3GPP TR 26.976 V0.3.0 (2001-06)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
AMR-WB Speech Codec Performance Characterisation;
(Release 5)

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

Keywords

AMR-WB

3GPP

Postal address

3GPP support office address

650 Route des Lucioles – Sophia Antipolis
Valbonne – FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

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:

6 presented to TSG for information;

7 presented to TSG for approval;

8 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.

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] "AMR-WB Feasibility Study Report v. 1.0.0", ETSI TC SMG Tdoc P-99-429, SMG meeting #29, 23rd – 25th June, 1999 (Miami, USA)

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

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

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

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

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

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

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

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

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

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

[12] 3G TR 21.905 Vocabulary for 3GPP Specifications

3 Definitions and abbreviations

3.1 Abbreviations

For the purposes of the present document, the following abbreviations 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
SNR	Signal To Noise Ratio
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

For abbreviations not given in this chapter, see TR 21.905 [12].

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 was the latest phase. During this phase the codec was tested in a more complete manner than in the selection phase. Characterisation was completed by TSG-SA#14 in December 2001.

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 participating listening test laboratories in any of the tested conditions. The codec fulfils all the design constraints.

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[1]:

- 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

The codec mode can be changed every 20 ms in 3G WCDMA channels and every 40 ms in GSM/GERAN channels.

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

4.3 Presentation of the following chapters

The following chapters provide a summary of the Selection, Verification and Characterisation Phase test results, including a review of the performance requirements and selection criteria. Chapter 5 defines the minimum performance requirements for speech quality. Chapter 6 will give short summary of the experiments performed (and to be performed) during the characterisation and verification phases of testing. Chapters 7-14 describe the results of the subjective listening tests undertaken during the characterisation phase. Chapters 15-25 contain results from the Verification Phase. Annex A contains detailed information about the AMR-WB selection 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 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.

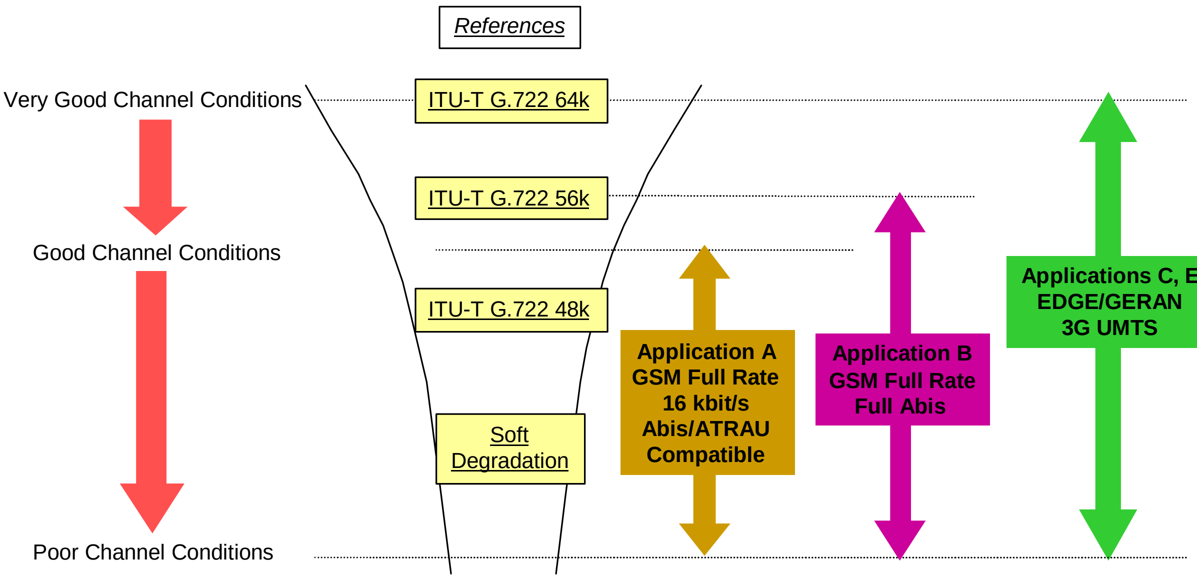


Figure 5.1: Quality requirements for the AMR-WB codec for the various applications [4].

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

6 Introduction to the testing of AMR-WB speech codec performance

6.1 AMR-WB Characterisation phase

The Characterisation Tests consist of 10 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 different characterisation phase experiments

Exp.	Characterises systems:	Test type	Title	Number of conditions	No. of 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 (Note 1)	ACR	The Effect of Static Errors under Clean Speech Conditions.	56	1
7b	3G (Note 1)	ACR	The Effect of Static Errors under Clean Speech Conditions.	56	1
8a	3G (Note 1)	DCR	The Effect of Background Noise 3 in Static C/I Conditions.	48	1
8b	3G (Note 1)	DCR	The Effect of Background Noise 4 in Static C/I Conditions.	48	1
8c	3G (Note 1)	DCR	The Effect of Background Noise 5 in Static C/I Conditions.	48	1
9a	EDGE 8-PSK (Note 2)	ACR	EDGE Characterisation, FR/HR/QR-channel The Effect of Static Errors under Clean Speech Conditions, set 1	-	1
9b	EDGE 8-PSK (Note 2)	ACR	EDGE Characterisation, FR/HR/QR-channel The Effect of Static Errors under Clean Speech Conditions, set 2	-	1
10	PS-systems (Note 2)	ACR?	Testing for Packet-switched (PS) conversational and streaming applications	-	1
			Total		18

Note 1: Experiments 7 and 8 will be performed in Phase 1B (after the error pattern simulations become available).
Note 2: Experiments 9 and 10 will be performed in Phase 2. The detailed test plan for these experiments is FFS.

Characterisation was divided between several laboratories using different speech databases and languages. Special laboratories were allocated for host lab and crosschecking functions. The work division is described in Table 6.2.

The Characterisation Phase was divided into two phases. Phase I covered characterisation of the AMR Wideband codec in error-free channel (all systems) and in GSM GMSK Full-Rate traffic channel. Characterisation in 3G WCDMA channel and EDGE 8-PSK channels and also characterisation in packet switched applications is to be carried out in phase II.

The phased approach made it possible to complete part of the testing by June 2001 even though the EDGE speech services would not be fully specified in time. Later during the test preparation it was found out that the 3G error patterns were not available for testing during spring 2001. Therefore, testing of AMR-WB in 3G channels was postponed into Phase Ib.

Chapters 7-14 contain the complete set of test results for the AMR-WB speech codec Characterisation Phase I, 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.

***Important Note 1**: 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 subjects…). 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. Finally, it should be noted that a difference of 0.2 MOS between two test results was usually found not statistically significant.*

***Important Note 2:** In the characterisation testing, experiments 1, 2 and 5 were conducted twice using different listening laboratories and languages. Tdoc S4-010393 from Dynastat (attached into this TR) 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.*

6.2 AMR-WB Verification phase

The following table lists the verification items relevant for performance characterisation and corresponding contributing organisations.

Table 6.3: Verification tasks and their allocation to the volunteering laboratories

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

7 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 7.1 and Figure 7.2 , 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.

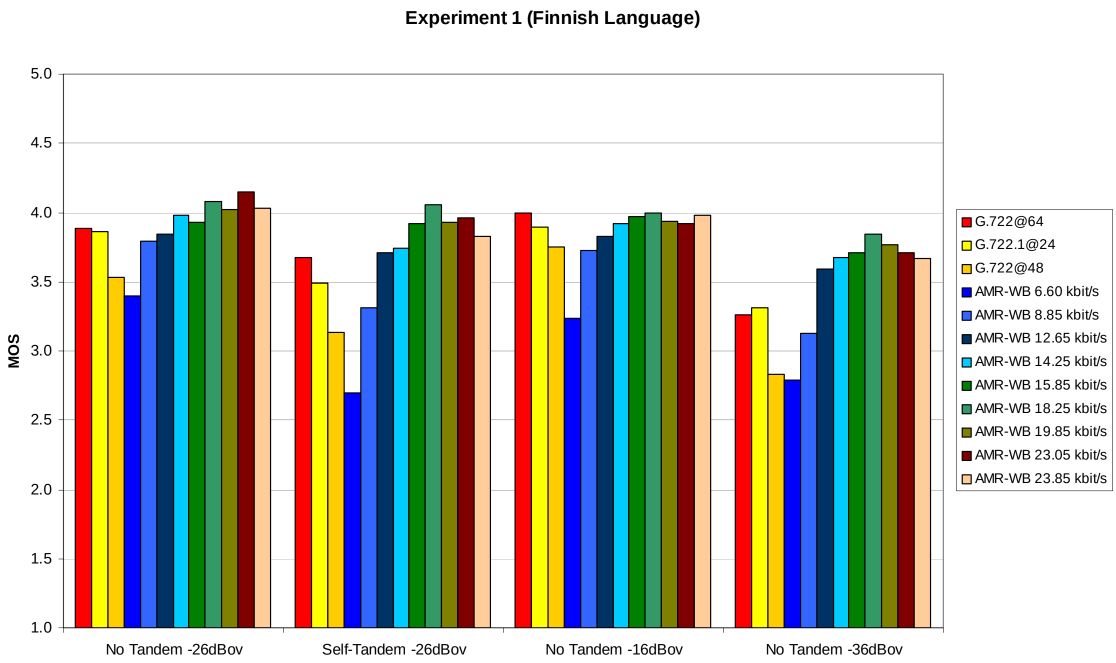


Figure 7.1: Experiment 1, testing Tandeming and input levels with Finnish language[2]

[2] *(The figures do not contain confidence intervals but they are planned to be added to the later versions.)*

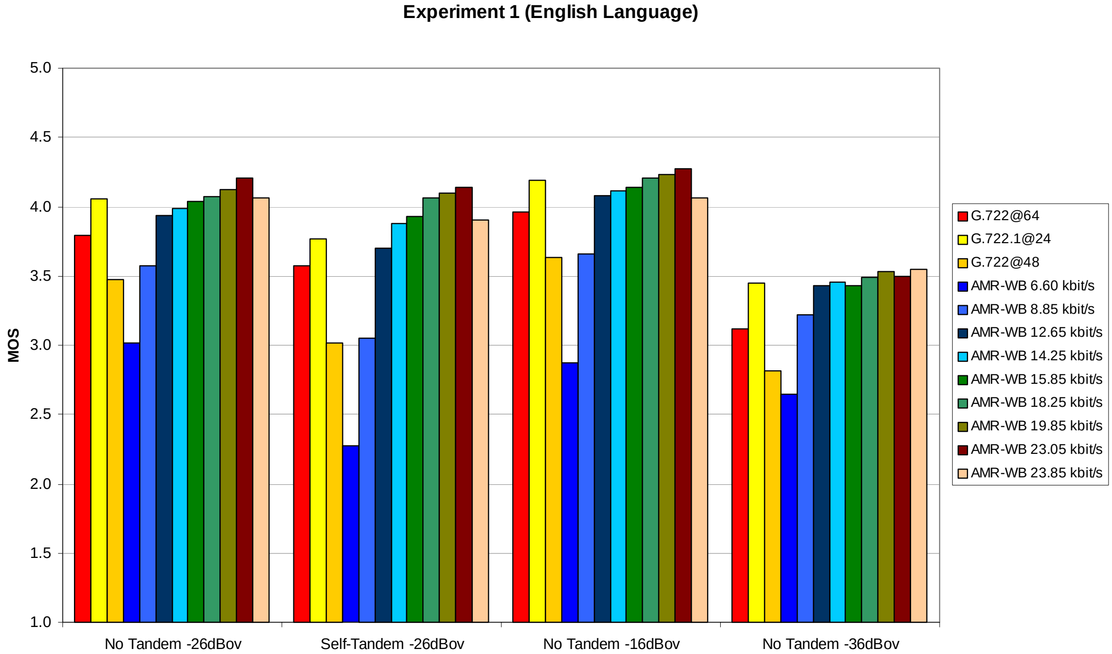


Figure 7.2: Experiment 1, testing tandeming and input levels with English language

8 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 the table below:

	Naming in the Figure 8.1
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).

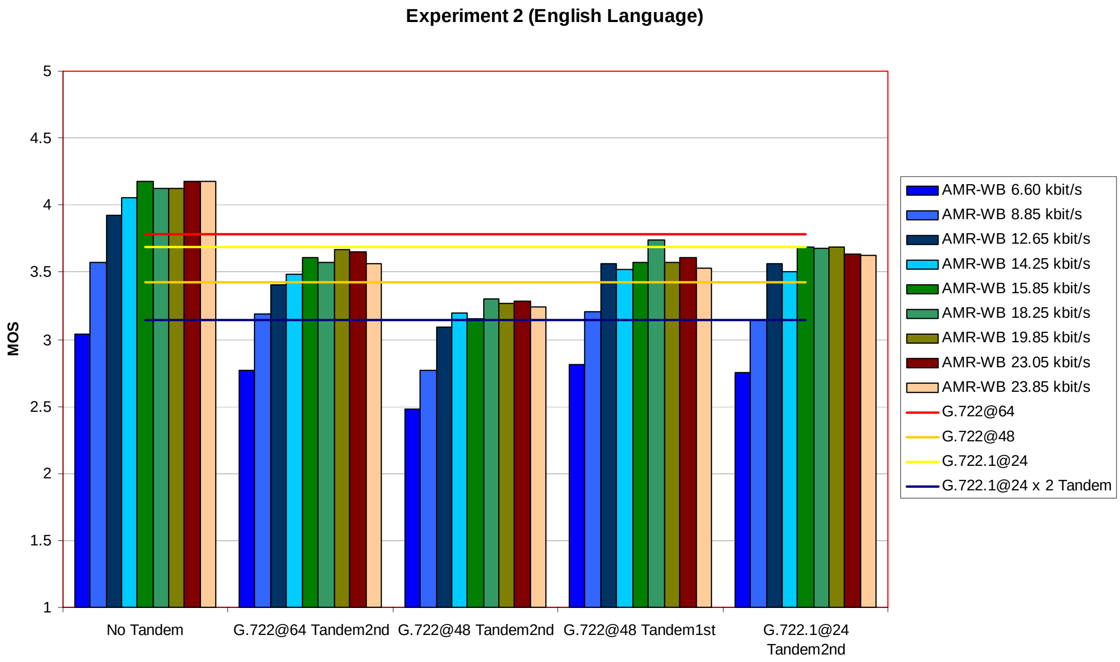


Figure 8.1: Experiment 2, testing tandeming with other standards with English language

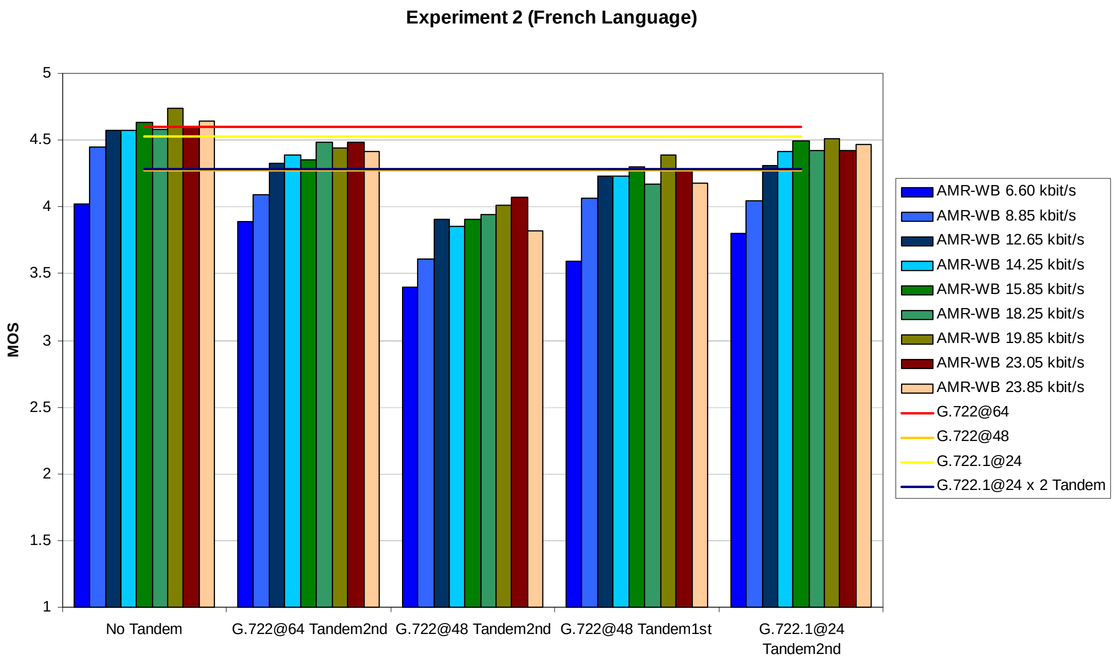


Figure 8.2: Experiment 2, testing tandeming with other standards with French language

9 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 G.729. English language was used in testing. All nine AMR-WB modes were tested with the following tandeming scenarios shown in the table below:

	Naming in the Figure 9.1
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 9.1, 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 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.

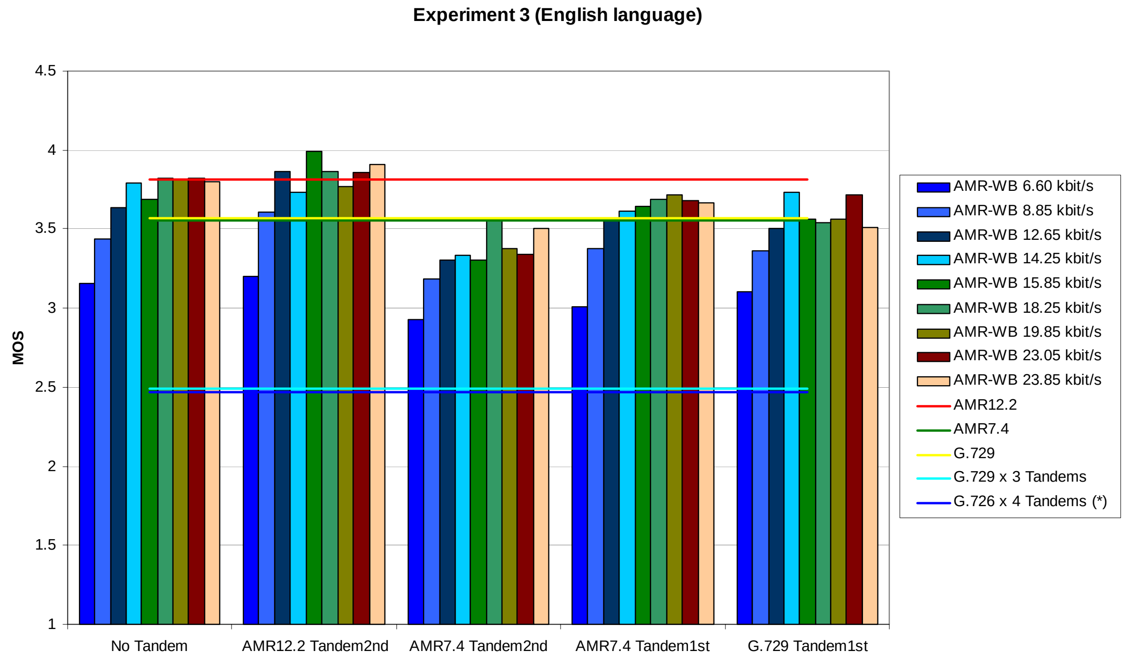


Figure 9.1: Experiment 3, testing tandeming with narrowband standards with English language

10 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 and 18.25 kbit/s. Both modes were tested with and without errors. ETSI GSM FR error profiles were used The following table describes the conditions in which the codec were tested 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 10.1, 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.

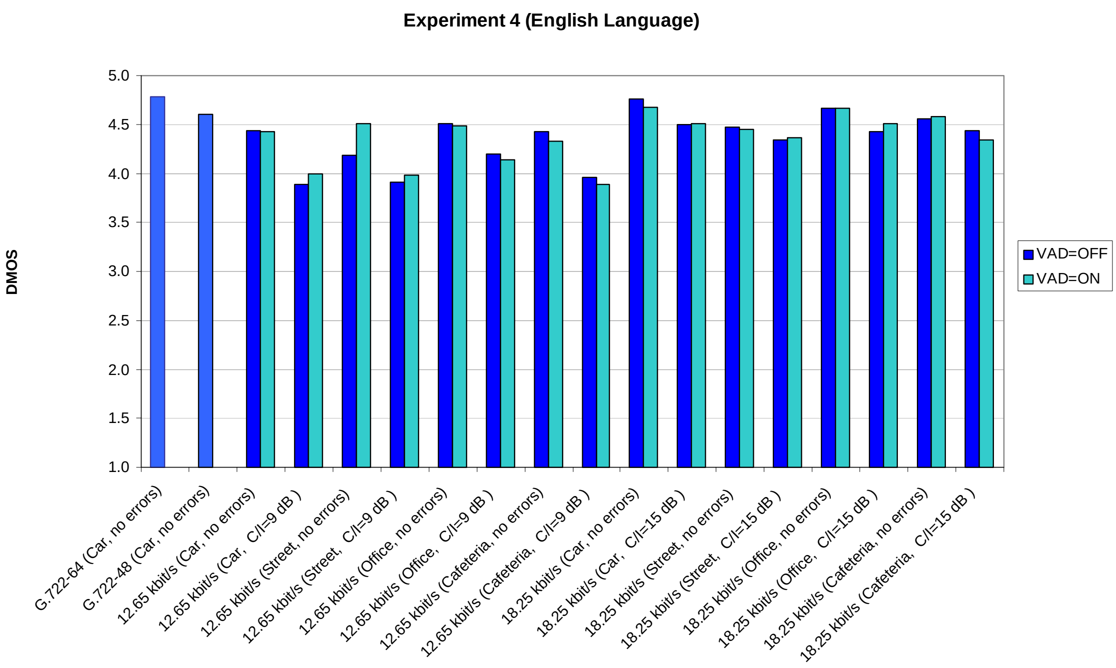


Figure 10.1: Experiment 4, testing VAD/DTX with English language

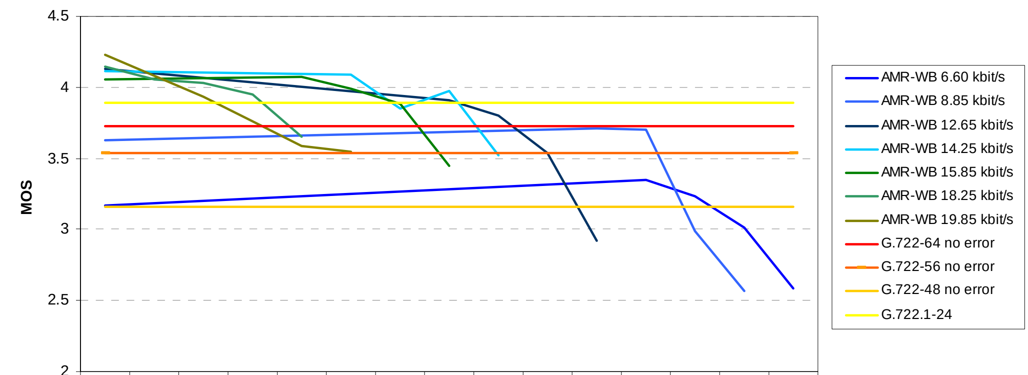
11 Performance in Static Errors under Clean Speech Conditions in GSM GMSK

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.

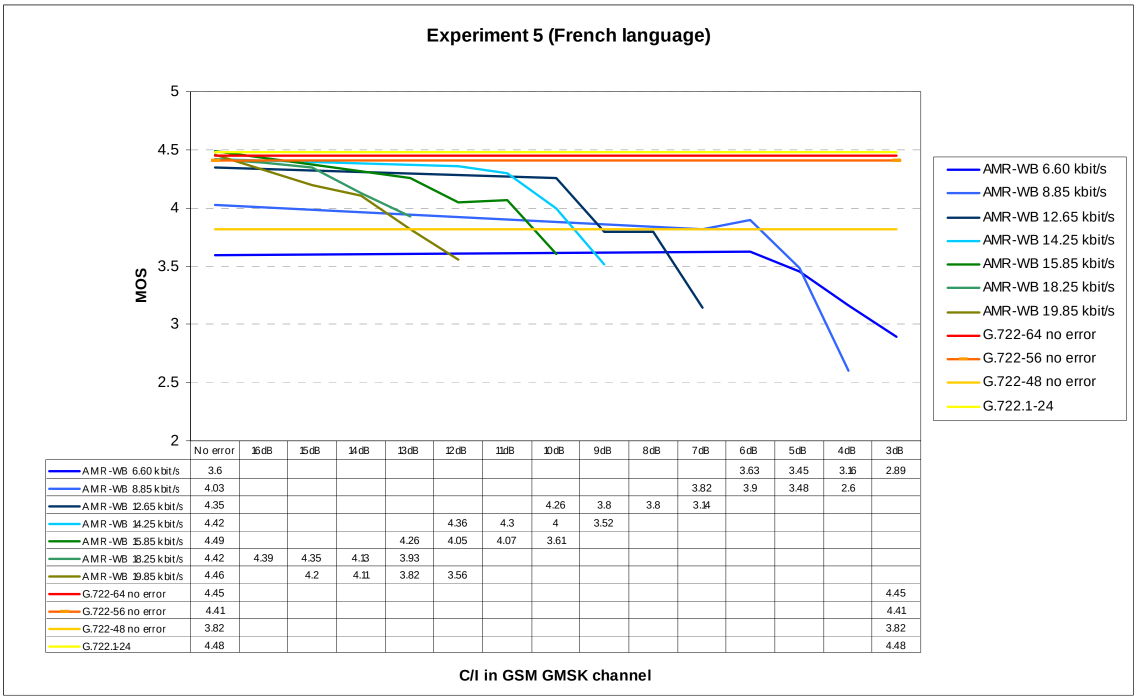
From both figures it can be seen that the quality of at least G.722 at 56 kbit/s can be achieved at about 10 dB C/I and above. The quality better or equal of at least G.722 at 64 kbit/s can be achieved at about 11 dB C/I and above.



	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				
AMR-WB 14.25 kbit/s	4.11					4.09	3.85	3.97	3.52						
AMR-WB 15.85 kbit/s	4.06				4.07	3.99	3.88	3.45							
AMR-WB 18.25 kbit/s	4.15	4.06	4.03	3.95	3.65										
AMR-WB 19.85 kbit/s	4.23		3.93	3.76	3.59	3.55									
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

C/I in GSM GMSK channel

Figure 11.1: 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				
AMR-WB 14.25 kbit/s	4.42					4.36	4.3	4	3.52						
AMR-WB 15.85 kbit/s	4.49				4.26	4.05	4.07	3.61							
AMR-WB 18.25 kbit/s	4.42	4.39	4.35	4.13	3.93										
AMR-WB 19.85 kbit/s	4.46		4.2	4.11	3.82	3.56									
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 11.2: Experiment 5, testing GSM FR channel with French language

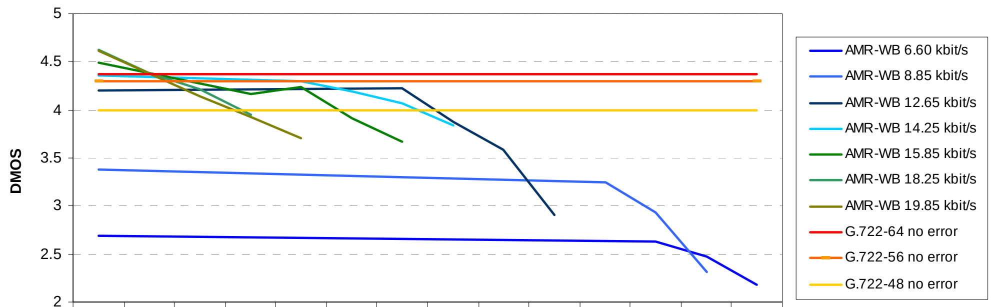
12 Performance in Background Noise in Static C/I Conditions in GSM GMSK

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 the table below:

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.

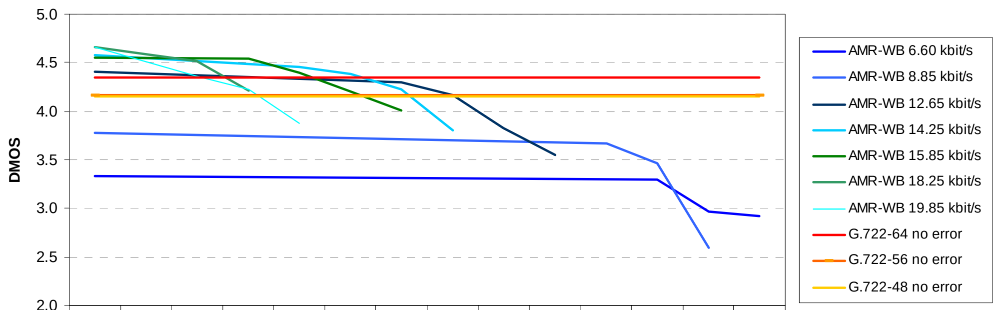
It seems, that both experiments give very similar results about the performance of the different AMR-WB modes in the presence of background noise. From both figures it can be seen that the quality of G.722 at 56 kbit/s can be achieved in C/I-ratios 10 dB and above. The quality better or equal to G.722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



	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				
AMR-WB 14.25 kbit/s	4.354				4.302	4.188	4.063	3.833						
AMR-WB 15.85 kbit/s	4.49			4.167	4.24	3.917	3.667							
AMR-WB 18.25 kbit/s	4.625	4.385	4.219	3.948										
AMR-WB 19.85 kbit/s	4.615		4.146	3.927	3.708									
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

C/I in GSM GMSK channel

Figure 12.1: 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				
AMR-WB 14.25 kbit/s	4.57				4.46	4.39	4.23	3.80						
AMR-WB 15.85 kbit/s	4.55			4.54	4.40	4.20	4.01							
AMR-WB 18.25 kbit/s	4.67	4.58	4.52	4.21										
AMR-WB 19.85 kbit/s	4.67		4.35	4.23	3.88									
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

C/I in GSM GMSK channel

Figure 12.2: Experiment 6b, testing GSM FR channel with Finnish language

13 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.

NOTE: This experiment will be conducted in the phase Ib of the testing when the relevant 3G error patterns will be available.

14 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 the table below:

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

NOTE: This experiment will be conducted in the phase Ib of the testing when the relevant 3G error patterns will be available.

15 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 15.1. The experiments were limited to error free conditions only.

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, 56 and 64 kbit/s, and the A-law codecs alone (direct condition). The DTMF signals were generated at the frequencies specified in ITU-T Rec. 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 Rec. Q.23. The duration of the individual DTMF digits was 80ms, with a 80ms 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 (20ms 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 15.1: Experimental conditions

Experiment	Low tone level (1)	High tone level (1)	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 1: The levels are given as measured at the input to the DTMF detector, however, since the DAC is calibrated according to ITU-T Rec. G.711, 0dBm in the analogue section is equivalent to -6.15dBov in the digital section.

The percentage of undetected digits measured for each codec mode in each experiment is given in Table 15.2. 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 15.2: 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 6dB of negative twist. It is noted that DTMF signals are often generated by PSTN telephones with negative twist, e.g. -2dB, to account for the characteristics of the local loop.

16 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 were evaluated by expert listening tests, whereas others studied the stability of the AMR-WB codec objectively using long speech and random files. 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

16.1 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 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.

16.2 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 +32767 and –32768 and the length of both files was 8s. 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 0.5s zero signal period:
2) Signal B: 2.0s random noise bursts separated by 1.0s zero signal period:

Time domain plots for the bursty random noise signals A & B is given in Figure 16.1.

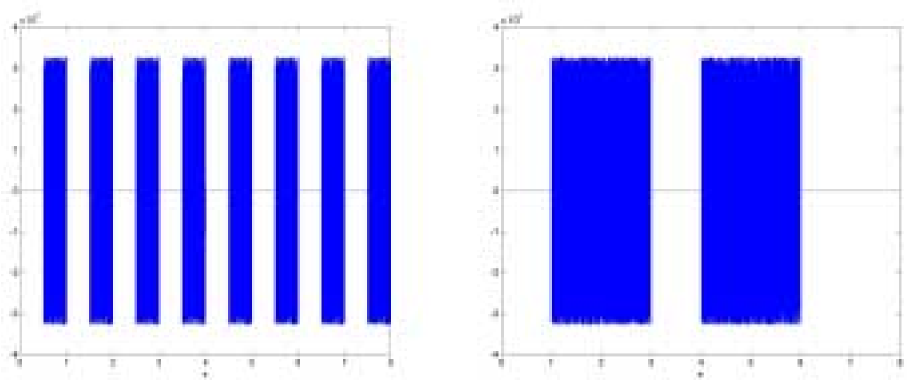


Figure 16.1: Time domain plots for the bursty random noise signals A&B respectively

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

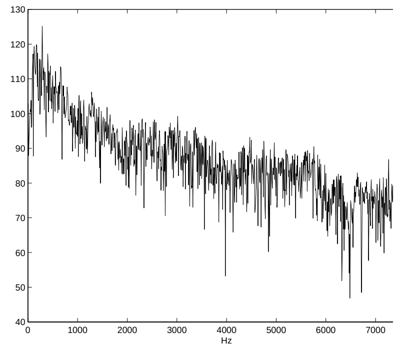
16.3 Background noise signals

Each mode was tested with many types of background noise signals. The noise types and their lengths are given in Table 16.1.

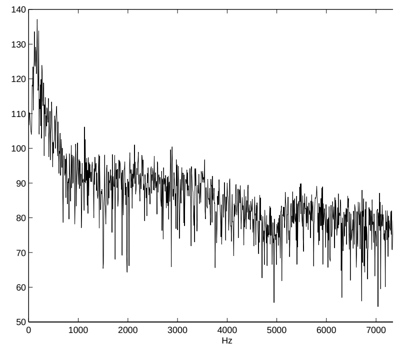
Table 16.1

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**

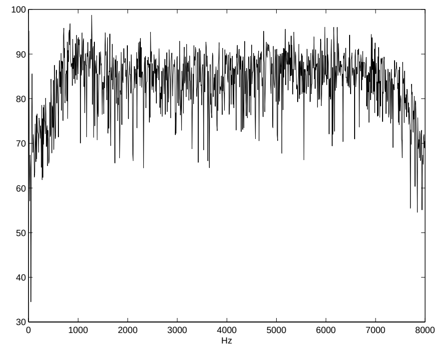
The frequency spectrum figures of the used noise signals are given in Figure 16.2. As a result, all the background noises coded with all the modes sounded normal and were recognised and no annoying artifacts were generated.



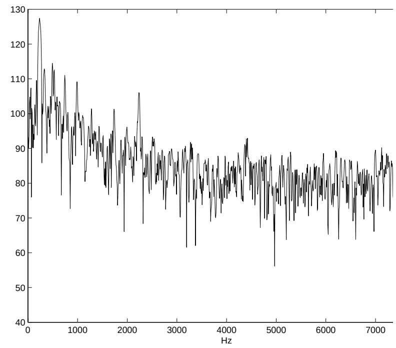
a) Frequency spectrum of the "car" noise



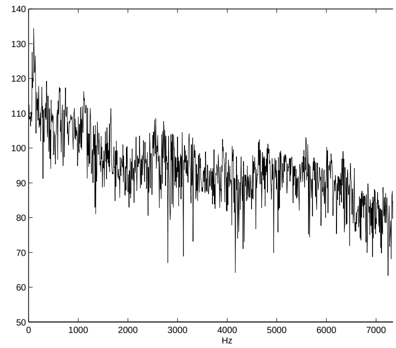
b) Frequency spectrum of the "cafeteria" noise



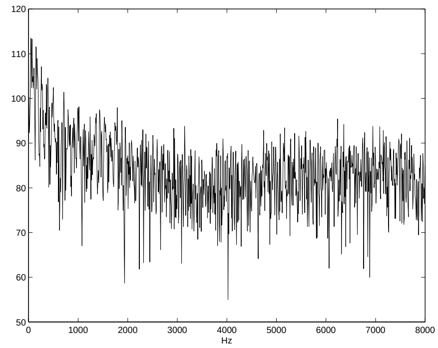
g) Frequency spectrum of the "rain" noise



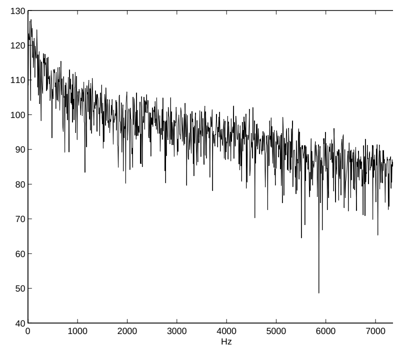
c) Frequency spectrum of the "Hoth" noise



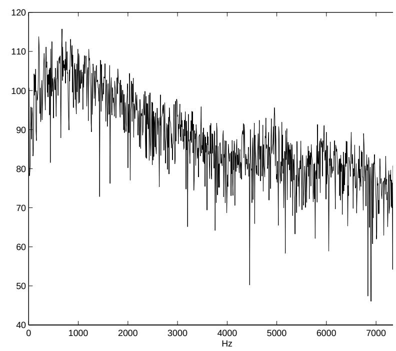
d) Frequency spectrum of the "motorbike" noise



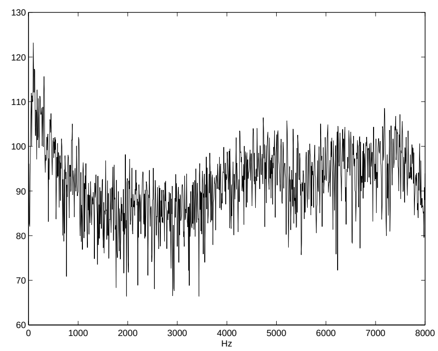
i) Frequency spectrum of the "wind" noise



e) Frequency spectrum of the "motorboat" noise



f) Frequency spectrum of the "railway station" noise



h) Frequency spectrum of the "thunder" noise

Figure 16.2. Frequency spectrums of the background noise files

16.4 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 0.5s 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 16.2 below.

Table 16.2: Frequency contents of different sinusoidal wave test signals

Test signal /	**Frequency [Hz]**									
(test type)	**300**	**500**	**700**	**1000**	**1500**	**2000**	**4000**	**5000**	**6000**	**8000**
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 6300 to 7000 Hz became noise-like because of artificial high band generation.

16.5 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 0.5s of all zero signal (test signal: 19). The length of the signals was about 8s. The frequency contents of different square test signals are given in Table 16.3 below.

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 16.3: Frequency contents of different square wave test signals

Test signal /	**Frequency [Hz]**									
(test type)	**300**	**500**	**700**	**1000**	**1500**	**2000**	**4000**	**5000**	**6000**	**8000**
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**								

Test signal / (test type)	Frequency [Hz] 300	500	700	1000	1500	2000	4000	5000	6000	8000
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	

16.6 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.

16.7 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 (2h 53min) 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.

16.8 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 chapter 16.4 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 seconds, two consecutive bad frames after 4 seconds and three consecutive bad frames after 6 seconds. The results were acceptable. (For one single sinusoidal tone of frequency 1500 Hz., temporary instability in the decoder.was observed).

16.9 Summary

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

17 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.

Figure 17.1 shows processing flow to prepare test files. The input level for AMR-WB coder was adjusted with '*sv56*' algorithm provided in the ITU-T 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.

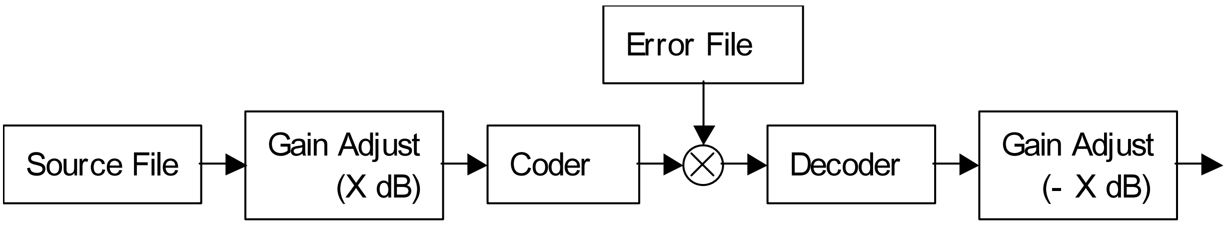


Figure 17.1: Test processing for overload performance

The processed files were up-sampled from 16kHz to 48kHz 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-seconds 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 26dB below overload. The gain was adjusted to X=0, 10, 20 or 30dB 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.

18 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.

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 8-9 frames. The decrease is smooth.

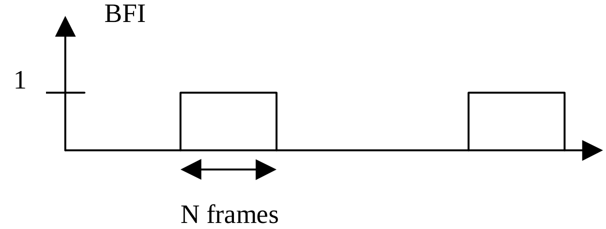


Figure 18.1: Test setup for test 1

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.

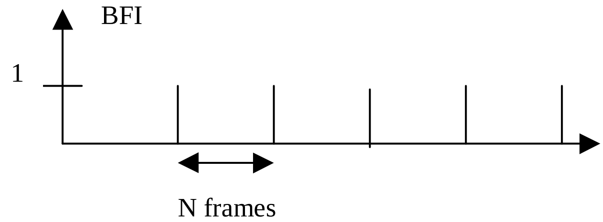


Figure 18.2: Test setup for test 2

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.

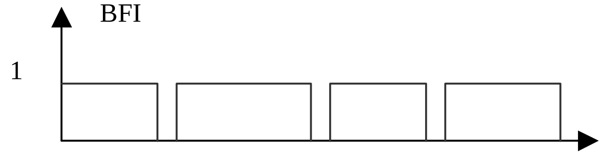


Figure 18.3: test setup for test 3

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.

19 Language Dependency

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

English (US & UK), Finnish, French, German, Japanese, Spanish

The results reported by the different laboratories were consistent.. Tests specially designed for language dependency testing, were not conducted.

20 Transmission Delay

The transmission delay of a GSM communication using AMR-WB has been evaluated using the same method as for the previous GSM speech codecs. The reference system delay distribution for the downlink and uplink directions are provided in Figure 20.1 and Figure 20.2 respectively. The speech transcoders are assumed to be remote located from the BTS (16 kbit/s or 32 kbit/s sub-multiplexing on the Abis & Ater Interfaces).

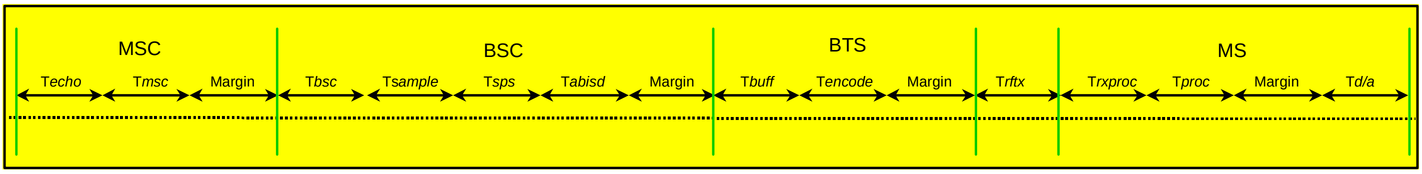


Figure 20.1: Reference Downlink delay distribution

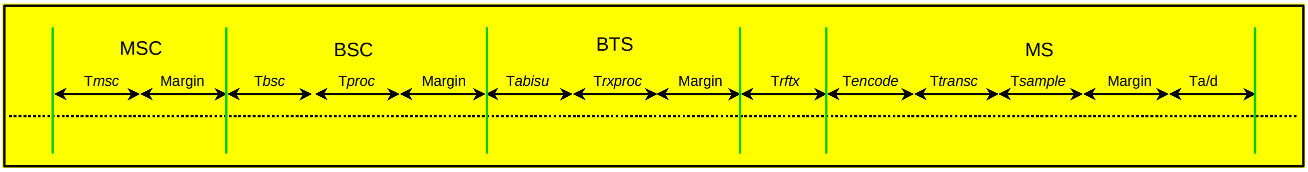


Figure 20.2: Reference Uplink delay distribution

The definition of the different delay parameters is given in the following table. The table also provides the value used for the parameter when not dependent of the type of speech codec or sub-multiplexing scheme over the Abis & Ater interfaces.

T*abisd* Time required to transmit the minimum number of speech data bits over the downlink Abis interface to start encoding a radio speech frame. Depends on the speech codec mode, the TRAU frame format and the Abis/Ater sub-multiplexing scheme. Note that most TRAU frame synchronization bits can ideally be transmitted by anticipation and are usually not included in this parameter.

T*abisu* Time required to transmit the minimum number of speech data bits over the uplink Abis interface to start decoding a speech frame. Depends on the speech codec mode, the TRAU frame format and the Abis/Ater sub-multiplexing scheme. Note that the TRAU frame synchronization bits can ideally be transmitted by anticipation and are usually not included in this parameter.

T*a/d* Delay in the analogue to digital converter in the uplink (implementation dependent). Set to 1ms.

T*bsc* Switching delay in the BSC (implementation dependent). Set to 0.5ms.

T*buff* Buffering time required for the time alignment procedure for the in-band control of the remote transcoder. Set to 1.25 ms.

T*d/a* Delay in the digital to analogue converter in the downlink (implementation dependent). Set to 1ms.

T*echo* Delay induced by the echo canceller (implementation dependent). Set to 1ms.

T*encode*: Processing delay required to perform the channel encoding (implementation dependent). Depends on the channel coding complexity of each codec mode.

T*msc* Switching delay in the MSC (implementation dependent). Set to 0.5ms.

T*proc* Processing delay required to perform the speech decoding (implementation dependent). Depends on the speech decoding complexity of each codec mode.

T*rftx* Time required for the transmission of a speech frame over the air interface. Derived from the radio framing structure and the interleaving scheme. Worst case is 37.5 ms in Full Rate mode.

T*rxproc* Processing delay required to perform the channel equalization, the channel decoding and SID-frame detection (implementation dependent). The channel decoding depends on the codec mode. The channel equalization part was set to 6.84 ms in Full Rate mode.

T*sample* Duration of the segment of PCM speech samples operated on by the speech transcoder: 25 ms in all cases corresponding to 20 ms for the processed speech frame and 5 ms of look ahead.

T*sps* Worst case processing delay required by the downlink speech encoder before an encoded bit can be sent over the Ater/Abis interface taking into account the speed on the Ater/Abis interface (implementation dependent). Depends on the speech coding complexity of each codec mode and on the sub-multiplexing rate on the Ater/Abis interface. Because of the priority given to the decoding, T*proc* is also added to the overall downlink transmission delay.

T*transc* MS speech encoder processing delay, from input of the last PCM sample to output of the final encoded bit (implementation dependent). For the evaluation of the transmission delay, it was assumed that the speech decoding has a higher priority than the speech encoding, i.e. this delay is artificially increased by the speech decoding delay.

Margin Implementation dependent margins in the different system components. Set as follows:
MSC Margin: 0.5 ms
BSC Margin: 0.5 ms
BTS Margin: 0.45 ms downlink, 0.3 ms uplink
MS Margin: 2 ms in Full Rate.

The processing delays were estimated using complexity figures for each codec mode. In addition, to take into account the dependence on the DSP implementation, the computation was based on the same methodology used for the previous GSM speech codecs.

The DSPs running the speech and channel codec are modeled with the 3 following parameters:

E represents the DSP Efficiency. This corresponds to the ratio tMOPS/wMOPS of the codec implementation on the DSP.

S represents for the speed of the DSP: Maximum Number of Operations that the DSP can run in 1 second. This number is expressed in MOPS.

P represents the percentage of DSP processing power assigned to the codec.

The processing delay of a task of complexity X (in wMOPS) can then be computed using the equation:

$$D = \frac{20X}{ESP} \text{ ms}$$

[To be completed]

.

.

.

21 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, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 & 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 chapters.

In the first method, tones signals have been generated in the range 10Hz – 7010 Hz with a frequency step of 20 Hz. Each tone had a duration of 10 seconds. The frequency response of the AMR codec has been evaluated by computing the logarithmic gain according to the following equation:

Logarithmic gain measure :

$$\mathrm{Gain}_{\mathrm{dB}} = 10\log_{10}\left[\sum_{k=1}^{M} \mathrm{out}(k)^2 / \sum_{k=1}^{M} \mathrm{inp}(k)^2\right]$$

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 21.1, Figure 21.2 and Figure 21.3. Figure 21.1 gives the results of the 1st method. Figure 21.2 and Figure 21.3 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 7kHz bandwidth. The 2 lowest modes are extremely limited and the 6 other modes present a bandwidth of 50 HZ – 5700 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 and 23.05 kbit/s present a bandwidth of 50 Hz – 6400 Hz. The highest bitrate has a bandwidth of 50 Hz – 6600 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 – 6000 Hz. The highest bitrate has a bandwidth of 50 Hz – 6600 Hz

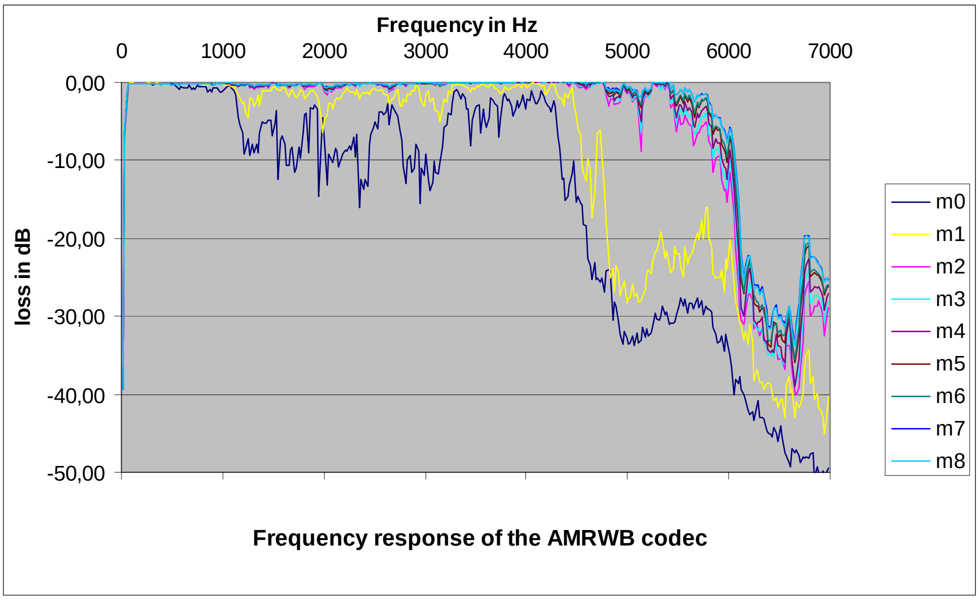


Figure 21.1: Frequency response of the AMR-WB codec (1st method)

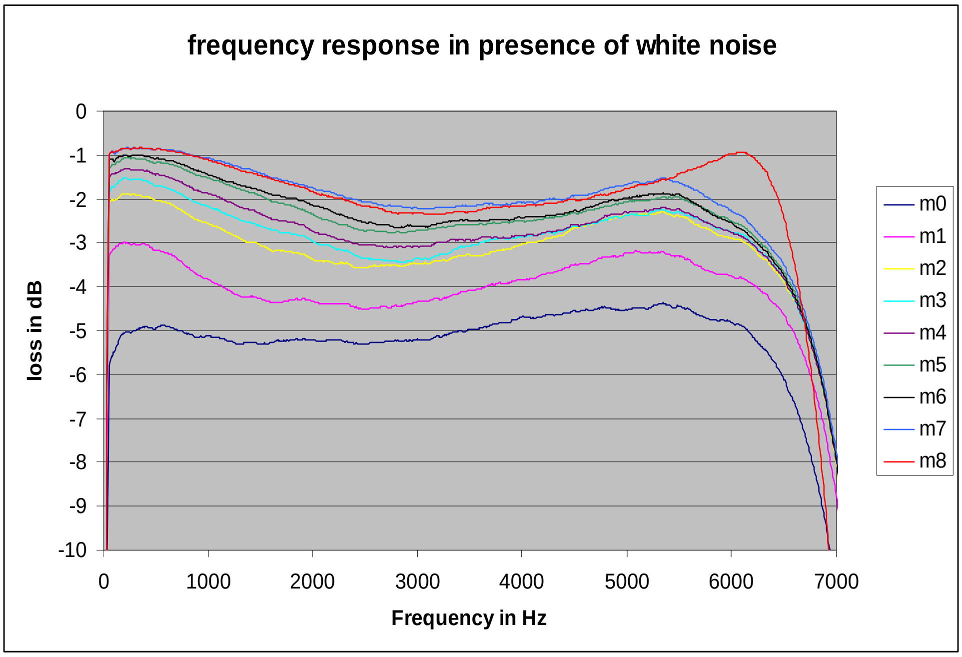


Figure 21.2: Frequency response of the AMR-WB codec (2nd method)

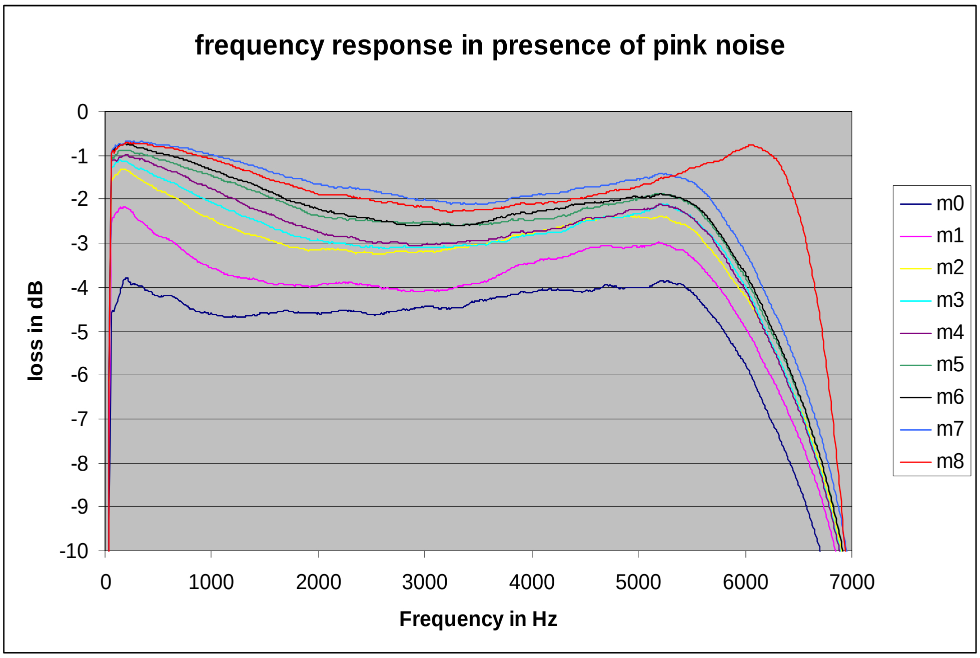


Figure 21.3: Frequency response of the AMR-WB codec (2nd method)

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.

22 Complexity Analysis

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

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 22.1 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 22.2 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 22.3, Table 22.4 and Table 22.5 provide the RAM, ROM Tables and Program ROM complexity figures for the speech and channel codecs.

Table 22.1

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 22.2

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 22.3

Data RAM (static + scratch)			
	static + scratch requirement	**static used**	**scratch used**
Speech Encoder + VAD+DTX	15000 + 149 Words	1381 Words	4389 Words
Speech Decoder + DTX		758 Words	
Channel Encoder (TCH/WFS)	3000 Words	229 Words	
Channel Decoder (TCH/WFS)		242 Words	
Link Adaptation		102 Words	
Total		**2712 Words**	**4389 Words**
		7101 Words	

Table 22.4

Data ROM Tables		
	requirement	**used**
Speech Codec + VAD + DTX	18000 + 513 Words	9929 Words
Channel Codec (TCH/WFS)	4500 Words	3075 Words
Link Adaptation	-	105 Words
Total	**23013 Words**	**13109 Words**

Table 22.5

Program ROM		
	requirement	**used**
Speech Codec + VAD + DTX	5821 + 491	3889 basic-ops
Channel Codec (TCH/WFS)	2013	418 basic-ops
Link Adaptation	-	48 basic-ops
Common (log2, oper32b)	-	35 basic-ops
Total	**8571 basic-ops**	**4390 basic-ops**

23 Comfort Noise Generation

This chapter 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.

23.1 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 –16dBov and –36dBov, 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 23.1.

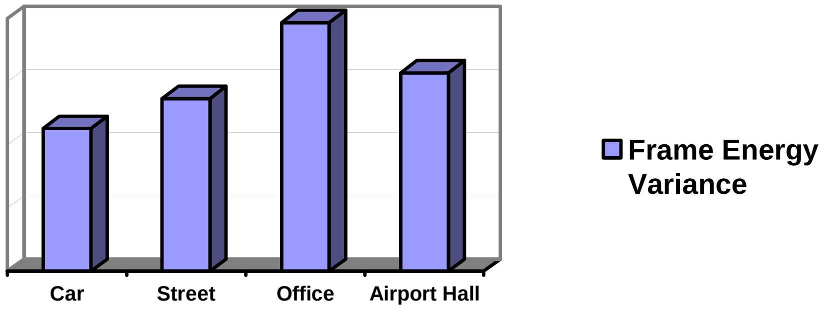


Figure 23.1 Stationarity of noises

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.

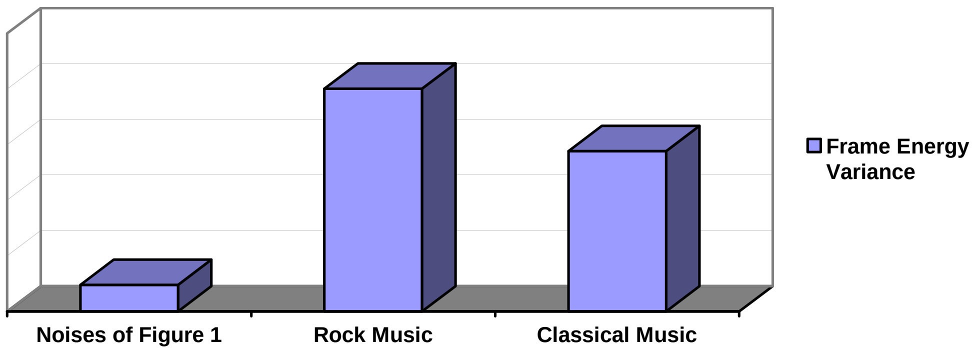


Figure 23.2 Stationarity of music files

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

23.2 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 23.3, Figure 23.4, Figure 23.5 and Figure 23.6.

In Figure 23.3 and Figure 23.4 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.

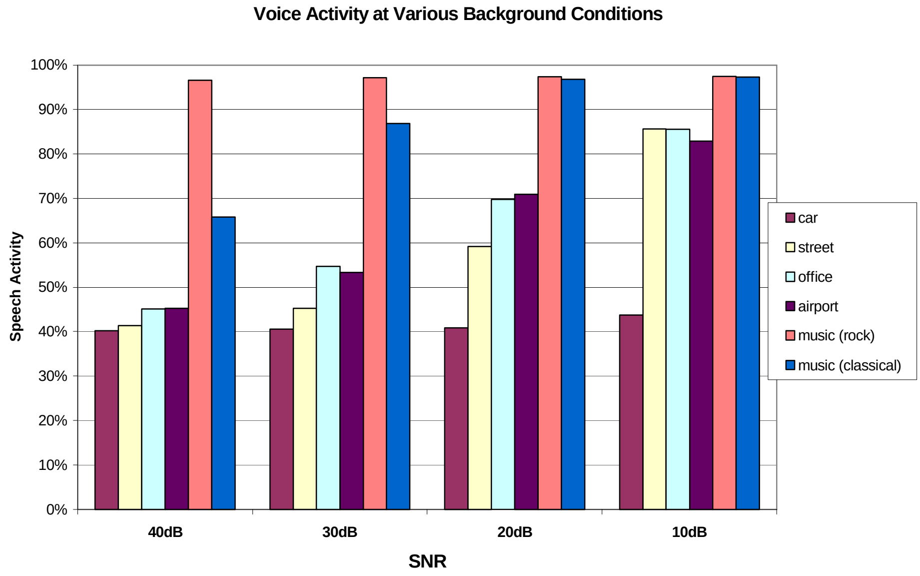


Figure 23.3 Voice activity for different background conditions, at speech level –26dBov. (Voice activity for clean speech is 40%)

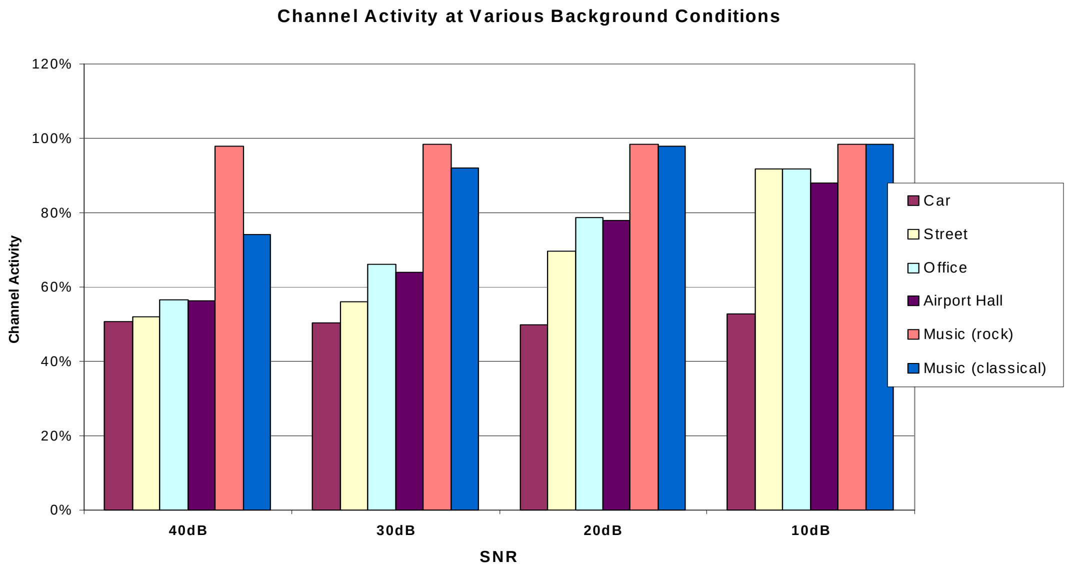


Figure 23.4 Channel Activity for different background conditions, input speech level = –26dBov (for clean speech; channel activity = 51 %)

Figure 23.5 and Figure 23.6 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.

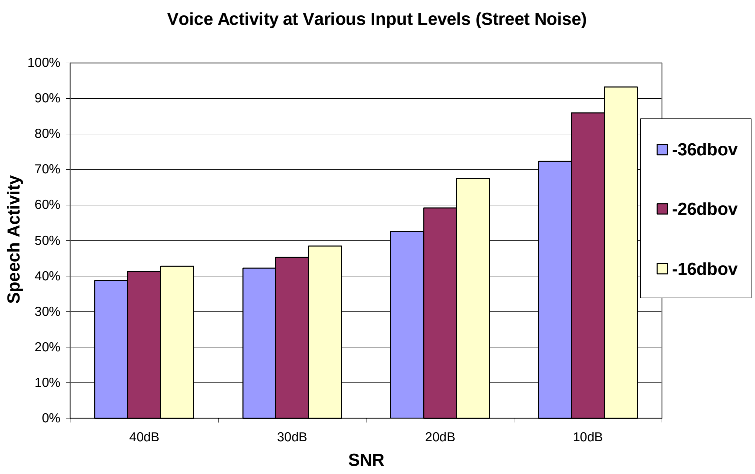


Figure 23.5 Voice Activity for different input levels (street noise)

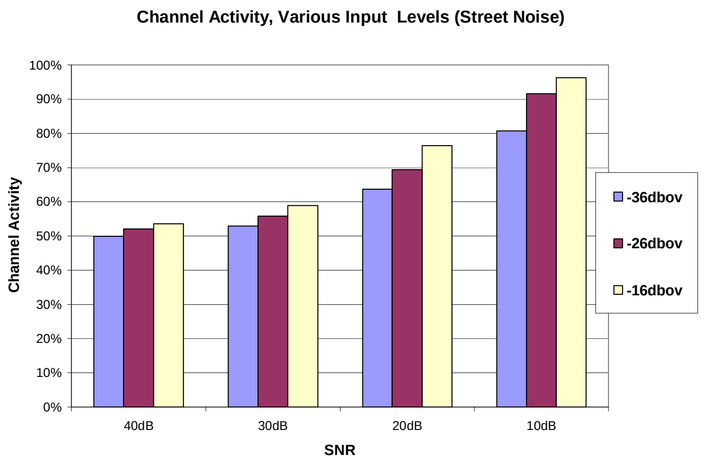


Figure 23.6 Channel Activity at different input levels (street noise)

23.3 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 23.7. Most clippings according to the measure applied are found for car background noise.

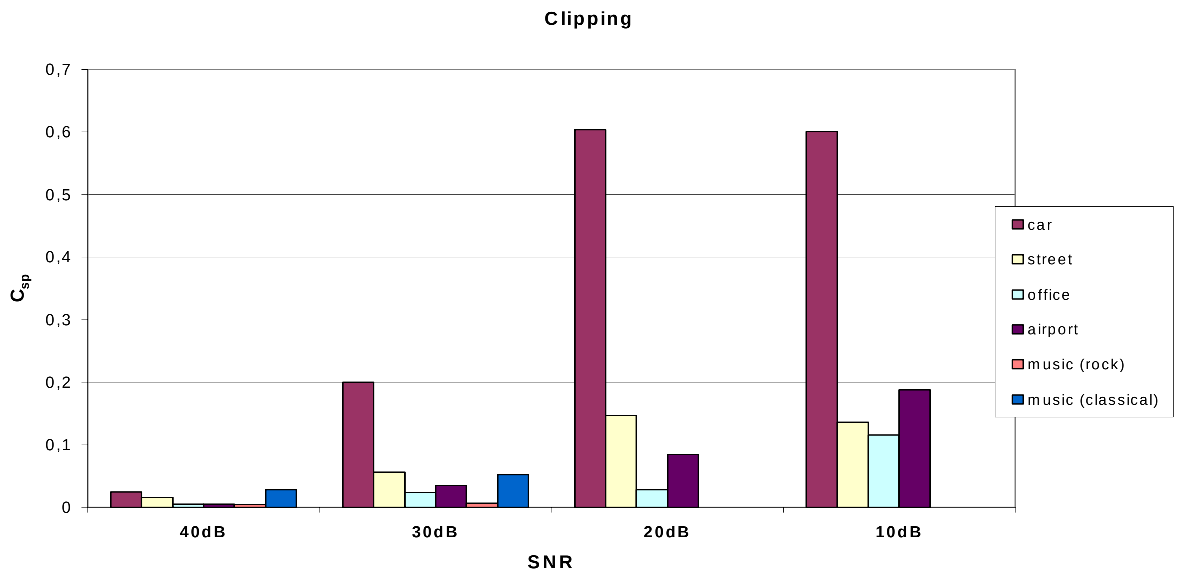


Figure 23.7: Clipping for different background conditions (clean case C_{sp} = 0.006)

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 10dB 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.

23.4 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 to 7000 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 paragraph 3.1 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.

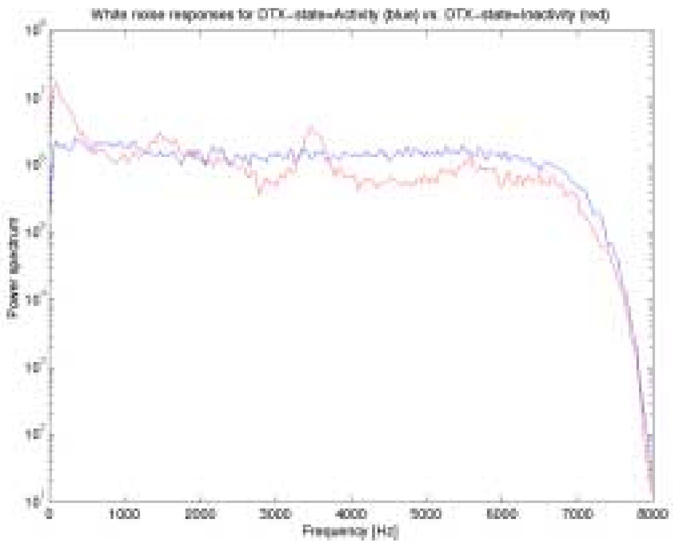


Figure 23.8: White noise responses for DTX-state=Activity (blue) and DTX-state=Inactivity (red)

23.5 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.

24 Performance with music signals (informal expert listening)

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

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

The following table lists the conditions that have been processed for each of the four long files:

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 9th symphony, Beatles), there are audible degradations, which led to the conclusion that G.722 is sometimes equivalent, sometimes slightly preferred to the WB-AMR 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 WB-AMR 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 WB-AMR 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.

25 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.

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 below, which contains the absolute number of choices (8 listeners).

	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.

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 1)		(see Note 1)	

Note 1: 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 13dB to 10dB, 13dB to 7dB, and 13dB to 4dB.

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

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[1].)

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 1 (in Table 3a)		See Note 1(in Table 1a)	

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

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.

---

1 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").

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 2a and for Application E in Table A.2b.

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

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.

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

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.

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

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

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

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

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

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 in Annex TBD. 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 as shown in Annex TBD. 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 Figures 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.

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

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

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[2] [10]:

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

1a: Applications A and B

2 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.)

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.

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

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

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.

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

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

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 mapping to particular candidates is:

- Codec 1 = Ericsson
- Codec 2 = FDNS consortium (consisting of France Télécom, Deutsche Telekom, Nortel Networks and Siemens)
- Codec 3 = Nokia
- Codec 4 = Motorola
- Codec 5 = Texas Instruments

During the selection process, Codec 4 was withdrawn.

The following sub-sections 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.

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

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

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

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-A.8c. The best FoM for each case is highlighted in the tables with a boldface font.

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

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

Table A.8b: FoM results for systematic failures.

Rule 3 FoM restricted to failures	Weighted ∆MOS					Weighted ∆dBQ				
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

Table A.8c: FoM results for weighted ∆MOS and weighted ∆dBQ when restricted to failures.

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).

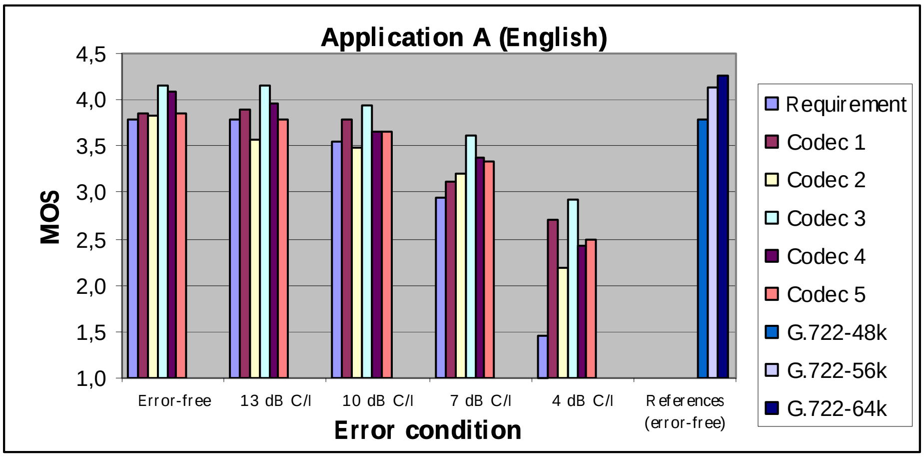
A.6 Key Selection Phase Documents in 3GPP FTP-site

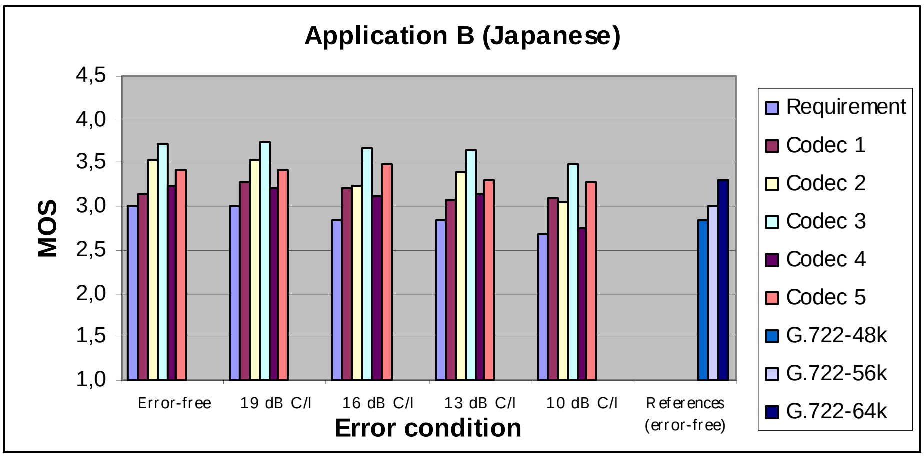
The standardisation of the WB-AMR codec is described in a series of permanent project documents. They contain the most important quidelines, rules and decisions. The following permanent project documents can be found in a specific location on the 3GPP FTP site:

ftp://ftp.3gpp.org/TSG_SA/WG4_CODEC/AMR-Wideband/Perm_Docs_Selection_Phase/

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





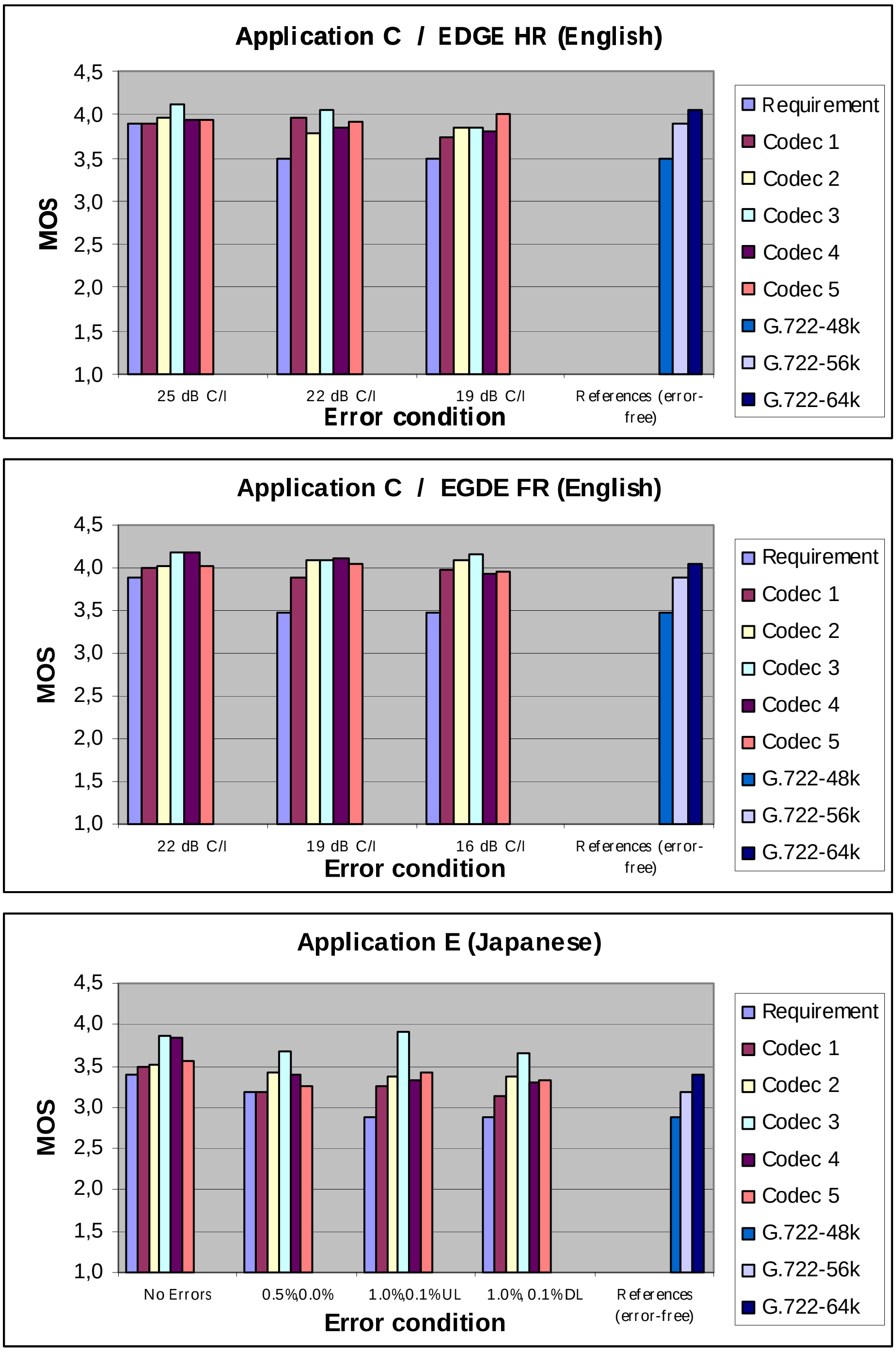


Fig 1: Experiment 2: Clean Speech performance with static errors (ACR)

a) Application A (English)

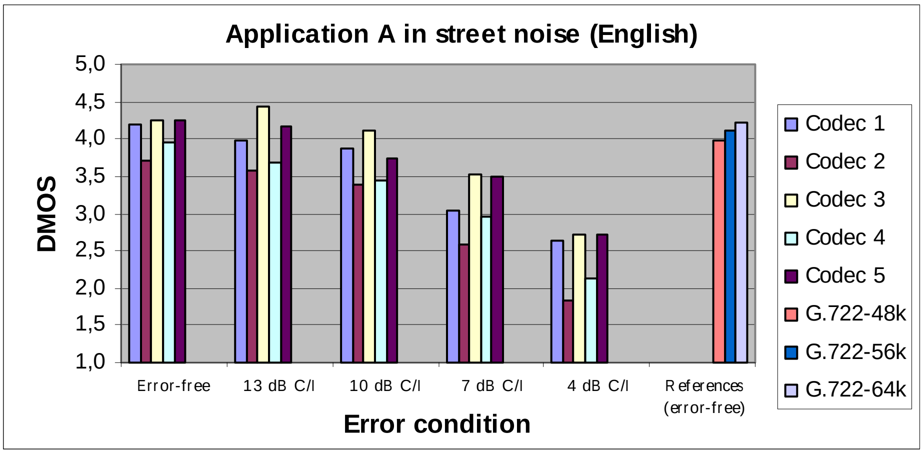
b) Application B (Japanese)

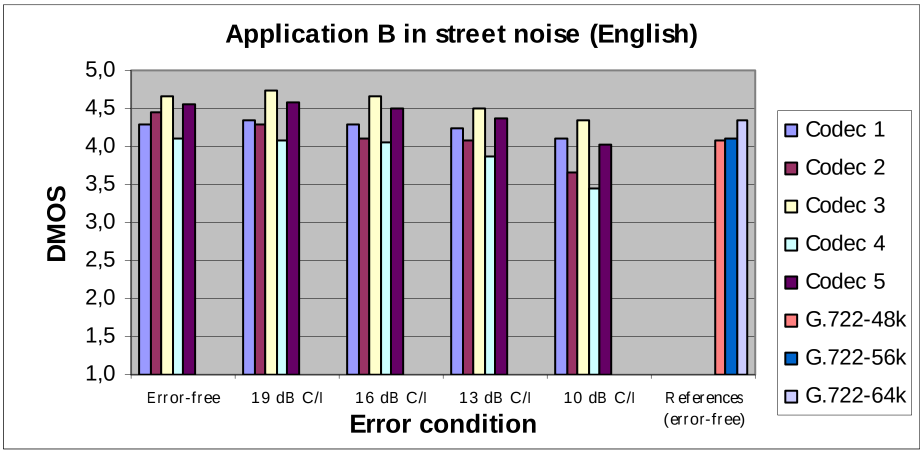
c) Application C / EDGE HR (English)

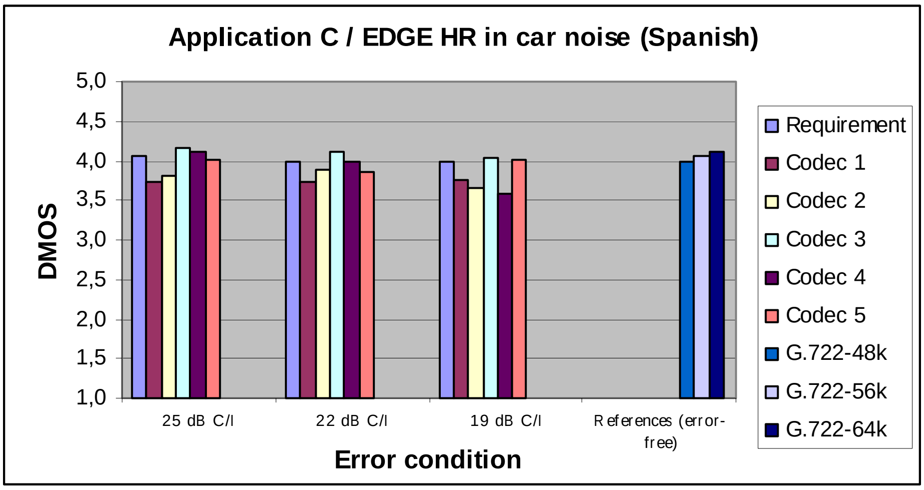
d) Application C / EDGE FR (English)

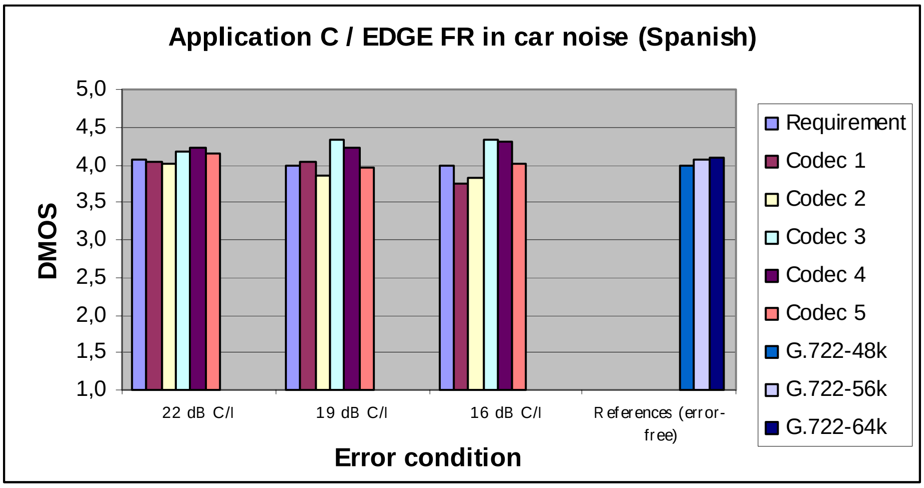
e) Application E (Japanese)

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


Application A in street noise (English)
DMOS
5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
Error-free
13 dB C/I
10 dB C/I
7 dB C/I
4 dB C/I
References (error-free)
Error condition
Codec 1
Codec 2
Codec 3
Codec 4
Codec 5
G.722-48k
G.722-56k
G.722-64k


Application B in street noise (English)
DMOS
5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
Error-free
19 dB C/I
16 dB C/I
13 dB C/I
10 dB C/I
References (error-free)
Error condition
Codec 1
Codec 2
Codec 3
Codec 4
Codec 5
G.722-48k
G.722-56k
G.722-64k


Application C / EDGE HR in car noise (Spanish)
DMOS
5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
25 dB C/I
22 dB C/I
19 dB C/I
References (error-free)
Error condition
Requirement
Codec 1
Codec 2
Codec 3
Codec 4
Codec 5
G.722-48k
G.722-56k
G.722-64k



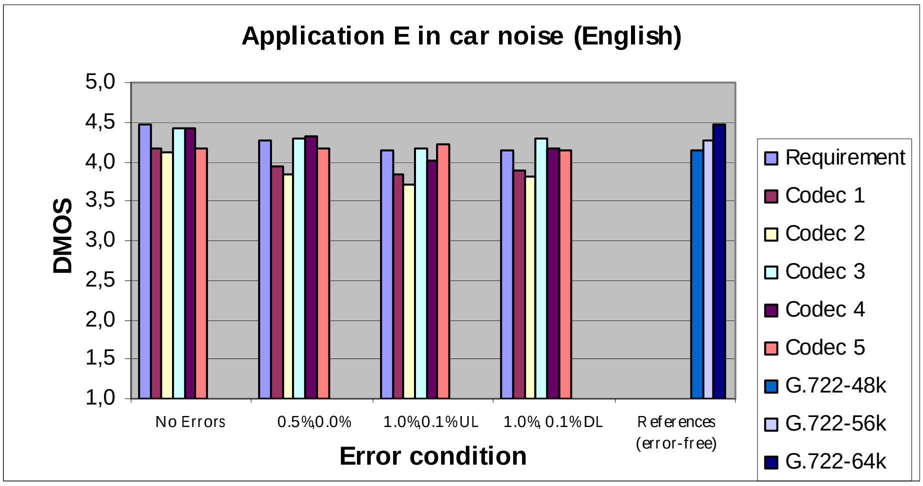


Fig 2: Experiment 3: Car and Street noise (15 dB SNR) performance for the GSM FR channel (DCR-test); and Experiment 4: Car and Street noise (15 dB SNR) performance for higher-rate channels (DCR-test)

a) 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.)

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 Annex gives estimates of the codec complexities (estimated by codec proponents)[3].The complexity was calculated as worst observed frame.

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.

[3] Codec 4 was withdrawn during the Selection Phase and no estimates for complexity were given for it.

Annex B: Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2001-06	12	SP-010302			Version 0.3.0 presented at TSG-SA#12 for information		0.3.0