asymptote at a DMOS of approx. 4.7 and the Nokia listeners (Finnish) approach an asymptote closer to the DMOS ceiling of 5.0.

Table 17.1: Summary Results for Experiment 1 (Dynastat - NAE and Nokia - Finish)

Coder/Condition	Dynastat - NAE		Nokia - Finnish	
	DMOS	StdDev	DMOS	StdDev
Codec@23.85kbit/s,-16dBov,VAD/DTX On	4.271	0.766	4.698	0.493
Codec@23.85kbit/s,-16dBov,VAD/DTX Off	4.323	0.766	4.703	0.512
Codec@15.85kbit/s,-16dBov,VAD/DTX On	4.245	0.885	4.531	0.622
Codec@15.85kbit/s,-16dBov,VAD/DTX Off	4.146	0.862	4.599	0.570
Codec@12.65kbit/s,-16dBov,VAD/DTX On	4.052	0.817	4.287	0.652
Codec@12.65kbit/s,-16dBov,VAD/DTX Off	3.891	0.923	4.438	0.636
Codec@23.85kbit/s,-26dBov,VAD/DTX On	4.406	0.753	4.651	0.530
Codec@23.85kbit/s,-26dBov,VAD/DTX Off	4.380	0.756	4.646	0.541
Codec@15.85kbit/s,-26dBov,VAD/DTX On	4.323	0.766	4.594	0.580
Codec@15.85kbit/s,-26dBov,VAD/DTX Off	4.313	0.797	4.490	0.605
Codec@12.65kbit/s,-26dBov,VAD/DTX On	4.125	0.815	4.333	0.642
Codec@12.65kbit/s,-26dBov,VAD/DTX Off	4.042	0.843	4.349	0.677
Codec@23.85kbit/s,-36dBov,VAD/DTX On	4.078	0.874	3.531	0.874
Codec@23.85kbit/s,-36dBov,VAD/DTX Off	4.141	0.872	3.542	0.867
Codec@15.85kbit/s,-36dBov,VAD/DTX On	4.234	0.845	3.432	0.822
Codec@15.85kbit/s,-36dBov,VAD/DTX Off	4.063	0.835	3.557	0.866
Codec@12.65kbit/s,-36dBov,VAD/DTX On	3.854	0.938	3.370	0.821
Codec@12.65kbit/s,-36dBov,VAD/DTX Off	3.922	0.949	3.464	0.843
G.722@48kbit/s,-26dBov	3.109	0.951	3.469	0.655
G.722@56kbit/s,-26dBov	4.068	0.892	3.990	0.731
G.722@64kbit/s,-26dBov	4.260	0.834	4.021	0.752
G.722.1@24kbit/s,-26dBov	3.563	1.021	4.089	0.692
G.722.1@32kbit/s,-26dBov	4.120	0.819	4.359	0.606
Direct	4.677	0.639	4.927	0.261
MNRU,Q=45dB	4.656	0.620	4.672	0.533
MNRU,Q=37dB	3.875	0.984	3.568	0.841
MNRU,Q=29dB	2.635	0.864	2.510	0.622
MNRU,Q=21dB	1.891	0.840	1.807	0.587
MNRU,Q=13dB	1.344	0.653	1.260	0.474
MNRU,Q=05dB	1.125	0.627	1.026	0.160

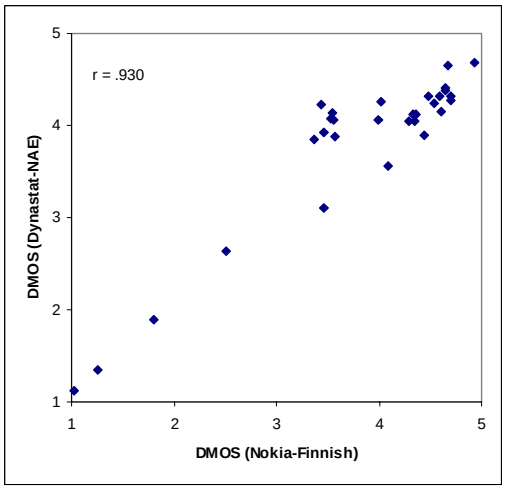


Figure 17.1: DMOS for Nokia vs. Dynastat Listening Labs for Exp.1

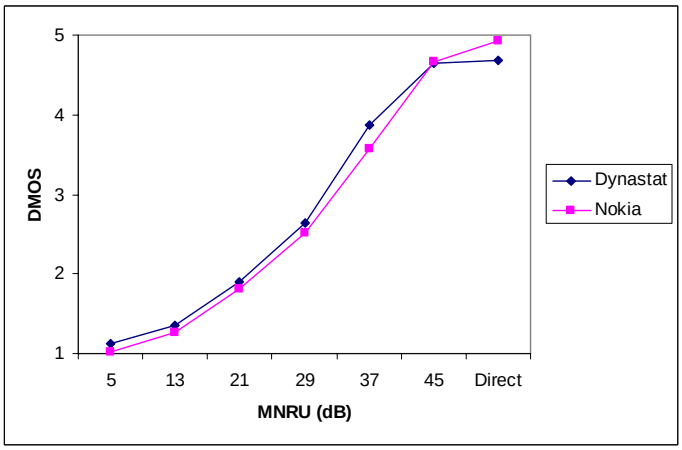


Figure 17.2: DMOS vs. MNRU by Listening Labs for Exp.1

Figure 17.3a shows the effects of *Bit Rate*, *Input Level*, and *VAD/DTX on/off* for the Wideband Coder in the Dynastat results for Exp.1. Figure 17.3b shows the corresponding results for the Nokia version of Exp.1. Also shown in the two figures are the scores for the G.722 and G.722.1 reference coders at various bit rates.

The results for the two LL's reveal the differences in the performance of the Wideband codec for the NAE and Finnish languages (and correspondingly for the Dynastat and Nokia LL's). In NAE, input level has little effect on DMOS while in Finnish the scores for *Low* input level (-36dBov) are markedly lower.

Analysis of Variance (ANOVA) was proposed as a method to examine the differences in the results obtained in the two LL's. Before an ANOVA can be used in this case, however, an initial analysis must be performed separately on the data from the two LL's to determine if an ANOVA is appropriate, i.e. a test for Homogeneity of Variance (HoV). For the two sets of LL results the Mean Square for *Test-Conditions* (N = 18) was .897 for Dynastat and 8.699 for Nokia. The resulting Cochran's statistic for the HoV test is .907, which is substantially higher than the criterion value (.581) for combining the data in a single AVOVA. Therefore, since it is not valid to combine the data for the two LL's into a single ANOVA, we will have to resort to comparisons of the summary results of separate analyses for each LL.

To examine the effects of *Bit Rate*, *Input Level* and *VAD/DTX*, separate ANOVA's were computed for the two LL's. Table 17.2a shows the results of the ANOVA for the Dynastat Exp.1, table 17.2b for the Nokia Exp.1.

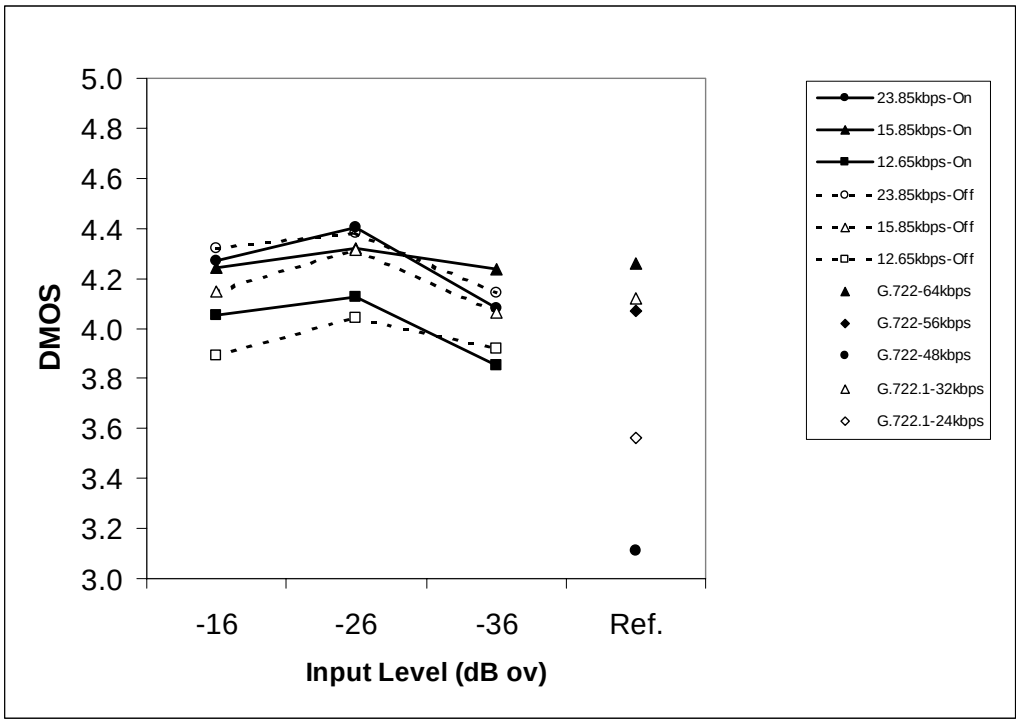


Figure 17.3a: Effects of Bit Rate, Input Level, and VAD/DTX 0n DMOS in the Dynastat Exp.1

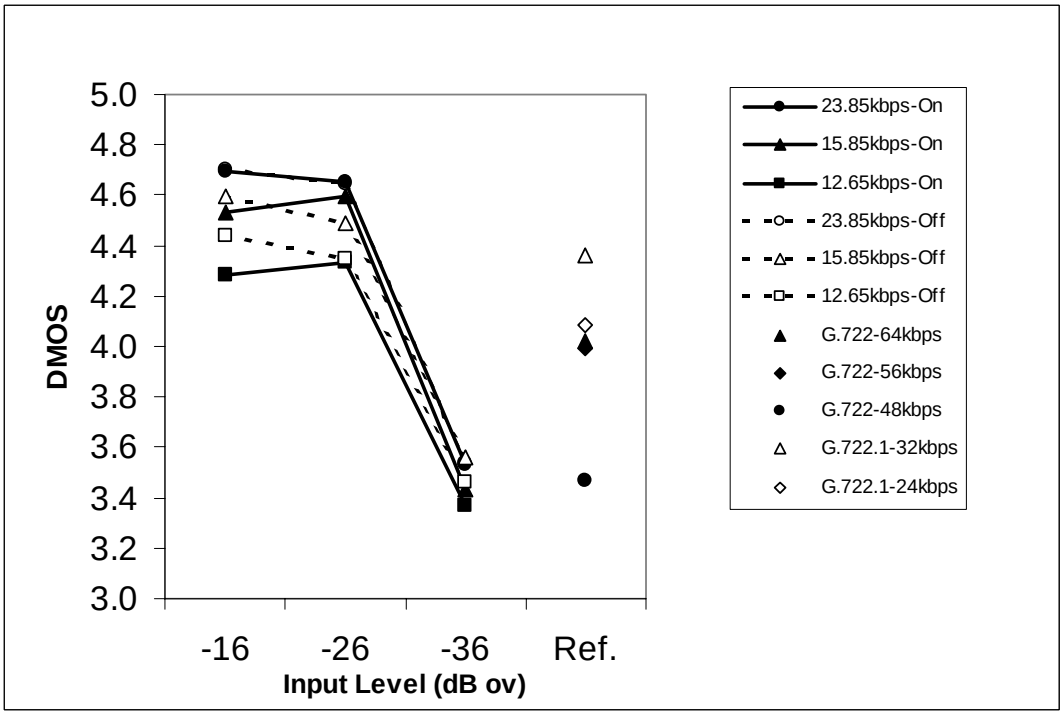


Figure 17.3b: Effects of Bit Rate, Input Level, and VAD/DTX on DMOS in the Nokia Exp.1

Table 17.2a-2b: Anova for input level x Bit Rate x VAD/DTX (on/off) for Exp. 1

Table 2a. - ANOVA for Input Level x Bit Rate x VAD/DTX (on/off) for Dynastat Exp. 1

Source of Variation	df	SS	MS	F	Prob
Input Level	2	4.48	2.242	**12.70**	0.000
BitRate	2	9.03	4.514	**36.25**	0.000
VAD/DTX (On/Off)	1	0.24	0.241	2.40	0.131
Subject	31	107.72	3.475		
Level x BitRate	4	0.39	0.097	1.17	0.327
Level x On/Off	2	0.08	0.038	0.49	0.615
BitRate x On/Off	2	0.39	0.195	**4.18**	0.020
Level x Subject	62	10.95	0.177		
BitRate x Subject	62	7.72	0.125		
On/Off x Subject	31	3.11	0.100		
Level x BitRate x On/Off	4	0.64	0.161	2.30	0.062
Level x BitRate x Subject	124	10.27	0.083		
Level x On/Off x Subject	62	4.84	0.078		
BitRate x On/Off x Subject	62	2.90	0.047		
Level x BitRate x On/Off x Subject	124	8.66	0.070		
Total	575	171.41			

Table 2b. - ANOVA for Input Level x Bit Rate x VAD/DTX (on/off) for Nokia Exp. 1

Source of Variation	df	SS	MS	F	Prob
Input Level	2	139.57	69.784	**84.14**	0.000
BitRate	2	6.39	3.197	**33.09**	0.000
VAD/DTX (On/Off)	1	0.23	0.230	**9.55**	0.004
Subject	31	66.31	2.139		
Level x BitRate	4	0.91	0.228	**5.74**	0.000
Level x On/Off	2	0.36	0.182	**4.08**	0.022
BitRate x On/Off	2	0.17	0.087	2.52	0.089
Level x Subject	62	51.42	0.829		
BitRate x Subject	62	5.99	0.097		
On/Off x Subject	31	0.75	0.024		
Level x BitRate x On/Off	4	0.24	0.060	1.96	0.105
Level x BitRate x Subject	124	4.93	0.040		
Level x On/Off x Subject	62	2.77	0.045		
BitRate x On/Off x Subject	62	2.14	0.034		
Level x BitRate x On/Off x Subject	124	3.82	0.031		
Total	575	286.00			

These ANOVA's included only the data for the 18 test conditions involving the Wideband codec (3 input levels x 3 bit rates x 2 VAD/DTX) but not the data for the reference conditions. Furthermore, the ANOVA's were conducted on the DMOS values averaged over the six talkers.

For the Dynastat data, the main effects for *Input Level* and *Bit Rate* were found to be significant as was the interaction of *Bit Rate x VAD/DTX*. For the Nokia data, the main effects for *Input Level*, *Bit Rate*, and *VAD/*DTX were significant as were the interactions of *Input Level x Bit Rate* and *Input Level x VAD/DTX*. Table17.3 shows the Mean scores for the main effects tested in the separate ANOVA's for the two Exp.1 LL's.

Table 17.3: Mean Scores for Main effects Tested in Exp 1 (* = significant p<.05)

Dynastat Results				Nokia Results			
Input Level *	*High* 4.155	*Nominal* 4.265	*Low* 4.049	Input Level *	*High* 4.543	*Nominal* 4.510	*Low* 3.483
BitRate *	*23.84k* 4.266	*15.85k* 4.220	*12.65k* 3.981	BitRate *	*23.84k* 4.295	*15.85k* 4.201	*12.65k* 4.040
VAD/DTX	*On* 4.177	*Off* 4.135		VAD/DTX *	*On* 4.159	*Off* 4.198	

In the ITU-WB Selection Test which preceded this Characterization Test, a number of *Requirements* and *Objectives* were specified for the candidate coders in the Terms of Reference (ToR) for Wideband Coders. Since several of the same test and reference conditions that were involved in those ToR *Requirements* and *Objectives* were included in Exp. 1, the GAL decided that it would be informative to perform the statistical comparisons where appropriate. Table 4 shows the results of those *Requirements* and *Objectives* comparisons for the Dynastat data; Table 5 shows the corresponding results for the Nokia data.

Of the 40 statistical comparisons shown in tables 17.4 and 17.5 there was only one failure (Dynastat, Req., C04 vs. C20). With a 95 % statistical criterion for pass/fail it would have been reasonable to expect at least two failures based on chance alone -- we could expect one significant result in 20 tests based on chance alone. The single "failed" comparison was a ToR *Requirement* that condition C04 (4.1458) score significantly "Better Than" C20 (4.0677). While the difference was in the right direction (+.0781), it wasn't large enough to be statistically significant. In summary, we believe that its safe to conclude that the Wideband coder successfully passed the ToR *Requirement* and *Objective* conditions included in Exp.1.

Table 17.4: Results of ToR Requirements and Objective Tests for DynaStat Exp.1.

												1.966		
File	Bit Rate	Inp Lvl	VAD/DTX	Req./Obj.	Reference Condition			Test Condition		Diff.	S.E.	t	Stat. Test	Result
					File	DMOS	StdDev	DMOS	StdDev					
C01	24	High	On	Req.	C21	4.2604	0.8344	4.2708	0.7655	-0.0104	0.0579	-0.1795	NWT	Pass
C02	24	High	Off	Req.	C21	4.2604	0.8344	4.3229	0.7657	-0.0625	0.0579	-1.0786	NWT	Pass
C03	16	High	On	Req.	C20	4.0677	0.8924	4.2448	0.8847	0.1771	0.0643	2.7545	BT	Pass
C04	16	High	Off	Req.	C20	4.0677	0.8924	4.1458	0.8620	0.0781	0.0635	1.2303	BT	**Fail**
C05	13	High	On											
C06	13	High	Off											
C07	24	Nom	On	Req.	C21	4.2604	0.8344	4.4063	0.7530	-0.1459	0.0575	-2.5371	NWT	Pass
C08	24	Nom	Off	Req.	C21	4.2604	0.8344	4.3802	0.7563	-0.1198	0.0576	-2.0792	NWT	Pass
C09	16	Nom	On	Req.	C20	4.0677	0.8924	4.3229	0.7657	0.2552	0.0602	4.2418	BT	Pass
C10	16	Nom	Off	Req.	C20	4.0677	0.8924	4.3125	0.7968	0.2448	0.0612	3.9993	BT	Pass
C11	13	Nom	On											
C12	13	Nom	Off											
C13	24	Low	On											
C14	24	Low	Off											
C15	16	Low	On											
C16	16	Low	Off											
C17	13	Low	On											
C18	13	Low	Off											
C01	24	High	On	Obj.	C21	4.2604	0.8344	4.2708	0.7655	-0.0104	0.0579	-0.1795	NWT	Pass
C02	24	High	Off	Obj.	C21	4.2604	0.8344	4.3229	0.7657	-0.0625	0.0579	-1.0786	NWT	Pass
C03	16	High	On	Obj.	C20	4.0677	0.8924	4.2448	0.8847	-0.1771	0.0643	-2.7545	NWT	Pass
C04	16	High	Off	Obj.	C20	4.0677	0.8924	4.1458	0.8620	-0.0781	0.0635	-1.2303	NWT	Pass
C05	13	High	On	Obj.	C19	3.1094	0.9509	4.0521	0.8170	-0.9427	0.0641	-14.6967	NWT	Pass
C06	13	High	Off	Obj.	C19	3.1094	0.9509	3.8906	0.9230	-0.7812	0.0678	-11.5216	NWT	Pass
C07	24	Nom	On	Obj.	C21	4.2604	0.8344	4.4063	0.7530	-0.1459	0.0575	-2.5371	NWT	Pass
C08	24	Nom	Off	Obj.	C21	4.2604	0.8344	4.3802	0.7563	-0.1198	0.0576	-2.0792	NWT	Pass
C09	16	Nom	On	Obj.	C20	4.0677	0.8924	4.3229	0.7657	-0.2552	0.0602	-4.2418	NWT	Pass
C10	16	Nom	Off	Obj.	C20	4.0677	0.8924	4.3125	0.7968	-0.2448	0.0612	-3.9993	NWT	Pass
C11	13	Nom	On	Obj.	C19	3.1094	0.9509	4.1250	0.8154	-1.0156	0.0641	-15.8464	NWT	Pass
C12	13	Nom	Off	Obj.	C19	3.1094	0.9509	4.0417	0.8428	-0.9323	0.0650	-14.3405	NWT	Pass
C13	24	Low	On											
C14	24	Low	Off											
C15	16	Low	On											
C16	16	Low	Off											
C17	13	Low	On											
C18	13	Low	Off											
C19	48	-26	-	-	-	-	-	3.1094	0.9509	-	-	-	-	-
C20	56	-26	-	-	-	-	-	4.0677	0.8924	-	-	-	-	-
C21	64	-26	-	-	-	-	-	4.2604	0.8344	-	-	-	-	-

Table 17.5: Results of ToR Requirements and Objective Tests for Nokia Exp.1.

												1.966		
File	Bit Rate	Inp Lvl	VAD/DTX	Req./Obj.	Reference Condition			Test Condition		Diff.	S.E.	t	Stat. Test	Result
					File	DMOS	StdDev	DMOS	StdDev					
C01	24	High	On	Req.	C21	4.0208	0.7517	4.6979	0.4933	-0.6771	0.0460	-14.7188	NWT	Pass
C02	24	High	Off	Req.	C21	4.0208	0.7517	4.7031	0.5120	-0.6823	0.0465	-14.6623	NWT	Pass
C03	16	High	On	Req.	C20	3.9896	0.7307	4.5313	0.6217	0.5417	0.0491	11.0356	BT	Pass
C04	16	High	Off	Req.	C20	3.9896	0.7307	4.5990	0.5703	0.6094	0.0474	12.8498	BT	Pass
C05	13	High	On											
C06	13	High	Off											
C07	24	Nom	On	Req.	C21	4.0208	0.7517	4.6510	0.5298	-0.6302	0.0471	-13.3934	NWT	Pass
C08	24	Nom	Off	Req.	C21	4.0208	0.7517	4.6458	0.5411	-0.6250	0.0474	-13.1889	NWT	Pass
C09	16	Nom	On	Req.	C20	3.9896	0.7307	4.5938	0.5803	0.6042	0.0477	12.6557	BT	Pass
C10	16	Nom	Off	Req.	C20	3.9896	0.7307	4.4896	0.6053	0.5000	0.0485	10.2992	BT	Pass
C11	13	Nom	On											
C12	13	Nom	Off											
C13	24	Low	On											
C14	24	Low	Off											
C15	16	Low	On											
C16	16	Low	Off											
C17	13	Low	On											
C18	13	Low	Off											
C01	24	High	On	Obj.	C21	4.0208	0.7517	4.6979	0.4933	-0.6771	0.0460	-14.7188	NWT	Pass
C02	24	High	Off	Obj.	C21	4.0208	0.7517	4.7031	0.5120	-0.6823	0.0465	-14.6623	NWT	Pass
C03	16	High	On	Obj.	C20	3.9896	0.7307	4.5313	0.6217	-0.5417	0.0491	-11.0356	NWT	Pass
C04	16	High	Off	Obj.	C20	3.9896	0.7307	4.5990	0.5703	-0.6094	0.0474	-12.8498	NWT	Pass
C05	13	High	On	Obj.	C19	3.4688	0.6545	4.2865	0.6522	-0.8177	0.0473	-17.2967	NWT	Pass
C06	13	High	Off	Obj.	C19	3.4688	0.6545	4.4375	0.6360	-0.9687	0.0467	-20.7460	NWT	Pass
C07	24	Nom	On	Obj.	C21	4.0208	0.7517	4.6510	0.5298	-0.6302	0.0471	-13.3934	NWT	Pass
C08	24	Nom	Off	Obj.	C21	4.0208	0.7517	4.6458	0.5411	-0.6250	0.0474	-13.1889	NWT	Pass
C09	16	Nom	On	Obj.	C20	3.9896	0.7307	4.5938	0.5803	-0.6042	0.0477	-12.6557	NWT	Pass
C10	16	Nom	Off	Obj.	C20	3.9896	0.7307	4.4896	0.6053	-0.5000	0.0485	-10.2992	NWT	Pass
C11	13	Nom	On	Obj.	C19	3.4688	0.6545	4.3333	0.6418	-0.8645	0.0469	-18.4325	NWT	Pass
C12	13	Nom	Off	Obj.	C19	3.4688	0.6545	4.3490	0.6773	-0.8802	0.0482	-18.2652	NWT	Pass
C13	24	Low	On											
C14	24	Low	Off											
C15	16	Low	On											
C16	16	Low	Off											
C17	13	Low	On											
C18	13	Low	Off											
C19	48	-26	-	-	-	-	-	3.4688	0.6545	-	-	-	-	-
C20	56	-26	-	-	-	-	-	3.9896	0.7307	-	-	-	-	-
C21	64	-26	-	-	-	-	-	4.0208	0.7517	-	-	-	-	-

Conclusions

a) Input Level - the Wideband codec shows a significant effect in both LL's with Nokia (Finnish) showing a marked drop in performance at the low level. The source of this degradation in performance at the low input level is not known at this time.

b) Bit rate - the Wideband codec shows a monotonic increase in performance with increasing bit rate; the effect is similar in both LL's.

c) VAD/DTX - there is no effect of VAD/DTX in the Dynastat LL but a small ($diff_{MOS}$ = .039 MOS) though significant effect in the Nokia LL.

d) ToR - of 40 ToR comparisons, a single ToR was failed (Dynastat LL, 15.85K bit/s, high input level, VAD/DTX off).

18 Effects of Bit Rate, Tandeming, and Background Noise (DCR)

The experiment in this clause was conducted by ITU. The purpose of experiment was to test additional background noise types and the tandeming with background noise.

The test methodology was Degradation Category Rating (DCR). Experiment was performed in two language: English and Finnish. The presentation of the results in this clause are extract from the ITU global analysis document [40].

Table 18.1 shows summary results for Exp.2 for the Dynastat and Nokia LL's. As in Table 17.1 results are presented for each of the 40 conditions (Mean and Standard Deviation) computed over the four talkers and 32 listeners involved in the experiment. The DMOS scores are even more strongly correlated across LL's (r = .971) than was the case in Exp.1. The Means across conditions for the two LL's are almost identical ($Mean_{Dyn}$ = 3.489, $Mean_{Nok}$ = 3.435) and the Nokia scores have slightly more variation ($StdDev_{Dyn}$ =1.077, $StdDev_{Nok}$ = 1.156). Figure 18.2 shows a scattergram of the Dynastat and Nokia DMOS scores for Exp.2 with separate symbols for the two background noise conditions involved in the experiment. Figure 18.2 shows DMOS for the MNRU reference conditions for the two background noises for each of the two LL's. The two functions, one for each background noise, for the Dynastat data are virtually identical while the functions for the Nokia data diverge in the midrange of MNRU. Moreover, the Dynastat functions show a steeper slope and a lower upper asymptote than the corresponding functions for the Nokia data.

Table 18.1: Summary Results for Experiment 2 (Dynastat - NAE and Nokia - Finnish)

File	Coder/Condition	Dynastat - NAE		Nokia - Finnish	
		MOS	StdDev	MOS	StdDev
C01	Codec@6.60k,1 tndm,Bab	2.6484	0.8837	2.6719	0.7221
C02	Codec@8.85k,1 tndm,Bab	3.5000	0.9222	3.5938	0.8078
C03	Codec@14.25k,1 tndm,Bab	4.1797	0.8078	4.2656	0.7041
C04	Codec@18.25k,1 tndm,Bab	4.3047	0.7893	4.5313	0.6140
C05	Codec@23.05k,1 tndm,Bab	4.3516	0.8093	4.5547	0.6736
C06	Codec@6.60k,2 tndm,Bab	1.5859	0.7688	1.7188	0.7092
C07	Codec@8.85k,2 tndm,Bab	2.4688	0.8689	2.6641	0.8256
C08	Codec@14.25k,2 tndm,Bab	3.4922	0.8962	3.7109	0.7649
C09	Codec@18.25k,2 tndm,Bab	3.8594	0.8204	4.0391	0.7777
C10	Codec@23.05k,2 tndm,Bab	4.1250	0.7736	4.3125	0.6611
C11	G.722@48k,Bab	3.8828	0.7799	4.1484	0.7434
C12	G.722@56k,Bab	4.1172	0.7697	4.2578	0.6669
C13	G.722@64k,Bab	4.2734	0.7707	4.3984	0.6317
C14	Direct,Bab	4.6719	0.5901	4.7422	0.4567
C15	MNRU,Q=45dB,Bab	4.6016	0.6063	4.6328	0.6625
C16	MNRU,Q=38dB,Bab	4.2969	0.7567	3.8828	0.8383
C17	MNRU,Q=31dB,Bab	3.5469	0.7194	2.8359	0.7401
C18	MNRU,Q=24dB,Bab	2.5234	0.8129	2.0469	0.4998
C19	MNRU,Q=17dB,Bab	1.5625	0.7503	1.4141	0.6087
C20	MNRU,Q=10dB,Bab	1.2656	0.8275	1.0703	0.2857
C21	Codec@6.60k,1 tndem,IntTlk	2.5859	0.9600	2.6094	0.7013
C22	Codec@8.85k,1 tndem,IntTlk	3.4766	0.9878	3.4141	0.7688
C23	Codec@14.25k,1 tndm,IntTlk	4.1406	0.8761	4.2188	0.7418
C24	Codec@18.25k,1 tndm,IntTlk	4.4688	0.7311	4.5234	0.6394
C25	Codec@23.05k,1 tndm,IntTlk	4.4922	0.7735	4.5781	0.5414
C26	Codec@6.60k,2 tndm,IntTlk	1.6953	0.8834	1.7422	0.7126
C27	Codec@8.85k,2 tndm,IntTlk	2.6094	0.9156	2.7109	0.8432
C28	Codec@14.25k,2 tndm,IntTlk	3.7500	0.8511	3.8281	0.7747
C29	Codec@18.25k,2 tndm,IntTlk	4.0156	0.8783	4.0547	0.7022
C30	Codec@23.05k,2 tndm,IntTlk	4.2891	0.8617	4.3750	0.6275
C31	G.722@48k,IntTlk	3.7656	0.9091	3.7734	0.7011
C32	G.722@56k,IntTlk	4.3047	0.7588	4.1328	0.7249
C33	G.722@64k,IntTlk	4.4063	0.7036	4.2578	0.6305
C34	Direct,IntTlk	4.6641	0.6309	4.7969	0.4412
C35	MNRU,Q=45dB,IntTlk	4.6719	0.6288	4.5000	0.5886
C36	MNRU,Q=38dB,IntTlk	4.3438	0.7780	3.5313	0.8412
C37	MNRU,Q=31dB,IntTlk	3.4141	0.9008	2.6016	0.7246
C38	MNRU,Q=24dB,IntTlk	2.4609	0.9041	1.9453	0.6558
C39	MNRU,Q=17dB,IntTlk	1.5703	0.7601	1.2969	0.4755
C40	MNRU,Q=10dB,IntTlk	1.1797	0.6573	1.0156	0.1245

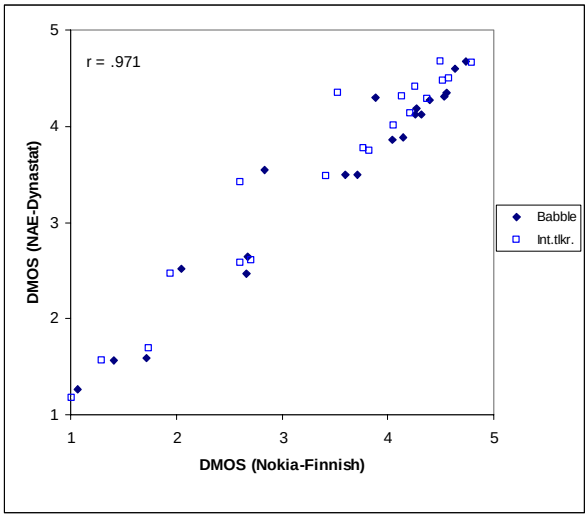


Figure 18.1: DMOS for Nokia vs. Dynastat Listening Labs for Exp.2

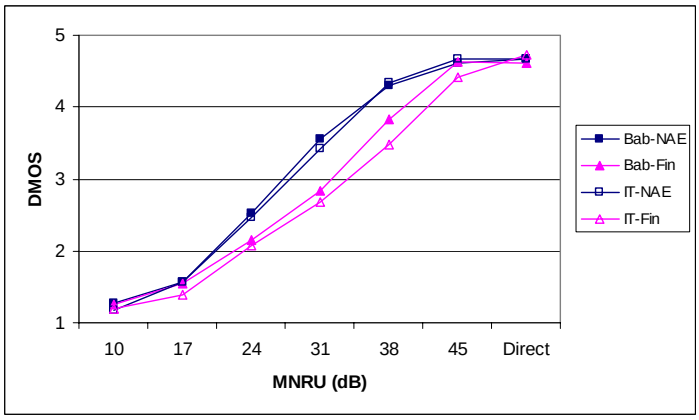


Figure 18.2: DMOS vs. MNRU by Background Noise and Listening Labs for Exp.2

The GAL performed Cochran's HoV test on the data for the two LL's in Exp.2. For the two sets of results the Mean Square for *Test-Conditions* (N = 20) was 26.909 for Dynastat and 27.823 for Nokia. The resulting Cochran's statistic for the HoV test is .508, well within the criterion value (.581) for combining the data into a single AVOVA. However, in light of the failure of the Exp.1 results to pass the HoV test, the GAL determined that it would be inconsistent to present combined results across LL's for Exp.2. Therefore, the presentation of results for Exp.2 will follow the same pattern as those for Exp.1.

Figure 18.3a shows the effects of *Bit Rate*, *Background Noise*, and *Tandeming (1 vs. 2)* on DMOS in the Dynastat results for Exp.2. Figure 18.3b shows the corresponding scores for the Nokia results for Exp.2. Also shown in the two figures are the scores (1 tandem only) for the G.722 reference coder at various bit rates. The results shown in the two figures are consistent except in the Nokia data where the G.722 reference coder scored higher in the *Babble Noise* than in the *Interfering Talker*.

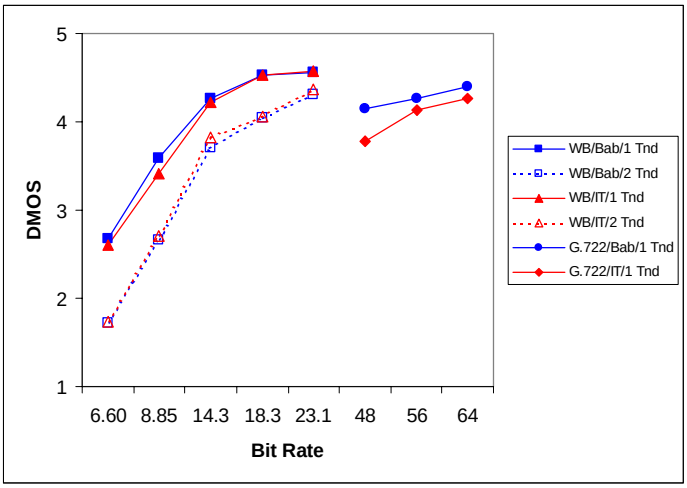


Figure 18.3a: Effects of Bit Rate, Background, and Tandeming on DMOS in the Dynastat Exp.2

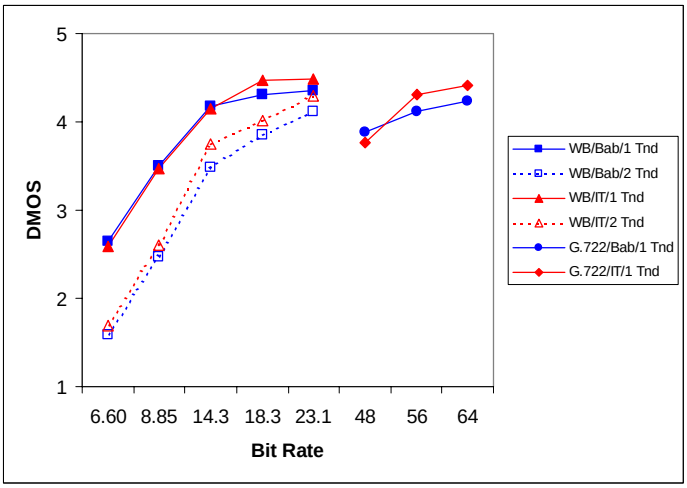


Figure 18.3b: Effects of Bit Rate, Background, and Tandeming on DMOS in the Nokia Exp.2

To test the effects of *Bit Rate*, *Background Noise* and *Tandeming* on DMOS, separate ANOVA's were computed for each of the two LL's. Table 18.7a shows the results of the ANOVA for Dynastat Exp.2, table 18.7b for Nokia Exp.2.

For the Dynastat data, the main effects for *Tandeming* and *Bit Rate* were found to be significant as was the interaction of *Tandeming x Bit Rate*. For the Nokia data, the main effects for *Tandeming* and *Bit Rate* were significant as were the interactions of *Tandeming x Bit Rate* and *Tandeming x Noise*. Table 18.8 shows the mean scores for the main effects tested in the ANOVA's for the results from the two Exp.2 LL's. These ANOVA's included only the data for the 20 test conditions involving the Wideband codec (2 tandems x 5 bit rates x 2 background noises) but not the data for the reference conditions. Furthermore, the ANOVA's were conducted on the DMOS values averaged over the four talkers.

Table 18.8: Mean Scores for Main Effects Tested in Exp. 2 (* = significant, p<.05)

Dynastat Results						Nokia Results					
Tandem *	*1 Tnd* 3.815	*2 Tnd* 3.189				Tandem *	*1 Tnd* 3.896	*2 Tnd* 3.316			
BitRate *	*6.6k* 2.129	*8.85k* 3.014	*14.25k* 3.891	*18.25k* 4.162	*23.05k* 4.314	BitRate *	*6.6k* 2.186	*8.85k* 3.096	*14.25k* 4.006	*18.25k* 4.287	*23.05k* 4.455
Noise	*Babble* 3.452	*Int Tlk* 3.552				Noise	*Babble* 3.606	*Int Tlk* 3.605			

Table 18.7a-7b: ANOVA for Tandeming x Bit Rate x Background Noise for Exp. 2

Table 7a. - ANOVA for Tandeming x Bit Rate x Background Noise for Dynastat Exp. 2

Source of Variation	df	SS	MS	F	Prob
Tandem	1	43.58	43.576	**144.59**	0.000
BitRate	4	399.68	99.920	**214.20**	0.000
Noise	1	0.10	0.100	0.38	0.542
Subject	31	69.28	2.235		
Tandem x BitRate	4	8.67	2.169	**20.01**	0.000
Tandem x Noise	1	0.09	0.088	0.62	0.437
BitRate x Noise	4	0.16	0.039	0.49	0.743
Tandem x Subject	31	9.34	0.301		
BitRate x Subject	124	57.84	0.466		
Noise x Subject	31	8.18	0.264		
Tandem x BitRate x Noise	4	0.40	0.099	0.99	0.416
Tandem x BitRate x Subject	124	13.44	0.108		
Tandem x Noise x Subject	31	4.37	0.141		
BitRate x Noise x Subject	124	9.97	0.080		
Tandem x BitRate x Noise x Subject	124	12.37	0.100		
Total	639	637.46			

Table 7b. - ANOVA for Tandeming x Bit Rate x Background Noise for Nokia Exp. 2

Source of Variation	df	SS	MS	F	Prob
Tandem	1	53.91	53.911	**361.05**	0.000
BitRate	4	463.72	115.930	**489.72**	0.000
Noise	1	0.00	0.000	0.00	1.000
Subject	31	69.58	2.245		
Tandem x BitRate	4	10.05	2.512	**28.84**	0.000
Tandem x Noise	1	0.46	0.465	**8.65**	0.006
BitRate x Noise	4	0.25	0.063	0.87	0.484
Tandem x Subject	31	4.63	0.149		
BitRate x Subject	124	29.35	0.237		
Noise x Subject	31	8.67	0.280		
Tandem x BitRate x Noise	4	0.24	0.059	1.06	0.379
Tandem x BitRate x Subject	124	10.80	0.087		
Tandem x Noise x Subject	31	1.67	0.054		
BitRate x Noise x Subject	124	9.03	0.073		
Tandem x BitRate x Noise x Subject	124	6.90	0.056		
Total	639	669.27			

Conclusions

a) Bit rate - the Wideband codec shows a monotonic increase in performance with increasing bit rate; the results are virtually identical in the two LL's.

b) Tandem – in both LL's there is a significant tandem effect and a significant "Bit rate x Tandem" interaction, i.e. the effects of tandeming (1 tandem vs. 2 tandems) decreases with increasing bit rate.

c) Noise – there was no significant difference in the performance of the Wideband codec across the two background noises (Babble and Interfering talker).

19 Effects of Wideband Coding and Test Method on Music Quality (ACR, DCR)

The experiment in this clause was conducted by ITU. The purpose of experiment was to test AMR-WB codec with additional background noise types and the tandeming with background noise.

Experiment 3 was performed in a single LL, Nokia, but consisted of two sub-experiments: Exp.3a used the ACR, Exp.3b the DCR. The same listeners were used in both sub-experiments to provide the most sensitive comparison of test methodology (ACR vs. DCR) for the evaluation of music quality. Appropriate experimental design procedures were employed to control for the effects of time and order of presentation.

In both the ACR and the DCR methods for evaluating the quality of *speech* signals, multiple talkers are used to sample the variance in performance due to *Talkers*. In Exp.3 *Music Classes* replaced the *Talkers* factor in the experimental design. The following six Music Classes were evaluated in the experiment:

- A1 Classical_1 (music only).
- A2 Classical_2 (music+vocal).
- A3 Modern_1 (music only).
- A4 Modern_2 (music+vocal).
- A5 VoiceOver_Classical.
- A6 VoiceOver_Modern.

Table 19.1 shows the results for Exps. 3a (MOS for the ACR) and 3b (DMOS for the DCR). The two sets of scores are almost perfectly correlated (r = .993) though they have different Means and variances across conditions ($Mean_{MOS}$ = 3.184, $Mean_{DMOS}$ = 3.575, $StdDev_{MOS}$ = 1.129, $StdDev_{DMOS}$ = 1.206).

Table 19.1: MOS (Exp.3a-ACR) and DMOS (Exp.3b-DCR) for Music Samples

File	Coder-Condition	Exp.3a - ACR		Exp.3b - DCR	
		MOS	StdDev	DMOS	StdDev
C01	Codec @23.85 kbit/s, -26 dB	3.8177	0.8641	4.3229	0.6631
C02	Codec @15.85 kbit/s, -26 dB	3.1354	0.9392	3.7865	0.9329
C03	Codec @12.65 kbit/s, -26 dB	2.6354	0.9280	3.1563	0.9635
C04	G.722 @56kbit/s, -26 dBov	3.8385	0.9095	4.1354	0.8328
C05	G.722.1@24kbit/s, -26 dBov	4.4167	0.7543	4.8333	0.4726
C06	Direct	4.4427	0.6763	4.8021	0.4716
C07	MNRU, Q = 45 dB	4.3125	0.7493	4.6823	0.5771
C08	MNRU, Q = 38 dB	3.7708	0.9320	4.2135	0.8869
C09	MNRU, Q = 31 dB	3.0521	1.0117	3.4479	0.9640
C10	MNRU, Q = 24 dB	2.1875	0.9302	2.6458	1.0128
C11	MNRU, Q = 17 dB	1.4792	0.6628	1.7344	0.6995
C12	MNRU, Q = 10 dB	1.1250	0.3316	1.1406	0.4293

Figure 19.1 shows the scattergram of MOS vs. DMOS for the 12 Music conditions evaluated in Exp.3. The high degree of correlation is evident. Figure 19.2 shows a similar plot with different symbols representing the six music classes. Figure 19.3 shows the performance, as measured by both the ACR and the DCR, of the Wideband codec over three bit rates relative to that of two reference codecs, G.722 (56 k bit/s) and G.722.1 (24 k bit/s).

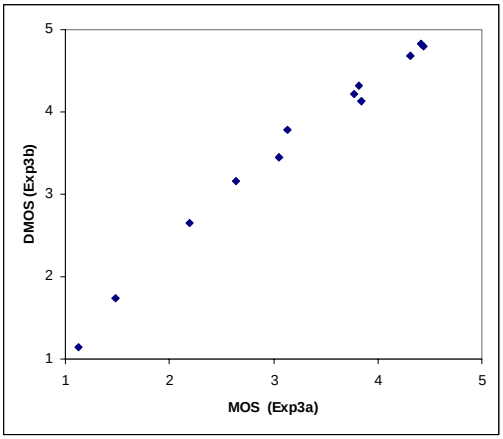


Figure 19.1: MOS (Exp3a – ACR) vs. DMOS (Exp.3b – DCR) for Music Samples

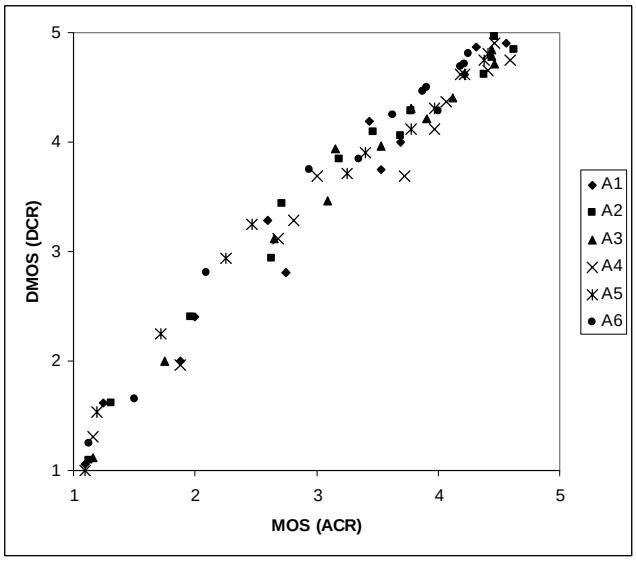


Figure 19.2: MOS (Exp3a – ACR) vs. DMOS (Exp.3b – DCR) by Music Classes

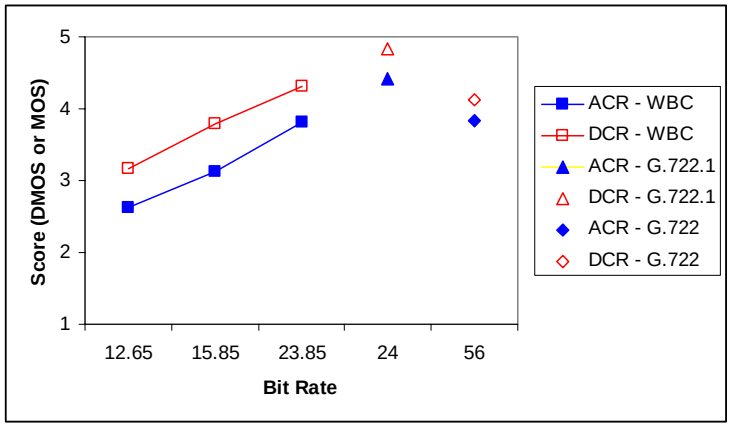


Figure 19.3: MOS (Exp3a – ACR) vs. DMOS (Exp.3b – DCR) for the Test and Reference Codecs

An examination of figure 17.3a/3b, figure 18.6a/6b and figure 19.3 reveal that for speech samples the Wideband Codec at 23.85 k bit/s performed better than the G.722.1 reference codec at 24 k bit/s. The opposite was the case for Music signals.

It's obvious from figures 19.1 to 19.3 that the MOS and DMOS are measuring the same underlying quantity. What is not obvious from these plots is the sensitivity or resolving power of the two methodologies, ACR vs. DCR. To answer this question the GAL performed separate ANOVA's for Exps. 3a and 3b for the five codecs (both test and reference) involved in the two sub-experiments. Table 19.2 shows the results of those ANOVA's.

Table 19.2: Comparison of ANOVA's for Test-conditions x Listeners for Exps. 3a and 3b

ANOVA for Test-conditions x Listeners for Exp.3a (ACR)

Source of Variation	df	SS	MS	F
Test-conditions	4	61.2	15.301	101.66
Listeners	31	24.6	0.794	
Cond. x Lsnrs.	124	18.7	0.151	
Total	159	104.5		

ANOVA for Test-conditions x Listeners for Exp.3b (DCR)

Source of Variation	df	SS	MS	F
Coders	4	50	12.509	91.18
Listeners	31	16.6	0.536	
Cond. x Lsnrs.	124	17	0.137	
Total	159	83.7		

The primary difference between Exp.3a and 3b was the "Test Methodology", ACR vs. DCR, used in the two experiments. The two experiments were conducted by the same LL and used the same music samples, the same experimental design, and some of the same listeners (11 of 32 listeners participated in both experiments). A comparison of the F-Ratios ("Conditions" / "Conditions x Listeners") for the two test methodologies in effect provides a comparison of the relative resolving power of the methodologies. The F-Ratio for the ACR (F=101.66) is in fact higher than that for the DCR (F=91.18). This result would indicate that the ACR has equivalent or possibly even better resolving power than the DCR for these experiments. This finding has important implications for the design of tests of Music quality and suggests additional research into the relative resolving power of various test methodologies, e.g. ACR, DCR, CCR, for a variety of test signals. Since the different methodologies require vastly different amounts of subject listening time (e.g. a typical DCR requires almost twice the amount of listening time as a corresponding ACR and the CCR almost four times as much time as the ACR) then the relative sensitivity of the test methodologies also has important implications in the cost of performing such subjective listening tests.

Conclusions

a) The results from the two methodologies (ACR vs. DCR) were virtually identical.

b) The ACR provided equivalent or better resolving power than the DCR for the test conditions.

c) Performance of the Wideband codec improved with increasing bit rate with the highest bit rate (23.85k bit/s) equivalent to the performance of G.722 at 56k bit/s. At bit rates below that highest rate, which was optimised for music, the codec showed substantially degraded performance for music signals. In particular, all scores were statistically equivalent for all music classes in "Direct" condition, while at 12.65 kbit/s "classical" music showed significant lower performance than "modern" music.

20 Performances with DTMF Tones

Six experiments were performed during the verification phase to evaluate the transparency of the AMR-WB codec modes to DTMF tones. The corresponding test conditions are listed in table 20.1. The experiments were limited to error free conditions only [16].

The frequency deviation was set for the duration of a digit, and was randomly chosen between -1.5 % and +1.5 %. The range of tone levels was chosen to avoid clipping in the digital domain and to exceed the minimum acceptable input level for the Linemaster™ unit used for the detection of DTMF tones.

A set of thirteen codecs was tested in each experiment, comprising the nine AMR-WB modes, G.722 at 48 kbit/s, 56 kbit/s and 64 kbit/s, and the A-law codecs alone (direct condition). The DTMF signals were generated at the frequencies specified in ITU-T Recommendation Q.23. In the DTMF generator, LSB idle noise was added to the test sequences to generate A-law idle noise between digits.

For each experiment, 20 test sequences were processed per codec under test. Each test sequence was produced by the DTMF generator, and comprised a header of *x* ms followed by each of the 16 DTMF digits as defined in ITU-T Recommendation Q.23. The duration of the individual DTMF digits was 80 ms, with a 80 ms gap between adjacent digits. The length of the header in sequence number *n*, was set to:

$x=200+n$ milliseconds; where $n=0..19$.

This approach was taken to exercise the speech codecs over the complete range of possible phase relationships between the start of a DTMF digit and a speech codec frame (20 ms in length). Thus each codec mode was subjected to 320 separate digits per experiment.

For each test sequence, the number of digits undetected by the DTMF detector was recorded. No specific attempt to identify falsely detected digits was made.

Table 20.1: Experimental conditions

Experiment	Low tone level (note)	High tone level (note)	Twist	Digit duration	Frequency deviation
1	-6 dBm	-6 dBm	0 dB	80 ms	none
2	-16 dBm	-16 dBm	0 dB	80 ms	none
3	-26 dBm	-26 dBm	0 dB	80 ms	none
4	-16 dBm	-16 dBm	0 dB	80 ms	±1.5 %
5	-19 dBm	-13 dBm	-6 dB	80 ms	none
6	-13 dBm	-19 dBm	6 dB	80 ms	none
NOTE: The levels are given as measured at the input to the DTMF detector, however, since the DAC is calibrated according to ITU-T Recommendation G.711, 0 dBm in the analogue section is equivalent to -6.15 dBov in the digital section.					

The percentage of undetected digits measured for each codec mode in each experiment is given in table 20.2. Inspection of the results for the AMR-WB speech codec reveals notably worse performance for DTMF signals generated with negative twist. It was noted that digits '2' and '4' were particularly likely to be missed. This was particularly noticeable with mode 1, when digit '4' was systematically not detected. On a one occasion, during Experiment 5, a single digit '7' was detected as two digit '7's for AMR-WB mode 2 (12.65kbit/s). No out of sequence digits observed during any of the Experiments.

Table 20.2: Percentage of DTMF digits undetected when passed through different codecs. The mean value is calculated over all six experiments

Codec mode	Rate (kbit/s)	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Mean
AMR mode 0	6.60	53.8 %	58.8 %	57.5 %	54.7 %	55.9 %	40.6 %	53.5 %
AMR mode 1	8.85	0.9 %	2.5 %	4.4 %	3.1 %	11.3 %	0.3 %	3.8 %
AMR mode 2	12.65	0.0 %	0.0 %	0.9 %	0.3 %	3.8 %	0.0 %	0.8 %
AMR mode 3	14.25	0.0 %	0.0 %	0.0 %	0.0 %	3.1 %	0.0 %	0.5 %
AMR mode 4	15.85	0.0 %	0.0 %	0.3 %	0.0 %	1.6 %	0.0 %	0.3 %
AMR mode 5	18.25	0.0 %	0.0 %	0.0 %	0.0 %	0.6 %	0.0 %	0.1 %
AMR mode 6	19.85	0.0 %	0.0 %	0.0 %	0.0 %	0.6 %	0.0 %	0.1 %
AMR mode 7	23.05	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
AMR mode 8	23.85	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
G.722	48.0	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
G.722	56.0	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
G.722	64.0	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
Direct (A-law)		0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %

No detection errors were measured for the reference A-law condition or the three G.722 modes. In all conditions except negative twist, the seven highest rate AMR-WB modes appear to be essentially transparent to DTMF signals under error free conditions, whereas the two lowest rate modes do not appear to be transparent. The two highest rate modes appear to be completely transparent to DTMF signals with 6 dB of negative twist. It is noted that DTMF signals are often generated by PSTN telephones with negative twist, e.g. -2 dB, to account for the characteristics of the local loop.

21 Performance with Special Input Signals

The purpose of this test was to verify the reliability and stability of the codec using different input signals. Each mode was tested separately in all the tests. The output of some tests was evaluated by expert listening tests, whereas others studied the stability of the AMR-WB codec objectively using long speech and random files [20]. Total of 8 different tests were performed. These tests contained the following signal types:

1) Arbitrary signal.
2) Bursty random noise signals.
3) Background noise signals.
4) Sinusoidal signals.
5) Square wave signals.
6) All zero signal.
7) Long speech signal (radio play).
8) Sinusoidal signals with bad frames.

21.1 Arbitrary signal

All the codec modes were tested with arbitrary signal (Windows DLL file). The main purpose of this test was not to study how well the codec reconstructs the test file but to test possible program failures created by this very untypical signal. Length of this signal was 4min. 39s and its frequency spectrum was relatively flat.

There were no overflows or crashes in any mode. Hence, all the modes passed this test.

21.2 Bursty random noise signals

In this test two signals having several bursts of random noise was used. Signal amplitude used the whole dynamic range from +32 767 and –32 768 and the length of both files was 8 s. The difference between the two signals was the length of the random noise and all zero signal bursts. Signals were the following:

1) Signal A: 0.5s random noise bursts separated by 0.5 s zero signal period.

2) Signal B: 2.0s random noise bursts separated by 1.0 s zero signal period.

Time domain plots for the bursty random noise signals A & B is given in figure 21.1.

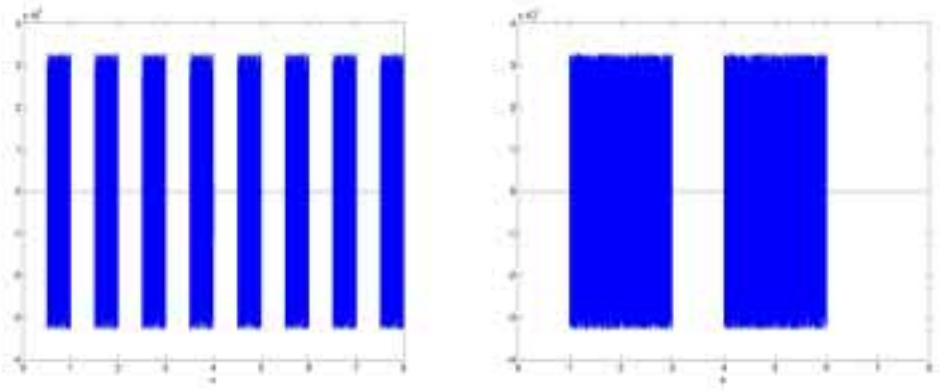


Figure 21.1: Time domain plots for the bursty random noise signals A&B respectively

All the modes produced random bursts. No overflows nor peculiar behaviour like instability was observed.

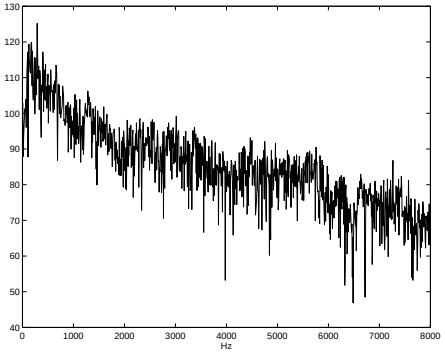
21.3 Background noise signals

Each mode was tested with many types of background noise signals. The noise types and their lengths are given in table 21.1.

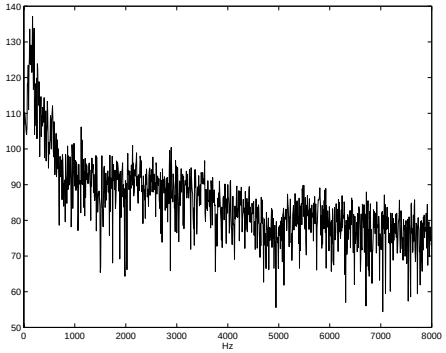
Table 21.1

Background noise type	Length [s]
Car	14.8
Cafeteria	8.5
Hoth	8.7
Motorbike	9.4
Motorboat	36.0
Railway station	46.1
Rain	40.0
Thunder	83.4
Wind	81.3

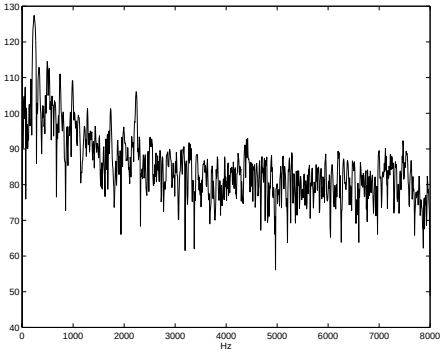
The frequency spectrum figures of the used noise signals are given in figure 21.2. As a result, all the background noises coded with all the modes sounded normal and were recognised and no annoying artifacts were generated.



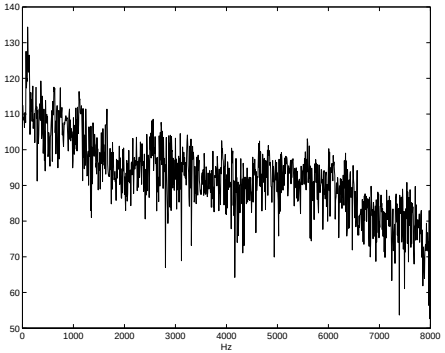
a) Frequency spectrum of the "car" noise



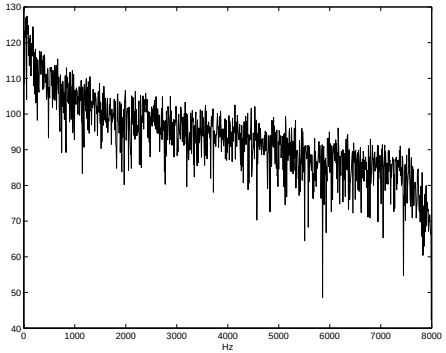
b) Frequency spectrum of the "cafeteria" noise



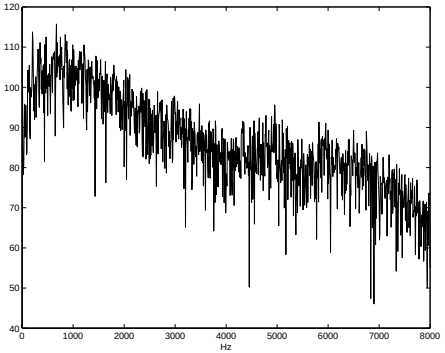
c) Frequency spectrum of the "Hoth" noise



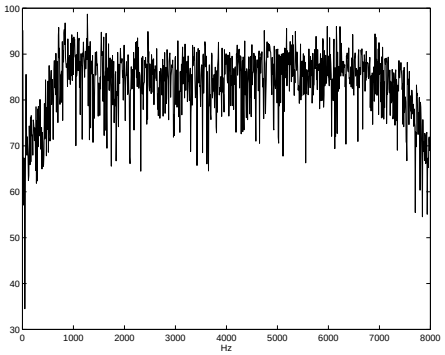
d) Frequency spectrum of the "motorbike" noise



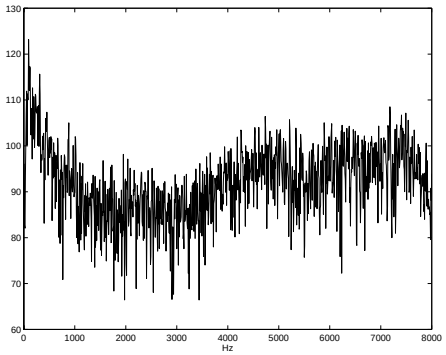
e) Frequency spectrum of the "motorboat" noise



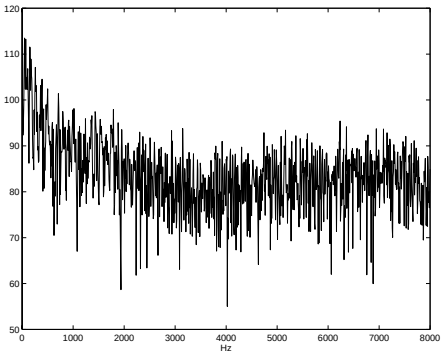
f) Frequency spectrum of the "railway station" noise



g) Frequency spectrum of the "rain" noise



h) Frequency spectrum of the "thunder" noise



i) Frequency spectrum of the "wind" noise

Figure 21.2: Frequency spectrums of the background noise files

21.4 Sinusoidal signals

Three types of sinusoidal signals were tested. [1)] Sinusoidal signal (test signals: 1..10), [2)] Sum of two sinusoidal signals (test signals: 11..18) and [3)] Sinusoidal signal bursts, where each burst were in different frequency and separated by 0.5 s of all zero signal (test signal: 19). The length of the signals was about 8s. The frequency contents of different sinusoidal test signals are given in table 21.2.

Table 21.2: Frequency contents of different sinusoidal wave test signals

Test signal /	Frequency [Hz]									
(test type)	300	500	700	1 000	1 500	2 000	4 000	5 000	6 000	8 000
1 (1)	X									
2 (1)		X								
3 (1)			X							
4 (1)				X						
5 (1)					X					
6 (1)						X				
7 (1)							X			
8 (1)								X		
9 (1)									X	
10 (1)										X
11 (2)	X	X								
12 (2)	X		X							
13 (2)	X			X						
14 (2)	X				X					
15 (2)		X	X							
16 (2)		X		X						
17 (2)		X			X					
18 (2)				X	X					
19 (3)	X	X	X	X	X	X	X	X	X	

The performance of the two lowest modes with sinusoidal tones (and also with DTMF signals) is relatively low. The power of the one frequency with dual frequency signals was in some cases decreased significantly. Also some single sinusoidal signals were degraded when two lowest modes were used. However, the two lowest modes are designed to be used only with mode adaptation in poor radio channel conditions only for a very limited time. For the higher modes, the outputs were acceptable. Frequencies from 6 300 Hz to 7 000 Hz became noise-like because of artificial high band generation.

21.5 Square wave signals

Three types of square wave signals with 50 % duty cycle were tested.

1) Square wave signal (test signals: 1..10);

2) Sum of two square wave signals (test signals: 11..18); and

3) Square wave signal bursts, where each burst were in different frequency and separated by 0.5 s of all zero signal (test signal: 19).

The length of the signals was about 8 s. The frequency contents of different square test signals are given in table 21.3.

The decoder output in this test was acceptable for the higher modes, but the output was distorted for two lowest modes, like in the case of sinusoidal signals.

Table 21.3: Frequency contents of different square wave test signals

Test signal /	Frequency [Hz]									
(test type)	300	500	700	1 000	1 500	2 000	4 000	5 000	6 000	8 000
1 (1)	X									
2 (1)		X								
3 (1)			X							
4 (1)				X						
5 (1)					X					
6 (1)						X				
7 (1)							X			
8 (1)								X		
9 (1)									X	
10 (1)										X
11 (2)	X	X								
12 (2)	X		X							
13 (2)	X			X						
14 (2)	X				X					
15 (2)		X	X							
16 (2)		X		X						
17 (2)		X			X					
18 (2)				X	X					
19 (3)	X	X	X	X	X	X	X	X	X	

21.6 All zero signal

An 8s long signal containing all zero samples was given as an input to each of the modes. Zero output was generated for all the modes and there were no problems.

21.7 Long speech signal (radio play)

The purpose of this test was to check possible overflows, for example, in the counters. The input file was very long (2 h 53 min) a radio play including speech and some music. Active speech level of the input was -26.305 dBov and the speech activity factor: 85.619 %. No problems were observed in any mode.

21.8 Sinusoidal signals with bad frames

The purpose of this test was to verify the behaviour of the codec during and after bad frames when the encoder input is sinusoidal or square wave signal. Same test sequences described in clause 17.4 were processed through the speech codec with all the modes with an exception that some frames were marked as "RX_TYPE=SPEECH_BAD" frames in the following way: One bad frame after 2 s, two consecutive bad frames after 4 s and three consecutive bad frames after 6 s. The results were acceptable. (For one single sinusoidal tone of frequency 1 500 Hz, temporary instability in the decoder was observed.)

21.9 Summary

The tests showed that the AMR-WB speech codec performs well with wide variety of signal types and no unexpected behaviour was observed.

22 Overload Performance

This test is designed to identify any significant problems exhibited under overload (high-level input signal) conditions. Errors were also included in the test. The test was carried out under informal expert listing [25].

Figure 22.1 shows processing flow to prepare test files. The input level for AMR-WB coder was adjusted with *'sv56'* algorithm provided in the ITU-T Recommendation G.191 software tool library (STL2000r3). The output level of decoder was also adjusted with *'sv56'*. A channel error was added in some conditions. An error insertion device adds the error to the code sequence according to the static error profile, provided with *'gen-pat'* in STL, as following: when an error occurs, the EID replaces RX_type to RX_SPEECH_LOST and fills NULL ('0') data to the body part.

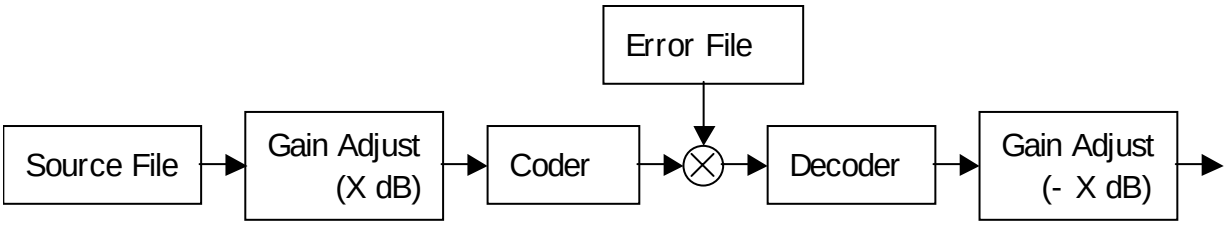


Figure 22.1: Test processing for overload performance

The processed files were up-sampled from 16 kHz to 48 kHz with STL's FIR filter and output digitally from workstation to D/A converter (dCS950) followed by headphone amplifier (TASCAM MH-40MkII) and headphone (AKG HD-25).

4 pairs (2 male and 2 female) of 8 s Japanese sentence were selected from NTT-AT database for the test process. P.341 filter was applied to the selected files with '*filter*' in STL. The mean active power of the source files were normalised to 26 dB below overload. The gain was adjusted to X = 0 dB, 10 dB, 20 dB or 30 dB for each condition. All 9 source coding rates of AMR-WB were tested for all 4 sentences and 4 input levels.

5 % random frame erasure was used as the worst case under 3G-channel. The error profile generated with STL was fed to the EID. The actual generated error rate was 4.5 %. 288 processed files (9 rates x 4 levels x 4 sentences x 2 channel conditions (error-free and 5 % random frame erasure)) were exposed to an expert listener.

In expert listening tests on overload input level, there was no evidence to identify any significant problems, such as gross instability.

23 Muting Behaviour

The error concealment of erroneous/lost frames was tested by setting the BFI flag to '1' (`RX_TYPE` = `RX_SPEECH_BAD` or `RX_TYPE` = `RX_LOST_FRAME`) and by setting the `RX_TYPE` flag to `RX_SID_BAD` if a SID update frame had been received. Several inputs were been tested: clean speech, noisy backgrounds (car and street) and male and female talkers. All the input files were processed in error-free condition; each speech coding rate with and without DTX was tested [24].

Test 1: The BFI flag is set to '1' during a time period of N speech frames. The erroneous/lost speech frames are substituted and the output level gradually decreases. Complete silence is reached after 8 or 9 frames. The decrease is smooth.

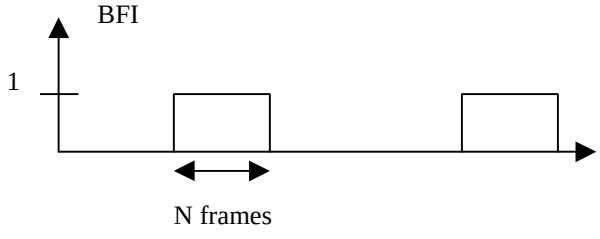


Figure 23.1: Test setup for test 1

Test 2 : The BFI flag is set to '1' every N speech frames. In this case, the erroneous/lost frames are substituted but there is no real cutting if N is large enough. If N = 10, speech is quite well synthesised, if N = 50, the difference is small, if N > 100, the difference is almost inaudible.

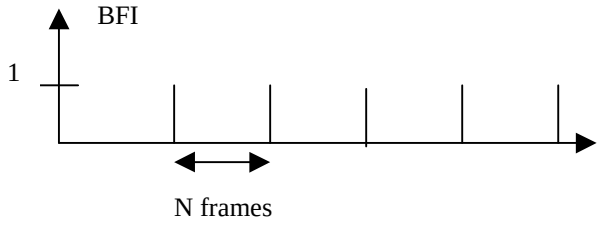


Figure 23.2: Test setup for test 2

Test 3 : The BFI flag is always set to '1' except sometimes for one speech frames. This profile tests the effect of isolated good speech frames. The decoder output is a silence cut by small burst of noise when a good speech frame is received; this noise is not loud but audible.

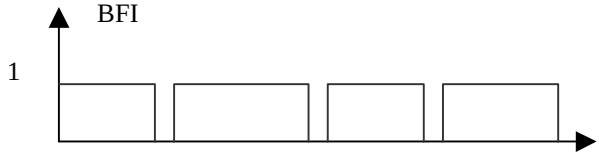


Figure 23.3: test setup for test 3

Test 4 : At the speech decoder input, a single SID update frame is classified as SID bad by modifying the flag `RX_SID_UPDATE` to `RX_SID_BAD`. In this case, this bad frame is substituted by the last valid SID frame information and the procedure for valid SID frames is applied.

Test5: At the speech decoder input, some first SID update frames are not modified and for all the followings, the flag `RX_SID_UPDATE` is changed to `RX_SID_BAD`. In this case of subsequent lost SID frames, the muting is applied, it gradually decreases the output level and complete silence is reached.

No artefacts in the muting behaviour of the AMR-WB were detected in any of the conducted tests. No annoying effects with isolated bad speech frames were detected and synthesis is completely muted after a reasonable period when receiving bad frames.

24 Language Dependency

The selection and characterization tests were performed by a large number of laboratories worldwide using different languages (see clause 6.1 and clause A.3.1). Tests were performed in:

- English (US & UK);
- Finnish;
- French;
- German;
- Japanese;
- Mandarin Chinese; and
- Spanish.

The results demonstrate the AMR-WB codec to perform well across different languages and show that the performance of the codec is not language dependent. The results reported by the different laboratories were consistent.

Tests specially designed for language dependency testing were not considered necessary and were not conducted.

25 Transmission Delay

During the AMR-WB Selection and Verification Phases, the algorithmic round trip delay of AMR-WB codec was estimated in the GSM FR channel (and was compared against the AMR narrowband codec). The algorithmic round trip delay of AMR-WB is very similar to the algorithmic round trip delay of the AMR narrowband codec with only slight increase of few milliseconds (about 3 ms).

Both AMR-WB and AMR narrowband codecs operate on the same frame length (20 ms) and with the same lookahead (5 ms) resulting is rather similar transmission delays. In the AMR-WB codec standardisation, some slight increase was allowed due to allowing the use of bandsplitting filters and also due to the inherently somewhat higher source coding bit-rates (resulting in some increase in GSM Abis-Ater delays). AMR-WB codec employs a bandsplitting filter but the delay of this filter is very low (one-way delay of only 0.9375 ms).

In the following, an overview of the MS-to-MS algorithmic round-trip delay assessment for AMR-WB codec is given. The estimation is taken from Selection Phase Deliverables Tdoc S4-000461. This estimation was verified during the Verification Phase (in Tdoc S4-010052). The delay assessment is given for application A (GSM full-rate channel with additional constraint of 16 kbit/s submultiplexing) and application B (GSM full-rate channel with higher submultiplexing than 16 kbit/s allowed).

The assessment is based on five codec dependent algorithmic delay contributors:

- **analysis frame length delay (T_{sample}):** duration of the segment of PCM speech operated on by the speech transcoder.
- **interleaving and de-interleaving delay (T_{rftx}):** time required for transmission of a speech frame over the air interface due to interleaving and de-interleaving.
- **uplink Abis delay (T_{Abisu}):** time needed to transmit the minimum amount of bits over the Abis interface that are required at the speech decoder to synthesise the first output sample.
- **downlink Abis delay (T_{Abisd}):** time required to transmit all the speech frame data bits over the Abis interface in the downlink direction that are required to encode one speech frame.
- **filter delay (T_{filter}):** total one-way delay of all time-invariant filters (e.g. band-splitting, band-limiting and re-composition filters) in encoder and decoder.

The algorithmic round trip delay without the Abis-Ater interface component (applications A and B):

The MS-to-MS algorithmic round-trip delay without the Abis-Ater interface components is defined as $D_{rt1} = 2(T_{sample} + T_{rftx})$.

For the AMR narrowband codec, $D_{rt1} = 2(T_{sample} + T_{rftx}) = 2(25 + 37.5) = 125$ ms (worst case: 12.2 kbit/s AMR mode).

For AMR-WB codec, for all modes in applications A and B the following applies: $T_{sample} = 25$ ms (duration of the 20 ms speech frame and 5 ms lookahead), $T_{rftx} = 37.5$ ms (same interleaving is used as in AMR narrowband FR channel mode). Therefore, the MS-to-MS algorithmic round-trip delay without the Abis-Ater interface component for AMR-WB is exactly the same as for AMR narrowband (125 ms).

The algorithmic round trip delay component over the Abis-Ater interface (note) (applications A and B):

NOTE: The AMR-WB TRAU frames were not known exactly during the time of the the above estimation resulting in some inaccuracy in the assessment.

The algorithmic round trip delay component over the Abis-Ater interface is defined as $D_{rt2} = T_{Abisu} + T_{Abisd}$.

For AMR narrowband codec, $D_{rt2} = 24.25$ ms (worst case: 12.2 kbit/s AMR mode).

For AMR-WB codec in application A, $D_{rt2} = 7.25$ ms +18.375 ms = 25.625 ms (worst case: highest mode applicable in application A, the 14.25 kbit/s mode).

For AMR-WB codec in application B, the Abis-Ater uplink and downlink delays are lower than for application A due to higher submultiplexing.

The overall algorithmic round trip delay with filter component [1] (applications A and B):

The overall MS-to-MS algorithmic round-trip delay is defined as $D_{round\text{-}trip} = 2(T_{sample} + T_{rftx}) + T_{Abisu} + T_{Abisd} + 2\,T_{filter} = D_{rt1} + D_{rt2} + 2\,T_{filter}$

For AMR narrowband codec, $D_{round\text{-}trip}$ = 149.25 ms (worst case: 12.2 kbit/s AMR mode).

For AMR-WB codec in application A, $D_{round\text{-}trip}$ = 152.5 ms (worst case: the 14.25 kbit/s mode). This exceeds AMR narrowband slightly (by about 3 ms).

For AMR-WB codec in application B, the Abis-Ater uplink and downlink delays are lower than for application A, and T_{filter} is the same for all codec modes in applications A and B. Therefore, the overall algorithmic round trip delay is lower for application B than for application A.

26 Frequency Response

This test is designed to test the frequency response of the AMR-WB codec. The AMR-WB codec has been tested at fixed bit rates (6.6 kbit/s, 8.85 kbit/s, 12.65 kbit/s, 14.25 kbit/s, 15.85 kbit/s, 18.25 kbit/s, 19.85 kbit/s, 23.05 kbit/s and 23.85 kbit/s) in error free condition. The DTX was switched off during the test. Three different methods were used to measure the frequency response and they are described in the following clauses [22].

In the first method, tones signals have been generated in the range 10 Hz to 7 010 Hz with a frequency step of 20 Hz. Each tone had a duration of 10 s. The frequency response of the AMR codec has been evaluated by computing the logarithmic gain according to the following equation:

Logarithmic gain measure: $$\mathrm{Gain_{dB}} = 10\log_{10}\left[\sum_{k=1}^{M}\mathrm{out}(k)^2 / \sum_{k=1}^{M}\mathrm{inp}(k)^2\right]$$

Where inp(k) and out(k) are the original and the processed signals and M is the number of processed samples.

In the second method, different types of noises have generated and processed. The frequency response has been evaluated by computing the spectra for input signal and processed signal. The considered noises are white noise and pink noise. Pink noise with an attenuation of 6dB per octave is a good representative of speech, so it is preferred way of measuring the frequency response of a speech codec designed specially for this type of signals.

The frequency responses of the 9 bit rates of the AMRWB codec are reported in figure 26.1, figure 26.2 and figure 26.3. Figure 26.1 gives the results of the 1st method. Figure 26.2 and figure 26.3 give the results of the 2nd method.

According to the 1st method, some limitations appear on all of the bit rates. When applying the definition of the 3 dB bandwidth, none of the bit rates has a 7 kHz bandwidth. The 2 lowest modes are extremely limited and the 6 other modes present a bandwidth of 50 Hz to 5 700 Hz.

According to the second method when the input signal is white noise, only the two lowest bit rates are really limited. The 5 bit rates between 12.65 kbit/s and 23.05 kbit/s present a bandwidth of 50 Hz to 6 400 Hz. The highest bitrate has a bandwidth of 50 Hz to 6 600 Hz. When the input signal is pink noise, the 2 lowest bit rates are limited, the 5 bit rates between 12.65 & 23.05 kbit/s present a bandwidth of 50 Hz to 6 000 Hz. The highest bitrate has a bandwidth of 50 Hz to 6 600 Hz

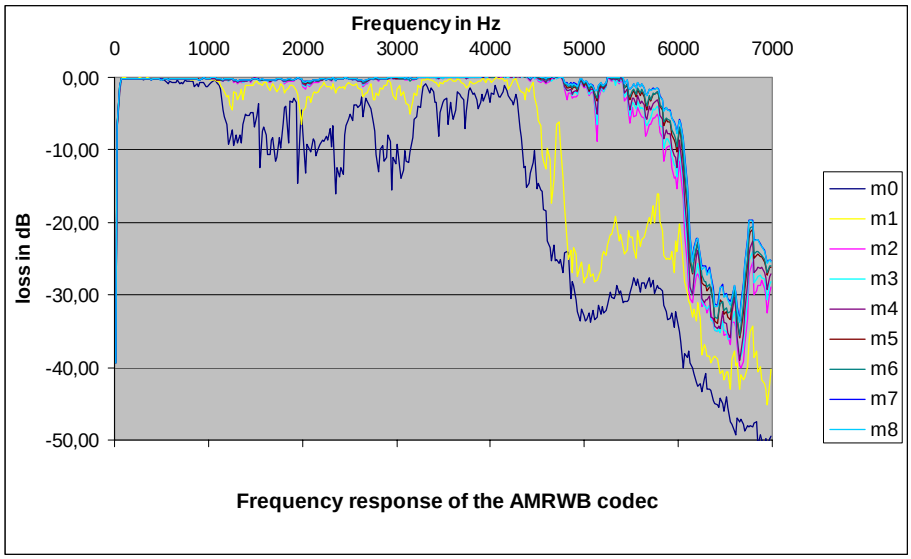


Figure 26.1: Frequency response of the AMR-WB codec (1st method)

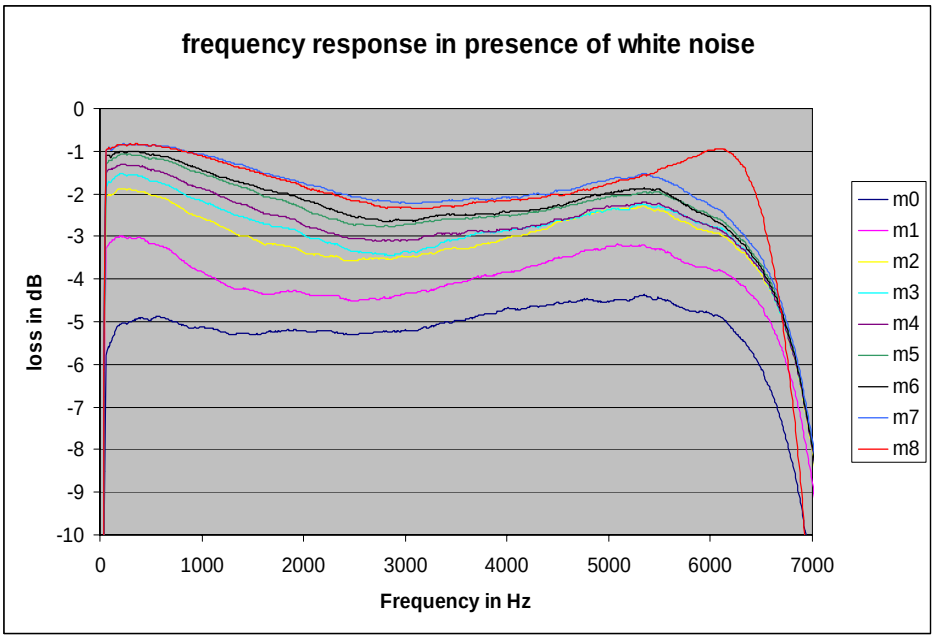


Figure 26.2: Frequency response of the AMR-WB codec (2nd method)

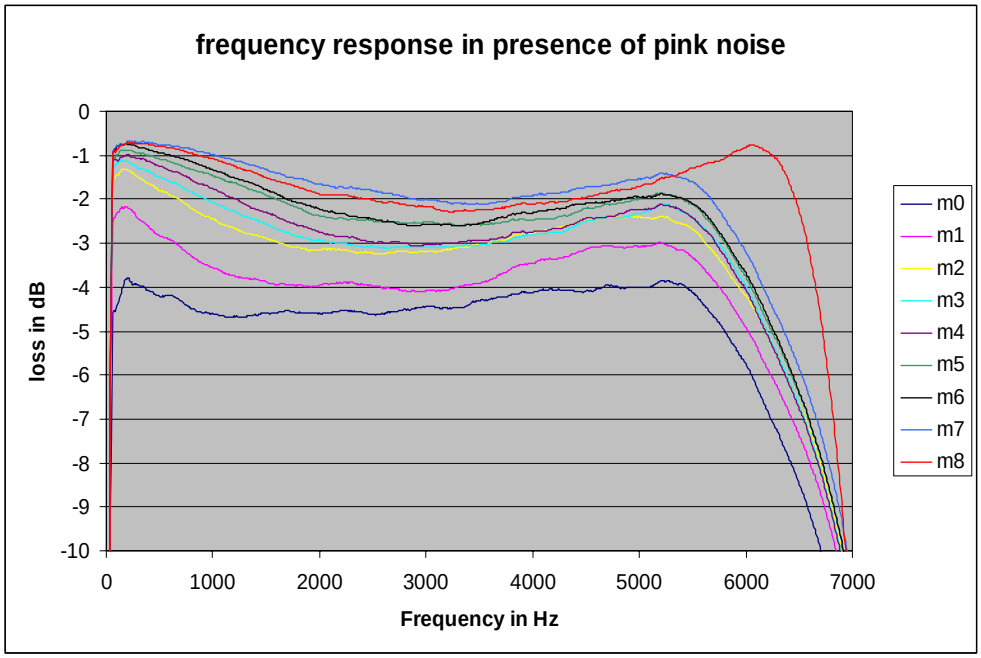


Figure 26.3: Frequency response of the AMR-WB codec (2nd method)

The AMR-WB codec is very dependent of the input signal. Considering that this codec is mainly to be used as a speech codec, the 2nd method seems to be more appropriated for computing the frequency response. The 2 lowest modes have somewhat limited frequency response but the 7 other modes are about compliant with the 7 kHz bandwidth.

27 Signalling Tones

This test checks the performance of the AMR-WB codec with signaling tones. The Software version was version 5.1.0 of the AMR-WB codec. Compilation and execution of the software was performed on PC platform using VisualC++ compiler [23].

Five different types of French network signaling tones have been tested: Two different dial tones, one ringing tone, a busy tone and a special information tone. The description of the different tones is given below:

1. Continuous DIAL TONE number 1 at 440 Hz, 10 s duration.
2. Continuous DIAL TONE number 2 at 330+440 Hz, 10 s duration.
3. RINGING TONE at 440 Hz with duration **1.5** – 3.5 and a total duration of 12.5 s.
4. BUSY TONE at 440Hz with duration **0.5** – 0.5 and a total duration of 12.5 s.
5. SPECIAL INFORMATION TONE at 950/1400/1800 Hz and duration (**3×0.3** – 2x0.03) – 1.0 and a total duration of 12.5 s.

The level of the signaling tones was set at -10 dBm0. Additionally, a set of signaling tones was generated at -15 dBm0 which is the lowest level recommended in ITU-T Recommendation E.180. They were used for testing at a subset of testing conditions. The signaling tones at a level of -10 dBm0 were tested under clean error conditions with no adaptation activated and fixing the codec mode to the 9 different possible modes. The test was run for DTX off and DTX on. The sampling frequency of 16 kHz and 8 kHz have been used.

The testing has been performed by informal listening involving trained listeners, their main concern being that the tones should be recognized.

The test results can be summarized in the following:

- No significant effect was perceived when listening with DTX ON or DTX OFF: the conclusions are the same.

- For the error free conditions: the decoded tones are clearly recognized. Yet the quality from the higher to the lower rate is decreasing and for the two lowest bit rates (6.6 and 8.85) the quality is rather poor.

Figure 27.1 shows the original special information tone (16 kHz) and the signal processed by the AMR-WB mode 0 (6.6 kbit/s). It is clear that the processed signal is severely degraded. When using 8 kHz sampling frequency as shown in figure 27.2, the test results are a little bit worse.

Though the quality of network signaling tones is decreasing audibly with lower bit rates, the signaling tones were clearly recognized under all testing conditions. The high recognition rate of the tones might be related to the fact that the user is expecting to hear a tone, and would be therefore recognizing the tone even at very poor quality.

The activation of DTX did not show any effect on the transparency of the AMR-WB codec towards signaling tones. This holds also for signaling tones at lower levels.

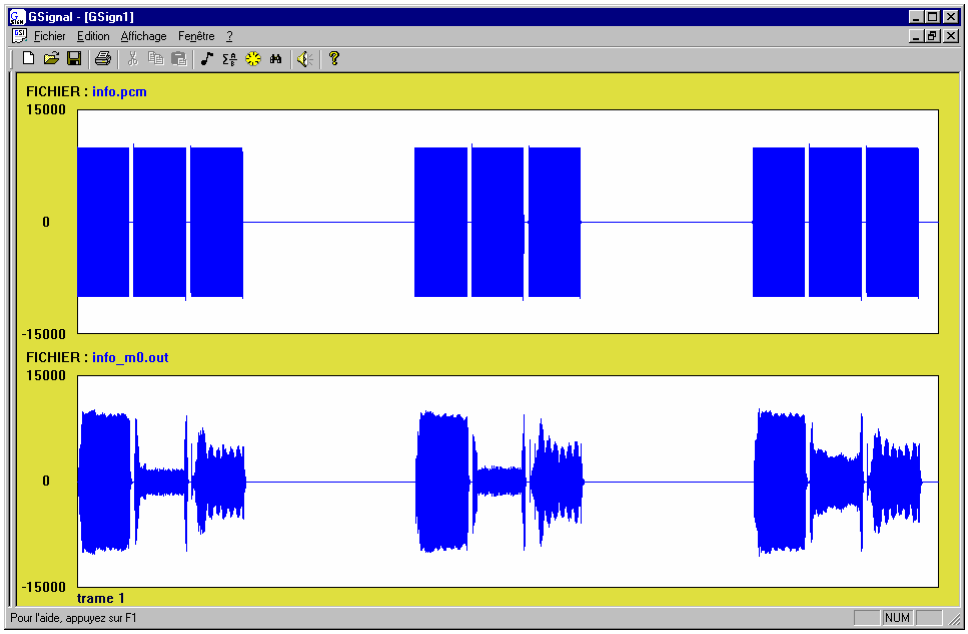


Figure 27.1

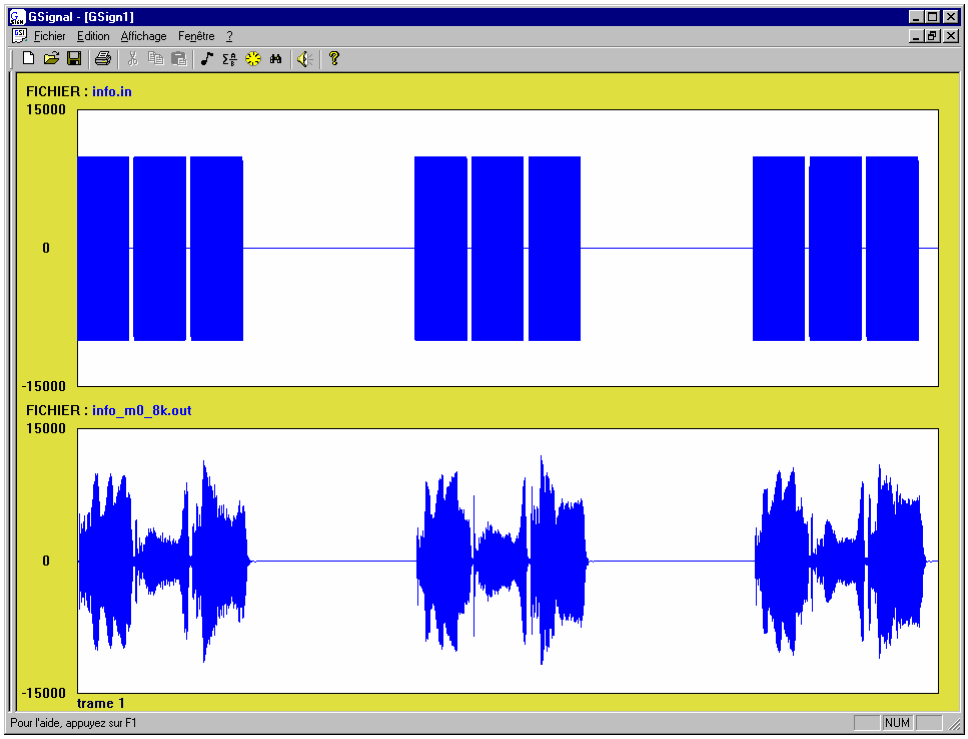


Figure 27.2

28 Complexity Analysis

The AMR-WB speech codec complexity was evaluated using the methodology previously agreed for the standardization of the AMR speech codec [14].

For each codec mode, the complexity is characterized by the following items:

- Number of cycles;
- Data memory size;
- Program memory size.

The actual values for these items will eventually depend on the final DSP implementation. The methodology adopted for the standardization of previous GSM speech codecs provides a way to overcome this difficulty.

In this methodology, the speech and channel coding functions are coded using a set of basic arithmetic operations. Each operation is allocated a weight representative of the number of instruction cycles required to perform that operation on a typical DSP device. The Theoretical Worst Case complexity (wMOPS) is then computed by a detailed counting of the worst case number of basic operations required to process a speech frame.

The wMOPS figure quoted is a weighted sum of all operations required to perform the speech and/or channel coding.

Note that in the course of the codec selection, the Worst Observed Frame complexity was also measured by recording the worst case complexity figure over the full set of speech samples used for the selection of the AMR-WB codec.

In the case of AMR-WB, the complexity was further divided in the following items:

- Speech coding complexity in terms of wMOPS, RAM, ROM Tables and Program ROM.
- GMSK Full Rate channel coding complexity in terms of wMOPS, RAM, ROM Tables and Program ROM.

The separation of the speech and channel complexity was motivated by the fact that these functions were generally handled by different system components in the network (speech transcoding functions in the TRAU and channel coding/decoding in the BTS).

Table 28.1 presents the Theoretical Worst Case (TWC) complexity (wMOPS) for the different AMR-WB speech codec modes in addition to the Worst Observed Frame (WOF) reported during the selection phase. According to the design constraints for the AMR-WB speech codec up to 41.6 wMOPS were allowed including the VAD/DTX system (see permanent document WB-4 [8]). The measured TWC figure of 38.97 wMOPS is clearly below this limit.

Table 28.2 provides the same parameters for the GSM GMSK Full Rate channel codec. According to the design constraints for the AMR-WB codec up to 5.7 wMOPS were allowed (see permanent document WB-4 [8]). Again, the measured TWC figure of 3.93 wMOPS is clearly below this limit.

Table 28.3, table 28.4 and table 28.5 provide the RAM, ROM Tables and Program ROM complexity figures for the speech and channel codecs.

Table 28.1

wMOPS / Speech Codec + VAD + DTX											
Mode	**23.85**	**23.05**	**19.85**	**18.25**	**15.85**	**14.25**	**12.65**	**8.85**	**6.60**	**TWC**	***WOF est***
Speech encoder	29.07	30.84	31.14	30.22	29.41	29.24	26.91	23.59	20.46	31.14	-
Speech decoder	6.90	6.89	6.83	6.82	6.79	**6.76**	6.73	7.47	7.83	**7.83**	-
Total Speech	35.97	37.73	37.97	37.04	36.20	36.00	33.64	31.06	28.29	**38.97**	*36.13*

Table 28.2

wMOPS / Channel Codec for TCH/WFS											
Mode	**23.85**	**23.05**	**19.85**	**18.25**	**15.85**	**14.25**	**12.65**	**8.85**	**6.60**	**TWC**	***WOF est***
Channel encoder	-	-	0.39	0.58	0.51	0.48	0.45	0.42	0.39	**0.58**	-
Channel decoder	-	-	1.32	3.35	2.95	**2.68**	2.42	1.85	1.53	**3.35**	-
Total Channel	-	-	1.71	3.93	3.46	3.16	2.87	2.27	1.92	**3.93**	*3.45*

Table 28.3

Data RAM (static + scratch)			
	static + scratch requirement	**static used**	**scratch used**
Speech Encoder + VAD+DTX	15 000 + 149 Words	1 381 Words	4 389 Words
Speech Decoder + DTX		758 Words	
Channel Encoder (TCH/WFS)	3 000 Words	229 Words	
Channel Decoder (TCH/WFS)		242 Words	
Link Adaptation		**102 Words**	
Total		2 712 Words	4 389 Words
		7 101 Words	

Table 28.4

Data ROM Tables		
	requirement	**used**
Speech Codec + VAD + DTX	18 000 + 513 Words	9 929 Words
Channel Codec (TCH/WFS)	4 500 Words	3 075 Words
Link Adaptation	-	105 Words
Total	**23 013 Words**	**13 109 Words**

Table 28.5

Program ROM		
	requirement	**used**
Speech Codec + VAD + DTX	5 821 + 491	3 889 basic-ops
Channel Codec (TCH/WFS)	2 013	418 basic-ops
Link Adaptation	-	48 basic-ops
Common (log2, oper32b)	-	35 basic-ops
Total	**8 571 basic-ops**	**4 390 basic-ops**

29 Comfort Noise Generation

This clause reports the results of the verification of the comfort noise generation system of the AMR-WB codec. For the purpose of verification an investigation of the VAD performance and its consequence both on the achievable voice/channel activity and speech quality has been made. Furthermore, it has been investigated if due to comfort noise generation noticeable artefacts are caused in the synthesised signal [21].

29.1 VAD

As a base for all experiments of the VAD performance a five minutes long file was used with conversational speech. This speech file is created from a database with Swedish speech material, containing two male and two female speakers. The material is concatenated so that it contains approximately 40 % speech time and 60 % time of silence. For the main part of the investigations the input level of the speech is set to –26 dBov. However, tests with different input levels of the speech material have also been made. In these cases, the input level was set to –16 dBov and –36 dBov, respectively.

Four different types of noises are added to the speech file. The noises are recordings from car, street, office and airport hall environments. The noises differ widely in stationarity. In order to give some idea of the stationarity of the noises, frame energy variances, i.e. the variances of frame-wise energy estimates, were calculated. The result of this computation is shown in figure 29.1.

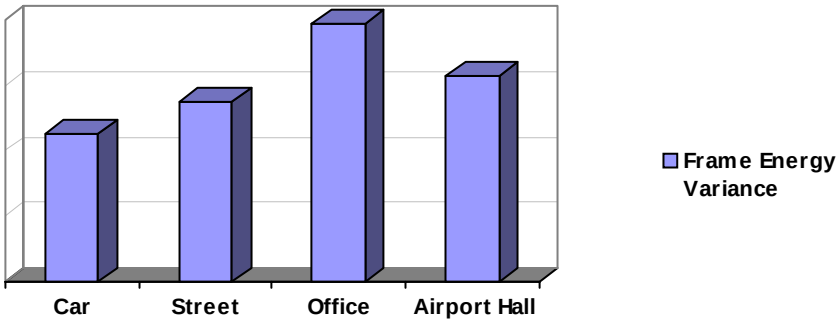


Figure 29.1: Stationarity of noises

In addition, two kinds of music are used as background noises. One file containing classical music (Bach) and one file containing rock music (Smashing Pumpkins). According to the stationarity measure from above, the file containing classical music is the more stationary one, and the music pieces are less stationary than the other noises.

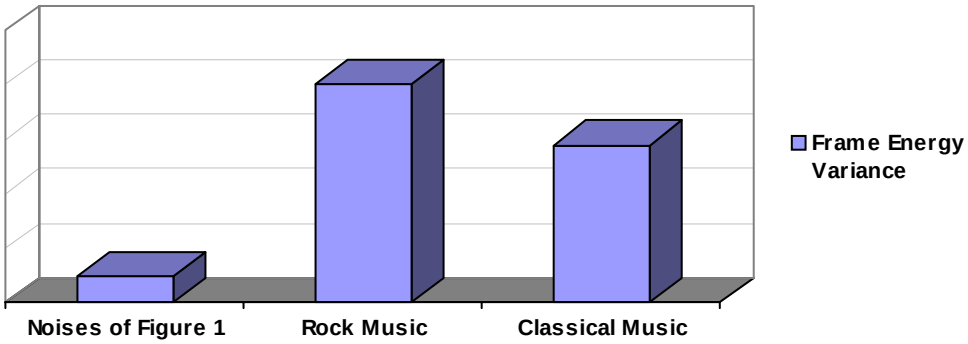


Figure 29.2: Stationarity of music files

The background files are added to the speech files at four different levels such that signal-to-noise ratios of 40 dB, 30 dB, 20 dB and 10 dB are obtained. The noise is scaled in the same way as in the AMR-WB selection tests, see [11].

29.2 Voice/Channel activity

To evaluate the performance of the voice activity detection we have observed the VAD-flag and calculated the voice activity and clipping for different background conditions. The voice activity is calculated as follows:

$$\text{voice activity} = \frac{\text{number of frames where VAD flag is "1"}}{\text{number of all frames}}$$

The voice activity obtained from the different background conditions is compared to the activity of the ideal case, i.e. the clean case without any background noise.

The channel activity is the relevant parameter for evaluating the gain of a DTX system. It is the ratio between the number of transmitted frames (SPEECH, SID_FIRST, SID_UPDATE) and the number off all frames including the NO_DATA frames. The channel activity is calculated as follows:

$$\text{channel activity} = \frac{\text{number of frames - number on NO\_DATA frames}}{\text{number of all frames}}$$

Voice activity and channel activity measurements for the different background cases and different input levels are shown in figure 29.3, figure 29.4, figure 29.5 and figure 29.6.

In figure 29.3 and figure 29.4 it can be seen that the achievable activity strongly depends on the type of noise (the stationarity). It is found that the activity levels for more stationary noises such as car are reasonably low, just above the corresponding activity levels for clean speech. For the less stationary noise and music background signals the activity levels approach 100 %.

Moreover, depending on the noise type, there is a lesser or stronger dependence on the SNR-ratio. For more stationary noise like car noise only a minor dependence of the achievable activity figures on the SNR-ratio was observed.

Comparing voice and channel activity figures, it can be stated that the channel activity figures at maximum are about 10 % higher than the corresponding voice activity figures. The biggest differences are found with 11 % for clean speech and the cases with low voice activity, as e.g. for car noise. Smaller differences occur for the cases with higher voice activity.

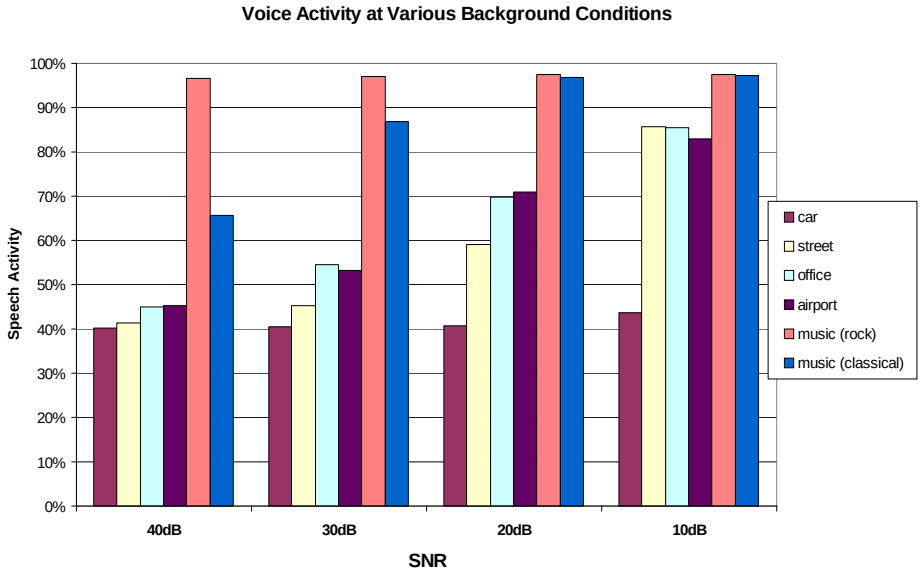


Figure 29.3: Voice activity for different background conditions, at speech level –26 dBov (Voice activity for clean speech is 40 %)

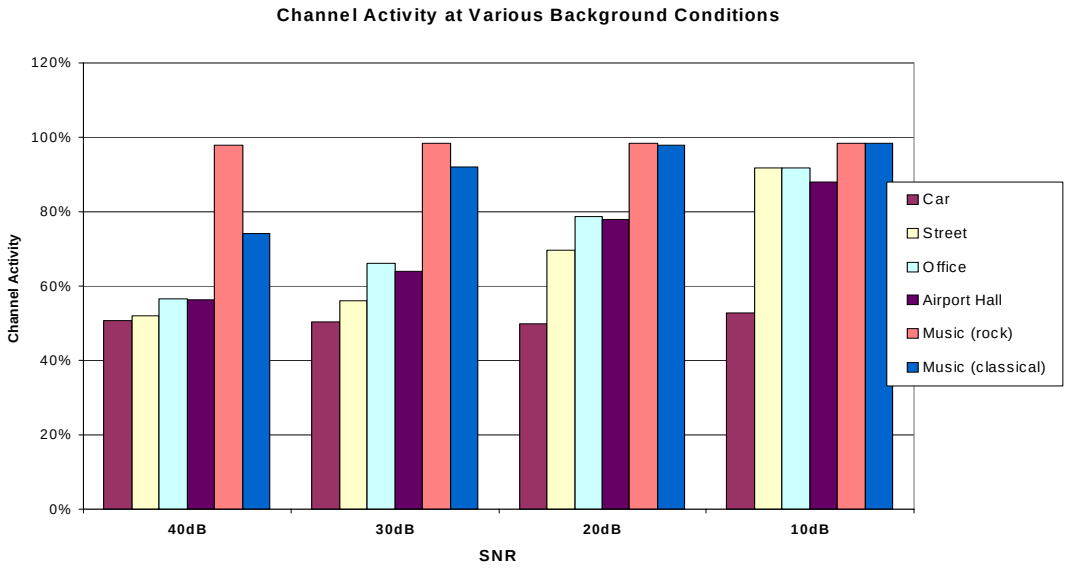


Figure 29.4: Channel Activity for different background conditions, input speech level = –26dBov (for clean speech; channel activity = 51 %)

Figure 29.5 and figure 29.6 show the dependence of the achievable voice and, respectively, channel activities on the input level for the example of street noise. It is found that the activities increase with the level. However, the dependence is not strong. For the more stationary car noise, this dependence is only minor.

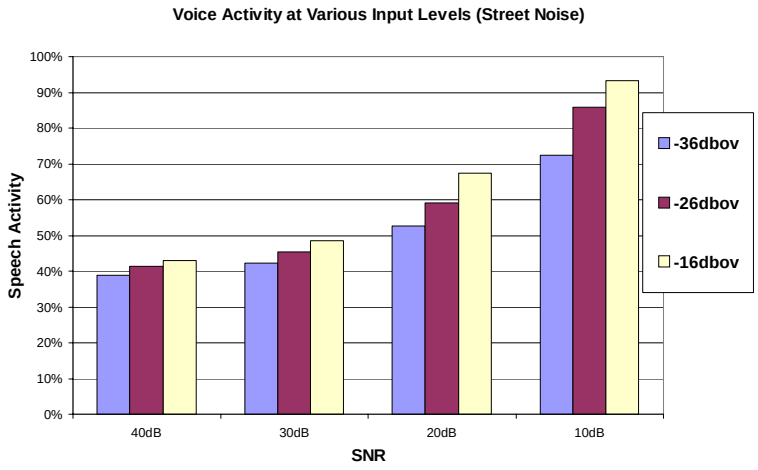


Figure 29.5: Voice Activity for different input levels (street noise)

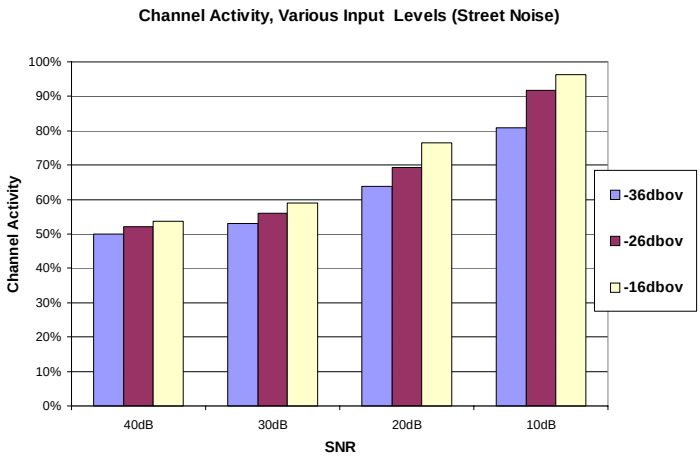


Figure 29.6: Channel Activity at different input levels (street noise)

29.3 Clipping

For speech clipping assessment, we first estimate how loudly speech is audible in each frame:

$$L_{sp}(n) = \left(\frac{\max(0, sp(n) - 0.25 * no(n))}{1 + (no(n)/sp(n))^2} \right)^{0.3},$$

where

sp(n): speech power of the frame n,

no(n): noise power of the frame n,

$L_{sp}(n)$: loudness of speech in frame n.

Speech and noise powers for each frame are calculated from the clean speech and noise files. The exponent of 0.3 is derived from the relation between loudness and intensity, i.e. an increase of 10 dB in the intensity causes the loudness to double. When speech power is 6 dB lower than noise power (see the 0.25 gain in the above equation), we assume that speech is not audible and loudness will be zero. Noise power in each frame is limited to below -55 dBm0, which is close to the noise level of the clean speech files. This limitation makes this equation applicable also for clean speech samples. Speech clipping is calculated as follows:

$$C_{sp} = \frac{\sum_n L_{sp}(n) * (1 - VAD\_flag(n))}{\sum_n L_{sp}(n)},$$

where VAD_flag(n) is the output of the VAD algorithm (1 for speech, 0 for noise).

As shown on the above equation, clipping is sum of loudness of the frames where VAD is "0" divided by sum of loudness of all frames.

The result of the investigations of the clipping with various background conditions can be seen in figure 29.7. Most clippings according to the measure applied are found for car background noise.

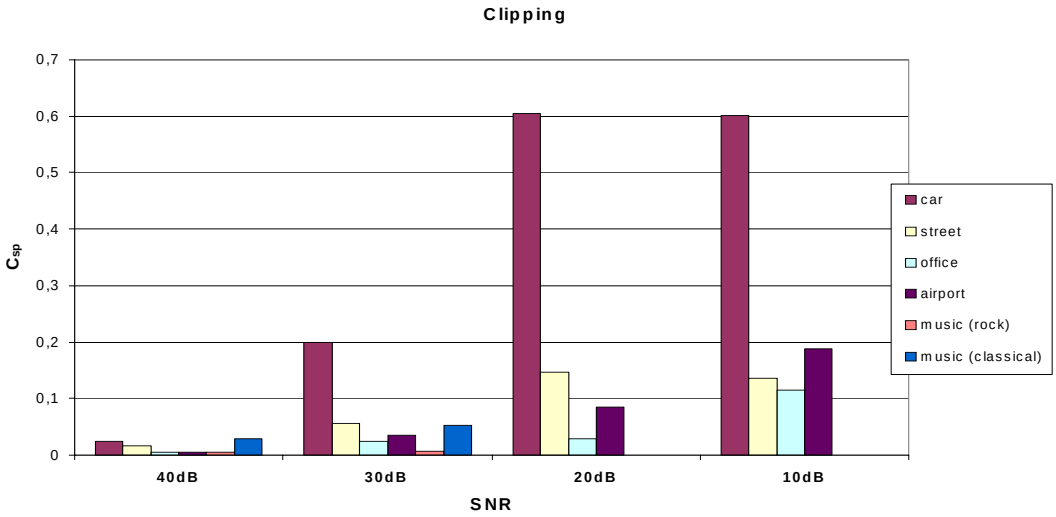


Figure 29.7: Clipping for different background conditions (clean case C_{sp} = 0.006)

For those speech samples for which severe clipping has been observed according to the clipping measure given above, careful expert listening has been carried out in order to check if the clipping is audible or annoying. For most cases no clipping was found. Only in extreme cases with car noise at 10 dB SNR, occasional slight clipping could be noticed. However, these effects were very minor and almost not audible.

Additionally, VAD performance for pure music files was tested. Ideally during music the VAD should detect everything as voice, and DTX-state should be activity. To test the system a wide range of diverse music files has been processed with the DTX turned on. The VAD-flag is printed out and the music files which contained frames with VAD-flag = 0 (i.e. no voice activity) are carefully examined by expert listeners.

The comfort noise system performs very well on most kinds of music. On most music files only a few sparse frames are classified as inactivity. However, this is hardly perceived as artifact. It has further been found that miss-classification can also occur after rapid decreases in intensity. Then the music is replaced by comfort noise for longer periods and this effect is clearly audible. In some specific kind of classical music with many large intensity changes (e.g. Carmina Burana by Orff), this effect is even annoying.

29.4 Comfort Noise Synthesis

The purpose of this investigation is to evaluate if the comfort noise synthesis generates a smoothly evolving comfort noise signal. It is assessed if there are situations where audible contrast effects occur either due to abrupt magnitude or due to abrupt spectral changes. The investigation is done in two parts, as follows.

In order to investigate the comfort noise synthesis during inactivity, coding is done with the VAD decision forced to 0. Input signals used in this test are:

- Car noise.
- Street noise.
- Office noise.
- Airport noise.
- Artificial white noise with slow random magnitude variations.
- Artificial narrow band noise with sweeping center frequency from 50 Hz to 7 000 Hz.

For all signals except the last, the synthesized comfort noise signal evolves smoothly and nothing remarkable can be reported.

For the narrow band noise with sweeping center frequency, the frequency of the synthesized signal seems to follow the input frequency somehow discontinuously or in steps. However, annoying artifacts are not produced.

This test was made with the original VAD decision enabled. The purpose was to test comfort noise contrast effects due to DTX state changes The input signals used are those listed in clause 29.1 but the level adjusted to such a value that the VAD decision is unstable. I.e. the VAD flag and in response to this, the DTX state toggles between activity and inactivity.

From all test signals it can be reported that slight differences in the synthesized signal are perceived when the DTX state changes. In some cases - even though not annoying - the effect is clearly audible as a contrast in the spectral characteristics of the synthesized signal.

The effect can be visualized by comparing the power spectra of the synthesized signals in response to a white noise input signal. While for DTX-state=Activity a spectrally flat signal (in the pass-band of the codec) is generated, this is not the case for DTX-state=Inactivity, i.e. during comfort noise synthesis. Clearly noticeable is a strong low-frequency component.

Figure 29.8: White noise responses for DTX-state=Activity (blue) and DTX-state=Inactivity (red)

29.5 Summary

In the tests we have found that the comfort noise system of the AMR-WB codec performs very well and that in general it does not cause quality degradations compared to operation without DTX.

The performance of the VAD is good for stationary types of background noise for which almost the same activity figures are measured as for clean speech. For more non-stationary kinds of noise and especially for low SNR ratios, the resulting voice and channel activity figures increase considerably, which may to some extent compromise the efficiency of the DTX system. On the other hand, however, speech quality is never degraded by clipping and only very few cases could be found where slight clipping was even noticeable. Furthermore, the VAD works satisfactorily most kinds of music.

The effect of comfort noise synthesis is audible but not annoying. For most types of input signals, the synthesis itself produces smoothly evolving comfort noise signals without any artefacts. However, audible noise contrast effects are caused by changes of the DTX-state between activity and inactivity. These effects increase with the signal level.

30 Performance with music signals (informal expert listening)

The results of this verification are based on the analysis of expert listeners [19]. Four different music signals have been used:

- classical, instrumental: Beethoven, Symphony No. 9, part 2 (49 s).
- classical, vocal: Beethoven, Fidelio (26 s).
- modern, instrumental: M. Knopfler (Guitar) (31 s).
- modern, vocal: Beatles, "Help" (31 s).

Table 30.1 lists the conditions that have been processed for each of the four long files.

Table 30.1

C01	Mode 8 (23.85 kbit/s)	DTX = 0
C02	Mode 5 (18.25 kbit/s)	DTX = 0
C03	Mode 2 (12.65 kbit/s)	DTX = 0
C04	Mode 0 (6.6 kbit/s)	DTX = 0
C05	Mode 8	DTX = 1
C06	Mode 5	DTX = 1
C07	Mode 2	DTX = 1
C08	Mode 0	DTX = 1
C09	g.722 @ 48 kbit/s	-
C10	Direct	-

The processed signals were analysed and compared by speech coding experts. For the listening, we did use binaural headphones (mono signal, binaural presentation) as well as loudspeakers. The complete list of conditions and the corresponding bit rates were known to all listeners from the file names being presented. All experts listened to the files in full length.

Using music as input signal, the intrinsic properties of the CELP speech coding algorithm become more obvious: Whenever speech (i.e. singing) is present, the coding quality seems to be better than the coding quality of instrumental music, because the speech is usually transmitted better than instrumental music. For instrumental parts of the music, degradations and distortions become more audible.

For the highest bit rate of 23.85 kbit/s (mode 8), the experts usually rated the quality of the music signal similar or very close to the quality of the G.722 codec at 48 kbit/s. For some music samples (Beethoven 9^{th} symphony, Beatles), there are audible degradations, which led to the conclusion that G.722 is sometimes equivalent, sometimes slightly preferred to the AMR-WB candidate. This high bit rate mode, however, was generally felt acceptable by all experts.

For medium bit rate at 18.25 kbit/s (mode 5), all experts agreed in preferring the subjective quality of the G.722@48 kbit/s. For music transmission, the quality of the AMR-WB candidate was felt acceptable by two experts, while three experts did consider the quality not acceptable.

After listening to the processed files at 12.65 kbit/s (mode 2), all experts agreed that the music signals are significantly distorted. It was felt, that the quality of the music signal is not sufficient for music transmission at this bit rate. At bit rates as low as 6.60 kbit/s (mode 0), we perceived very strong degradation. However, the processed signals are still recognizable as music.

The experiments indicate, that DTX on or off does not have a relevant influence on the perceived music's quality. In fact, it is generally inaudible whether DTX was set to 0 or 1.

The AMR-WB Codec performance with music signals is satisfactory at the highest bit rate of 23.85 kbit/s. During the listening, we did not observe any clicks or instabilities in the processsed samples of any bit rate of the AMR-WB candidate codec. The processed signals were always recognizable as music.

The highest bit-rate mode (23.85 kbit/s) is intended also for music and other non-speech signals. For music signals, this mode was generally felt acceptable by all experts.

31 Switching Performance between AMR and AMR-WB modes

This verification item is meant to investigate the perceived speech quality in possible switching scenarios between AMR-WB and AMR. Although it is not expected that such switching appears on a frame-by-frame basis, it can happen e.g. once per call because of handover or TFO negotiation [17].

An A-B-listening test was conducted to compare the subjective quality of two different wideband / narrowband switching schemes: The first without using a bandwidth extension scheme, the second one employing one. Both schemes were evaluated under three conditions: clean speech, car noise (SNR=15 dB), and street noise (SNR=15 dB). The number of sample pairs presented to the subjects for their preference decision was 24 samples = 2 orderings * 4 speakers (2 male, 2 female) * 3 background noises. All input samples are in German language. The test was carried out with 8 native German expert listeners.

Three different types of signals were generated in the processing phase for each speaker and background noise: A wideband signal (**WB**), i.e. AMR-WB coded and decoded speech with mode 19.85 kbps. A narrowband signal (**NB**), i.e. AMR coded and decoded speech with mode 12.2 kbps. A wideband signal (**EXT**) generated from the "NB" signal by subsequent bandwidth extension.

These samples were artificially cut and pasted in a way that in each sentence a switch from WB to NB or a switch from WB to EXT is performed. The cutting procedure was done in a way that no discontinuities were left in the signal - visually and audibly verified.

Scheme A: **WB – NB – WB - NB**

Scheme B: **WS – EXT – WB -EXT**

The results are shown in table 31.1, which contains the absolute number of choices (8 listeners).

Table 31.1

	A	B
all	63	129
CLEAN	20	44
CAR	20	44
STREET	23	41

The results show an approximately 2:1 preference score of the switching scenario with the artificially extended bandwidth of the NB signal versus the plain NB signal. Please note that in practical switching scenarios also switching delay effects and effects from the AMR coder starting from zero-state may occur.

Annex A:
Detailed information about the AMR-WB selection phase

A.1 Performance requirements

A.1.1 GSM FR channel (applications A and B)

For clean speech, at 19 dB C/I and above, the AMR-WB codec is required to provide in Application A quality better than (error-free) G.722-48k, and in Application B quality equal to G.722-56k. At 13 dB C/I, quality should still be equal to (error-free) G.722-48k in both applications. Under 13 dB C/I, graceful degradation comparable to the performance demonstrated by GSM EFR (Enhanced Full Rate) codec is required. Table A.1a shows the requirements for clean speech.

Table A.1a: Clean speech requirements under static error conditions for Applications A and B

Clean speech	Application A: GSM FR with 16 kbit/s submultiplexing		Application B: GSM FR	
C/I	Performance requirement	Performance objective	Performance requirement	Performance objective
no errors	better than G.722-48k	G.722-56k	G.722-56k	G.722-64k
19 dB	better than G.722-48k		G.722-56k	
16 dB	G.722-48k		G.722-48k	
13 dB	G.722-48k		G.722-48k	
< 13dB	(see note)		(see note)	
NOTE: The degradation in subjective performance shall not be greater than the degradation in subjective performance demonstrated by EFR over the same C/I interval. The specific intervals of interest are 13 dB to 10 dB, 13 dB to 7 dB, and 13 dB to 4 dB.				

For background noise conditions (speech in background noise), the requirements are given in Table A.1b. The requirements are the same as for clean speech except that quality equal to G.722-48k is required for Application A at C/I ≥ 19 dB. (Also, a different testing methodology, Poor or Worse, considered more suitable for background noise testing, was adopted (note).)

NOTE: Poor or Worse methodology is employed, where "with 10 % PoW" is interpreted as no more than 10 additional percentage points of annoying degradation with respect to the reference codec (in terms of annoying or very annoying quality scores in the listening tests: "1" and "2" out of votes ranging from "1" to "5").

Table A.1b: Background noise requirements under static error conditions for Applications A and B.

Speech in background noise	Application A: GSM FR with 16 kbit/s submultiplexing		Application B: GSM FR	
C/I	Performance requirement	Performance objective	Performance requirement	Performance objective
no errors	G.722-48k (with 10 % PoW)	G.722-56k	G.722-56k (with 10 % PoW)	G.722-64k
19 dB	G.722-48k (with 10 % PoW)		G.722-48k (with 10 % PoW)	
16 dB	G.722-48k (with 10 % PoW)		G.722-48k (with 10 % PoW)	
13 dB	G.722-48k (with 10 % PoW)		G.722-48k (with 10 % PoW)	
< 13dB	See note		See note	
NOTE: The degradation in subjective performance shall not be greater than the degradation in subjective performance demonstrated by EFR over the same C/I interval. The specific intervals of interest are 13 dB to 10 dB, 13 dB to 7 dB, and 13 dB to 4 dB.				

In tandem (2 asynchronous encodings), the requirement for AMR-WB for both clean speech and background noise is to be equal to G.722-48k in tandem for Application A and equal to G.722-56k in tandem for Application B. For input level

dependency, for clean speech, the general requirement is to be better than G.722-48k for Application A and equal to G.722-56k for Application B. For talker and language dependency, the requirement is to provide in Application A the same quality as G.722-48k and in Application B the same quality as G-722-56k.

For Applications A and B, requirements were set also for dynamic conditions (codec operated with mode adaptation on). Under typical dynamic error conditions, the requirement is to be better than EFR under the same error conditions. For difficult error conditions (6 dB worse than typical C/I-conditions), the requirement is to be at least as good as the EFR codec in the same conditions.

A.1.2 Higher rate channels (applications C and E)

In the EDGE half-rate channel, for clean speech and speech in background noise, AMR-WB should give at 25 dB C/I and above quality equal to (error-free) G.722-56k. At 19 dB C/I, quality should still be equal to (error-free) G.722-48k. In the EDGE full-rate channel, the same quality as in the HR-channel should be obtained at 3 dB worse C/I conditions.

In the 3G UTRAN channel, AMR-WB should give in error-free transmission quality equal to (error-free) G.722-64k. Quality equal to (error-free) G.722-48k is required at FER = 1.0 % / RBER = 0.1 %.

The requirements for Application C are given in table A.2a and for Application E in table A.2b.

Table A.2a: Requirements for clean speech and background noise under static test conditions for Application C

Clean speech and speech in background noise	Application C: Half-Rate Circuit Switched EDGE Phase II channel	Application C: Full-Rate Circuit Switched EDGE Phase II channel)
C/I	Performance requirement	Performance requirement
25 dB	G.722-56k	
22 dB	G.722-48k	G.722-56k
19 dB	G.722-48k	G.722-48k
16dB		G.722-48k

Table A.2b: Requirements for clean speech and background noise under static test conditions for Application E

Clean speech and speech in background noise	Application E: 3G UTRAN channel	
Error Condition [FER, RBER]	Performance requirement	Performance objective
No errors	G.722-64k	
[0.5 %, -]	G.722-56k	
[1.0 %, 0.1 %], Uplink (note 1)	G.722-48k	
[1.0 %, 0.1 %], Downlink (note 1)	G.722-48k	
[1.0 %, 0.1 %], Uplink (note 2)		G.722-48k
NOTE 1: The least significant bits shall be subjected to the residual error profile. The number of bits in this class shall be 25 % of the total bits per frame. NOTE 2: The least significant bits shall be subjected to the residual error profile. The number of bits in this class shall be 50 % of the total bits per frame.		

Application E includes all bit rates. The requirements are however only tested for the highest modes. The error performance for Application E is specified and evaluated using error protection schemes from the UTRAN toolbox. Each error condition is defined using two error profiles, one FER profile (single indicator per frame) and one residual BER profile (bit-level residual error channel). The requirement for the no error case applies to modes with higher bit rates, i.e., those not tested in Applications A and B.

For both Application C and E, in tandem (2 asynchronous encodings), the requirement for clean speech is to be equal to G.722-64k in tandem, and in background noise to be equal to G.722-56 in tandem. For input level dependency, for clean speech, the general requirement is to be equal to G.722-64k. For talker and language dependency, equal performance to G.722-64k is required.

A.1.3 Other requirements and objectives

The following tables summarise some additional requirements set for the AMR-WB codec: source controlled operation in the DTX mode (discontinuous transmission), non-speech inputs and music.

Table A.3a: Additional performance requirements for speech signals in source controlled operation (all applications)

Condition	Requirement
Switching between different AMR-WB bit-rates	No annoying artefacts
Clean speech with DTX enabled	Performance with DTX disabled
Speech and background noise with DTX enabled	Performance with DTX disabled

Table A.3b: Requirements and objectives for speech codec performance with non-speech inputs (all applications)

Condition	Requirement	Objective
DTMF		Transparent transmission of DTMF
Information tones	Recognisable as given information tone.	
Idle noise	-66dBm0 (unweighted)	

Table A.3c: Requirements and objectives with music for Applications C and E.

Condition	Requirement	Objective
Music	No annoying effects	G.722-56k

A.1.4 Testing of performance requirements in the selection tests

The selection tests were extensive consisting of altogether 6 experiments and 19 sub-experiments and covering all the four applications defined for AMR-WB. All above mentioned performance requirement conditions were included in the testing except only a few ones considered less critical for the selection (e.g. testing in tandem under background noise, switching between different AMR-WB bit-rates, and testing with non-speech signals and music). These were excluded for practical reasons to keep the selection tests within a reasonable size and will be covered during the post-selection phases: the verification phase and the characterisation phase.

A.2 Selection procedure and methodology for comparison of candidates

The selection procedure consisted of comparing the performances of the candidate codecs against a set of performance requirements and ranking the candidate performances using a number of Figures of Merit. Technical descriptions and other deliverables from the proponents were also reviewed and compliance with a set of mandatory design constraints was analysed.

The Selection Procedure followed the pre-defined selection rules described in Permanent AMR-WB Project Document: Selection Rules [7]. The selection procedure consisted of the following steps:

1. The selection test results will be presented and analysed while keeping secret the identity of the candidates. Each candidate will be informed of the code used for its own solution and its solution only. (The selection rules 2a, 2b and 3 will be applied at this stage.)

2. After the review and discussion of the test results (as specified for rule 3), TSG-SA4 will try to reach a consensus on a quality ranking of the candidates.

3. Each candidate will then present its solution and show the compliance with the design constraints. All candidates not compliant with all design constraints will be excluded (according to the selection rule 1).

4. The test results obtained by each candidate will then be revealed.

5. A final discussion and review of the solution characteristics and test results will take place.

6. SA4 will then try to reach a consensus on a single candidate to serve as the basis for the AMR-WB standardisation.

The first two selection rules are eliminating rules. The first rule excludes all candidates failing to demonstrate full compliance with the AMR-WB design constraints. The second rule excludes all candidates with test results too far below the expected performance level. The third rule consists of a direct comparison between candidates using a set of Figures of Merit.

A.2.1 Design constraints (Rule 1)

Design constraints are a set of mandatory requirements that the AMR-WB codec needs to fulfil. Any candidate codec not compliant with all design constraints is excluded from selection. The design constraints include constraints, e.g. for implementation complexity and transmission delay.

The computational complexity of the speech codec (without channel coding) was limited below 40 wMOPS for all applications. For speech coding and channel coding (Applications A and B), the detailed complexity limits are given. For Application C, the definition of the channel is carried out in TSG-GERAN. However, for the purposes of AMR-WB selection tests, the codec proponents had to provide an example channel codec solution complying with a number of constraints. Application E was tested with residual error patterns (impacting the bit-stream from/to speech codec), and the proponents did not therefore need to provide channel codec as part of the proposal.

The algorithmic transmission delay requirement was set for the GSM FR channel, where the same delay as in AMR narrowband codec was required but with 6.5 ms relaxation. The relaxation is needed because of the increased Abis/Ater delay (caused by the higher speech coding bit-rates) and also due to allowing the use of band-splitting and re-composition filters in the solutions, as felt necessary for wideband coding.

The proponents were required to provide for the Selection Phase, a fixed-point C-code implementation of the proposed AMR-WB codec. This consisted of speech codec (including voice activity detection and source controlled rate mechanism) for all applications, channel coding for the GSM FR channel, and example channel codings for EGDE FR and EDGE HR channels.

The same codec mode and channel measurement signalling scheme as used in AMR narrowband was required to be used. Also, the same source controlled rate scheme with regard to transport format and update frequency as in AMR narrowband was a requirement.

The design constraints are explained in detail in Permanent AMR-WB Project Document: Design Constraints [8].

For the analysis the codec proponents were required to deliver detailed information of their codec proposal as described in Permanent AMR-WB Project Document: Selection Deliverables [9].

A.2.2 Speech quality

A.2.2.1 Failures in meeting performance requirements (Rule 2)

This rule is an eliminating rule to exclude all candidates with performance too far below the expected performance level. The rule consists of two parts: Rule 2a checks that more than 50 % of the performance requirements were met for various subsets of the tests. Rule 2b checks that there were no more than 10 % of severe failures for each of the subsets.

Selection Rule 2a: Any candidate failing 50 % or more of the test conditions contained in any of the following test sets will be excluded. A test is failed if the codec performance (measured MOS score or PoW) does not meet the requirement specification at the 95 % confidence level.

List of test sets for Rule 2a:

- Set #1: all conditions (90 conditions), including the CCR Tests;

- Set #2: all clean conditions (47);

- Set #3: all background noise conditions (43), including the CCR Tests;

- Set #4: all conditions of application A (30);
- Set #5: all conditions of application B (26), including the CCR Tests;
- Set #6: all conditions of application C, E (34).

Selection Rule 2b: Any candidate severely failing more than 10 % of the test conditions contained in any of the following test sets will be excluded.

List of test sets for Rule 2b:

- Set #1: all conditions (87), excluding the CCR Tests;
- Set #2: all clean conditions (47);
- Set #3: all background noise conditions (40), excluding the CCR Tests;
- Set #4: all conditions of application A (30);
- Set #5: all conditions of application B (23), excluding the CCR Tests;
- Set #6: all conditions of application C, E (34).

A.2.2.2 Direct comparison of candidates (Rule 3)

A number of Ffigures of Merit (FoM) were identified to be used to analyse and compare the performance of the candidates. See table A.4. None of the Figures of Merit was intended to serve as single selection criteria.

Table A.4: List of FoMs selected for the evaluation of the test results.

Metric (FoM)	Ranking Provided
Weighted ΔdBq	Per experiment and across all experiments Per lab and across labs Full set of test results (Preferred FoM) and restricted to the failed tests only (ΔdBq computed with reference to the requirement in this case)
Weighted ΔMOS	Per experiment and per lab (cannot be computed across labs and experiments) Full set of test results and restricted to failed tests
Number of systematic failures in meeting performance requirements (2 failures out of 2 tests)	Per experiment and across all experiments Across labs
Unweighted ΔPoW percentages (for the relevant conditions)	Per experiment and across all relevant experiments
Unweighted ΣCMOS (for the relevant conditions)	Per experiment and across all relevant experiments
NOTE: ΔMOS = Codec MOS - Reference MOS, ΔdBq = Codec dBq - Reference dBq.	

Details on the FoMs and on how rules 2 and 3 are applied can be found in [7].

A.3 Selection phase listening tests

The five candidate codecs were tested in a variety of test conditions in six independent test laboratories. The tests took place during a period from September to October 2000. The test plan is described in detail in Permanent AMR-WB Project Document: Selection Test Plan [10]. The processing of speech samples in the selection tests is described in Permanent AMR-WB Project Document: Processing Functions [11].

A.3.1 Overview of the test plan

The tests covered all the four applications (A, B, C and E) specified for the AMR-WB codec. The performances of the candidate codecs were evaluated in multiple of test conditions consisting of 6 experiments and 19 sub-experiments. Testing was carried out using 5 languages (French, Japanese, Mandarin Chinese, North American English, and Spanish).

The experiments and sub-experiments included in the selection tests are as follows (note) [10]:

NOTE: Experiments 1, 2 and 5 are Absolute Category Rating (ACR) tests, experiments 3 and 4 are Degradation Category Rating (DCR) tests, and experiment 6 is a Comparison Category Rating (CCR) test. The results are given as Mean Opinion Scores (MOS), Differential MOS (DMOS), or Comparison MOS (CMOS), respectively. ACR tests ask the listeners to assess the quality of each speech sample under test while DCR and CCR tests ask the listeners to assess the quality differences between two samples. The difference between DCR and CCR tests is that in DCR tests the listeners assess the degradation in the second sample compared to the first one, while in CCR tests the listeners assess the quality difference between the samples. (ACR, DCR and CCR tests are all well-established and recognised speech quality testing methodologies. These methodologies are used within the experiments, depending on which is the most suitable one for each test.)

Experiment 1: Input Level and tandeming performance for clean speech (ACR-test)

1a: Applications A and B.

1b: Applications C and E.

Experiment 2: Clean Speech performance with static errors (ACR)

2a: Clean Speech and in Static Errors for GSM FR Channel (Application A).

2b: Clean Speech and in Static Errors for GSM FR Channel (Application B).

2c: Clean Speech and in Static Errors for Higher-Rate Channels (Application C).

2d: Clean Speech and in Static Errors for Higher-Rate Channels (Application E).

2e: Clean Speech and in Static Errors for GSM EFR and wideband to narrowband tandeming.

Experiment 3: Car and Street noise (15 dB SNR) performance for the GSM FR channel (DCR-test)

3a: GSM FR channel (Application A) in Car noise.

3b: GSM FR channel (Application A) in Street noise.

3c: GSM FR channel (Application B) in Car noise.

3d: GSM FR channel (Application B) in Street noise.

3e: GSM EFR performances in Car and Street noise.

Experiment 4: Car and Street noise (15 dB SNR) performance for higher-rate channels (DCR-test)

4a: Higher-rate channels (Application C) in Car noise.

4b: Higher-rate channels (Application C) in Street noise.

4c: Higher-rate channels (Application E) in Car noise.

4d: Higher-rate channels (Application E) in Street noise.

Experiment 5: Performance in Dynamic Conditions (ACR-test)

5a: Performance in Dynamic Conditions for AMR-WB (Application A).

5b: Performance in Dynamic Conditions for EFR.

Experiment 6: VAD/DTX in GSM FR channel for Application B (CCR-test)

The listening test laboratories participating into the AMR-WB selection tests were: ARCON (North American English), AT&T (Mandarin Chinese, North American English, Spanish), Dynastat (North American English, Spanish), France Télécom (French), Lockheed-Martin Global Telecommunications (North American English, Spanish), and NTT-AT (Japanese). Each experiment in the tests was carried out with two languages to avoid any bias due to a particular language. The allocation of experiments to listening laboratories, and the languages used for each experiment, are shown in table A.5.

Table A.5: Allocation of Experiments to the Listening Laboratories.

Experiment	ARCON	AT&T	Dynastat	FT	LMGT	NTT-AT	Total of languages
1a	NAE			FR			2
1b	NAE			FR			2
2a			NAE			JP	2
2b			NAE			JP	2
2c			NAE			JP	2
2d			NAE			JP	2
2e			NAE			JP	2
3a		SP			NAE		2
3b		SP			NAE		2
3c		MCH			NAE		2
3d		MCH			NAE		2
3e			SP		NAE		2
4a		NAE			SP		2
4b		NAE			SP		2
4c			NAE		SP		2
4d			NAE		SP		2
5a		NAE		FR			2
5b		NAE		FR			2
6	NAE					JP	2
Total of sub-experiments	3	8	8	4	9	6	38

NOTE: NAE: North American English;
MCH: Mandarin Chinese;
SP: Spanish;
FR: French;
JP: Japanese.

Processing of speech samples through the candidate algorithms was carried out by the candidate organisations themselves and was crosschecked for correctness by other candidates. Two host laboratories, ARCON and Lockheed-Martin Global Telecommunications processed the samples through reference codecs. A blind procedure was followed to ensure that the listening test laboratories and the test subjects had no knowledge of the codec algorithms. The test results from the individual laboratories were combined by a Global Analysis Laboratory (ARCON) and were presented at SA4#13 in October 2000.

A.3.2 Schedule of the selection tests and related activities

The processing of speech samples was carried out during August and early September 2000. Listening tests started in mid-September. The listening test results and deliverables from the codec proponents (technical descriptions of the codec algorithms) were reviewed at SA4#13 in October 2000.

Before the processing of speech samples started the candidates had to deliver, in early August, an executable of their codec software to ETSI freezing the algorithm development.

The key milestones of the listening tests and the relating selection phase activities are shown in table A.6.

Table A.6: Key milestones of the AMR-WB Selection Phase Tests

Responsible	Action Description	Deadline (2000)
Test laboratories	Delivery of the speech samples to the host laboratories for processing	July 31st
Candidates	Receipt of executables for AMR-WB candidates by ETSI	August 6th
Candidates	Send executables, processed material etc to the crosschecking candidate, and to the host laboratory (without the executable).	August 24th
Candidates	Completion of processing and verification of correctness	August 28th
Host Laboratories	Sending of final set of speech material to test laboratories	September 13th
Candidates	Delivery of all remaining Selection Deliverables (technical descriptions of candidate algorithms, analysis of compliance to design constraints etc.) to ETSI	October 18th
Candidates	Delivery of complete IPR declaration to ETSI	October 8th
Test laboratories	End of listening tests	October 9th
Test laboratories	Delivery of test results (test raw data) to ETSI and Global Analysis Laboratory	October 9th
Global Analysis Laboratory	Preparation and delivery of test results summary / technical report to the SA4-reflector	October 16th
Host and listening laboratories	Presentation of test results to SA4	SA4#13 (October 23rd –27th)
SA4	Review of the selection test results, recommendations for the codec to be chosen	SA4#13 (October 23rd –27th)
SA4	Review of draft specifications and first verification results	SA4#14 (Nov 27th – Dec 1st)
SA4	Presentations of Selection Test results and AMR-WB codec selection for approval. Presentation of AMR-WB draft specifications for information.	TSG-SA#10, Dec 2000
SA4	Presentation of AMR-WB specifications for approval.	TSG-SA#11, March. 2001

Nortel Networks provided the error patterns required in the testing for Applications A, B and C. the error patterns for testing of Application E were provided by Ericsson (Uplink) and Nokia (Downlink). The seed-values of the error patterns were kept secret during testing.

A.4 Results of the selection tests

The codec candidates were referred to as Codec 1...Codec 5 during the analysis. The candidate selected as the AMR-WB standard is shown in the results as a Codec 3 (Nokia).

The following subclauses give analysis results for the codec candidates.

Annex TBD gives graphical representation of some extracts from the selection phase tests. Annex TBD contains the complete spreadsheet of selection phase results. This is the full record of the results achieved from the subjective listening tests.

A.4.1 Comparison against performance requirements

The candidate performances were analysed in accordance to the selection Rule 2. The number of failures for each subset of conditions is given in tables A.7a and A.7b.

Table A.7a: Number of failures for sets #1 - #3

Rule 2A	Candidate Failures in Set#1					Candidate Failures in Set#2					Candidate Failures in Set #3				
Codec #	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Number of failures	17	29	0	13	11	6	5	0	3	3	11	24	0	10	8
Failure-%	10,6	18,1	0,0	8,1	6,9	8,1	6,8	0,0	4,1	4,1	12,8	27,9	0,0	11,6	9,3
Pass / Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

Table A.7b: Number of failures for sets #4 - #6

Rule 2A	Candidate Failures in Set#4					Candidate Failures in Set#5					Candidate Failures in Set#6				
Codec #	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Number of failures	4	8	0	5	3	2	3	0	4	4	11	18	0	4	4
Failure-%	9,1	18,2	0,0	11,4	6,8	4,5	6,8	0,0	9,1	9,1	16,7	27,3	0,0	6,1	6,1
Pass / Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

All candidates met the requirement of Rule 2a requiring less than 50 % failures in each set. For Codec 3, no failures against the performance requirements were found at all in any of the tests.

All codec candidates met Rule 2b requiring 10 % or less severe failures in each set. None of the candidate codecs had severe failures in any of the sets.

A.4.2 Direct comparison of candidates

A number of pre-defined figures of Merit were used to analyse and compare the performance of the candidates. The results are given in tables A.8a to A.8c. The best FoM for each case is highlighted in the tables with a boldface font.

Table A.8a: FoM results for weighted ΔMOS, weighted ΔdBQ and unweighted %ΔPOW

Rule 3 FoM	Weighted ΔMOS					Weighted ΔdBQ					Unweighted %ΔPOW				
Codec #	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Total	19.0	6.8	**60.4**	19.6	32.0	146.9	47.6	**787.6**	217.7	353.4	36,5 %	68,8 %	**10,4 %**	49,0 %	19,8 %

Table A.8b: FoM results for systematic failures

Rule 3 FoM	Number of systematic failures				
Codec #	1	2	3	4	5
Total	3	7	**0**	4	3

Table A.8c: FoM results for weighted ΔMOS and weighted ΔdBQ when restricted to failures.

Rule 3 FoM restricted to failures	Weighted ΔMOS					Weighted ΔdBQ				
Codec #	1	2	3	4	5	1	2	3	4	5
Total	-2.1	-5.6	**0,0**	-1,4	-1.3	-30.4	-65.7	**0,0**	-13,9	-17.0

The comparison shows that Codec 3 is the best quality codec in all the total FoMs.

A.4.3 Conclusions on the AMR-WB codec candidates

On basis of the analysis of the codec algorithms and their speech quality performance, the following can be concluded:

- All candidate algorithms fulfil the mandatory design constraints (Rule 1).
- All candidate algorithms meet the Rule 2 requirements for the amount of failures and severe failures. Codec 3 is the only codec candidate that meets all the performance requirements in all of the laboratories in the selection tests. It has no failures at all.
- The Figures of Merit show that Codec 3 has the best quality of the candidates. Codec 3 is ranked as the best codec with regard to speech quality. (Quality ranking for the remaining codecs was not performed.)
- Taking into account the listening test results, technical descriptions and other relevant information, Codec 3 is the best candidate.

Based on the results of the Selection Phase, SA4#13 recommended in October 2000 Codec 3 to be chosen to the AMR-WB codec standard. The selection of Codec 3 was approved at the following TSG-SA#10 meeting in December 2000.

A.5 Highlights of the best candidate codec (Codec 3) based on the selection tests

Based on the Selection Phase results the speech quality performance of AMR-WB codec (Codec 3) can be characterised as follows:

Applications A and B (GSM FR channel):

- For clean speech, the codec provides in Application A error-free quality exceeding G.722-48k and in Application B quality equal to G.722-56k.
- Under background noise, the codec provides in Application A error-free quality equal to G.722-48k and in Application B quality equal to G.722-56k.
- In both Applications A and B, at 13 dB C/I, quality is still equal to the quality of error-free G.722-48k, for both clean speech and in background noise. Below 13 dB C/I, smooth degradation (comparable to degradation for GSM EFR) is provided.

Applications C and E (GSM EDGE, 3G UTRAN):

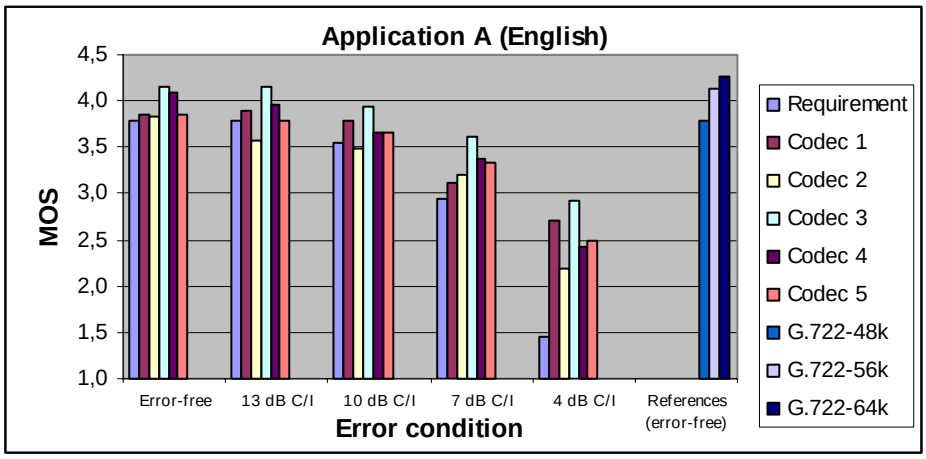
- In the EDGE FR-channel, for clean speech and speech in background noise, at 22 dB C/I and above quality equal to error-free G.722-56k is provided. At 16 dB C/I, quality equal to error-free G.722-48k is still produced.
- In the EDGE HR-channel, for clean speech and speech in background noise, at 25 dB C/I and above quality equal to error-free G.722-56k is provided. At 19 dB C/I, quality equal to error-free G.722-48k is still produced.
- In the 3G UTRAN channel, for clean speech and speech in background noise, quality equal to G.722-64k is provided for error-free transmission. Under transmission errors at FER=1.0 % / RBER=0.1 %, quality equal to G722-48k is given. (The least significant bits are subjected to the residual error profile with the number of bits in this class 25 % of the total bits per frame).

A.6 Key Selection Phase Documents in 3GPP FTP-site

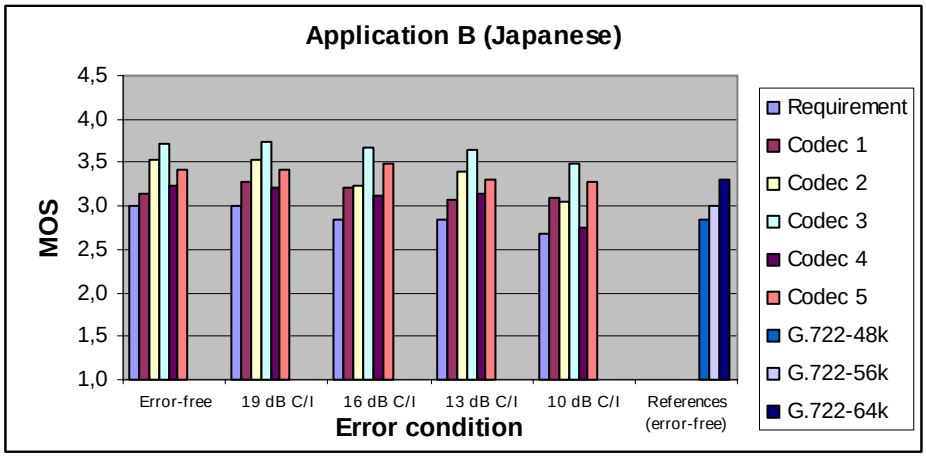
The standardisation of the AMR-WB codec is described in a series of permanent project documents. They contain the most important quidelines, rules and decisions. The following permanent project documents can be found in a specific location on the 3GPP FTP site:

Project Plan	S4-000526_WB2_pplan_v0.4.zip. ..
Overview of AMR-WB development	S4-000410_AMR-WB-1_overview...
Performance Requirements	S4-000321_Performance_requireme...
Selection Test Plan	S4-000382_AMR-WB-8b Selection T...
Selection Test Processing Functions	S4-000389_AMR-WB-7b Selection P...
Selection Deliverables	S4-000427_AMR-WB-6b_SelectionDe...
Selection Rules	S4-000508_AMR-WB-5b_SelRulesv1...

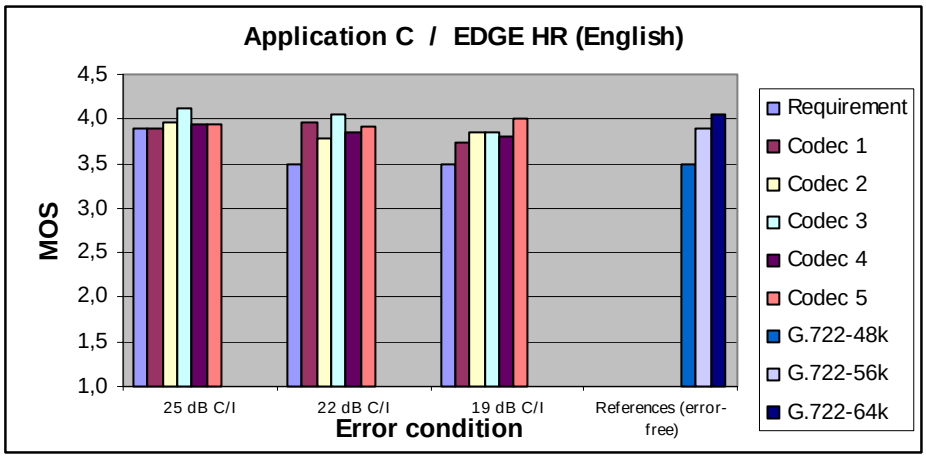
A.7 Extracts from the AMR-WB Selection Test Results



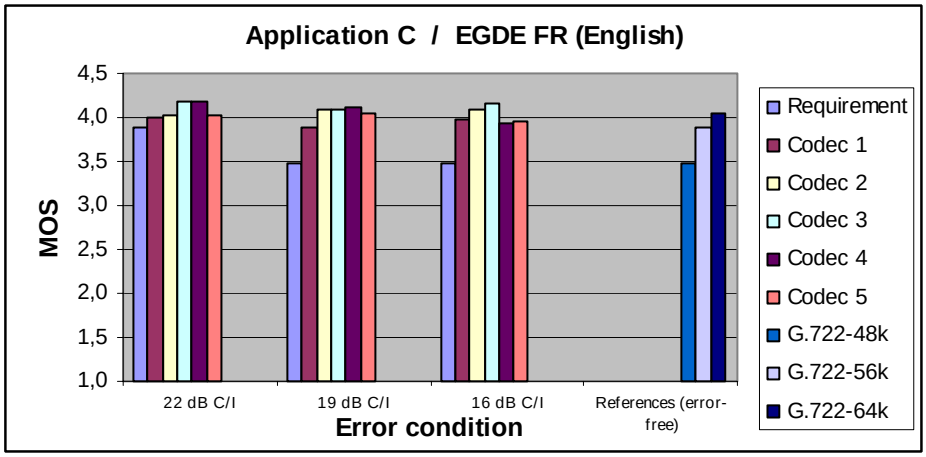
a) Application A (English)



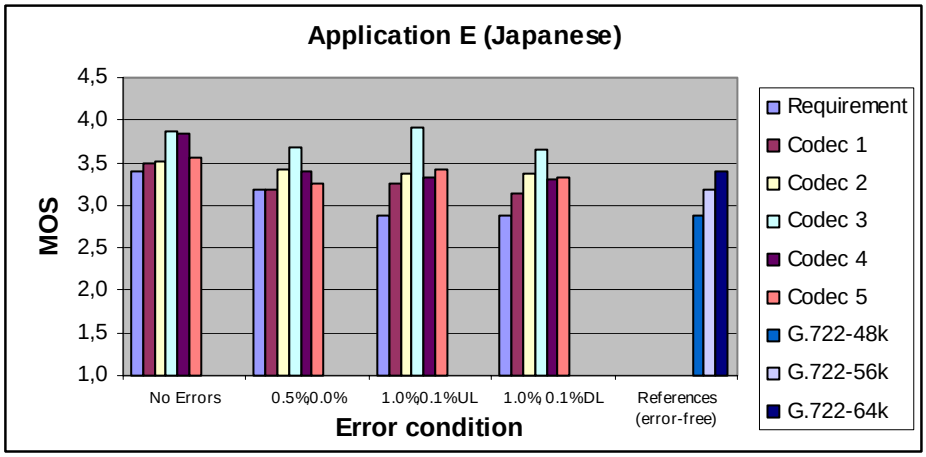
b) Application B (Japanese)



c) Application C / EDGE HR (English)



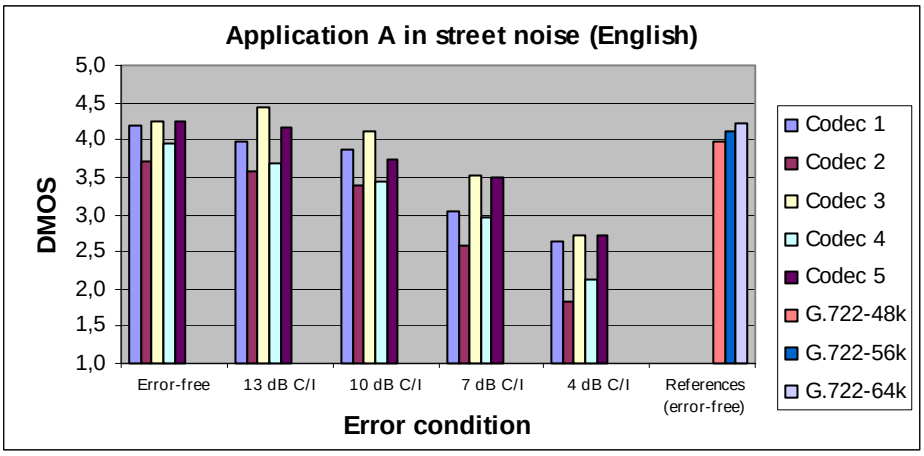
d) Application C / EDGE FR (English)



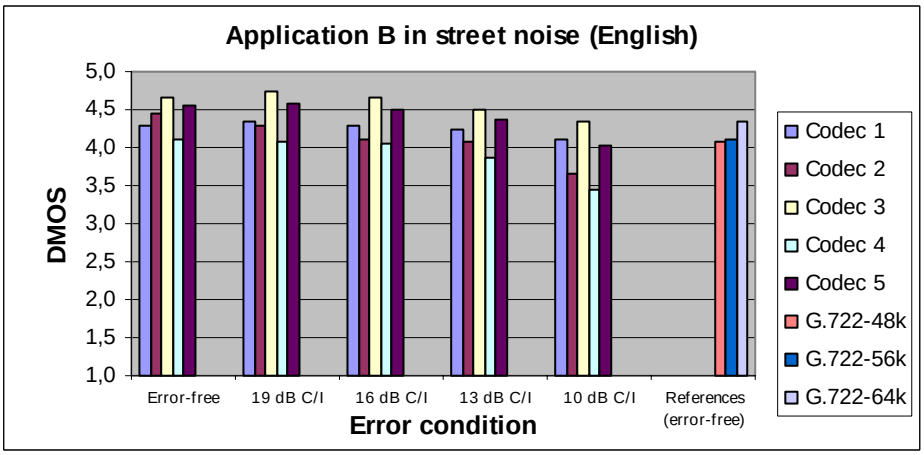
e) Application E (Japanese)

NOTE: The absolute MOS values depend on the test setting and conditions and are not directly comparable between the sub-experiments.

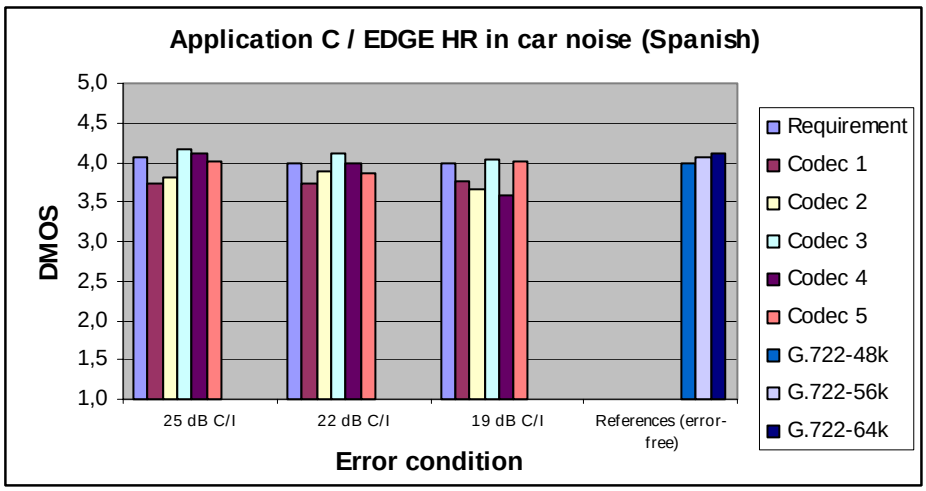
Figure A.1: Experiment 2: Clean Speech performance with static errors (ACR)



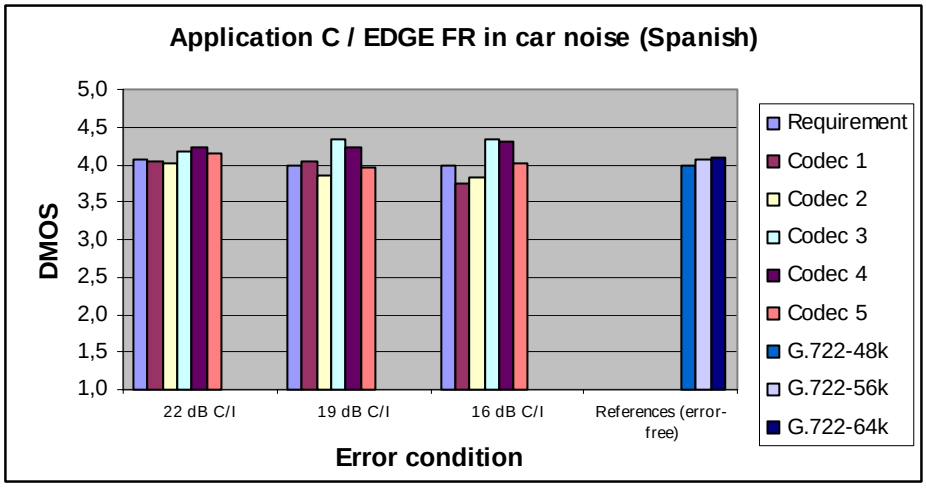
a) Application A in street noise (English)



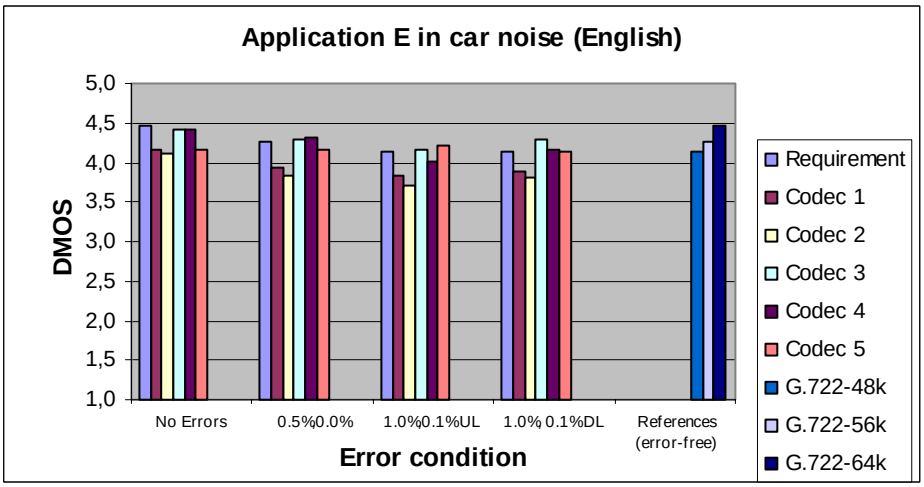
b) Application B in street noise (English)



c) Application C / EDGE HR in car noise (Spanish)



d) Application C / EDGE FR in car noise (Spanish)



e) Application E in car noise (English)

NOTE: The absolute DMOS values depend on the test setting and conditions and are not directly comparable between the sub-experiments. (Note also that the requirements are not drawn in figures 2a and 2b since they are not given as DMOS-values, but instead as 10 % PoW measures.)

Figure A.2: Experiment 3: Car and Street noise (15 dB SNR) performance for the GSM FR channel (DCR-test); and Experiment 4: Car and Street noise (15 dB SNR) performance for higher-rate channels (DCR-test)

A.8 Global Analysis Spreadsheet

See the Excel-spreadsheet in the attached file "AMRWB_GAL.zip" (contained also in SA4 document S4-000485).

This 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.

A.9 Complexity of the AMR-WB Candidate Codecs

This clause gives estimates of the codec complexities (estimated by codec proponents) (note).The complexity was calculated as worst observed frame.

NOTE: Codec 4 was withdrawn during the Selection Phase and no estimates for complexity were given for it.

Table A.9

COMPLEXITY	Requirement	Codec 1	Codec 2	Codec 3	Codec 5
Speech codec complexity A: wMOPS B: RAM C: ROM D: Program ROM	A: wMOPS ≤ 40 wMOPS B: RAM ≤ 15 kwords C: ROM ≤ 18 kwords D: Prog. ROM ≤ 5821 basic operators	A: 38.63 wMOPS B: 13.415 kwords C: 16.279 kwords D: 4798 basic ops	A: 37.09 wMOPS B: 12.066 kwords C: 7.332 kwords D: 5481 basic ops	A: 35.4 wMOPS B: 6.42 kwords C: 9.94 kwords D: 3771 basic ops	A: 38.9 wMOPS B: 5.94 kwords C: 16.02 kwords D: 5512 basic ops
Additional complexity for source controlled rate operation (over speech coding complexity limits) E: wMOPS F: RAM G: ROM H: Program ROM	E: wMOPS ≤ 1.6 wMOPS F: RAM ≤ 149 words G: ROM ≤ 513 words H: Program ROM ≤ 491 basic operators	E: 0.833 wMOPS F: B includes this G: C includes this H: D includes this	E: 0.479 wMOPS F: 107 words G: 7 words H: 131 basic ops	E: 0.73 wMOPS F: 75 words G: 0 words H: 268 basic ops	E: 0.36 wMOPS F: 65 words G: 0 words H: 314 basic ops
Channel codec complexity for Applications A and B: I: wMOPS J: RAM K: ROM ***L: Program ROM***	I: wMOPS ≤ 5.7 wMOPS J: RAM ≤ 3.0 kwords K: ROM ≤ 4.5 kwords L: Program ROM ≤ 2013 basic operators	I: 4.51 wMOPS J: 2722 kwords K: 4075 kwords L: 1346 basic ops	I: 5.42 wMOPS J: 2.359 kwords K: 4.242 kwords L: 360 basic ops	I: 3.45 wMOPS J: 2.88 kwords K: 3.18 kwords L: 579 basic ops	I: 5.5 wMOPS J: 2.787 kwords K: 2.985 kwords L: 910 basic ops
Constraints for channel codec in Application C (example solution used in testing)	Only the polynomials denoted G1-G7 in 05.03 can be applied. Recursive Systematic Codes as used in TCH/AFS and TCH/AHS can be used. Constraint length K=7 can be used in all modes. Use of a single CRC is allowed up to 16 parity bits. 24 bits should be reserved to an inband channel in FR and 12 bits in HR.	Requirement is met.	Requirement is met.	Requirement is met.	Requirement is met.

Annex B:
AMR-WB Floating-Point Verification

This annex contains the verification results for the AMR-WB floating-point codec 3GPP TS 26.204. This floating-point codec specification is targeted to be used in multimedia applications and in packet-based applications. (The floating-point codec may be used instead of the fixed-point codec when the implementation platform is better suited for a floating-point implementation.) However, the fixed-point specification of 3GPP TS 26.173 is the only allowed implementation of the AMR-WB codec for the speech service, and the use of the floating-point code is limited to other services. The bit-exact fixed-point C-code also remains the preferred implementation for all services.

The floating-point ANSI-C code in the present document is the only standard conforming non-bit-exact implementation of the Adaptive Multi Rate speech transcoder (3GPP TS 26.190), Voice Activity Detection (3GPP TS 26.194), comfort noise generation (3GPP TS 26.192), and source controlled rate operation (3GPP TS 26.193). The floating-point code also contains example solutions for substituting and muting of lost frames (3GPP TS 26.191).

The floating-point encoder in the present document is a non-bit-exact implementation of the fixed-point encoder producing quality indistinguishable from that of the fixed-point encoder. The decoder in the present document is functionally a bit-exact implementation of the fixed-point decoder, but the code has been optimized for speed and the standard fixed-point libraries are not used as such.

B.1 Subjective test results

This clause presents subjective test results of AMR-WB floating-point codec. The test has been conducted according to the test plan found in S4-010667. The processing of the material has been performed according to the AMR-WB characterisation processing plan S4-010464 [35].

The codec used in this study is the AMR-WB floating-point codec V0.2.0 (converted from fixed-point 5.3.0). The fixed-point ETSI reference codec was V5.3.0. All 9 AMR-WB bit-rates were tested with DTX off and subset of the modes was tested also with DTX on. The test was split into 4 Experiments listed in the table B.1.1.

Table B.1.1

Exp. No.	Title	Listening lab	Language
1	CCR-test, Clean speech and input levels for the 5 modes (6.60 kbit/s, 8.85 kbit/s, 14.25 kbit/s, 18.25 kbit/s, 23.05 kbit/s)	Nokia	Finnish
2	CCR-test, Clean speech and input levels for the 5 modes (6.60 kbit/s, 12.65 kbit/s, 15.85 kbit/s (*), 19.85 kbit/s, 23.85 kbit/s)	Ericsson, RCDCT (note)	Chinese
3	CCR-test, Background noise, Noise type: car noise	Nokia	Finnish
4	CCR-test, Background noise, Noise type: babble noise	Ericsson, RCDCT	Chinese
NOTE: Research Center of Digital Communications Technology (Beijing, China): * dBov = 10 * log ($\frac{V_{max}^2}{2}$)			

Summary of the results

Over all the experiments shown in figure B.1, most of the conditions were showing that fixed-point and floating-point performance is equal. See the table B.1.2.

Altogether, the results show that the performance of the AMR-WB floating-point is equal to that of the AMR-WB fixed-point. There are some individual test cases (2) where the floating-point codec gets slightly worse scores but on the other hand, there are more cases (3) where AMR-WB fixed-point gets slightly worse scores. Also, no systematic rule can be found between these single instances, which are evenly distributed over different experiments and codec modes.

Table B.1.2

Exp.	Condition	Preference	Notes
Exp 1	AMR@23.85 Fixed – AMR@23.85 Float	Floating-point is better in male talkers	Equal in all talkers
Exp 2	AMR@6.6 Fixed - AMR@6.6 Float	Floating-point is better in male talkers	Equal in all talkers
Exp 3	AMR@15.85 Fixed – AMR@15.85 Float	Floating-point is better in female talkers	Also better in all talkers
Exp 3	AMR@6.6 Fixed - AMR@6.6 (DTX ON)	Fixed-point is better in male talkers	Equal in all talkers
Exp 4	AMR@23.05 Fixed – AMR@23.05 Float	Fixed-point is better in male talkers	Also worse in all talkers
NOTE: There were total of 52 different conditions in the tests.			

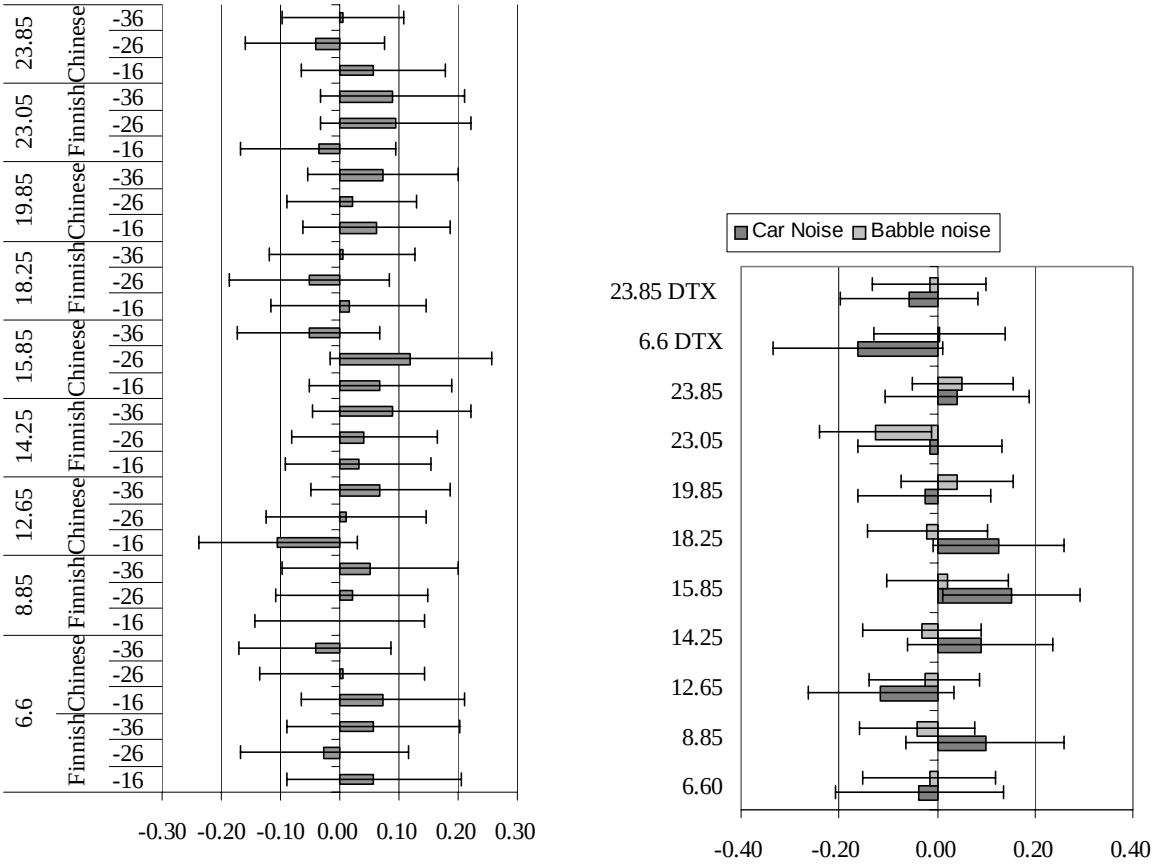


Figure B.1: Left: Experiments 1 and 2, Right: Experiments 3 and 4. The 95 % confidence intervals are plotted in the pictures as error bars

B.2 Non-speech signals

This clause reports the results of the verification of the floating point version of the AMR-WB codec. The V5.3.0 of AMR-WB Codec was used as reference during the verification. All processing were done on a Windows NT4 platform using Microsoft Visual C++ compiler. The purpose of the verification was to test the behaviour of the floating point version AMR-WB codec on non speech signals as well as the bit exactness of the floating point decoder versus the fixed point decoder [33].

Several types of non speech signals were used during the verification, tones, signalling tones and music.

Each input signal was processed by the fixed point encoder and by the floating point encoder. It resulted in two bit stream files: a fixed point bit stream and a floating point bit stream.

The fixed point bit stream was processed by the fixed point decoder. The fixed point bit stream was converted by the interface module and decoded by the floating point decoder. These two output files were compared to test the bit-exactness of the decoder.

On the same way, the floating point bit stream was processed by the floating point decoder. The floating point bit stream was converted by the interface module and decoded by the fixed point decoder. These two output files were compared to test the bit-exactness of the decoder. This was repeated for each mode. The test was limited to error free condition. The test was run with DTX switched off.

Tones signals have been generated in the range 10 Hz to 7 010 Hz with a frequency step of 20 Hz. Each tone had a duration of 10 s. The DTX was switched off during the test.

Signalling tones

Five different types of French network signalling tones have been tested: Two different dial tones, one ringing tone, a busy tone and a special information tone. The description of the different tones is given below:

- Continuous DIAL TONE number 1 at 440 Hz, 10 s duration.
- Continuous DIAL TONE number 2 at 330 + 440 Hz, 10 s duration.
- RINGING TONE at 440 Hz with duration 1.5 – 3.5 and a total duration of 12.5 s.
- BUSY TONE at 440 Hz with duration 0.5 – 0.5 and a total duration of 12.5 s.
- SPECIAL INFORMATION TONE at 950 Hz/1 400 Hz/1 800 Hz and duration (3 × 0.3 – 2 x 0.03) - 1.0 and a total duration of 12.5 s.

The level of the signalling tones was set at -10 dBm0. The test has been performed by informal listening involving trained listeners. The test methodology was pair comparison test. The DTX was switched off during the test. The result of the test was that the floating point V0.2.1 (note version 0.2.1 is algorithmically identical to the V0.2.0 used in some other verification items, except the error in the I/O-interface was corrected) did not perform worse than the fixed point V5.3.0 of AMR-WB. For each mode and each signalling tone, the bit exactness of the fixed point decoder and the floating point decoder has been verified.

Music signals

Some music signals were taken as input signals, the music items were classical music, modern music, single instruments, singer and singer with music. The different music items have been processed using the floating point V0.2.1 of AMR-WB and also using the fixed point V5.3.0 of AMR-WB. In order to have a comparison, G.722.1 at 24 kbps was included in the test. The test has been performed by informal listening including trained listeners. The result of the test was that the floating point V0.2.1 did not perform worse than the fixed point V5.3.0 of AMR-WB. The G722.1 at 24 kbps was scored better than AMR-WB for most of the music files. For each mode and each music signal, the bit exactness of the fixed point decoder and the floating point decoder has been verified.

Conclusion

No exception of bit exactness between fixed point decoder V5.3.0 and floating point decoder V0.2.1 has been found during the test. The floating point V0.2.1 of AMR-WB did not perform worse than the fixed point version of AMR-WB.

B.3 Bit-Exactness, Idle-Channel Behaviour and Long-Term Stability Performance

For all the tests, the V5.3.0 of the AMR-WB fixed-point code and the V0.2.1 of the AMR-WB floating-point code were used. The compilation was on Linux workstation and GNU C compiler [31].

Idle channel behavior (output signal when low noise input signal)

4 different low noise input signals (car, wind, bells, train) were encoded and decoded by the AMR-WB floating point coder in all 9 modes. The output files were listened by experts and no strange behavior or annoying artefacts was recognized. The outputs were also compared to those of AMR-WB fixed-point coder and no difference was noticed.

Stability of the codec over time

The purpose of this test was to check possible overflows when using very long input file. Speech signal of 2 hours 37 minutes was used as input. The speech activity of the file was 78 % and active speech level –26 dBov. The file contained German and English languages. The input file was encoded and decoded by the floating-point coder. That was repeated using all 9 AMR-WB modes (DTX and no DTX). No stability problems were observed in any mode.

Bit-exactness of the decoder

Bit-exactness of the decoder was tested with AMR Wideband Speech Codec test sequences 3GPP TS 26.174 V5.2.0. All encoded files .cod (both DTX and no DTX) were decoded by the 02.1. decoder and compared to the V5.2.0 output files .out. All test sequences passed the test. The synchronization frames were not tested.

B.4 Music Performance (Expert Listening Tests)

For all the tests, the v5.3.0 of the AMR-WB fixed-point code and the V0.2.1 of the AMR-WB floating-point code were used. The compilation was on Linux workstation and GNU C compiler [30].

Four music signals were used for this test:

- Classical, instrument: Beethoven, Symphony No. 9, part 2.
- Classical, vocal: Beethoven, Fidelio.
- Modern, instrumental: Radiohead, Karma Police (Piano+Guitar).
- Modern, vocal: Depeche Mode, Dream on.

All signals were encoded and decoded first using AMR-WB fixed-point C code. The same was repeated using AMR-WB floating-point C. Different modes and DTX ON/OFF were varied according to the table B.4.1.

Table B.4.1

C01	Mode 8 (23.85 kbit/s)	DTX = 0
C02	Mode 5 (18.25 kbit/s)	DTX = 0
C03	Mode 2 (12.65 kbit/s)	DTX = 0
C04	Mode 0 (6.6 kbit/s)	DTX = 0
C05	Mode 8	DTX = 1
C06	Mode 5	DTX = 1
C07	Mode 2	DTX = 1
C08	Mode 0	DTX = 1

Afterwards, those output files were compared in a informal expert listening test. The floating-point V0.2.1 performed equal to the fixed-point V5.3.0.

B.5 Overload Performance

This clause reports verification results of overload performance (high-level input signal conditions) of the AMR-WB floating-point C-code [29].

The C-code of V0.2.1 was compiled successfully with MS Visual C++ on a PC platform under Windows98, gcc (egcs-2.91.60) on a PC platform under Linux (kernel 2.2.9) and gcc (2.95.2) on SUN Ultra-60 workstation.

Four Japanese sentences (2 males and 2 females) from NTT-AT database were used as input sources. Each sentence has 8 s duration and its mean active power is normalized to 26 dB below overload with P.56 algorithm provided as 'sv56demo' in the ITU-T Recommendation G.191 software tool library.

Four kinds of input levels for AMR-WB coder (-26 dB, -16 dB, -6 dB, +4 dB to overload) were tested. The levels were set by using 'sv56demo'. All of 9 coding rates were used without Source Controlled Rate (SCR) operation. Two kinds of channel conditions (error free and 5 % random frame erasure) were simulated at the decoder. When testing the frame erasure condition, frame type was set to 'SPEECH_LOST' in the frames erased at the decoder. The level of decoded signals was adjusted again to 26 dB below overload in order to listen to them. The fixed-point coder of V5.3.0 was also used as reference for subjective quality evaluation.

All of 288 processed files (4 sentences x 4 levels x 9 rates x 2 channel conditions) were presented to an expert listener.

As the result of listening test, any significant problems were not found for all conditions. It was also shown that AMR-WB floating-point coder has subjectively equal quality compared to the fixed-point coder.

B.6 Transparency of Codec for DTMF signals

This clause describes the test for verifying the transparency of AMR-WB Floating-point (AMR-WBFl) speech codec for DTMF signals. This verification is performed in digital domain using software DTMF detector [32].

The objective of the activity is to generate DTMF test sequences corresponding to different scenarios (like different high frequency and low frequency power levels, DTMF duration and frequency deviation) and measure the percentage of detected DTMF digits for these sequences using AMR-WBFl under error free conditions.

The configuration that is being used by Hughes Software Systems (HSS) to verify the transparency of AMR-WB Floating-point speech codec for DTMF signals is in figure B.6.1. It essentially consists of DTMF generator, AMR-WBFl encoder & decoder, pre and post processing components (A-law compression and expansion, up and down sampling) and DTMF detector. Currently all the components are being done in software i.e. DTMF generation, A-law coding, sample-rate conversation, speech coding and DTMF detection are all performed using software simulations itself. The setup for using hardware DTMF detector for this activity is also shown in the figure.

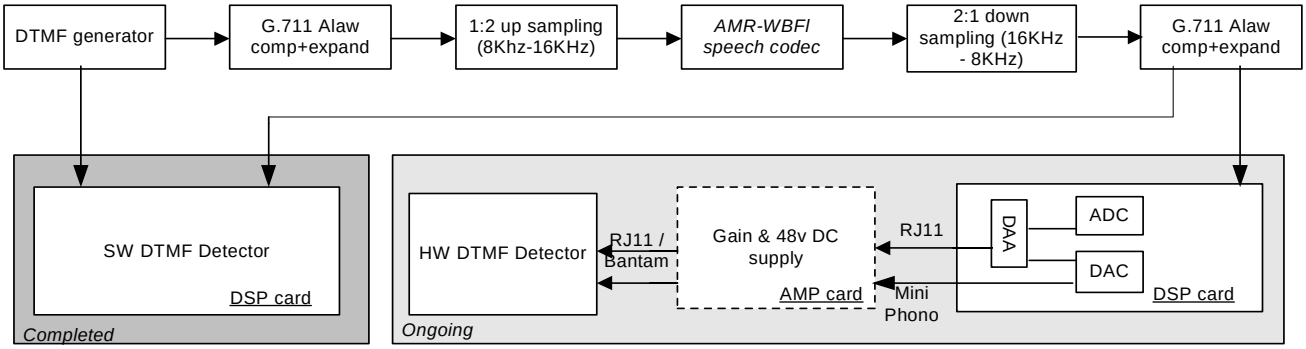


Figure B.6.1: DTMF Test Setup

The low and high frequency groups defined in ITU-T Recommendation Q.23 were used in generating the DTMF signals by the software generator. The other DTMF parameters like power levels (including twist), timing criteria and frequency bandwidths were generated as per specifications defined in ITU-T Recommendation Q.24. All the DTMF signals generated will be at 8KHz sampling rate.

Two types of DTMF generators are being used in the current activity, HSS software DTMF generator and Mitel Test Sequences.

HSS DTMF Generator was used to generate DTMF signals of different characteristics like power levels (including twist), timing criteria and frequency bandwidths. A second-order digital sinusoidal oscillator was used for generating the high and low frequency tones of DTMF signal.

Mitel test sequences, which are typically used for testing the performance of DTMF detector, have also been used as DTMF input source in our testing activity. If AMR-WBFl speech codec is transparent to DTMF signals, then a Mitel compliant DTMF detector should pass all the tests even after passing through the speech codec.

The verification activity has been planned with both hardware and software DTMF detectors and the test setup for this is shown in figure B.6.1. As hardware and software detectors may be used in a typical network, it is important to verify the codec's performance in both the domains.

A software detector does all the processing like reading samples from a file, DTMF detection and writing samples back to file in digital domain itself. HSS software DTMF detector is being used for DTMF detection in the current activity. This detector uses Goertzel algorithm to extract the spectral information of the DTMF signal by means of recursive digital filters. Once the spectral information is calculated for high and low frequencies, a number of checks are done to determine the validity of signal before declaring a digit as detected. The software is implemented in assembly and runs on a DSP card (TMS320c542). The detector software is well tested and is being used by HSS customers in satellite based systems.

The AMR-WBFl speech codec was simulated using *V0.2.2* of reference floating-point 'C code provided by Nokia. Encoder and Decoder reference executables are built from the reference code and have been used in the performance testing.

The other components used in the test setup are A-law codec, up and down sampling blocks. The A-law codec's compression and expansion modules present at pre-processing and post-processing stages of the setup simulate the effect of A-law narrowband digital connection in a typical network. These modules were simulated using the ITU-T Recommendation G.711 software provided in ITU-T Recommendation G.191 Software Tool Library (STL).

As the AMR-WBFl codec works at 16 kHz sampling rate, the test signals have to be up-sampled before passing to encoder and down-sampled back to 8 kHz after the decoder operation is completed. The sample rate conversion was performed with the high-quality FIR filter provided in ITU-T Recommendation G.191.

A number of test sequences corresponding to different DTMF signal properties have been generated for the activity using DTMF Generator apart from the standard Mitel sequences. Two categories of vectors have been generated using DTMF generator namely, signals with 80ms duration and DTMF signals with 50 ms duration.

80ms DTMF

A set of 13 experiments (HG1 to HG13) corresponding to DTMF signals of 80 ms duration have been used in testing activity and is shown in table B.6.1. The inter-digit silence for these sequences is also of 80 ms duration. The power level of both high and low frequency signals is represented as dB value with reference to the overload point (*dBov**), where the level of a sine-wave with peak amplitude of 1.0 corresponds to –3.01 dBov. All the testing has been done under ideal transmission conditions and no error patterns were simulated before sending the signal to detector. To avoid clipping when a DTMF signal is formed from high and low frequency tones the minimum power level used in experiment is –10 dBov.

Each experiment consists of a 16-digit DTMF frame (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, *, #, A, B, C, D) repeated 10 times with silence of finite duration inserted between frames.

Table B.6.1: DTMF Experiments used in AMR-WBFl (80ms)

Exp #	Low Frequency Power level (dBov)	High Frequency Power level (dBov)	Signal Duration (ms)	Frequency Deviation (±1.5 %)	Count (Total Digits)	Comments
HG1	-10	-10	80	0	160	0123456789*#ABCD sequence used in frame
HG2	-12	-12	80	0	160	
HG3	-12	-10	80	0	160	Standard Twist (2dB)
HG4	-12	-14	80	0	160	Reverse Twist (2dB)
HG5	-16	-16	80	0	160	
HG6	-16	-16	80	1	160	
HG7	-16	-13	80	0	160	Standard Twist (3dB)
HG8	-16	-19	80	0	160	Reverse Twist (3dB)
HG9	-16	-10	80	0	160	Standard Twist (6dB)
HG10	-16	-22	80	0	160	Reverse Twist (6dB)
HG11	-18	-18	80	0	160	
HG12	-22	-22	80	0	160	
HG13	-26	-26	80	0	160	

50ms DTMF

A set of 13 experiments (HG14to HG26) corresponding to DTMF signals of 50 ms duration have been used in testing activity and is shown in table B.6.2. The parameters used in these sequences are similar to the ones described in clause 0 except that the signal and silence durations are of 50 ms. These sequences have been generated so as to have commonality with the duration of DTMF signals in Mitel test sequences (which are also of 50 ms duration). So when doing the DTMF detection in hardware domain, the performances with both the vectors can be compared.

Table B.6.2: DTMF Experiments used in AMR-WBFl (50ms)

Exp #	Low Frequency Power level (dBov)	High Frequency Power level (dBov)	Signal Duration (ms)	Frequency Deviation (±1.5 %)	Count (Total Digits)	Comments
HG1	-10	-10	50	0	160	0123456789*#ABCD sequence used in frame
HG2	-12	-12	50	0	160	
HG3	-12	-10	50	0	160	Standard Twist (2dB)
HG4	-12	-14	50	0	160	Reverse Twist (2dB)
HG5	-16	-16	50	0	160	
HG6	-16	-16	50	1	160	
HG7	-16	-13	50	0	160	Standard Twist (3dB)
HG8	-16	-19	50	0	160	Reverse Twist (3dB)
HG9	-16	-10	50	0	160	Standard Twist (6dB)
HG10	-16	-22	50	0	160	Reverse Twist (6dB)
HG11	-18	-18	50	0	160	
HG12	-22	-22	50	0	160	
HG13	-26	-26	50	0	160	

Mitel Vectors

Additionally the transparency of AMR-WBFl speech codec has been tested by Mitel test vectors also. Although Mitel test sequences are typically used for testing the performance of DTMF detector, these have been used in our testing to verify the transparency of AMR-WBFl speech codec for DTMF signals and also to ensure that the DTMF detector being used in the testing is Mitel compliant. With a Mitel compliant detector one can measure the degradation provided by AMR-WBFl by checking the performance with that of normal scenario (input fed directly to detector). This testing is required as not all DTMF detectors in a typical network are going to be hardware based.

Although the testing has been done on the whole set of Mitel test sequences, a selected set of these (named MT1 to MT4) has been captured in the document. These test sequences are for basic digit sequence test, amplitude ratio test (twist), dynamic range tests and signal to noise ratio test (with noisy scenarios) and are given below in table B.6.3. All the Mitel test vectors are of 50 ms signal duration.

Table B.6.3: Mitel Experiments used in AMR-WBFl

Exp #	Test	Description	Pass Criteria
MT1	DTMF Decode Check	All 16 digits each of 50 ms duration	160
MT2	Amplitude Ratio (Twist) Test	8 sections of Standard twist (to 20 dB) and Reverse twist (0 to –20 dB) for Digits 1,5,9 and D with each section containing 200 pulses with 50ms duration/pulse	Standard Twist ≥ 4 dB Reverse Twist ≥ 8 dB (in each section)
MT3	Dynamic Range Test	35 tone pair pulses with 50 ms duration/pulse attenuated to –35 dB below from the nominal level in steps of 1dB	≥ 25 dB
MT4	Signal to Noise Ratio Test	3 sections with 1000 pulses/section with different white noise level for each section. The first level is at 24dB below the tone level, second at 18dB below and the third at12 dB below	1 000 (in each section)

Test Results

80 ms DTMF

The test sequences of 80 ms duration is given in table B.6.4. For all test sequences the count of number of digits detected was stored and the percentage of successful detection is calculated against the actual number of digits in a test vector (which is 160 digits in our experiments).

The tables given below show the percentage of successful detection only for all the codec modes. The output of the detector with direct input and with A-law companding codec (compression and expansion) is also provided for reference. Both these scenarios should be transparent to all test sequences.

Table B.6.4: Transparency of AMR-WBFl for DTMF Experiments (80 ms)

Mode-Exp	HG1	HG2	HG3	HG4	HG5	HG6	HG7	HG8	HG9	HG10	HG11	HG12	HG13
Mode 0	90	86.25	83.75	67.5	76.88	61.88	93.75	76.25	73.75	50.63	78.75	73.75	81.25
mode 1	100	100	100	100	100	100	100	100	96.25	93.75	100	100	100
mode 2	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 3	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 4	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 5	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 6	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 7	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 8	100	100	100	100	100	100	100	100	100	100	100	100	100
Direct Input	100	100	100	100	100	100	100	100	100	100	100	100	100
A-law	100	100	100	100	100	100	100	100	100	100	100	100	100

50 ms DTMF

The test sequences of 50 ms duration is given in table B.6.5. For all test sequences the count of number of digits detected was stored and the percentage of successful detection is calculated against the actual number of digits in a test vector (which is 160 digits in our experiments).

The tables given below show the percentage of successful detection only for all the codec modes. The output of the detector with direct input and with A-law companding codec (compression and expansion) is also provided for reference. Both these scenarios should be transparent to all test sequences.

Table B.6.5: Transparency of AMR-WBFl for DTMF Experiments (50 ms)

Mode-Exp	HG1	HG2	HG3	HG4	HG5	HG6	HG7	HG8	HG9	HG10	HG11	HG12	HG13
mode 0	58.75	59.38	66.88	57.5	56.88	43.13	65	56.25	33.75	36.88	51.88	60.63	58.13
mode 1	100	100	100	100	100	98.13	100	100	86.25	91.88	100	100	100
mode 2	100	100	100	100	100	99.38	100	100	100	98.13	100	100	100
mode 3	100	100	100	100	100	100	100	100	100	98.75	100	100	100
mode 4	100	100	100	100	100	100	100	100	100	98.75	100	100	100
mode 5	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 6	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 7	100	100	100	100	100	100	100	100	100	100	100	100	100
mode 8	100	100	100	100	100	100	100	100	100	100	100	100	100
Direct Input	100	100	100	100	100	100	100	100	100	100	100	100	100
A-law	100	100	100	100	100	100	100	100	100	100	100	100	100

Mitel Vectors

The test results for AMR-WBFl speech codec using Mitel test sequences is given in table B.6.6. For all test sequences the pass/fail criteria is decided based on the count of number of digits detected. This check ensures that the software DTMF Detector being used in Mitel Test Compliant and the transparency of AMR-WB Floating-point speech codec. If the AMR-WBFl is transparent to DTMF signals, the performance of detector with direct DTMF signals and that after passing through the speech codec should be similar (if not same i.e. meet the pass criteria).

Table B.6.6 Transparency of AMR-WBFl for Mitel Experiment

Mode- Exp	MT1	MT2	MT3	MT4
mode 0	Fail (117 instead of 160)	For Digit D, fails in both standard and reverse twist. All other Passed	Pass	Fail (998, 992 and 995 for the 3 sections instead of 1000)
mode 1	Pass	Pass	Pass	Pass
mode 2	Pass	Pass	Pass	Pass
mode 3	Pass	Pass	Pass	Pass
mode 4	Pass	Pass	Pass	Pass
mode 5	Pass	Pass	Pass	Pass
mode 6	Pass	Pass	Pass	Pass
mode 7	Pass	Pass	Pass	Pass
mode 8	Pass	Pass	Pass	Pass
Direct Input	Pass	Pass	Pass	Pass
A-law	Pass	Pass	Pass	Pass

Conclusion

All the test sequences (HG1 to HG26) used for testing the AMR-WB Floating-point speech codec are detected by software DTMF detector. The G.711 (A-law) codec is transparent to DTMF signals and all the digits have been detected (for all experiments). With DTMF signals of 80ms duration, all the vectors were successfully detected except for mode 0 and mode 1.

The lowest codec mode (mode 0) is not transparent to DTMF signals (for 50 ms and 80 ms duration). The output of codec in modes 0 and 1 for DTMF signals of 50 ms duration is degraded compared to signals of 80 ms duration. Also performance of speech codec with standard twist is relatively better compared to reverse twist.

Only the last four modes (modes 5, 6, 7 and 8) appear to be completely transparent to DTMF signals of 50 ms duration and with reverse twist of 6 dB.

For Mitel test sequences (MT1 to MT4) also, mode 0 of speech codec is not meeting the pass criteria, which indicates that the speech codec is definitely not transparent for this mode. Even the experiment MT1, which is the basic decode check, is failing for mode 0. The direct input and output of A-law codec pass the criteria for all test cases.

B.7 Perceptual Evaluation of Speech Quality (PESQ)

This clause presents verification results for the floating-point implementation of AMR-WB using a wideband version of the ITU-T Recommendation P.862 Perceptual Evaluation of Speech Quality (PESQ) algorithm [26] and [27].

Narrowband PESQ (P.862) was standardised by the ITU-T as Recommendation P.862 in February 2001 after winning the ITU-T competition to find a replacement for PSQM (P.861). The algorithm passed all of the ITU's performance requirements in independent verification procedures, which were based on the results of thirty subjective experiments.

As the name suggests, Wideband PESQ (WB-PESQ) extends the operation of PESQ to the assessment of wideband speech systems. The algorithm was presented to the ITU-T in October 2001, and a complete description can be found in an ITU-T white contribution COM12-36 [1].

The verification was divided into the four experiments described in table B.7.1. All nine AMR-WB modes were tested in each experiment in addition to a case where the mode was selected at random every 20ms. The background noise types and signal to noise (SNR) ratios used are consistent with those used in Experiment 6 of the AMR-WB Characterisation Phase.

Table B.7.1: Verification experiments

Exp	Noise	SNR	DTX
1	Clean	-	no
2	Vehicle	15dB	no
3	Office	20dB	no
4	Office	20dB	yes

A set of 32 files was processed for each test condition, comprising four samples from two male and two female talkers in two languages (British English and French). Each sample was a standard 8-second sentence pair of the type commonly used in subjective experiments.

The fixed-point ANSI C code was taken from V5.3 of 3GPP TS 26.173. Both codecs were compiled under Microsoft Visual C/C++ 6.0 with the /O2 optimisation level. The codecs were executed on a 600Mhz Dual Processor Pentium 3 running Windows NT 4.0.

PESQ is an intrusive speech quality measurement algorithm, and as such requires a reference and degraded signal pair to measure the performance of a speech transmission system (see figure B.7.1). For this validation, the reference signal used was the speech signal after the addition of background noise. This configuration is consistent with the Degradation Category Rating method of subjective testing.

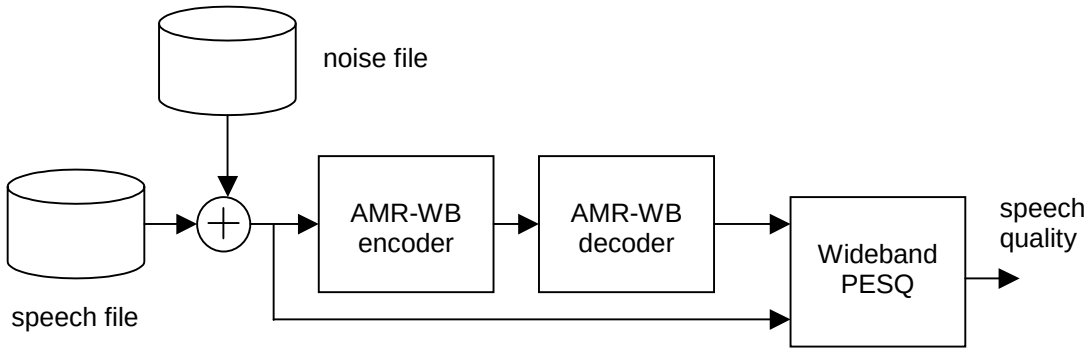


Figure B.7.1: PESQ configuration.

The input signals were pre-processed according to the procedures defined in the AMR-WB Characterisation Processing Plan [2]. Rounding to 14-bits was not implemented, in order to allow for any differences in handling the least significant bits of the input signal. Each 8-second file was processed separately without a preamble.

Results

The floating-point implementation of the AMR-WB decoder is designed to provide bit-identical operation with the fixed-point decoder. Bit-stream files were generated using the fixed-point encoder for all 1280 test files (32 speech samples x 10 modes x 4 experiments). The outputs of the two decoders were compared for each test condition, and were found to be identical in all cases.

The performances of the fixed-point and floating-point encoders were measured using the Wideband PESQ algorithm for each test file. For this evaluation, each encoder was used with its corresponding decoder: the fixed-point encoder was used with the fixed-point decoder, and the floating-point encoder was used with the floating-point decoder. For each experiment, we show the following graphs:

(a) WB-PESQ scores for fixed-point encoder

This plots the condition average WB-PESQ score for each mode for the fixed encoder. The error bars plot the minimum and maximum WB-PESQ scores observed. In addition to modes 0–8, results are also given for a switched-rate condition in which the mode was changed randomly for each frame.

(b) WB-PESQ scores for floating-point encoder

This plot is equivalent to (a), but shows the results for the floating-point encoder.

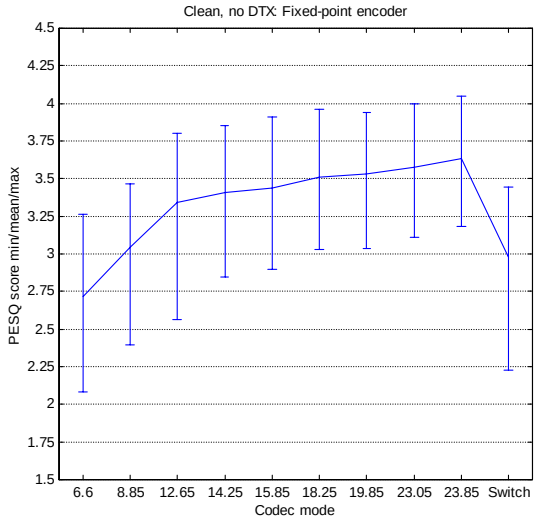
(c) Difference in WB-PESQ scores between encoders

The condition average difference between the WB-PESQ scores given to each encoder are shown in this plot. The minimum and maximum differences for a given original speech file are shown by the error bars.

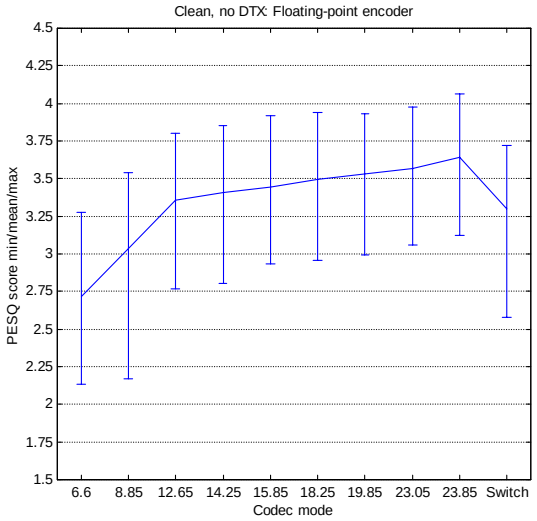
(d) Distribution of differences in WB-PESQ scores

This plots the histogram of the file-by-file differences between the two encoders. The histogram bins used are separated by 0.05 and centred on 0.0.

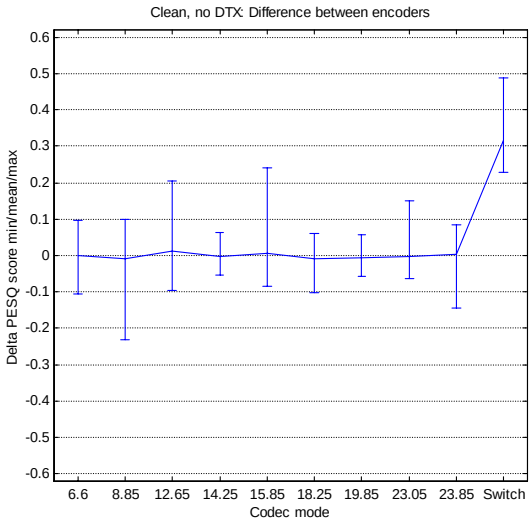
Experiment 1: Clean speech



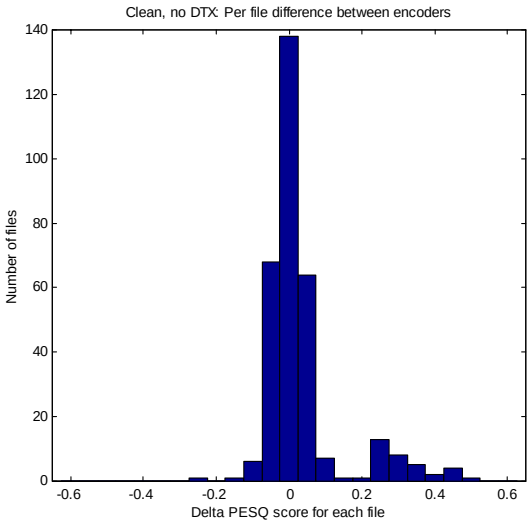
(a) WB-PESQ scores for fixed-point encoder



(b) WB-PESQ scores for floating-point encoder

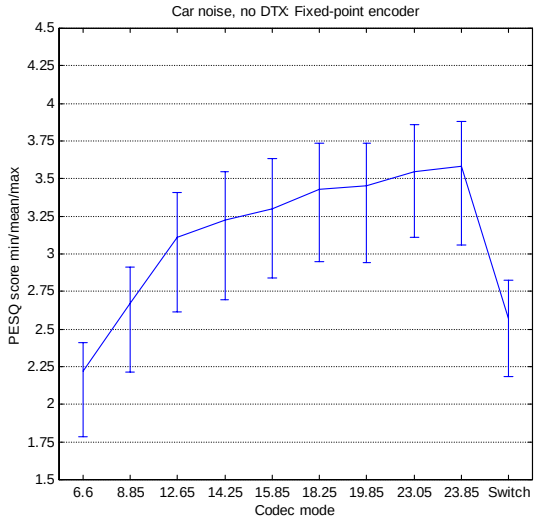


(c) Difference in WB-PESQ scores between encoders

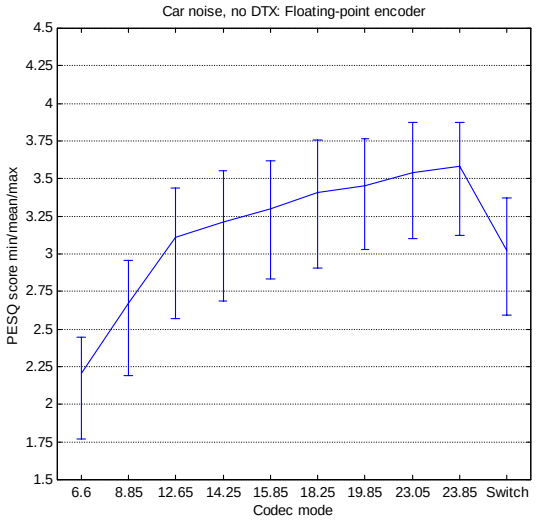


(d) Distribution of differences in WB-PESQ scores

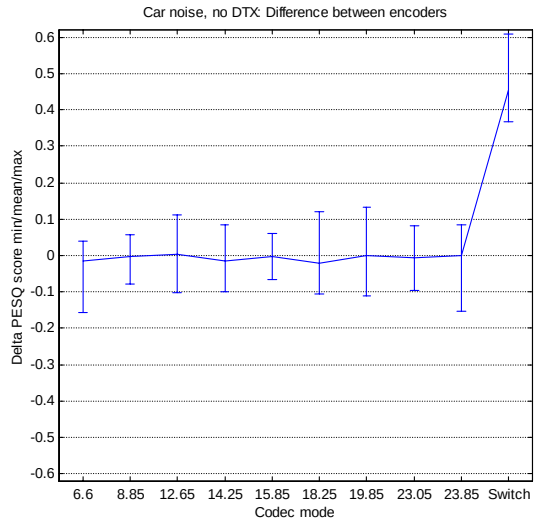
Experiment 2: Vehicle noise at 15dB SNR



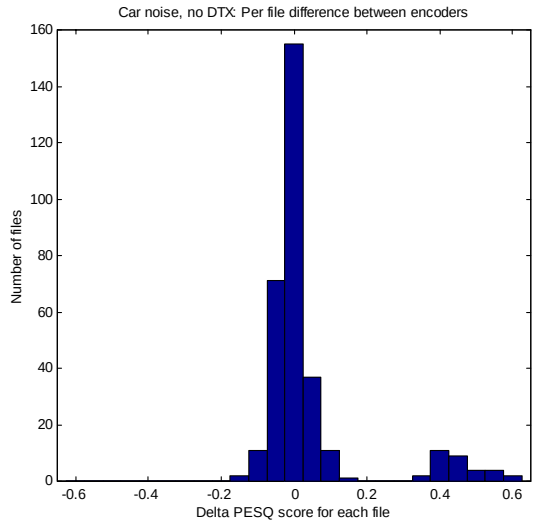
(a) WB-PESQ scores for fixed-point encoder



(b) WB-PESQ scores for floating-point encoder

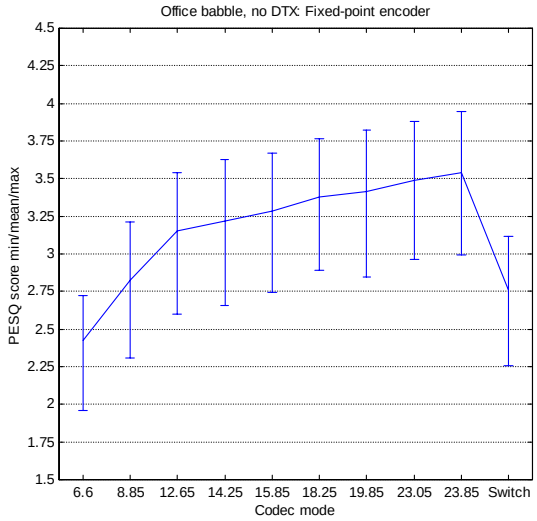


(c) Difference in WB-PESQ scores between encoders

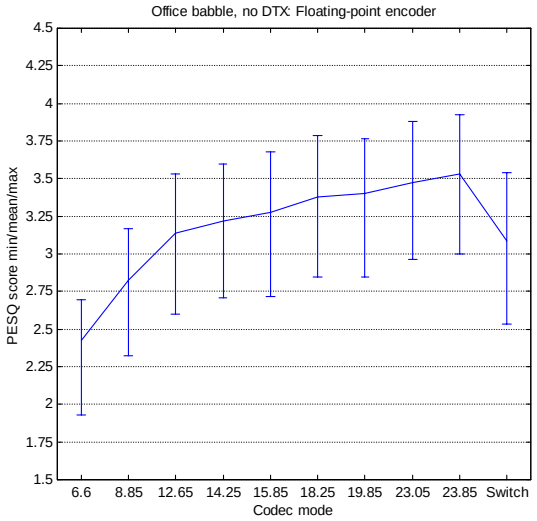


(d) Distribution of differences in WB-PESQ scores

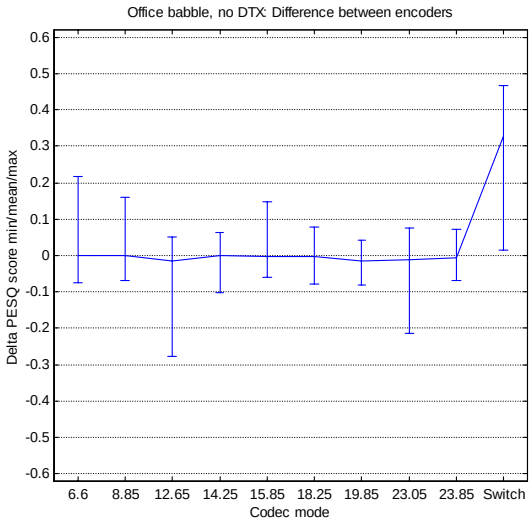
Experiment 3: Office noise at 20dB SNR



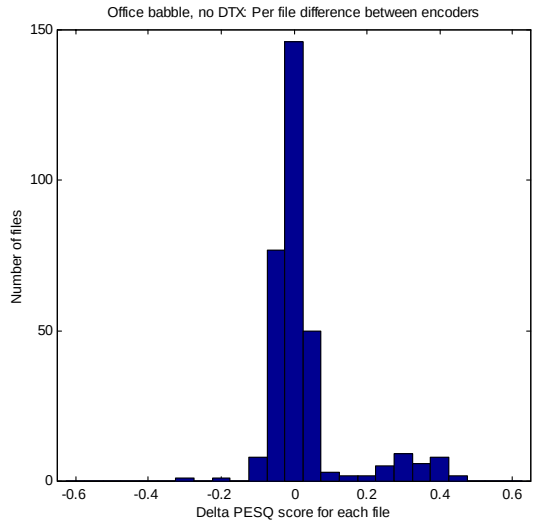
(a) WB-PESQ scores for fixed-point encoder



(b) WB-PESQ scores for floating-point encoder

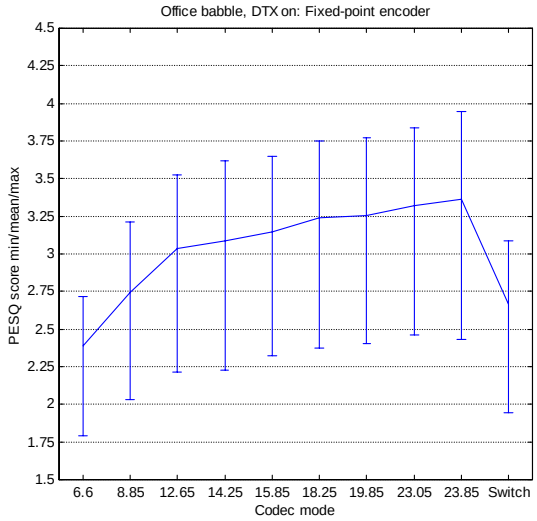


(c) Difference in WB-PESQ scores between encoders

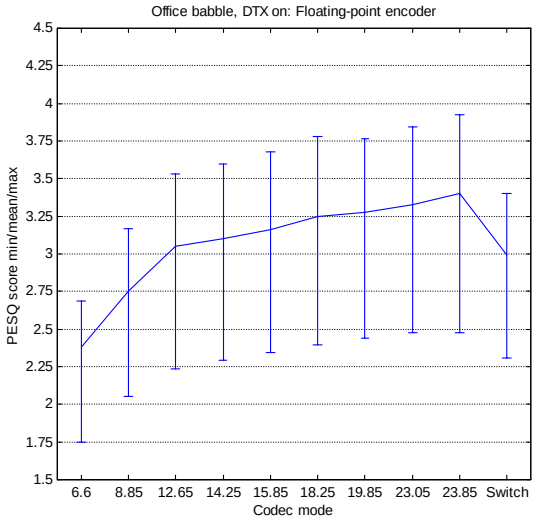


(d) Distribution of differences in WB-PESQ scores

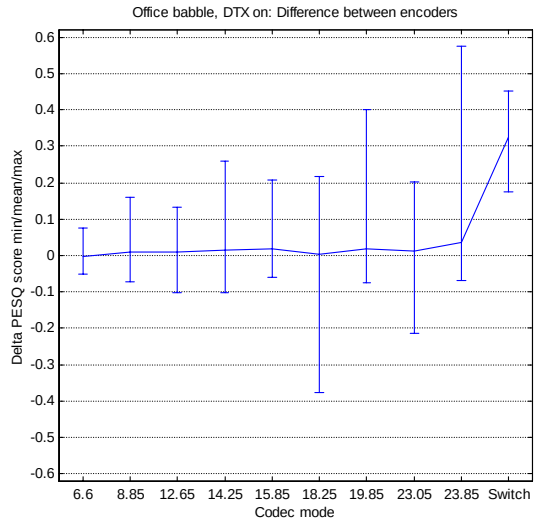
Experiment 4: Office noise at 20dB SNR, with DTX



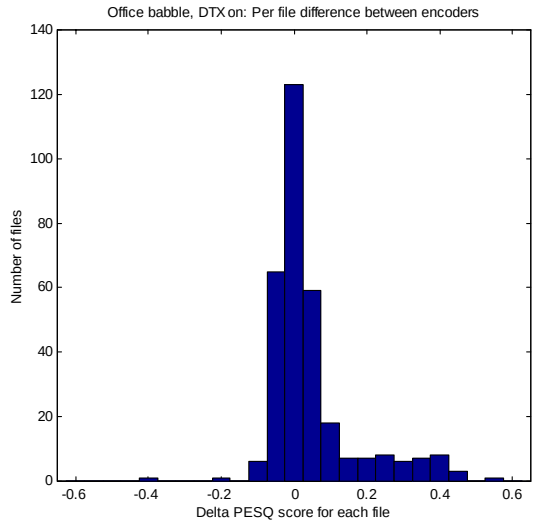
(a) WB-PESQ scores for fixed-point encoder



(b) WB-PESQ scores for floating-point encoder



(c) Difference in WB-PESQ scores between encoders

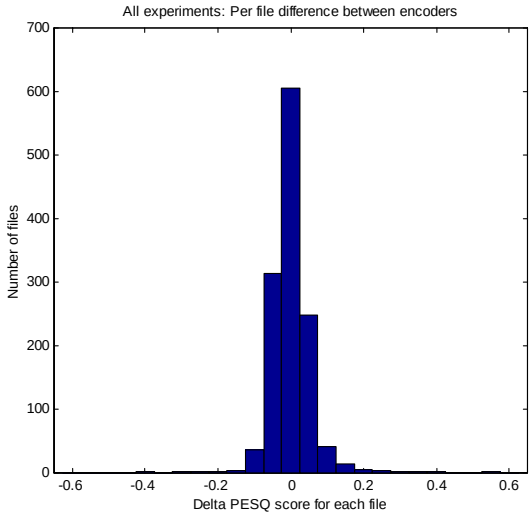


(d) Distribution of differences in WB-PESQ scores

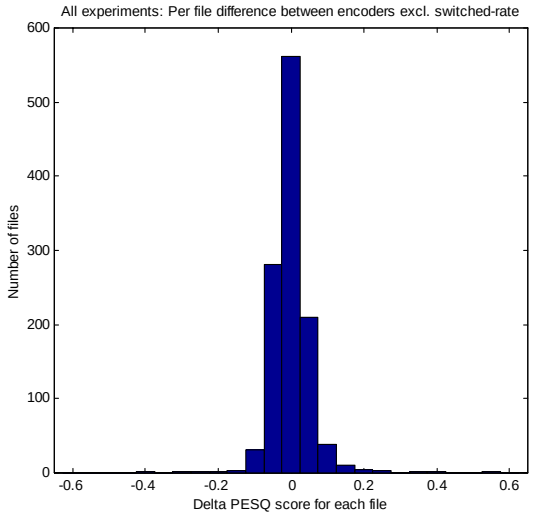
Combined results

The distribution of differences between the encoders, across all experiments and all modes, is shown in figure (a). The distribution of differences for all fixed-rate modes (excluding the switched-rate conditions) are shown in figure (b). The distribution of differences for only the switched-rate conditions are shown in figure (c).

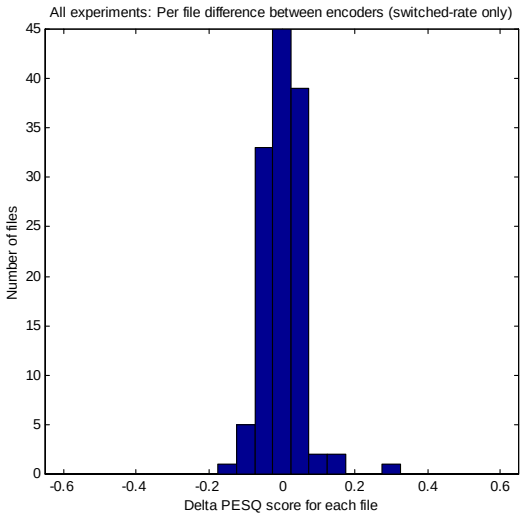
As before, the histogram bins used are separated by 0.05 and centred on 0.0.



(a) All combined results



(b) All fixed-rate modes (excluding switched-rate conditions)



(c) All switched-rate conditions

Table B.7.2: Significance tests

Sample	Sample mean	Sample std. dev	Min/max range	Data points	Approx. symm. 99 % CI	Significant?
All conditions	0.0001	0.054	-0.38 0.58	1 280	0.0039	No
All modes (excl. switched-rate)	-0.0002	0.054	-0.38 0.58	1 152	0.0041	No
Switched-rate conditions only	0.0026	0.052	-0.14 0.29	128	0.0118	No

Conclusions

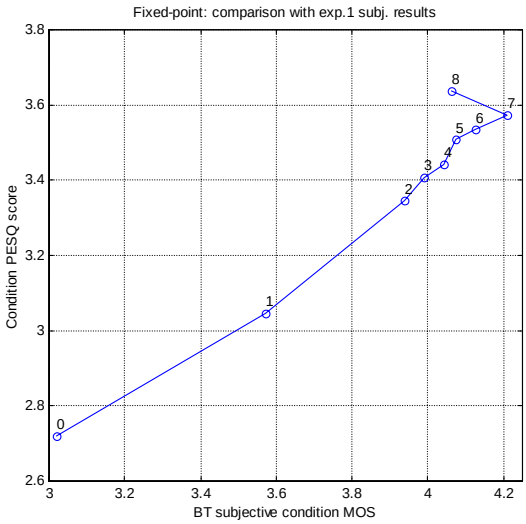
It is most likely, from the data, that there is no significant subjective difference between V5.3.0 of the fixed-point AMR-WB encoder with CR011 implemented and V0.2.2 of the floating-point AMR-WB encoder.

Appendix: Comparison between WB-PESQ and subjective MOS

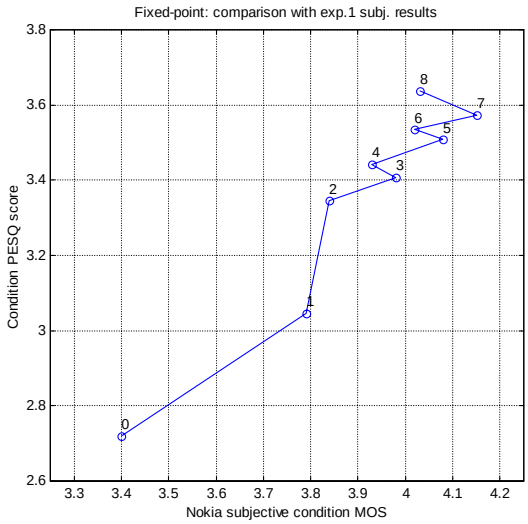
Just as there are normal variations in MOS from one subjective test to another, and between subjective listening laboratories, so there are variations between PESQ score and subjective MOS. However, before we can be satisfied about the results of the validation described in the present document, we need to know the relationship between PESQ and subjective MOS. This also makes it possible to understand the results of the validation: for example, is a change in PESQ score of 1.0 comparable to a change in MOS of 1.0?

For the four subjective test results reported in [1], WB-PESQ has an average correlation with MOS, measured per condition after monotonic 3rd-order polynomial mapping, of 96.5 %. However WB-PESQ had not previously been validated with the AMR-WB codec. In this annex we present a comparison with MOS for experiment 1 of the fixed-point AMR-WB characterisation tests.

Because it was not possible to replicate the more complex subjective test conditions given the limited data made available to us, we present results only for the clean speech conditions, with no tandeming and at nominal levels. The following graphs compare the subjective MOS reported by BT [4] and Nokia [5] for Experiment 1 with WB-PESQ, for the fixed-rate codec modes from 0 to 8. Condition averages are used both for subjective MOS and PESQ score. The linear correlation coefficients for these data sets are 97.4 % and 95.8 % respectively.



(a) BT results



(b) Nokia results

Given these results, our conclusions are as follows.

- WB-PESQ scores are monotonically increasing with bit-rate for this codec.
- There are small deviations from a smooth curve. It is difficult to account for these deviations without reference to the subjective test data, but they may be due to subtle differences in background noise processing or to subjective factors such as randomisation or material dependence.
- WB-PESQ appears to give scores that are slightly lower overall than subjective MOS for these tests.
- A range of WB-PESQ scores of about 0.9 (2.72 to 3.64) corresponds to a range of MOS of about 1.2 (3.02 to 4.21) for the BT test, and 0.75 (3.40 to 4.15) for the Nokia test. Differences in WB-PESQ score are clearly of similar magnitude to differences in MOS.
- WB-PESQ is applicable to the AMR-WB codec and appears to have a high correlation with MOS.

B.8 Operation of the VAD and comfort noise

This clause reports the results of the verification of the comfort noise generation system of the AMR-WB Floating point codec [28]. A comparative investigation with the AMR-WB Fixed-Point codec was made. The investigation compares the performance of the respective VADs and the behavior of the comfort noise generation. The study is organized similarly to the verification comfort noise generation system of the AMR-WB Fixed-Point codec [36]. In the course of the verification a bug causing a floating point exception was encountered. The bug was fixed after communication with Nokia and the verification was carried out with the accordingly modified codec implementation.

Test Conditions

In accordance to verification of the AMR-WB Fixed-Point VAD [36], as a base for all experiments regarding VAD performance a five minutes long file was used containing conversational speech. This speech file was created from a database with Swedish speech material, comprising two male and two female speakers. The material was concatenated so that it contained approximately 40 % speech time and 60 % time of silence. For the main part of the investigations the input level of the speech was set to –26 dBov. However, tests with different input levels of the speech material have also been made. In these cases, the input level was set to –16 dBov and –36 dBov, respectively.

Four different types of noises are added to the speech file. The noises are recordings from car, street, office and airport hall environments. The noises differ widely in stationarity. To get In order to give some idea of how the stationarity of the noises, frame energy variances, i.e. the variances of frame-wise energy estimates, were calculated. are, we have computed how the energy-variance of the signal changes in-between the frame, i.e. the variance of the energy-variance. The result of this computation is shown in figure B.8.1.

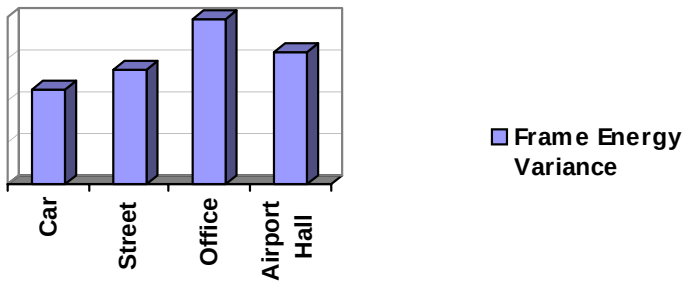
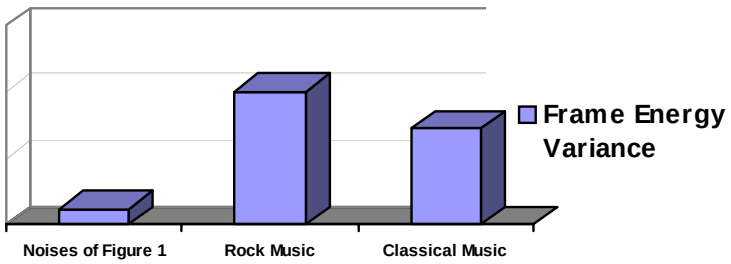


Figure B.8.1: Stationarity of noises

In addition, are two kinds of music are used as background noises. One file containing classical music (Bach) and one file containing rock music (Smashing Pumpkins). According to the stationarity measure from above, Then, the measure of stationarity above is used, does the file containing classical music show to be theis the more stationary one, and the music pieces are less stationary than the other noises.



FigureB.8.2: Stationarity of music files

The background files are added to the speech files at four different levels such that signal-to-noise ratios of 40 dB, 30 dB, 20 dB, and 10 dB are obtained. The noise is scaled in the same way as in the processing for the AMR-WB selection tests [37].

Voice/Channel activity

To evaluate the performance of the voice activity detection we have observed the VAD-flag and calculated the voice activity and clipping for different background conditions. The voice activity is calculated as follows:

Equation 1:

$$\text{voice activity} = \frac{\text{number of frames where VAD flag is "1"}}{\text{number of all frames}}$$

The voice activity obtained from the different background conditions is compared to the activity of the ideal case, i.e. the clean case without any background noise.

The channel activity is the relevant parameter for evaluating the gain of a DTX system. It is the ratio between the number of transmitted frames (SPEECH, SID_FIRST, SID_UPDATE) and the number off all frames including the NO_DATA frames. The channel activity is calculated as follows:

Equation 2:

$$\text{channel activity} = \frac{\text{number of frames - number on NO\_DATA frames}}{\text{number of all frames}}$$

Results

Measured Voice- and Channel Activity Factors for clean speech are given in Table B.8.1. It is seen that the differences are only minor.

Table B.8.1: Activity factors for clean speech

	Floating-Point VAD	Fixed-Point VAD
Voice Activity Factor [%]	40.1307	40.1477
Channel Activity Factor [%]	50.7216	50.7386

Voice activity and channel activity measurements for the different background cases and different input levels are shown in figures B.8.3 to B.8.6. Bars representing the respective activity figures for Floating-Point VAD and Fixed-Point VAD measured for a given condition are depicted next to each other in different patterns.

In figures B.8.3 and B.8.4 it can be seen that the achievable activity is very similar for the different VADs. In total, the Floating-Point VAD leads to a slightly higher activity than the Fixed-Point VAD.

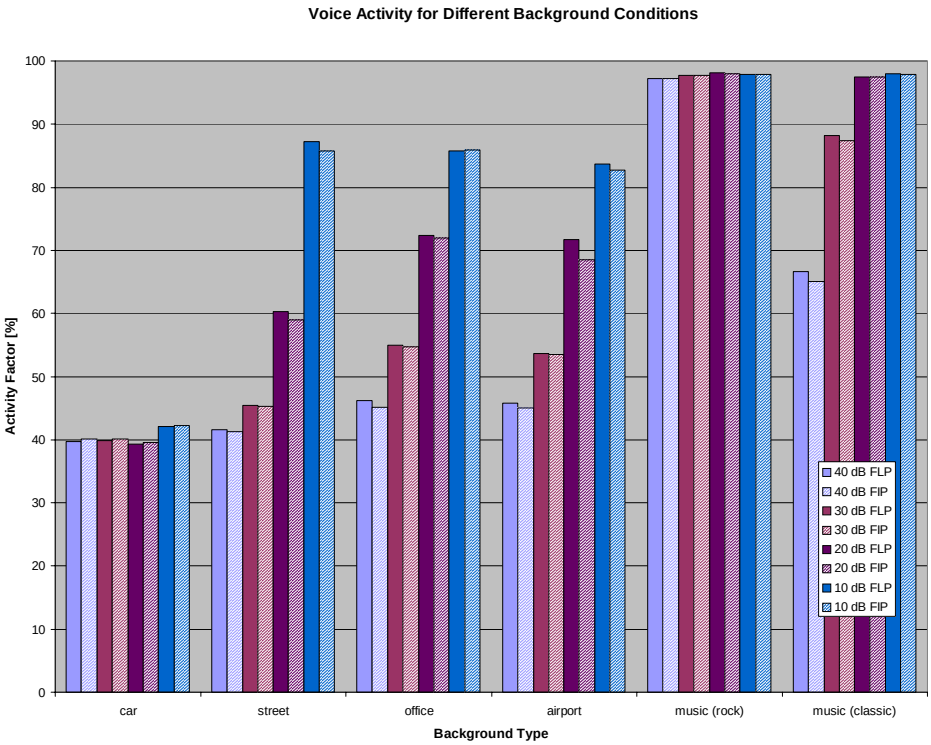
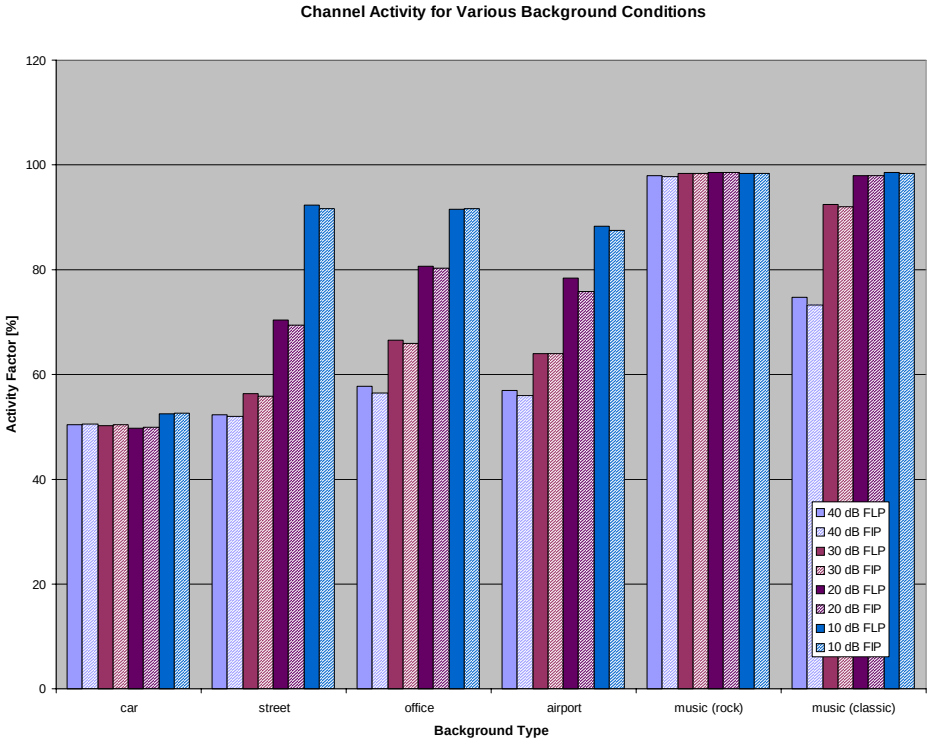


Figure B.8.3: Voice activity for different background conditions, input speech level -26 dBov



FigureB.8.4: Channel Activity for different background conditions, input speech level = -26 dBov

Figures B.8.5 and B.8.6 show the dependence of the achievable voice and, respectively, channel activities on the input level for the example of street noise. It is again found that the measured activity factors are very similar. However, the following tendencies are visible:

- The Floating-Point VAD leads to relatively higher activity factors for poor SNR conditions.

- For low input levels, the Floating-Point VAD leads to relatively lower activity factors.

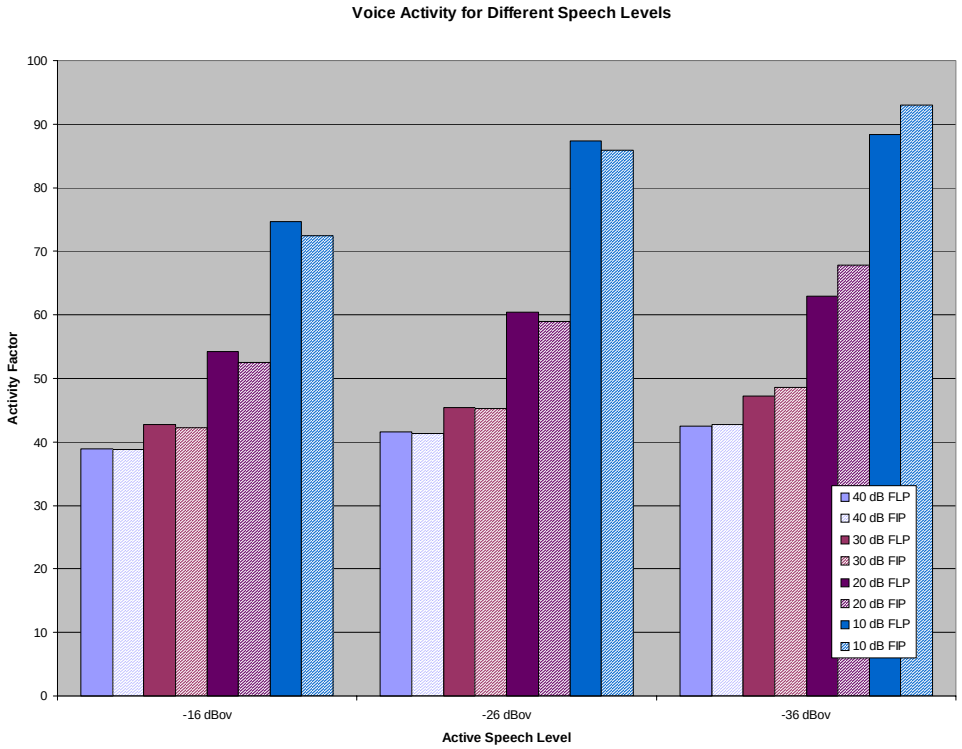


Figure B.8.5: Voice Activity for different input levels (street noise)

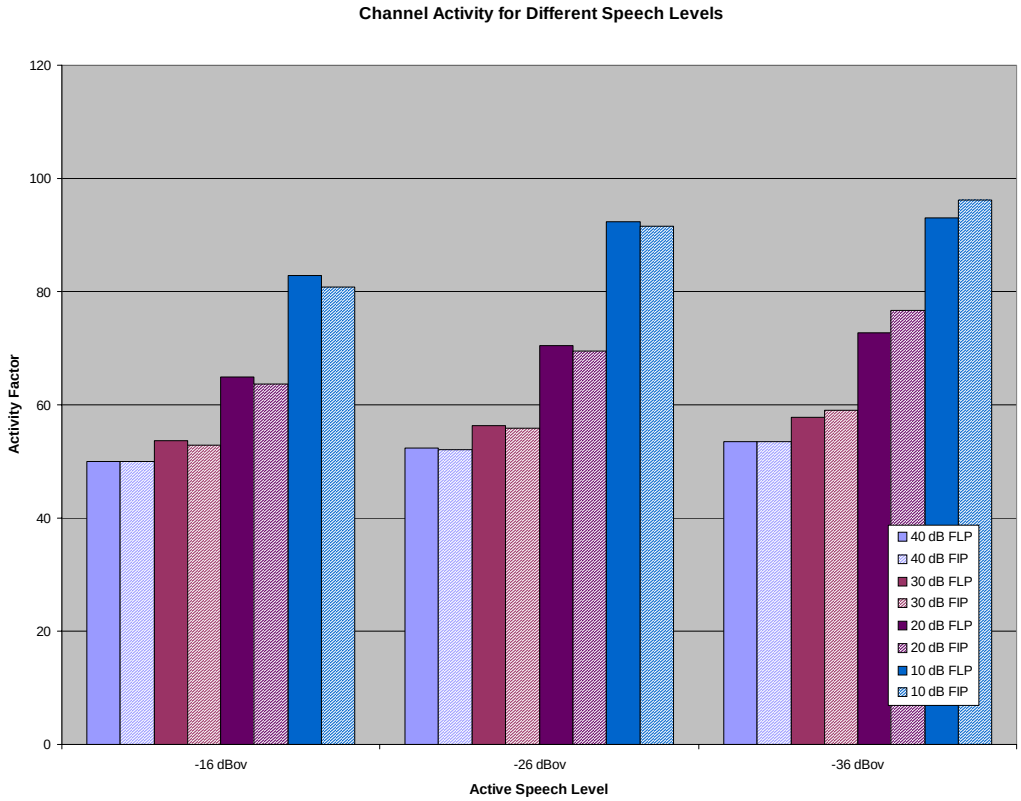


Figure B.8.6: Channel Activity at different input levels (street noise)

Clipping

For speech clipping assessment, the methodology described in [38] was taken over. This methodology is restated as follows: We first estimate how loudly speech is audible in each frame:

$$L_{sp}(n) = \left(\frac{\max(0, sp(n) - 0.25 * no(n))}{1 + (no(n)/sp(n))^2} \right)^{0.3},$$

Equation 3:

where:

sp(n): speech power of the frame n.

no(n): noise power of the frame n.

$L_{sp}(n)$ loudness of speech in frame n.

Speech and noise powers for each frame are calculated from the clean speech and noise files. The exponent of 0.3 is derived from the relation between loudness and intensity, i.e., an increase of 10 dB in the intensity causes the loudness to double. When speech power is 6 dB lower than noise power (see the 0.25 gain in the above equation), we assume that speech is not audible and loudness will be zero. Noise power in each frame is limited to below -55 dBm0, which is close to the noise level of the clean speech files. This limitation makes this equation applicable also for clean speech samples. Speech clipping is calculated as follows:

$$C_{sp} = \frac{\sum_n L_{sp}(n) * (1 - VAD\_flag(n))}{\sum_n L_{sp}(n)},$$

Equation 4:

where VAD_flag(n) is the output of the VAD algorithm (1 for speech, 0 for noise).

As shown on the above equation, clipping is sum of loudness of the frames where VAD is "0" divided by sum of loudness of all frames.

Clipping measurements according to Eq. 4 for the different background cases and different SNRs are shown in figures B.8.7. The two VADs behave very similar and no consistent tendency can be observed according to which the VADs perform significantly differently. For clean speech, for both VADs a clipping figure of 0.0060 is measured.

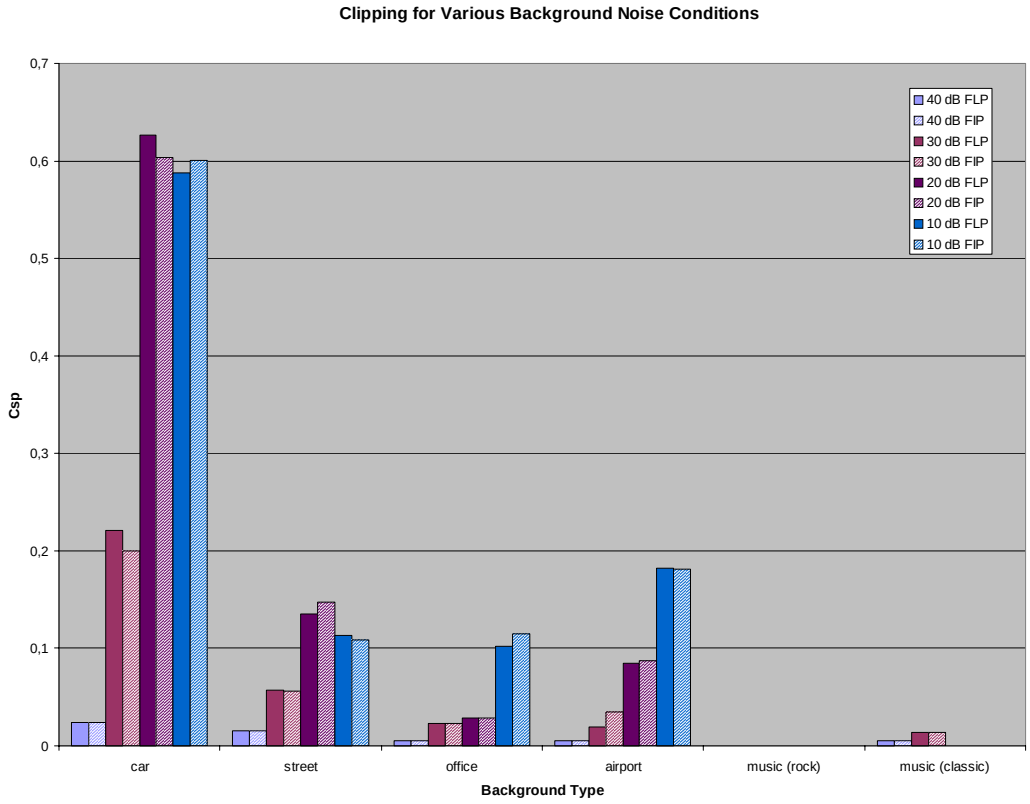


Figure B.8.7: Clipping for different background conditions, input speech level –26 dBov

For those speech samples for which severe clipping has been observed according to the clipping measure given above, careful expert listening has been carried out in order to check if the clipping is audible or differences between the two VAD implementations can be perceived. For most cases no clipping was found. In cases of slight clipping (car noise, low SNR) no significant differences could be noticed.

Additionally, VAD performance for pure music files was tested. Both VADs perform very similarly. On most music files only a few sparse frames are classified as inactivity, which does not affect the quality significantly. For certain problematic pieces (such as Carmina Burana by Orff) where the VAD switches to inactivity for longer periods, the quality is degraded. However, there are only minor differences between the two VADs.

Comfort Noise Synthesis

The purpose of this investigation is to evaluate if there are noticeable differences between the comfort noise syntheses of fixed and floating point implementations of AMR-WB, which would result from different comfort noise parameter calculations in the encoder. The investigation is done in two parts, as follows.

Comfort Noise Contrast Effects During Inactivity

In order to investigate the comfort noise synthesis during inactivity, coding is done with the VAD decision forced to 0. Input signals used in this test are:

- Car noise.
- Street noise.
- Office noise.
- Airport noise.
- Artificial white noise with slow random magnitude variations.
- Artificial narrow band noise with sweeping center frequency from 50 Hz to 7 000 Hz.

For none of the signals remarkable differences between Fixed-Point and Floating-Point implementations of AMR-WB can be reported.

Comfort Noise Contrast Effects due to DTX state changes

This test was made with the respective VADs enabled. The input signals used are those listed in the beginning of this clause but the level adjusted to such a value that the VAD decision is unstable. I.e. the VAD flag and in response to this, the DTX state toggles between activity and inactivity.

For none of the test signals significant qualitative differences can be reported between the two AMR-WB implementations. However, it is noticeable that the two VAD implementations slightly differ in sensitivity. This causes activity-inactivity transitions in the decoded signals to be located differently.

Conclusion

VAD and comfort noise generation of the AMR-WB Floating-Point codec perform very similar to the corresponding fixed-point implementation. The floating-point VAD has a slightly different sensitivity which may lead to small differences between the achievable activity figures for a given signal condition. Even though for certain input signals this may result in slightly different decoded signals, no characteristic differences in perceived signal quality are to be reported.

Annex C:
Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2002-12	18	SP-020682			Version 2.0.0 presented at TSG-SA#18 for approval		5.0.0
2003-09	21	SP-030450	001		Reference to incorrect test results	5.0.0	5.1.0

History

Document history		
V5.0.0	December 2002	Publication
V5.1.0	September 2003	Publication