ETSI TR 126 952 V13.1.0 (2016-04)

TECHNICAL REPORT

Universal Mobile Telecommunications System (UMTS);
LTE;
Codec for Enhanced Voice Services (EVS);
Performance characterization
(3GPP TR 26.952 version 13.1.0 Release 13)

Reference

RTR/TSGS-0426952vd10

Keywords

LTE,UMTS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Intellectual Property Rights

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

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

Foreword

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

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

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

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Foreword

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

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

Version x.y.z

where:

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

1 Scope

The present document provides information on the Enhanced Voice Services (EVS) codec Selection, Verification and Characterization Phases which were run using the fixed-point code (3GPP TS 26.442). Experimental test results from the subjective quality testing are reported to illustrate the behaviour of the EVS codec. Additional information is provided on implementation complexity of the EVS codec and objective test results. Also the verification results for the floating-point version of the EVS codec (3GPP TS 26.443) are presented.

2 References

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

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

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 26.441: "Codec for Enhanced Voice Services (EVS); General overview".

[3] 3GPP TS 26.442: "Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point)".

[4] 3GPP TS 26.443: "Codec for Enhanced Voice Services (EVS); ANSI C code (floating-point)".

[5] 3GPP TS 26.444: "Codec for Enhanced Voice Services (EVS); Test Sequences".

[6] 3GPP TS 26.445: "Codec for Enhanced Voice Services (EVS); Detailed algorithmic description".

[7] 3GPP TS 26.446: "Codec for Enhanced Voice Services (EVS); Adaptive Multi-Rate - Wideband (AMR-WB) backward compatible functions".

[8] 3GPP TS 26.447: "Codec for Enhanced Voice Services (EVS); Error concealment of lost packets".

[9] 3GPP TS 26.448: "Codec for Enhanced Voice Services (EVS); Jitter buffer management".

[10] 3GPP TS 26.449: "Codec for Enhanced Voice Services (EVS); Comfort Noise Generation (CNG) aspects".

[11] 3GPP TS 26.450: "Codec for Enhanced Voice Services (EVS); Discontinuous Transmission (DTX)".

[12] 3GPP TS 26.451: "Codec for Enhanced Voice Services (EVS); Voice Activity Detection (VAD)".

[13] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[14] 3GPP TS 26.131: "Terminal acoustic characteristics for telephony; Requirements".

[15] 3GPP SP-100202: "EVS Work Item Description".

[16] 3GPP TR 22.813: "Study of use cases and requirements for enhanced voice codecs for the Evolved Packet System (EPS) ".

[17] EVS-3 Permanent Document: "EVS Performance Requirements".

[18] EVS-4 Permanent Document: "EVS Design Constraints".

[19] EVS-5b Permanent Document: "EVS Selection Rules".

[20] EVS-6b Permanent Document: "EVS Selection Deliverables".

[21] EVS-7b Permanent Document: "Processing Plan for the EVS Selection Phase".

[22] EVS-8b Permanent Document: "Test Plan for the EVS Selection Phase".

[23] EVS-7c Permanent Document: Processing Plan for the EVS Characterization Phase

[24] EVS-8c Permanent Document: "Test Plan for the EVS Characterization Phase".

[25] Recommendation ITU-T P.800: "Methods for subjective determination of transmission quality".

[26] 3GPP TR 22.105: "Services and service capabilities".

[27] Recommendation ITU-T G.191: “Software tools for speech and audio coding standardization”, 03/2010, electronic attachment: STL2009 Software Tool Library.

[28] Recommendation ITU-T G.100.1: “The use of the decibel and of relative levels in speechband telecommunications”, 11/2001.

[29] 3GPP TS 26.132: “Speech and video telephony terminal acoustic test specification (Release 12)”.

[30] Recommendation ITU-T P.501: “Test signals for use in telephonometry”, 01/2012.

3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

ACELP	Algebraic Code-Excited Linear Prediction
ACR	Absolute Category Rating
AMR	Adaptive Multi-Rate
AMR-WB	Adaptive Multi-Rate Wideband
CCR	Comparison Category Rating
CI	Confidence Interval
CMOS	Comparison MOS
CNG	Comfort Noise Generation
CS	Circuit Switched
CuT	Codec under Test
DCR	Degradation Category Rating
DMOS	Differential MOS
DS	Direct Source
DTMF	Dual Tone Multi Frequency
DTX	Discontinuous Transmission
EDGE	Enhanced Data rates for GSM Evolution
EFR	Enhanced Full-Rate
ETSI	European Telecommunication Standards Institute
EVS	Enhanced Voice Services
FB	Fullband
FEC	Frame Erasure Concealment
FER	Frame Erasure
FR	Full-Rate
GAL	Global Analysis Laboratory
GERAN	GSM/EDGE Radio Access Network
GSM	Global System for Mobile communications
HD	High Definition
HR	Half-Rate
IO	Interoperable
ITU-T	International Telecommunication Union – Telecommunications Standardisation Sector
IP	Internet Protocol
JICO	Jitter Induced Concealment Operation
JBM	Jitter Buffer Management

LL	Listening Laboratory
MNRU	Modulated Noise Reference Unit
MOPS	Million of Operation per Seconds
MOS	Mean Opinion Score
MSB	Most Significant Bit
MTSI	Multimedia Telephony Service for IMS
NB	Narrowband
PS	Packet Switched
PSTN	Public Switched Telephone Network
REF	Reference
TSG-SA	Technical Specification Group - Service and System Aspects
SA4	Service and System Aspects Working Group 4 (TSG-SA WG4)
SAD	Sound Activity Detection
SC-VBR	Source Controlled - Variable Bit Rate
SID	Silence Insertion Descriptor
SNR	Signal To Noise Ratio
SWB	Super Wideband
TFO	Tandem Free Operation
TSG	Technical Specification Group
UMTS	Universal Mobile Telecommunication System
UTRAN	Universal Terrestrial Radio Access network
VAD	Voice Activity Detection
WID	Work Item Description
WB	Wideband
wMOPS	weighted Million of Operations per Second

4 General

4.1 Project History

In 2010, 3GPP finalized the Enhanced Voice Services (EVS) study item with the publication of TR 22.813 [16]. This study focused on how 3GPP could maintain the high value and competitiveness of its voice services and whether the new Evolved Packet System (EPS) with LTE (Long Term Evolution) access could open up new opportunities for a major voice service enhancements. Mobile use cases pertinent to LTE access and that may benefit from improved audio quality were studied. Part of the study included examining any potential need for enhanced codecs beyond AMR and AMR-WB, the codecs now used in 3GPP voice services. Envisioned use cases for enhanced voice services included improvements beyond classical telco-grade telephony (typically realized as IMS Multimedia Telephony), high-quality multi-party conferencing, call on hold or audio-visual communication, offering a 'being-there' quality of experience. Additional aspects of the study included how enhanced voice services could complement the existing voice service. Even streaming voice and audio as well as offline voice and audio delivery were also considered as an application scenario using the EVS codec.

Based on the conclusions of the study item in TR 22.813, 3GPP immediately launched a work item targeting the standardization of a new speech codec for Enhanced Voice Services, the EVS codec. The goal of the work item with its WID objectives was to provide clear benefit in terms of overall service quality, service efficiency and interoperability in 3GPP LTE networks. As a result of the study item, it is anticipated that enhanced voice services based on the new EVS codec will become the dominant voice service in 3GPP LTE networks. It is further envisioned that enhanced voice services with EVS will extend beyond 3GPP LTE system scope, ranging from deployments in circuit switched, to other mobile and wireless (WiFi) networks, fixed networks and the Internet. In that context not only the performance of the EVS codec in comparison to existing 3GPP and ITU-T codecs is of interest but even to other state-of-the art codecs.

Thirteen companies declared their intention to submit codecs to the Qualification Phase. Each codec was evaluated in 12 subjective experiments, each conducted twice; once in the candidates' own laboratory and once in a laboratory selected at random from the other 12, see EVS-7b [21] and EVS-8b [22]. Tests were blinded with all of the processing being conducted by a dedicated Host laboratory (Dynastat Inc.). Each of the candidates was evaluated against the requirements by an independent (non-codec proponent) Global Analysis Laboratory (GAL, Dynastat Inc.). At 3GPP SA4#72bis meeting in March 2013, the top five candidates were judged to have qualified although all 13 codecs had passed more than 95 % of the 296 requirements tested duplicated in two languages. After the qualification process, companies declared several collaborations around the qualified candidates. Note that the test results of the Qualification Phase are not included in the present document because they reflect different coders than the final standard.

As a result of examining the codec high level descriptions provided by each candidate at the Qualification meeting, it became clear to the various collaboration groups that all of the qualified candidates were based upon very similar coding principles.

In September 2013, 12 companies (Ericsson, Fraunhofer IIS, Huawei, Nokia, NTT, NTT DOCOMO, Orange, Panasonic, Qualcomm, Samsung, VoiceAge and ZTE Corporation) that confirmed their intent to submit a codec in selection declared their intention to work together and to develop a single jointly-developed candidate for the Selection Phase by merging the best elements of the codecs from each of the different collaboration groups.

Even though only a single codec entered the Selection Phase the strict 3GPP process for codec selection was maintained. The subjective Selection testing comprised 24 experiments, each conducted in two languages. Independent (non-codec proponent) Host Lab (Dynastat Inc.), Cross-check Lab (Audio Research Labs, LLC), Listening Labs (Dynastat Inc., DELTA, and Mesaqin.com s.r.o. (Ltd.)) and Global Analysis Lab (Dynastat Inc.) were used. This testing allowed the codec to be evaluated in 389 requirements, duplicated in two languages. The codec exhibited only two systematic failures (in both languages) at the 95 % confidence level. One of these failures was subsequently addressed as it was found to be the result of a software bug. Objective testing was also performed.

The single joint candidate was selected at 3GPP SA4#80bis meeting in August 2014 and the EVS codec specifications were approved at 3GPP TSG-SA#65 in September 2014. The selected EVS codec fulfils the project targets.

Verification Phase was launched and several organizations volunteered to verify that the code supplied to 3GPP complied with the design constraints and requirements.

The Characterization Phase is the latest phase. During this phase the codec was tested in a more complete manner than in the selection phase. In order to evaluate the selected codec in the broadest possible way a further set of 17 subjective experiments have been designed. Five of these experiments have been conducted in two different languages, for a total of 22 tests. The aim of these additional experiments, and other objective evaluations, was to evaluate features of the codec which remained untested in previous phases or to highlight areas of interest to 3GPP such as tandeming cases, fullband cases, and multi-bandwith comparisons. The same listening laboratories used for selection were again employed in characterization.

3GPP has also specified a floating-point version of the EVS codec (3GPP TS 26.443). This work was completed by 3GPP TSG-SA#66 in December 2014.

4.2 Overview of the EVS Codec Work Item

This clause provides an overview of the objectives before the actual work started, as a historical background. The standardized EVS codec fulfilled all project objectives [15].

With the advent of increasingly compact yet powerful mobile devices and the proliferation of high-speed wireless access to telecommunications networks around the globe, users of mobile devices expect and demand growing sophistication in the communication services being offered. Multi-modal interfaces supporting rich multimedia services for content and conversation are commonplace on the desktop, with demand for smart mobile devices with similar functionality steadily growing.

The identification of this potential was the background for 3GPP to launch a study investigating and defining the use cases and requirements for an Enhanced Voice Service in the Evolved Packet System leading to TR 22.813 [16]. The present document defines a new set of high-level technical recommendations and recommended requirements for a new codec for the Enhanced Voice Service and concludes that substantially enhanced voice services will become possible with a codec meeting them. The present document recommends starting an EVS codec development work item with the target to meet the requirements and recommendations set in it.

The overall objective of this work item is to develop a codec suitable for the Enhanced Voice Service in the EPS. The following objectives should be achieved with the new codec:

- Enhanced quality and coding efficiency for narrowband (NB) and wideband (WB) speech services, leading to improved user experience and system efficiency. This should also be achieved in interoperation with 3GPP pre-Rel-10 systems and services employing WB voice.

- Enhanced quality by the introduction of super-wideband (SWB) speech, leading to improved user experience.

- Enhanced quality for mixed content and music in conversational applications (for example, in-call music), leading to improved user experience for cases when selection of dedicated 3GPP audio codecs is not possible.

- Robustness to packet loss and delay jitter, leading to optimized behaviour in IP application environments like MTSI within the EPS.

- Backward interoperability to the 3GPP AMR-WB codec by having some WB EVS modes supporting the AMR-WB codec format used throughout 3GPP conversational speech telephony service (including CS). The AMR-WB interoperable operation modes of the EVS codec may be either identical to those in the AMR-WB codec or different but bitstream interoperable with them.

These are the project objectives while meeting all design constraints and performance requirements set forth in 3GPP TR 22.813 [16]. It is further desirable that the codec fulfills needs for enhanced voice services in other 3GPP systems, such as CS. The developments under this work item should lead to a set of new specifications defining among others textual description of the coding algorithm and the VAD/DTX/CNG scheme.

Following 3GPP practice, fixed-point and floating-point C code and associated test vectors should also be part of this set of specifications. The included AMR-WB interoperable coding format may become an alternative implementation for AMR-WB operation, provided that the enhancements are consistently significant. Jitter buffer management and packet loss concealment should be specified as part of the set of EVS specifications.

The EVS codec enhances coding efficiency and quality for NB and WB for a large bit rate range, starting from 5.9 kbps VBR. It further provides a significant step in quality over these traditional telephony bandwidths with SWB and FB operation starting from 9.6 and 16.4 kbps, respectively. Maximum bit rate is 128 kbps with support for WB, SWB, and FB. The ability to switch the bit rate at every 20-ms frame allows the codec to easily adapt to changes in channel capacity. The codec features discontinuous transmission (DTX) with algorithms for voice/sound activity detection (VAD) and comfort noise generation (CNG). An error concealment mechanism mitigates the quality impact of channel errors resulting in lost packets. A system for jitter buffer management (JBM) is included. The codec also features a channel-aware mode to further improve frame/packet error resilience. Enhanced interoperation with AMR-WB is provided over all nine bit rates between 6,6 kbps and 23,85 kbps.

4.3 Presentation of the Following clauses

Clause 5 outlines the Terms of reference for the EVS project. In clause 6, the selection process in 3GPP is presented. An overview of selection and characterization tests can be found in clause 7. The subjective tests provide statistical data which are subject to variations; important notes about interpretation of results are described in clause 8.

The actual test results are presented in clause 9 (narrowband), clause 10 (wideband), and clause 11(super-wideband). Clause 12 contains the results of mixed-bandwidth and full-band test, while clause 13 presents the results of objective evaluations.

5 Terms of Reference

3GPP sets the codec Terms of Reference as Design Constraints and Performance Requirements.

The design constraints specified in the EVS-4 Permanent Document [18] set the framework for the EVS codec in terms of capability and resource usage. As such they list functionalities that are divided into mandatory, recommended and optional features to be provided by EVS codec candidates. In the final standard, all modes have an equal status and they together form the EVS codec.

Codec features were defined as follows:

- support for input-output sampling at 8, 16, 32, 48 kHz independent of coded audio bandwidth;

- support of narrowband (NB), wideband (WB), superwideband (SWB) and fullband (FB) coded bandwidths;

- support for bit-rates of 5.9 (VBR), 7.2, 8, 9.6, 13.2, 16.4, 24.4, 32, 48, 64, 96, and 128 kbps for the *EVS primary modes*;

- support of the 9 AMR-WB bit-rates for the *EVS AMR-WB interoperable modes*;

- a jitter buffer management (JBM) solution conforming to TS 26.114 [13];

- rate switching at arbitrary frame boundaries; packet loss concealment; and

- discontinuous transmission (DTX) operation for rates up to 24.4 kbps.

Table 5.1 shows the EVS primary modes and the signal bandwidth supported at each codec bit rate. Discontinuous transmission (DTX) operation is supported at each bit rate of the standardized codec (primary modes and interop modes).

Table 5.1: Source codec bit-rates for the EVS Primary Modes

Source codec bit-rate (kbit/s)	Signal bandwidths supported
5,9 (VBR)	NB, WB
7,2	NB, WB
8,0	NB, WB
9,6	NB, WB, SWB
13,2	NB, WB, SWB
13,2 (channel aware)	WB, SWB
16,4	NB, WB, SWB, FB
24,4	NB, WB, SWB, FB
32	WB, SWB, FB
48	WB, SWB, FB
64	WB, SWB, FB
96	WB, SWB, FB
128	WB, SWB, FB

Table 5.2: Source codec bit-rates for the EVS AMR-WB IO Modes

Source codec bit-rate (kbit/s)
6,6
8,85
12,65
14,25
15,85
18,25
19,85
23,05
23,85

The EVS-4 Permanent Document [18] also sets constraints on maximum algorithmic delay (32 ms); frame length (20ms); maximum computational complexity (88 WMOPS); memory limits; and limit of the output gain. As a recommended feature, 5.9 kbps operation with source controlled variable bit-rate (VBR) is included. Further constraints are set for optional features.

The minimum performance of the EVS codec was defined across relevant operating points in the EVS-3 Permanent Document [17]. This document reflects the performance required for an enhanced voice service, following the recommendations specified in TR 22.813 [16]. The EVS-3 Permanent Document [17] lists subjective performance requirements in the form of statistical tests (e.g. not worse than, better than), as well as objective performance requirements on VAD, background noise attenuation, and JBM. The subjective requirements cover operating points in clean speech, noisy speech (car, street, office noise), music and mixed content, including clean and noisy channel (0%, 3%, 6% FER, delay-loss conditions) and several input levels (-16, -26, and -36 dBov), for all operation modes of EVS.

A full description of the performance requirements can be found in EVS-3 Permanent Document: Performance Requirements [17].

6 Selection Process

3GPP runs codec selection as a rigorous process, outlined below.

Codec selection in 3GPP follows pre-defined procedures. Proponents are obliged to provide certain information about their candidate to facilitate selection, and strict rules are set prior to selection to provide guidance on selecting the candidate to be standardized. Verification serves the purpose of cross-check and provision of additional (technical) information.

Selection Deliverables are specified in EVS-6b Permanent Document [20].

Proponents were required to provide the following information about their candidate for selection (named selection deliverables):

- High-level description and draft codec specifications
- Report of compliance to Design Constraints
- Funding payment (proponents paid for selection testing)
- IPR declaration
- Objective evaluation results
- Candidate codec fixed-point source code

Selection rules are specified in EVS-5b Permanent Document [19].

The strict 3GPP selection process involved the following rules (which were agreed before selection) to determine the candidate to be standardized:

- Provision of a full set of selection phase deliverables
- Compliance with design constraints
- Fulfilment of objective performance requirements
- Codec performance analyzed in sets according to EVS WID [15]:
    - Enhanced quality and coding efficiency for NB and WB speech services
    - Enhanced quality by SWB speech
    - Enhanced conversational music quality
    - Robustness to packet loss and delay jitter
    - Backward interoperability to AMR-WB

In the 3GPP SA4#80bis meeting the selection deliverables and selection test results were reviewed and based on this information, 3GPP SA4 selected the jointly developed EVS codec candidate as the 3GPP EVS standard. Sub-sequently the SA#65 plenary meeting approved the EVS codec selection and the set of EVS specifications [2], [3], [4], [5], [6], [7], [8], [9], [10], [11] and [12].

7 Introduction to the Testing of the EVS codec

7.0 General methodology

The fixed-point EVS codec was rigorously tested using the ITU-T P.800 [25] methodology with naïve listeners, demonstrating fulfillment of all testable EVS WID objectives (see clause 4.2 for more details). The extensive Selection and Characterization testing required a budget exceeding 1 Million €. The tests were conducted in independent (non-codec proponent) laboratories to guarantee a transparent process.

7.1 EVS Selection Phase Testing

The selection phase test details can be found in EVS-8b Permanent Document, Test Plan for Selection Phase [22], and in EVS-7c Permanent Document, Processing Plan for Selection Phase [21]. In the selection phase, an executable based on the 16-bit fixed-point C code developed by the proponents of the single joint candidate was used. The proponents submitted the selection executable for selection testing and it is included in the electronic attachment to 3GPP TS 26.442 v.12.0.0.

The EVS codec Selection Tests are split into 24 experiments listed in Table 7.1. Each experiment is performed twice and this results 48 listening tests in total. Table 7.1 shows the allocation of experiments.

Table 7.1: List of experiments in the EVS codec Selection Tests

#	Exp.	Group	Content/Description of Test Conditions
1	n1	NB	NB clean speech under clean channel condition including input level dependency
2	n2	NB	NB clean speech under impaired channel conditions including delay/jitter profiles
3	n3	NB	NB noisy speech under clean channel condition and impaired channel conditions
4	n4	NB	NB mixed content and music under clean channel condition and impaired channel conditions including delay/jitter profiles
5	w1	WB	WB clean speech under clean channel condition including input level dependency
6	w2	WB	WB clean speech under impaired channel conditions including delay/jitter profiles
7	w3	WB	WB noisy speech under clean channel condition
8	w4	WB	WB noisy speech under impaired channel conditions including delay/jitter profiles
9	w5	WB	WB mixed contents and music under clean channel condition
10	w6	WB	WB mixed contents and music under impaired channel conditions
11	w7	WB	WB mixed contents and music under impaired channel conditions including delay/jitter profiles
12	i1	IO	AMR-WB IO clean speech under clean channel condition including input level dependency
13	i2	IO	AMR-WB IO clean speech under impaired channel conditions
14	i3	IO	AMR-WB IO noisy speech under clean channel condition
15	i4	IO	AMR-WB IO noisy speech under impaired channel conditions
16	i5	IO	AMR-WB IO mixed contents and music under clean channel condition
17	i6	IO	AMR-WB IO mixed contents and music under impaired channel conditions
18	s1	SWB	SWB clean speech under clean channel condition including input level dependency
19	s2	SWB	SWB clean speech under impaired channel conditions including delay/jitter profiles
20	s3	SWB	SWB noisy speech under clean channel condition
21	s4	SWB	SWB noisy speech under clean channel condition
22	s5	SWB	SWB noisy speech under impaired channel conditions
23	s6	SWB	SWB mixed contents and music under clean channel condition
24	s7	SWB	SWB mixed contents and music under impaired channel conditions including delay/jitter profiles

The selection test plan defined 24 P.800 experiments consisting of 7 ACR and 17 DCR tests and containing 389 conditions for the codec under test. A total of 6 talkers/language (3 male + 3 female) and 6 categories (e.g. classical, modern, movie trailer, ...) were used for the speech experiments and the music experiments, respectively. Each experiment was conducted twice (i.e. by 2 different listening laboratories in different languages). In total, 48 listening tests were performed with 10 languages. Each test involved the use of 32 naïve listeners. The 778 ToR conditions were tested against performance requirements by the dependent groups Students T-test with 95 % confidence interval. Additional evaluation against performance objectives were performed using the independent groups T-test wherever available.

The selection test plan also defined numerous objective evaluations, including gain, JBM compliance, active frame ratio, attenuation in inactive region, bit rate and complexity.

To evaluate the EVS codec under well-defined and reproducible conditions, SA4 developed a selection processing plan in EVS-7b Permanent Document [21] defining processing steps for subjective and objective tests. Most methods are based on well-established procedures already used in other standardization efforts, for example, AMR-WB. Additional methods address novel features of the EVS codec, e.g. evaluation of the jitter buffer manager. The processing methods were implemented and crosschecked by two independent entities, ensuring that the audio material was processed error-free for the subjective evaluations.

Table 7.2 shows a list of the 24 Experiments (48 tests) involved in the EVS Selection Phase. For each Experiment, the table shows the Experiment Label, subjective test methodology (ACR or DCR), the Source Materials (Speech or Music/Mixed Content), and the number of test-conditions. Also shown is information on the two LLs conducting the Tests for the Experiment, including: Test-Label, Listening Lab, and Language. The Test Label is a three-character code (xy#), where:

- x is the LL designator - a=Delta, b=Dynastat, c=Mesaqin.com;
- y is the Experiment group designator - n=NB, w=WB, i=IO, s=SWB;
- # is the specific Experiment within the Group - 1, 2, 3, 4, 5, 6, 7.

Table 7.2: Allocation of listening laboratories and languages in selection

Exp.	Method	Source Materials	# Test Conds	Test#1			Test#2		
				Label	LL	Language	Label	LL	Language
n1	ACR	Speech	42	bn1	Dynastat	NA English (1)	cn1	Mesaqin	Chinese
n2	ACR	Speech	36	an2	Delta	Finnish	bn2	Dynastat	NA English (2)
n3	DCR	Speech	36	an3	Delta	Swedish	cn3	Mesaqin	French
n4	ACR	Music/Mixed	48	an4	Delta	Danish	bn4	Dynastat	LA Spanish (m)
w1	ACR	Speech	48	bw1	Dynastat	NA English (3)	cw1	Mesaqin	Slavic
w2	ACR	Speech	48	bw2	Dynastat	LA Spanish	cw2	Mesaqin	German
w3	DCR	Speech	30	aw3	Delta	Finnish	bw3	Dynastat	NA English (1)
w4	DCR	Speech	36	aw4	Delta	Japanese	bw4	Dynastat	NA English (2)
w5	DCR	Music/Mixed	30	bw5	Dynastat	NA English (m)	cw5	Mesaqin	French (m)
w6	DCR	Music/Mixed	36	aw6	Delta	Swedish (m)	cw6	Mesaqin	German (m)
w7	DCR	Music/Mixed	24	aw7	Delta	Danish (m)	cw7	Mesaqin	Chinese (m)
i1	ACR	Speech	48	ai1	Delta	Finnish	bi1	Dynastat	LA Spanish
i2	ACR	Speech	42	ai2	Delta	Japanese	ci2	Mesaqin	Slavic
i3	DCR	Speech	36	ai3	Delta	Danish	ci3	Mesaqin	French
i4	DCR	Speech	36	bi4	Dynastat	NA English (3)	ci4	Mesaqin	Chinese
i5	DCR	Music/Mixed	36	ai5	Delta	Swedish (m)	bi5	Dynastat	LA Spanish (m)
i6	DCR	Music/Mixed	36	bi6	Dynastat	NA English (m)	ci6	Mesaqin	German (m)
s1	DCR	Speech	36	bs1	Dynastat	NA English (1)	cs1	Mesaqin	French
s2	DCR	Speech	36	as2	Delta	Japanese	bs2	Dynastat	LA Spanish
s3	DCR	Speech	24	as3	Delta	Swedish	bs3	Dynastat	NA English#1
s4	DCR	Speech	24	bs4	Dynastat	NA English (2)	cs4	Mesaqin	Chinese
s5	DCR	Speech	36	as5	Delta	Finnish	bs5	Dynastat	NA English (3)
s6	DCR	Music/Mixed	24	as6	Delta	Danish (m)	cs6	Mesaqin	Chinese (m)
s7	DCR	Music/Mixed	36	bs7	Dynastat	NA English (m)	cs7	Mesaqin	German (m)

For the evaluation of test results, the Test Plan specified that Requirement ToR tests would use Dependent Groups T-tests (DGTT) to statistically evaluate the performance of a CuT condition relative to that of one or more REF conditions. The randomization-playlists were designed so that such comparisons would employ the highest-precision comparison available while also providing a valid and unconfounded test statistic. The GAL developed two independent procedures for conducting the DGTT ToR tests. One procedure was an Excel spreadsheet tool that accessed the raw voting data directly from the data-delivery spreadsheets and computed the DGTT T-test statistic for each ToR. For the second procedure, the GAL developed a FORTRAN program that read ASCII data files derived from the data-delivery spreadsheets and computed the appropriate DGTT T-test statistic. Across the 24 Experiments there were 389 Requirement ToR tests. With each ToR evaluated in two subjective Tests, there were a total of 778 Requirement ToR tests. However, since many of the ToR tests required comparisons of the CuT against two or three REFs, there were a total of 1018 DGTT T-test statistics to be computed. The GAL cross-checked all of the T-test statistics provided by the two ToR computation processes, one set of statistics from the Excel spreadsheet tool and one set from the FORTRAN program. All 1018 T-test statistics were successfully cross-checked.

The Test Plan specified that Objective ToR tests would use Independent Groups T-tests (IGTT) to statistically evaluate the performance of a CuT condition relative to that of one or more REF conditions. The IGTT was specified for the Objective ToR tests because randomization-playlists could not be designed to accommodate both Requirement and Objective ToR tests and still maintain acceptable "Balance" in the allocation of samples to CuT and REF conditions (a requirement for DGTT ToR tests). The GAL used the same two procedures for cross-checking Objective ToR tests as was described for Requirements in the previous clause. Across the 24 Experiments there were 295 Objective ToR tests. With each ToR evaluated in two tests, there were a total of 590 Objective ToR tests. Since a few of the Objective ToR tests require comparisons of the CuT against two REFs, there were a total of 612 IGTT T-test statistics to be computed. The GAL cross-checked all of the T-test statistics provided by the two ToR computation processes. All 612 T-test statistics were successfully cross-checked.

7.2 EVS Characterization Phase Testing

The characterization phase test details can be found in EVS-8c Permanent Document, Test Plan for Characterization Phase [24], and in EVS-7c Permanent Document, Processing Plan for Characterization Phase [23]. In the characterization phase, the standardized 16-bit fixed-point C code (electronic attachment to 3GPP TS 26.442, September 15, 2014, v.12.0.0) was used.

The purpose of the Characterization phase experiments is to evaluate the performance of the EVS codec candidate algorithm in three different categories:

- Conditions that were not tested in selection.
- Conditions specifically deferred to characterization.
- Additional interesting use cases not specifically defined in EVS-3 performance requirements [17].

Table 7.3 shows allocation of LLs and their languages to be used so that each experiment is conducted twice by different LLs with different languages.

Table 7.3: Overview of characterization experiments with allocation of LLs and languages

Experiment	# Tests	Content	# Cond	Test Type	Lab-a	Lab-b	Lab-c
N1	1	Clean Speech	48	ACR	FIN		
N2	1	Noisy Speech	36	DCR		NAE2	
N3	1	Noisy Speech	36	DCR			FRN
N4	1	Music/mixed	48	ACR	DANm		
W1	2	Clean Speech	48	ACR		NAE1	CHNm
W2	2	Noisy Speech	36	DCR		SPN	SLV
W3	2	Clean Speech	48	ACR		NAE3	SLV
W4	1	Music/mixed	36	DCR		NAEm	
W5	1	Clean Speech tandeming	48	ACR	DAN		
S1	2	Clean Speech	36	DCR	DAN	NAE1	
S1(noisy)	1	Noisy Speech	36	DCR		NAE1	
S2	2	Noisy Speech	36	DCR	FIN		FRN
S3	1	Music/mixed	36	DCR		SPNm	
M1	1	Speech	36	DCR		NAE2	
M2	1	Noisy Speech	36	DCR	FIN		
M3	2	Music/mixed	36	DCR		NAEm	CHNm
F1	1	Clean Speech	36	DCR			GER
F2	1	Music/mixed	36	DCR	DANm		
Total	24	-	-		7	10	7

8 Important Notes about the Interpretation of Test Results

Mean Opinion Scores can only be representative of the test conditions in which they were recorded (speech/music material, processing, listening conditions, language, and cultural background of the listening subject). Listening tests performed with other conditions than those used in the testing could lead to a different set of MOS results. On the other hand, the relative performances of different codecs under test is considered more reliable and less impacted by cultural difference between listening subjects than absolute MOS values. When looking at the relative differences of the codecs in the same test, it should be noted that a difference of typically 0.15-0.2 MOS between two test results would not usually be found statistically significant; appropriate statistical significance tests such as Student's T-test should be used to get an accurate figure of statistically significant difference between conditions within an experiment.

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

Furthermore, in a number of experiments both clean and erroneous channel conditions were presented in the same experiment. It can be expected that the separation of the different clean channel conditions is less in those experiments compared to experiments where only clean channel conditions are presented. During the setup of the listening experiments SA4 experts made every effort to minimize effects like scale saturation and alike. However, the large number of conditions to be tested and the limited number of experiments that could be conducted made certain compromises unavoidable. It should be noted that the testing effort in the selection and characterization of the EVS codec with 48 and 22 P.800 tests in the selection phase and characterization phase, respectively, is un-precedented.

The resolution of the testing is limited. The listeners only use a scale from 1 to 5 to rank the different codecs. However, during the tests presented in the present document, we are characterising a large number of different EVS modes, most of which are very high quality codecs and this may cause sometimes a "saturation" effect in the test, i.e. the listeners cannot discriminate the different codecs because of the limited range and scale.

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

Throughout the present document, test results are presented in two different graphical forms.

- **Test Profiles** - For each Selection and Characterization Experiment, a *Test Profile* presents the results for the test in a column graph showing the Mean scores and 95 % Confidence Intervals for each test condition involved in the experiment. The scores are grouped into three major categories: Reference Conditions (Direct Source and MNRU conditions), Reference Codec Conditions, and EVS Codec Conditions. Furthermore, the Codec conditions are color-coded according to the major parameters under test in the Experiment. Each value shown in the Profiles are based on 192 votes. The score Profiles are intended to give the reader a snapshot of the results for the Experiment and no conclusions on the statistical significance of the results should be inferred from the profiles. Instead, appropriate statistical tests such as Students T-test should be used to obtain a valid and accurate figure of statistically significant differences between conditions within an experiment.

- **Line-graphs** - Summary results are presented in line-graphs which compare the Reference codecs and the EVS codec for various test parameters, (e.g. Bit-rate, Frame Error Rate, DTX on/off, etc.). The line-graphs only include conditions from within a test, i.e., no comparisons are made across tests. Confidence intervals are not shown in the line-graphs as they tend to clutter the graphs and obscure the general trends for which the line-graphs are intended. Again, statistically significant differences should not be inferred from the graphical results except where specified in the text.

9 EVS Performance in Narrowband

9.1 NB Selection Tests

In selection phase, four experiments, N1, N2, N3, N4 were designed to evaluate the performance of the EVS codec in narrowband.

- Experiment N1 (ACR): NB clean speech under clean channel condition including input level dependency.

- Experiment N2 (ACR): NB clean speech under impaired channel conditions including delay/jitter profiles.

- Experiment N3 (DCR): NB noisy speech under clean channel condition and impaired channel conditions (Car noise at 15 dB SNR).

- Experiment N4 (ACR): NB Mixed content and music under clean channel condition and impaired channel conditions including delay/jitter profiles.

When testing under impaired channel conditions, AMR reference used the informative packet loss concealment technologies specified in the AMR specification 3GPP TS 26.091, however, implementations in mobile devices may use more advanced PLC algorithms. Further reference was the more recent G.718 codec that also includes a more advanced PLC technology than AMR. Packet loss concealment in EVS is normative.

Furthermore, a network simulator [21] is used to mimic the delay jitter/loss impaired channel characteristic for EVS-NB conditions and subsequently tested under the -voip mode of EVS. On the other hand, the delay/loss profiles are mapped to an error pattern using the tool (dlyerr_2_errpat.exe [21] [23]) that is applied to the AMR bit stream to mimic the delay jitter/loss impaired channel characteristic. The JBM profiles used in the EVS Selection and Characterization testing includes JBM Profiles 1 through 10 covering different cases of delay jitter/loss characteristics [13].

9.1.1 Experiment N1

This ACR test was conducted to evaluate EVS codec in narrowband mode under clean channel conditions. Dynastat and Mesaqin.com conducted the North American English and Chinese language tests respectively.

EVS performance was compared against AMR at 3 different input levels, i.e. low (-36 dBov), nominal (-26 dBov) and high (-16 dBov). Both DTX on (Figure 9.1) and DTX off (Figure 9.2) conditions of EVS were compared to AMR DTX off case. EVS bit rates 5.9, 7.2, 8.0, 9.6 and 13.2 kbps that support narrowband bandwidth were compared against AMR narrowband mode bitrates 7.95, 10.2 and 12.2 kbps.

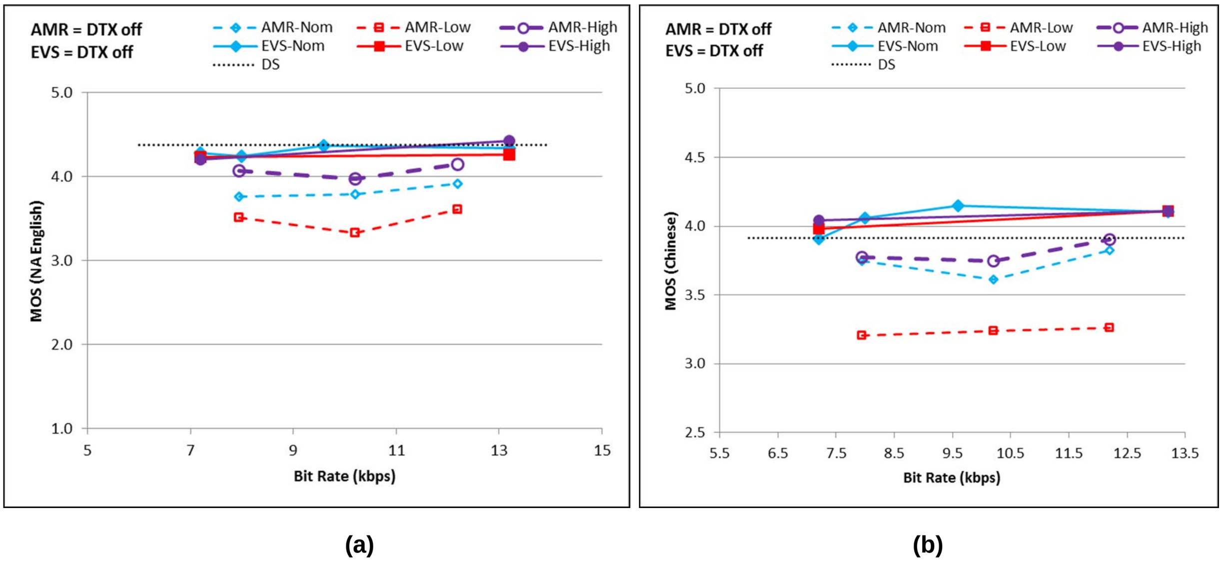
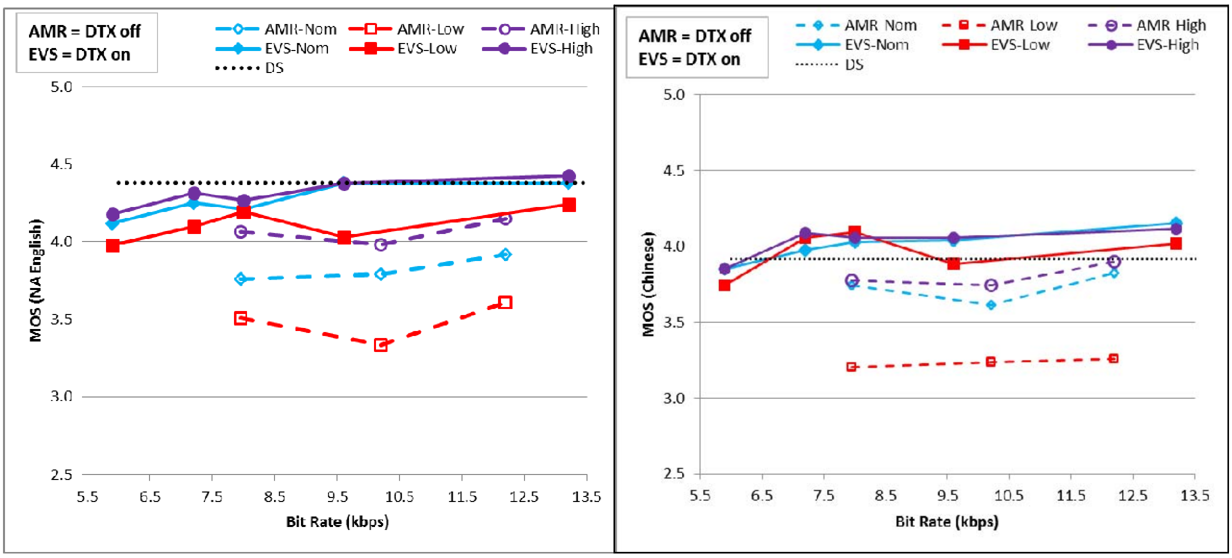


Figure 9.1: Experiment N1, testing EVS DTX off with NB clean speech under clean channel condition with level dependency with (a) North American English language and (b) Chinese language



(a) **(b)**

Figure 9.2: Experiment N1, testing EVS DTX on with NB clean speech under clean channel condition with level dependency with (a) North American English language and (b) Chinese language

Experiment N1 has shown that in narrowband mode, EVS meets or exceeds all performance requirements set forth in [17] under clean channel conditions. Moreover, Figures 9.1 and 9.2 indicate that EVS exhibits statistically significant performance improvement over AMR, for NB clean speech under clean channel conditions. This clear improvement over AMR is present across all narrowband bit rates, and across all input speech levels tested.

Subjective quality variation of EVS over bit rate does not appear significant. When EVS uses DTX on or off, it has no significant impact on the performance.

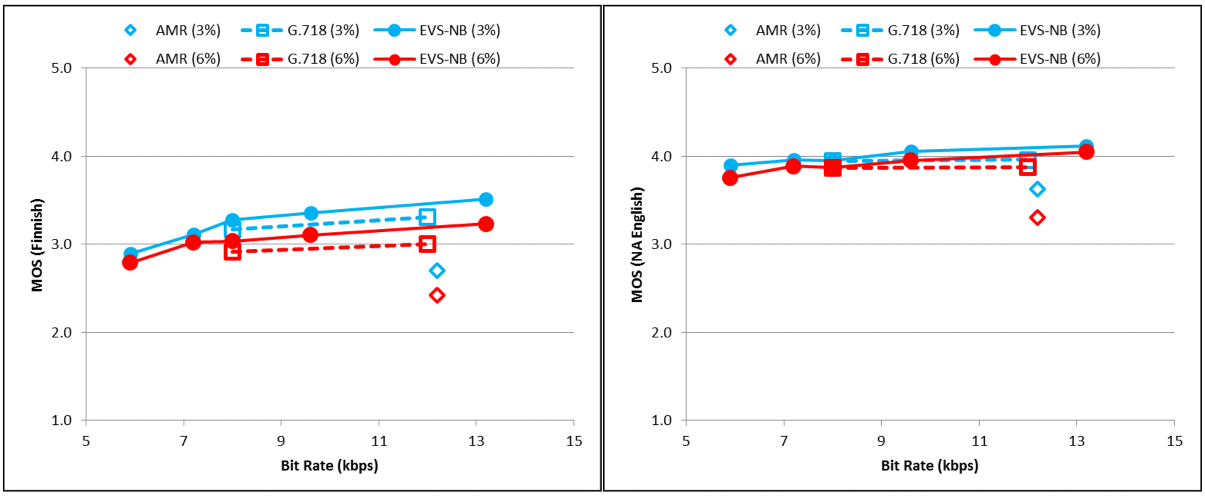
It is important to note that for low level inputs (-36 dBov), EVS consistently achieves more than 0.4 MOS performance gain over corresponding AMR bitrates, across both languages.

At most bit rates, EVS offers transparent NB quality.

9.1.2 Experiment N2

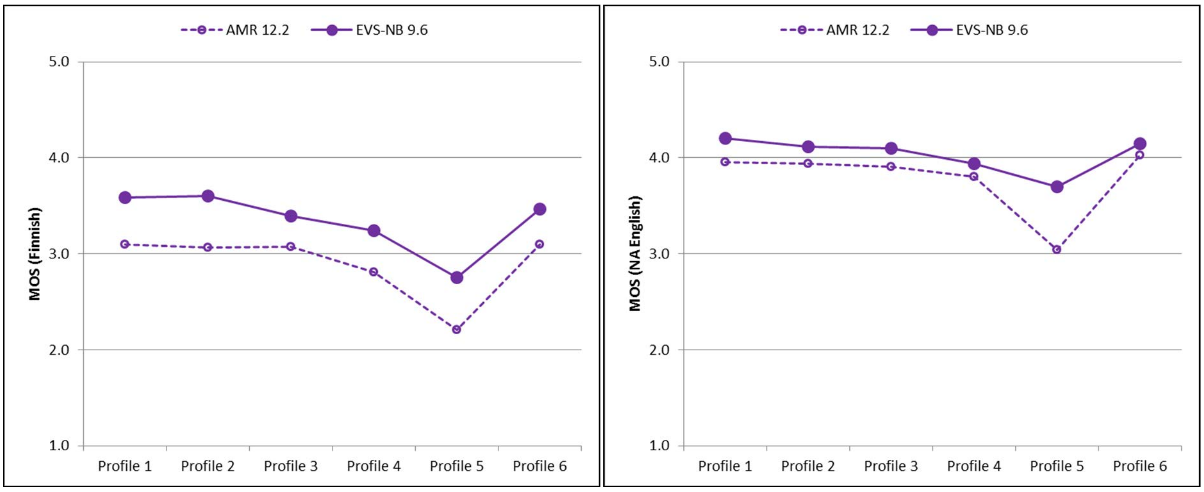
To evaluate performance of EVS codec in narrowband mode under various impaired channel conditions with the EVS JBM, Experiment N2 was conducted as an ACR test. This included comparisons of EVS against both AMR and G.718 codecs. Both 3 % and 6 % FER channel errors were used to simulate impaired channel conditions, at nominal input signal level. Also as part of this experiment, EVS bitrate 9.6 kbps in narrowband mode was compared against AMR bitrate 12.2 kbps, under MTSI delay-loss profiles 1…6 defined in 3GPP TS 26.114 [13]. Nominal level (-26 dBov) inputs were used.

Delta and Dynastat labs conducted the Finnish language and North American English language tests, respectively.



(a) (b)

Figure 9.3: Experiment N2, testing EVS-NB clean speech under impaired channel conditions (a) with Finnish language and (b) with North American English language



(a) (b)

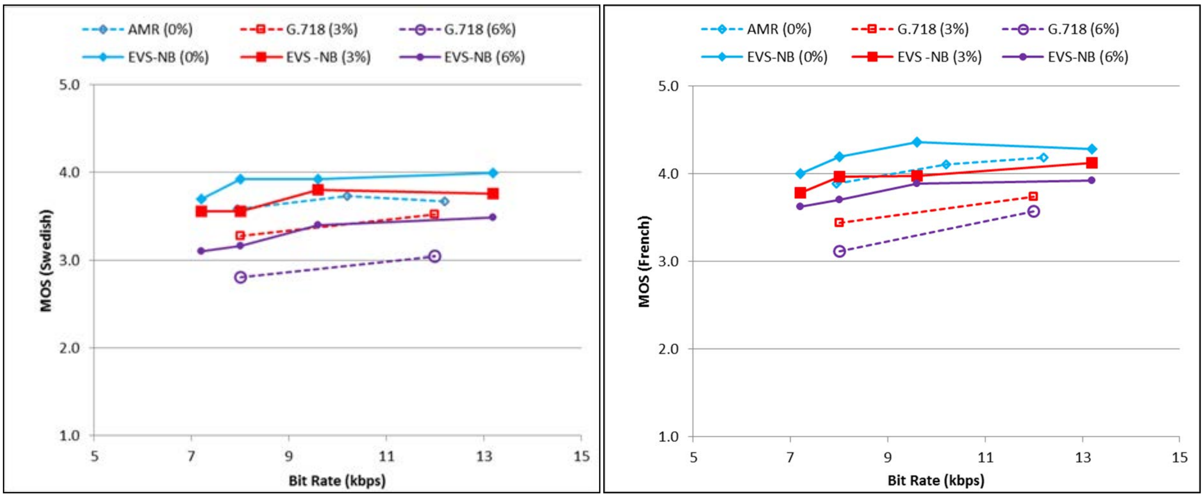
Figure 9.4: Experiment N2, testing EVS-NB clean speech under impaired channel conditions including delay/jitter profiles (a) with Finnish language and (b) with North American English language

The test results in Figure 9.3 show that the performance of EVS narrowband mode is NWT G.718 codec across multiple bit rates, under both 3% and 6% FER channel error conditions. Furthermore, EVS-NB mode at its lowest bit rate of 5.9 kbps already demonstrates improvement over AMR 12.2 kbps.

The test results of Experiment N2 in Figures 9.3 and 9.4 show a consistent performance advantage for EVS in impaired channel conditions especially when compared to AMR. Figure 9.4 shows that EVS, running at 9.6 kbps, is especially better for delay-loss profile 5 which contains the most severe channel impairments, compared to AMR running at 12.2 kbps. There is a noticeable difference in absolute MOS grades between Finnish and North American English content.

9.1.3 Experiment N3

A DCR test was conducted in Experiment N3 to evaluate performance of EVS for noisy speech signal inputs, under both clean and impaired channel conditions of 3% and 6% FER. Car noise mixed with speech at 15 dB SNR level (see [21] for details of noisy input signal generation) was used as input. Comparison was done against both AMR and G.718 codecs. Nominal level (-26 dBov) inputs were used for this test. Delta and Mesaqin.com labs conducted the tests for Swedish language and French language, respectively.



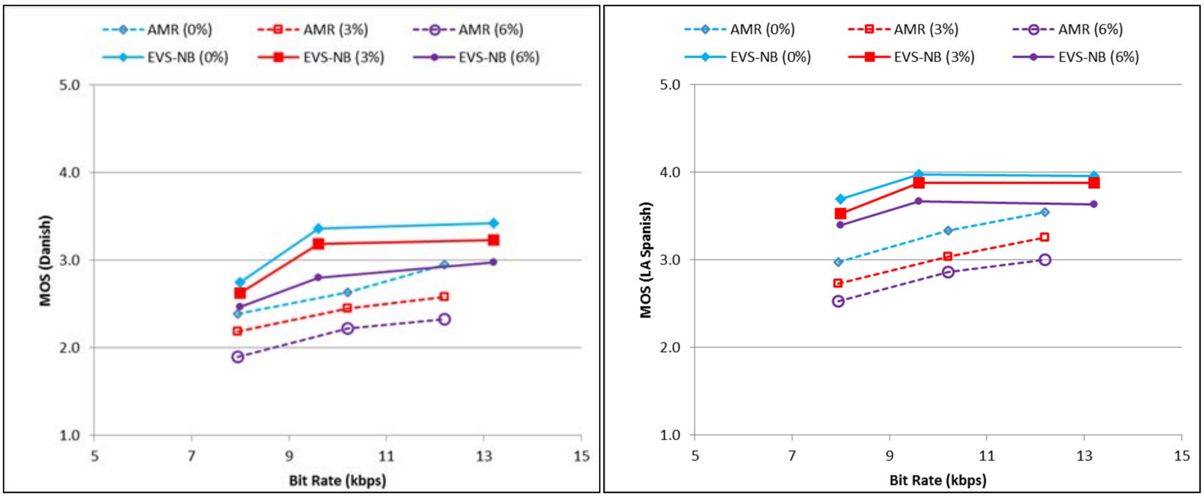
(a) (b)

Figure 9.5: Experiment N3, testing EVS-NB noisy speech (car noise at 15 dB SNR) under clean channel condition and impaired channel conditions (a) with Swedish language and (b) with French language

Figure 9.5 shows consistently superior performance of EVS-NB for noisy speech inputs over both AMR and G.718 codecs. The performance improvement is clearly evident across all bitrates tested and across all channel conditions (0 %, 3% and 6% FER channel errors). Furthermore, it is important to note that the performance gain achieved by EVS is more pronounced under impaired channel conditions as shown in Figure 9.5.

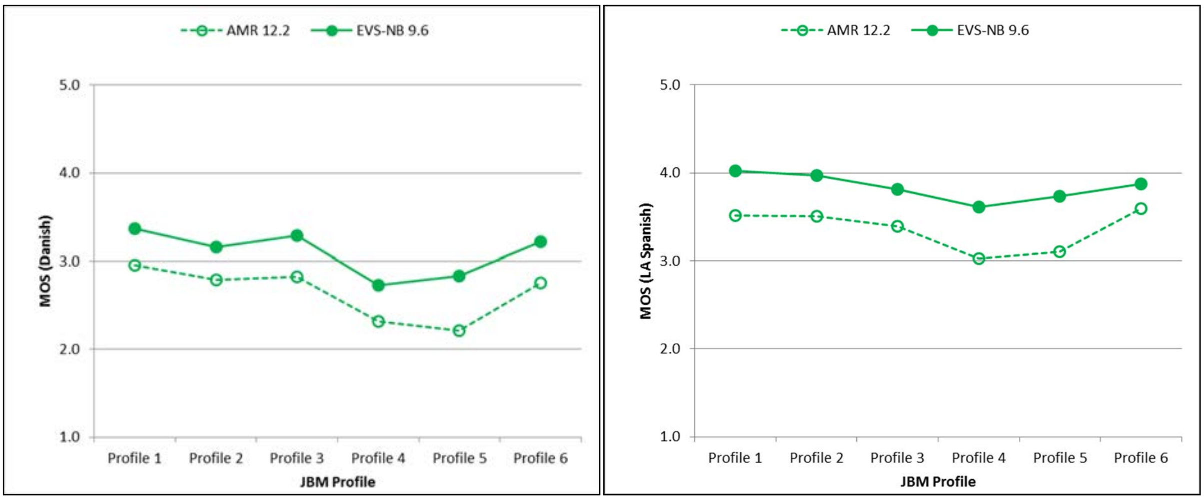
9.1.4 Experiment N4

Another ACR test was conducted in Experiment N4 to evaluate and compare the performance of narrowband mode of EVS codec against AMR for mixed speech/music and music content inputs. This test included both clean channel and impaired channel conditions, as well as delay-loss profiles that simulate the network conditions in a packet switched network. Nominal level (-26 dBov) inputs were used. Delta and Dynastat labs conducted the Danish and Latin American Spanish language tests, respectively.



(a) (b)

Figure 9.6: Experiment N4, testing EVS-NB Mixed content and music under clean channel condition and impaired channel conditions including delay/jitter profiles (a) with Danish language and (b) with LA Spanish language



(a) (b)

Figure 9.7: Experiment N4, testing EVS-NB Mixed content and music under clean channel condition and impaired channel conditions including delay/jitter profiles (a) with Danish language and (b) with LA Spanish language

Results of Experiment N4 in Figure 9.6 show superior performance of EVS narrowband modes compared to AMR across all bitrates. Under impaired channel conditions, the subjective performance gain offered by EVS over AMR for mixed and music content input is consistently more than 0.4 MOS across all bitrates and both languages tested in Experiment N4. Note that in certain cases, the improvement offered by EVS is as high as 0.8 MOS as shown in Figure 9.6.

In Figure 9.7, performance of EVS narrowband at 9.6 kbps bitrate (with EVS JBM) is compared to AMR 12.2 kbps bitrate (with JBM simulation) under impaired channel conditions, using MTSI delay-loss profiles defined in 3GPP TS 26.114 [13]. This indicates the superior performance gain of EVS narrowband mode over AMR across all six delay-loss profiles, for mixed and music content inputs.

Note that Figure 9.7 compares EVS at 9.6 kbps with AMR at 12.2 kbps. EVS at 9.6 and 13.2 kbps gets close to transparency and performs as well as G.711 (G.711 results are not shown in the graphs, see attachment).

9.2 NB Characterization Tests

9.2.0 List of experiments in the narrowband telephone bandwidth

In characterization phase, four experiments, N1, N2, N3, N4 were designed to evaluate the performance of the EVS codec in narrowband:

- Experiment N1 (ACR): NB clean speech in Finnish language to evaluate rate switching, tandeming and JBM.
- Experiment N2 (DCR): NB speech in North American English language under street background noise at 20 dB SNR to evaluate rate switching, untested conditions in selection testing, and tandeming.
- Experiment N3 (DCR): NB speech in French language under street background noise at 25 dB with impaired channels at high FER.
- Experiment N4 (ACR): NB music and mixed content in Danish language to evaluate rate switching and untested conditions in selection phase.

9.2.1 Experiment N1

The purpose of this experiment was to evaluate the performance of the EVS codec in narrowband mode with respect to AMR codec, with clean speech inputs of different levels, under clean and impaired channel conditions, rate switching, tandeming and JBM conditions. ACR test was conducted in Finnish Language by Delta lab.

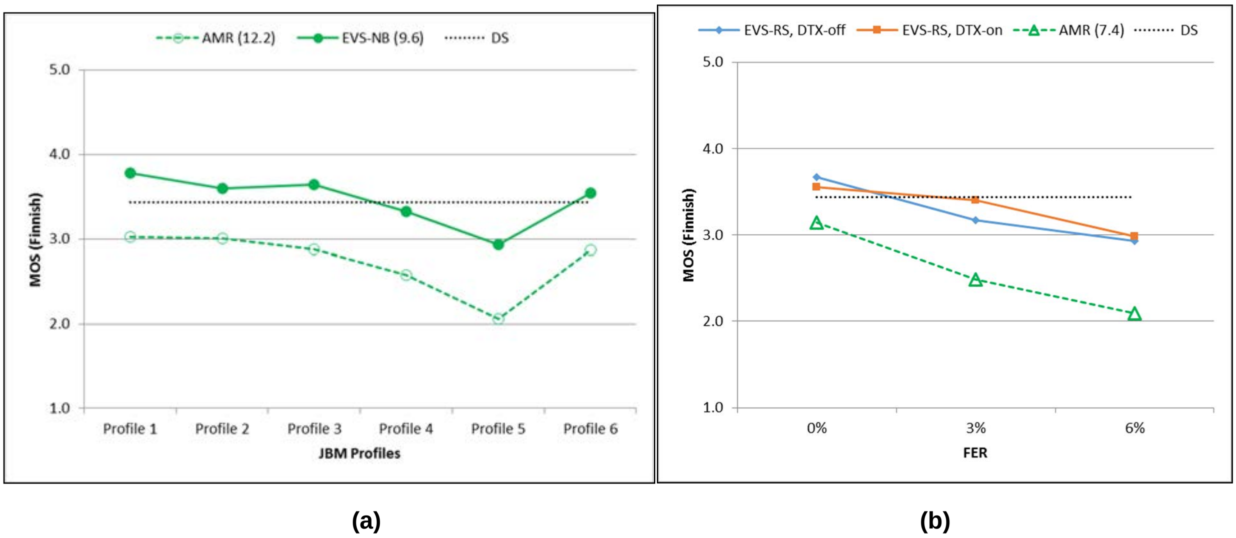


Figure 9.8: Experiment N1, testing EVS-NB clean speech under impaired channel condition with delay-loss profiles and 3% and 6% FER, with Finnish language

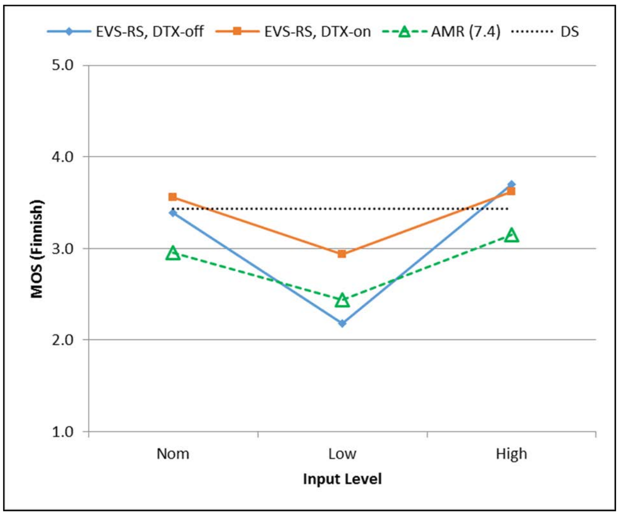


Figure 9.9: Experiment N1, testing EVS-NB clean speech at various input levels with Finnish language

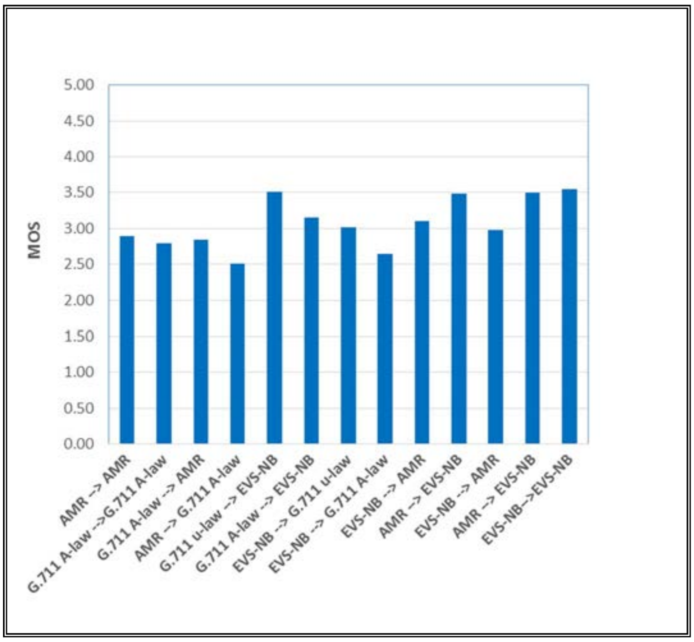


Figure 9.10: Experiment N1, testing EVS-NB clean speech with Finnish language

Figure 9.8(a) shows results similar to Figure 9.4 where EVS narrowband mode scores consistently over AMR under impaired channel conditions simulated by MTSI delay-loss profiles 1…6. In case of profile 5, the subjective performance gain offered by EVS-NB over AMR is close to 0.9 MOS.

Comparison of EVS rateswitching conditions (switching among EVS narrowband rates 7.2, 8.0. 9.6 and 13.2 kbps) with AMR 7.4 kbps DTX off is shown in Figure 9.8(b), under both clean and impaired channel conditions 3 % and 6 % FER. This comparison is shown for different levels of clean input speech, i.e. low (-36 dBov), nominal (-26 dBov) and high (-16 dBov). EVS-NB rate switching in both DTX on and DTX off cases offer statistically similar subjective performance. However, these EVS cases far exceed the subjective performance of AMR 7.4 kbps across all channel conditions shown in Figure 9.9. In case of 6% FER channel errors, EVS-NB rate switching conditions (7.2-13.2 kbps) demonstrate more than 0.8 MOS subjective performance gain as compared to AMR 7.2 kbps.

For low input level, performance of EVS-NB rate switching subjective performance is similar to or NWT AMR 7.4 kbps. This is in contrast to the results shown in clause 9.1.1 (Figures 9.1, 9.2) where EVS-NB performance for low level clean speech inputs far exceed that of AMR. However, a similar trend is noticed here for nominal and high level clean speech inputs.

9.2.2 Experiment N2

In this experiment, EVS-NB rate switching conditions were evaluated with noisy speech signal input against AMR 7.95 and 10.2 kbps rates. Street background noise mixed with speech at 20 dB SNR level [21] at nominal level -26 dBov was used as input. This DCR test was conducted by Dynastat using North American English language.

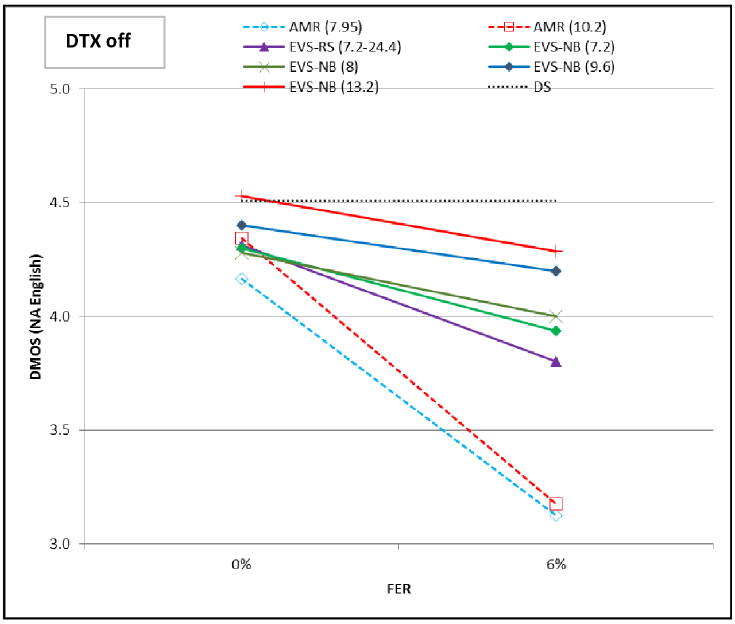


Figure 9.11: Experiment N2, testing EVS-NB noisy speech (street noise at 20 dB SNR), DTX off, with North American English language

Results in Figure 9.11 show statistically similar or better performance of EVS-NB mode (both rateswitching and non rateswitching conditions) compared to AMR under clean channel conditions, with noisy speech input at nominal levels. In contrast, subjective quality of EVS-NB is significantly higher compared to AMR 7.95 kbps and 10.2 kbps rates, under impaired channel with 6% errors.

9.2.3 Experiment N3

In this experiment, EVS-NB performance was evaluated with respect to AMR under channel error conditions as high as 10 %. Nominal level noisy speech input was used with street background noise at 25 dB SNR. This DCR test was conducted in French language by Mesaqin.com lab.

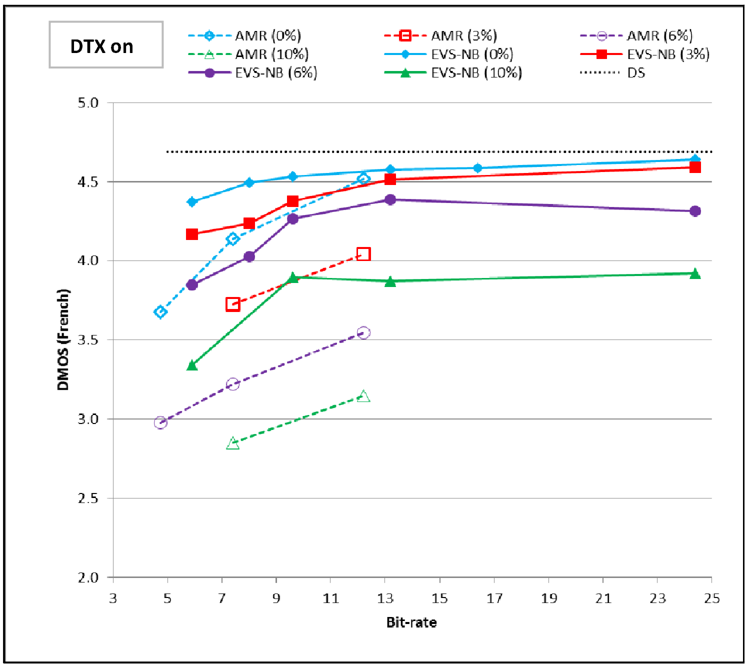


Figure 9.12: Experiment N3, testing EVS-NB noisy speech (street noise at 25 dB SNR), DTX on, under impaired channel conditions with French language

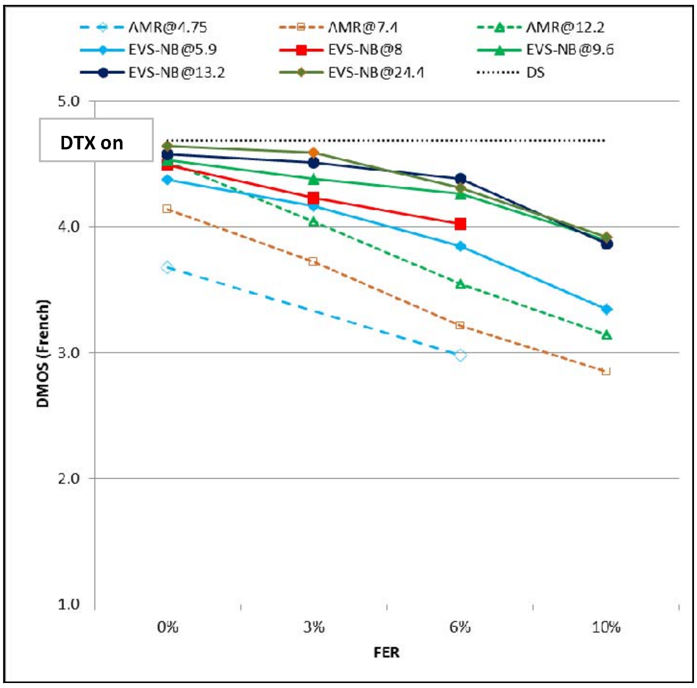


Figure 9.13: Experiment N3, testing EVS-NB noisy speech (street noise at 25 dB SNR), DTX on, under impaired channel conditions at 3 %, 6 %, and 10 % FER with French language

Figure 9.13 shows the gradual degradation of subjective quality of EVS-NB with increasing channel error percentage, as expected, under noisy input conditions. However, in all impaired channel conditions with noisy input, EVS-NB achieves significantly high MOS scores compared to AMR, with highest increases shown at 6 % and 10 % channel error rates.

At higher bitrates towards 24.4 kbps, EVS-NB mode approaches the saturation region near the Direct Source quality for noisy speech inputs, under clean channel conditions.

9.2.4 Experiment N4

Experiment N4 in characterization phase was conducted to evaluate performance of EVS for narrowband, nominal level mixed and music content under both clean and impaired channel conditions. Delta lab conducted this test in Danish language. VBR mode is designed to achieve the average data rate (ADR) of 5.9 kbps for active speech. In order to further evaluate and confirm the performance of the VBR mode in music/mixed content, this experiment included the VBR condition in NB. While achieving the ADR of 5.9 kbps for active speech, the VBR mode may result in a different ADR between 5.9 and 8 kbps for music/mixed content; the ADR value was in this experiment N4 7.07 kbps for music/mixed content.

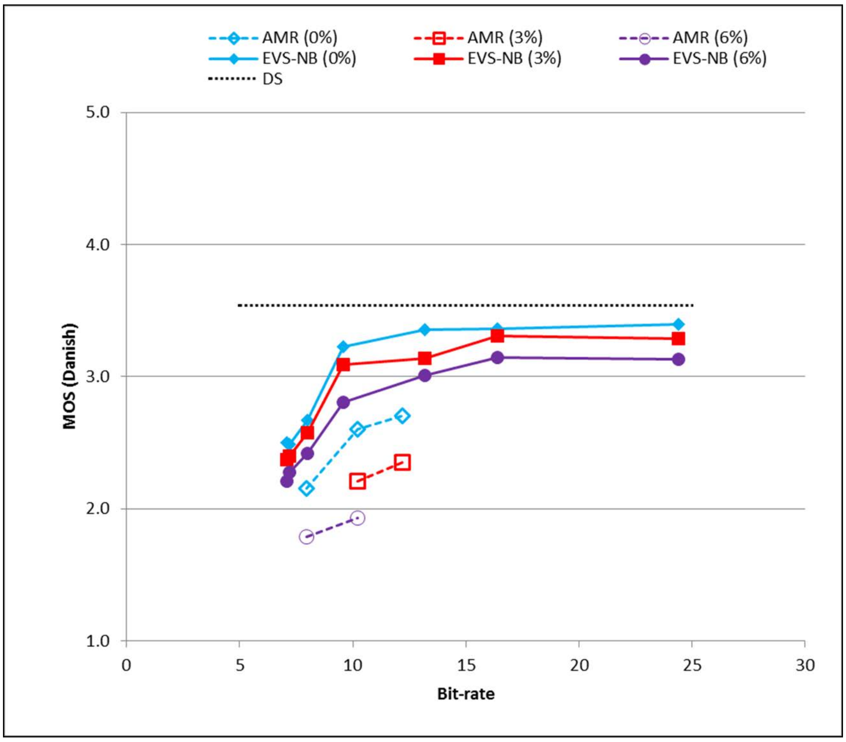


Figure 9.14: Experiment N4, testing EVS-NB music and mixed content with Danish language

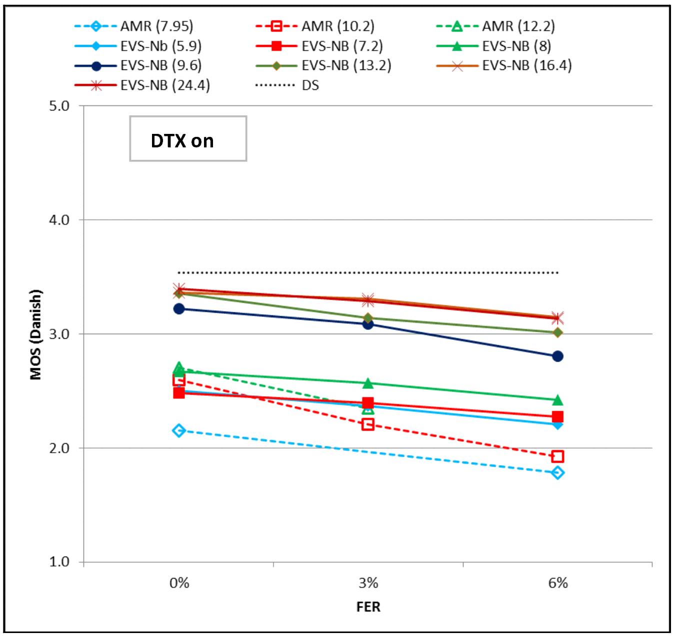


Figure 9.15: Experiment N4, testing EVS-NB music and mixed content with Danish language

Figure 9.14 shows a clear performance improvement of EVS-NB over AMR even at lowest bitrates of EVS for mixed and music content input. This is trend is significantly evident when channel errors are introduced, where AMR 7.95 kbps MOS scores at 0 % errors (clean channel) are statistically similar or below that of EVS-NB under impaired channel with 6% FER errors.

Trends shown in Figure 9.15 indicate the gradual degradation of subjective speech quality in both EVS-NB and AMR when the channel conditions are degraded. It is noticeable that AMR, in general, presents a more steeper drop in MOS scores in the face of channel impairments. Under clean channel conditions, EVS-NB VBR and 7.2 kbps mode achieve similar MOS scores as AMR 10.2 kbps. Same is observed for EVS-NB 8.0 kbps and AMR 12.2 kbps.

For music and mixed content, EVS starting with 9.6 kbps provides better performance than any rate of the AMR codec.

9.3 Conclusions on EVS Performance in Narrowband

As discussed in the results shown in clauses 9.1 and 9.2 above, EVS codec in narrowband mode demonstrates a significantly improved performance over prior codecs including AMR, G.711, and G.718.

The improvement is evident across all bitrates, all tested languages, various input speech levels, different input content spanning speech / music / mixed content, as well as clean vs noisy inputs. This performance improvement in EVS-NB is more pronounced under impaired channel conditions, emphasizing the improved robustness of EVS against transmission errors, as compared to AMR.

The EVS codec demonstrates a significantly improved performance in NB over the AMR codec at all bit rates, with clean and noisy speech and mixed content and music, both for clean and impaired channels. The differences are especially large for mixed and music content, where EVS at 9.6 and 13.2 kbps gets close to transparency and performs as well as G.711, and for high FER values which emphasizes the improved robustness of EVS against transmission errors, as compared to AMR.

10 EVS Performance in Wideband

10.1 WB Selection Tests

In selection phase, seven experiments, W1…W7 were designed to evaluate the performance of the EVS Primary Modes in wideband:

- Experiment W1 (ACR): WB clean speech under clean channel condition including input level dependency: The purpose of this experiment is to evaluate the performance of the EVS candidate algorithm with respect to well-known references in WB clean speech (free of background noise), clean (unimpaired) channel condition and different input levels.

- Experiment W2 (ACR): WB clean speech under impaired channel conditions including delay/jitter profiles: The purpose of this experiment is to evaluate the performance of the EVS candidate algorithm with respect to well-known references in WB clean speech (free of background noise) and impaired channel conditions including delay/jitter profiles.

- Experiment W3 (DCR): WB noisy speech under clean channel condition (Car noise at 15 dB SNR): The purpose of this experiment is to evaluate the performance of the EVS candidate algorithm with respect to well-known references in WB noisy speech and clean channel condition.

- Experiment W4 (DCR): WB noisy speech under impaired channel conditions (Street noise at 20 dB SNR)**:** The purpose of this experiment is to evaluate the performance of the EVS candidate algorithm with respect to well-known references in WB noisy speech and impaired channel conditions.

- Experiment W5 (DCR): WB mixed contents and music under clean channel condition: The purpose of this experiment is to evaluate the performance of the EVS candidate algorithm with respect to well-known references in WB mixed content and music and clean channel condition.

- Experiment W6 (DCR): WB mixed contents and music under impaired channel conditions including delay/jitter profiles: The purpose of this experiment is to evaluate the performance of the EVS candidate algorithm with respect to well-known references in WB mixed content and music and impaired channel conditions included delay/jitter profiles. This experiment is focused on the bitrate at 13.2 kbps or lower.

- Experiment W7 (DCR): WB mixed contents and music under impaired channel conditions: The purpose of this experiment is to evaluate the performance of the EVS candidate algorithm with respect to well-known references in WB mixed content and music and impaired channel conditions. This experiment is focused on the bitrate at 16.4 kbps or higher.

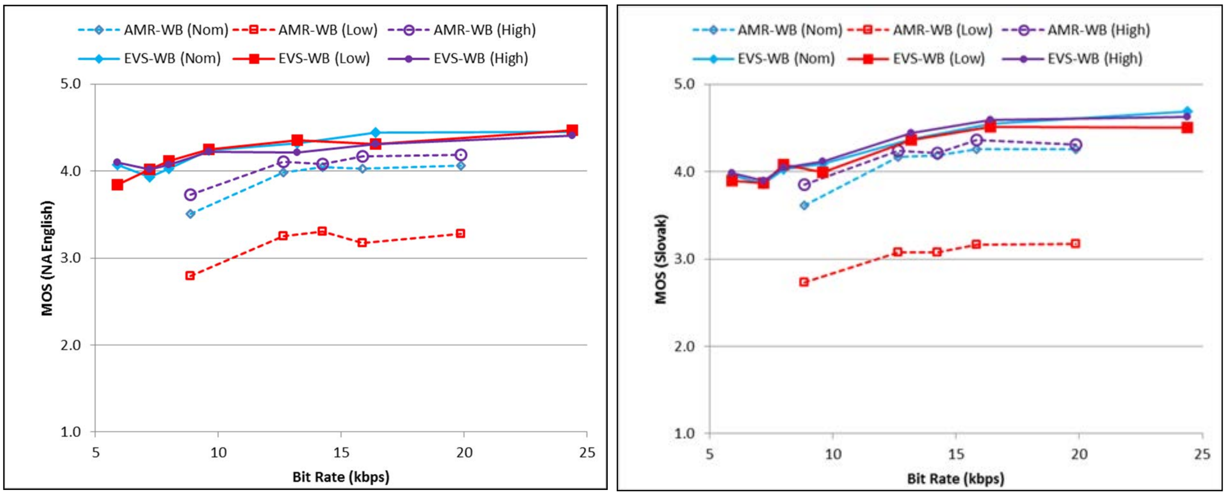
In selection phase, six experiments, I1…I6 were designed to evaluate the performance of the EVS AMR-WB IO Modes in wideband. The EVS AMR-WB IO mode is evaluated in three different configurations, namely: case A: EVS AMR-WB IO encoding, AMR-WB decoding, Case B: AMR-WB encoding, EVS AMR-WB IO decoding, and Case C: EVS AMR-WB IO encoding/decoding.

- Experiment I1 (ACR): EVS AMR-WB IO clean speech under clean channel condition including input level dependency
- Experiment I2 (ACR): EVS AMR-WB IO clean speech under impaired channel conditions
- Experiment I3 (DCR): EVS AMR-WB IO noisy speech under clean channel condition
- Experiment I4 (DCR): EVS AMR-WB IO noisy speech under impaired channel conditions
- Experiment I5 (DCR): EVS AMR-WB IO mixed contents and music under clean channel condition
- Experiment I6 (DCR): EVS AMR-WB IO mixed contents and music under impaired channel conditions

In WB Selection tests W2, W4, I2, I4, and I6, when testing impaired channel conditions with FER, AMR-WB decoding was replaced by G.718 IO decoding, which provides backward compatible operation with improved error concealment over the informative concealment specified in the AMR-WB standard in 3GPP TS 26.191. Actual AMR-WB implementations may or may not use such improved concealment technology. PLC in EVS is normative.

Furthermore, a network simulator [21] is used to mimic the delay jitter/loss impaired channel characteristic for EVS conditions and subsequently tested under the -voip mode of EVS. On the other hand, the delay/loss profiles are mapped to an error pattern using the tool (dlyerr_2_errpat.exe [21] [23]) that is applied to the AMR-WB bit stream to mimic the delay jitter/loss impaired channel characteristic. The JBM profiles used in the EVS Selection and Characterization testing includes JBM Profiles 1 through 10 covering different cases of delay jitter/loss characteristics [13].

10.1.1 Experiment W1

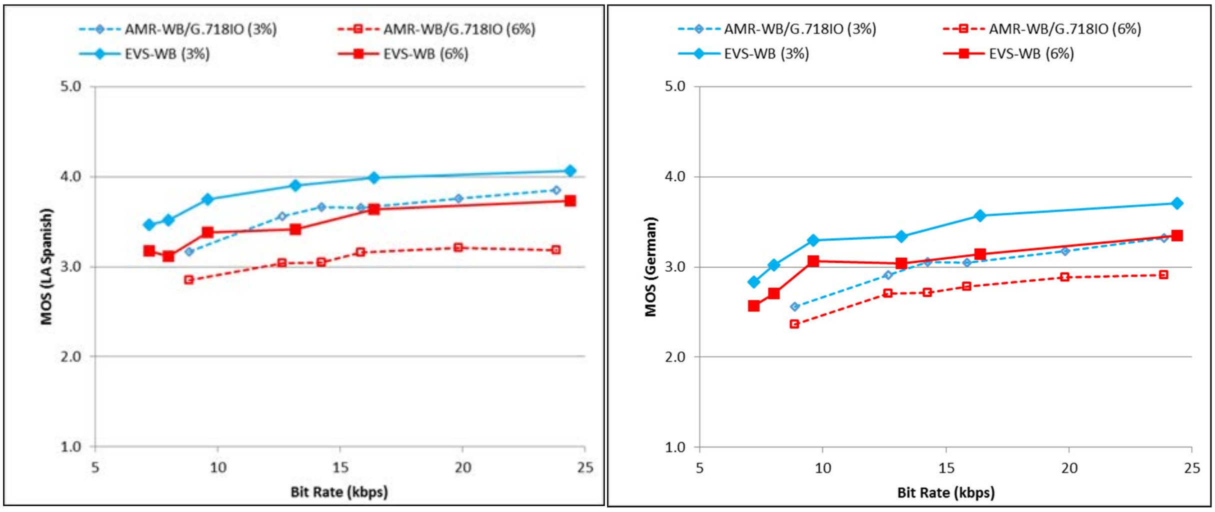


(a) (b)

Figure 10.1: Experiment W1, testing EVS-WB clean speech under clean channel condition including input level dependency (a) with North American English language and (b) with Slovak language

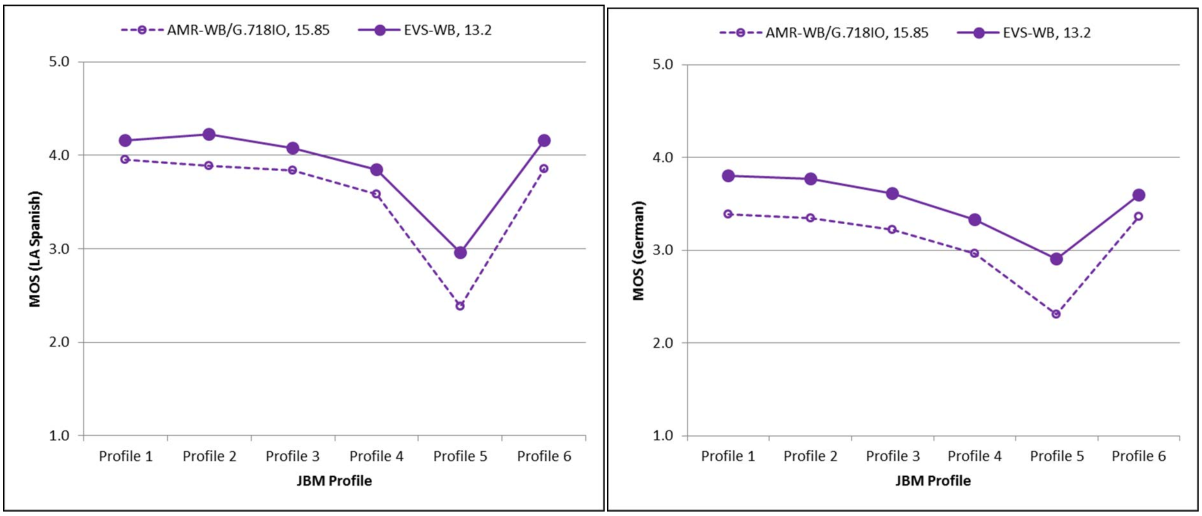
The test results of Experiment W1 in Figure 10.1 show improved performance of EVS over AMR-WB across all bit-rates and input levels. The improvement is even more pronounced for low input level at all bit-rates and across all levels for the lower bit-rates.

10.1.2 Experiment W2



(a) (b)

Figure 10.2: Experiment W2, testing EVS-WB clean speech under impaired channel conditions including delay/jitter profiles (a) with LA Spanish language (a) and (b) with German language

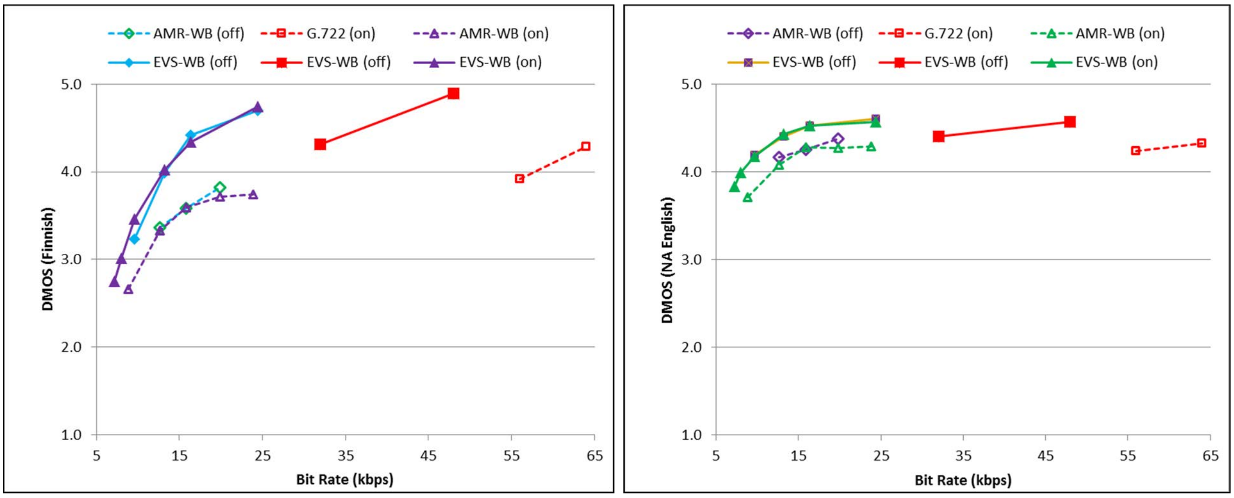


(a) (b)

Figure 10.3: Experiment W2, testing EVS-WB clean speech under impaired channel conditions including delay/jitter profiles (a) with Spanish language and (b) with German language

Experiment W2 was conducted in LA Spanish and German languages, to evaluate the EVS codec performance for WB clean speech under impaired channel conditions. As shown in Figure 10.2, the EVS-WB codec demonstrates improved performance over AMR-WB at frame erasure rates (FERs) of 3 % and 6 % for each bit-rate. Figures 10.3 illustrates the improved performance of the EVS codec at 13.2 kbps over AMR-WB at 15.85 kbps under varying delay/jitter conditions simulated by delay-loss profiles 1…6. A network simulator [21] is used to mimic the delay jitter/loss impaired channel characteristic for EVS conditions and subsequently tested under the -voip mode of EVS. On the other hand, the delay/loss profiles are mapped to an error pattern using the tool (dlyerr_2_errpat.exe [21] [23]) that is applied to the AMR-WB bit stream to mimic the delay jitter/loss impaired channel characteristic. AMR-WB as shown in Figure 10.2 uses improved packet loss concealment specified in G.718 IO mode. The delay-loss profiles 1…6 are defined in [13].

10.1.3 Experiment W3



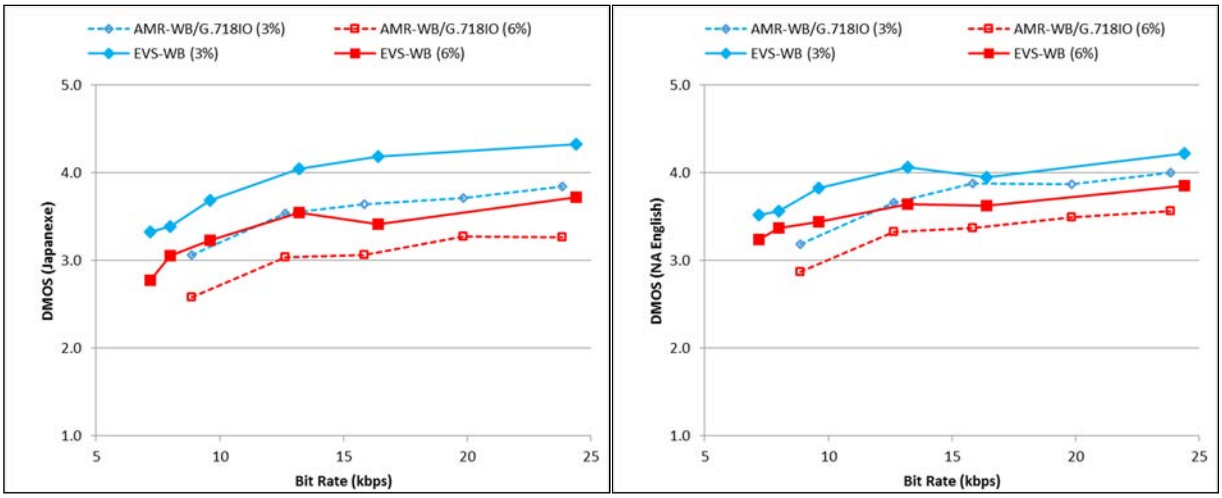
(a) (b)

Figure 10.4: Experiment W3, testing EVS-WB noisy speech (car noise at 15 dB SNR) under clean channel condition, (a) with Finnish language and (b) with North American English language

The WB noisy speech (car noise at 15 dB SNR) test results of Experiment W3 in Figure 10.4 show that EVS-WB at a particular bit rate offers quality which is either equivalent or better than AMR-WB at a higher bit rate for both DTX on and off cases. Furthermore, EVS-WB performance at 32 and 48 kbps is improved over the corresponding G.722 reference codec at 56 and 64 kbps, respectively.

EVS codec data point at 32 kbps and 64 kbps had a fixed point implementation bug in selection testing that was corrected in characterization testing (See clause 10.2.2 and clause 12.1.2).

10.1.4 Experiment W4

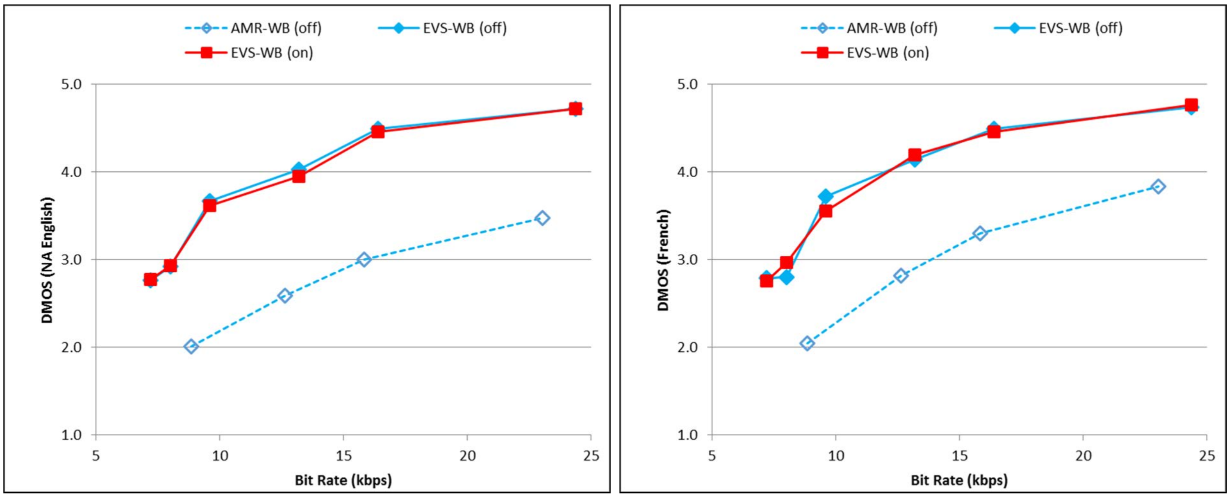


(a) (b)

Figure 10.5: Experiment W4, testing EVS-WB noisy speech (street noise at 20 dB SNR) under impaired channel conditions (a) with Japanese language and (b) with North American English

The WB noisy speech (street noise at 20 dB SNR) test results of Experiment W4 in Figure 10.5 show improved performance of EVS-WB over AMR-WB at each bit rate and FER rate, for both languages. AMR-WB uses improved packet loss concealment specified in G.718 IO mode.

10.1.5 Experiment W5

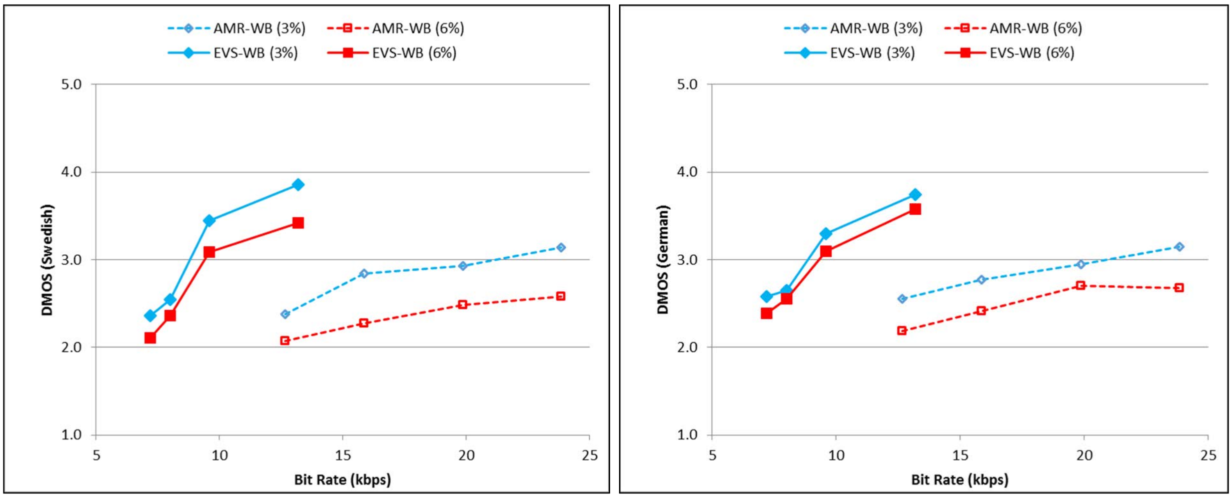


(a) **(b)**

Figure 10.6: Experiment W5, testing EVS-WB mixed contents and music under clean channel condition (a) with North American English language and (b) with French language

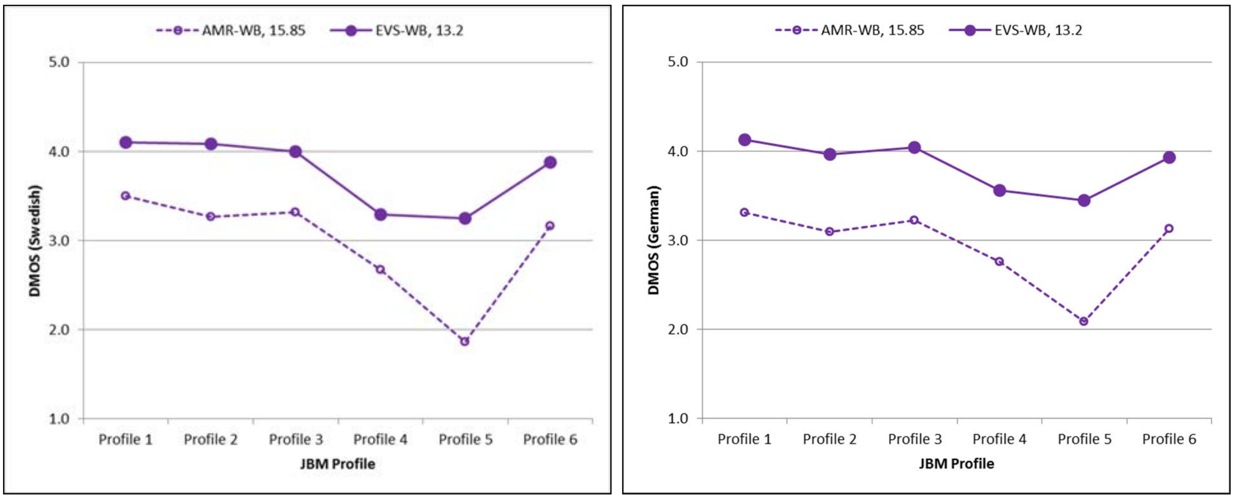
The WB mixed content and music test results of Experiment W5 in Figure 10.6 show largely improved performance of EVS-WB over AMR-WB at each bit rate, for both languages. EVS also scales towards much higher quality than AMR-WB for high rates and reaches transparency at 24.4 kbps. In case of EVS, DTX on and DTX off cases make no significant difference.

10.1.6 Experiment W6



(a) **(b)**

Figure 10.7: Experiment W6, testing EVS-WB mixed contents and music under impaired channel conditions including delay/jitter profiles (a) with Swedish language and (b) with German language

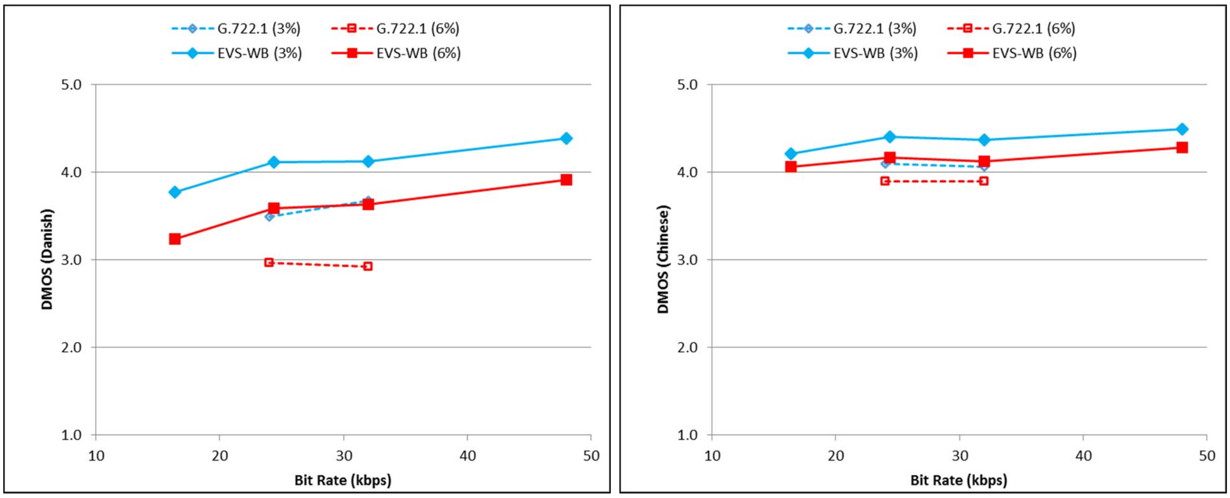


(a) (b)

Figure 10.8: Experiment W6, testing EVS-WB mixed contents and music under impaired channel conditions including delay/jitter profiles (a) with Swedish language and (b) with German language

As illustrated in Figure 10.7, EVS-WB at a particular bit-rate offers quality which is either equivalent or better than AMR-WB at a higher bit rate for both 3% and 6% FERs. Figures 10.8 shows the significantly improved performance of EVS-WB at 13.2 kbps over AMR-WB at 15.85 kbps under varying delay/jitter conditions simulated by delay-loss profiles 1…6.

10.1.7 Experiment W7

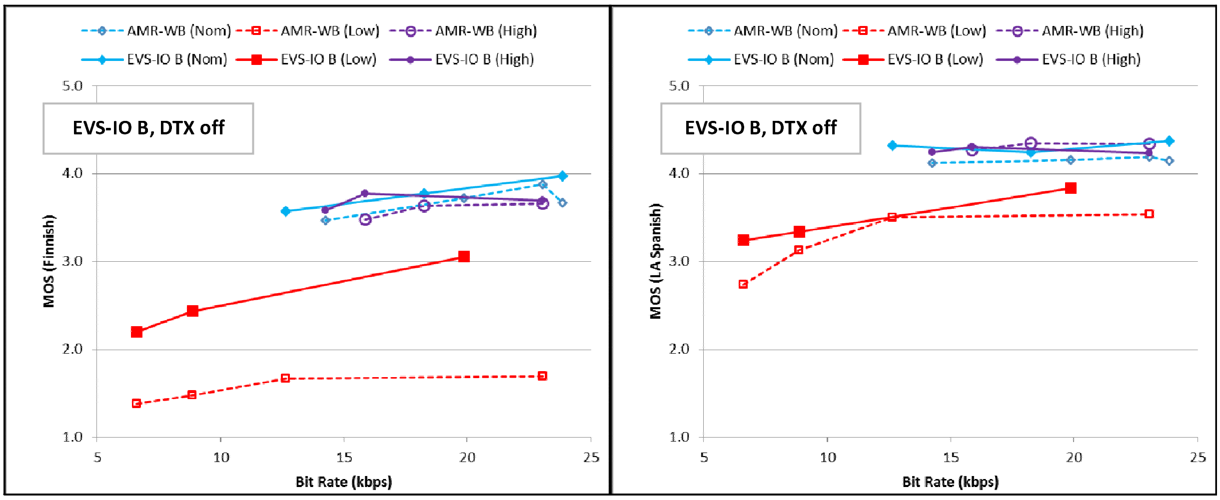


(a) (b)

Figure 10.9: Experiment W7, testing EVS-WB mixed contents and music under impaired channel conditions (a) with Danish language and (b) with Chinese language

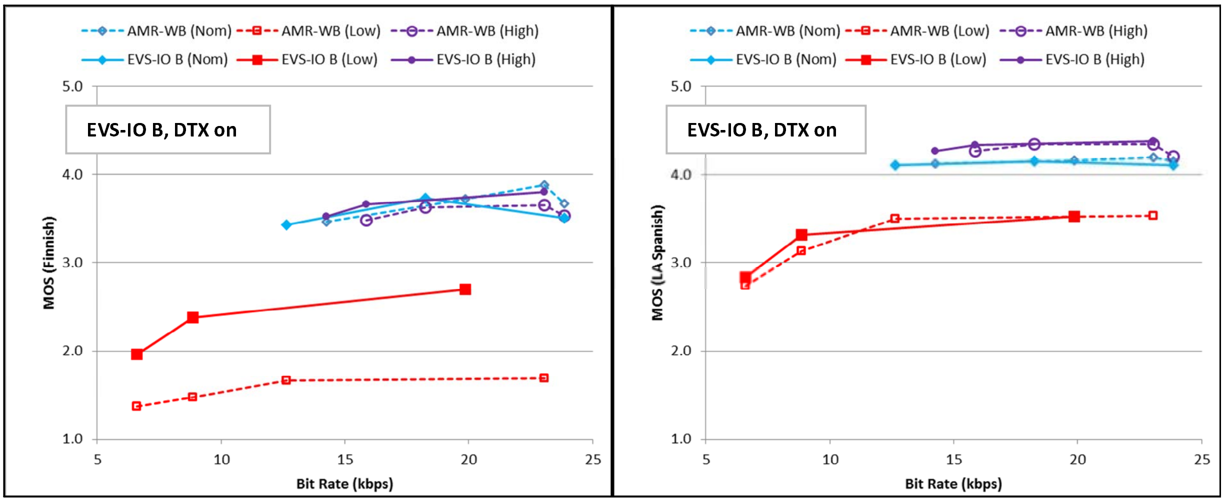
The WB mixed content and music test results of Experiment W7 in Figure 10.9 show an improved EVS performance over G.722.1 at 3% and 6% FER rates. The improvement is much more significant for Danish language than for Chinese language.

10.1.8 Experiment I1



(a) (b)

Figure 10.10: Experiment I1, testing EVS AMR-WB IO case B with clean speech under clean channel condition including input level dependency, DTX off, (a) with Finnish language and (b) with LA Spanish language



(a) (b)

Figure 10.11: Experiment I1, testing EVS AMR-WB IO case A with clean speech under clean channel condition including input level dependency, DTX on, (a) with Finnish language and (b) with LA Spanish language

The test results of Experiment I1 in Figures 10.10 and 10.11 show statistically equal performance of EVS AMR-WB IO and AMR-WB at nominal and high levels. The performance of the EVS AMR-WB IO mode and its reference, AMR-WB have shown a significant performance drop for low level speech compared to the performance for nominal and high level speech. EVS AMR-WB IO offers improved performance over AMR-WB at low level for Finnish language. However, this improvement is not observed for LA Spanish language.

10.1.9 Experiment I2

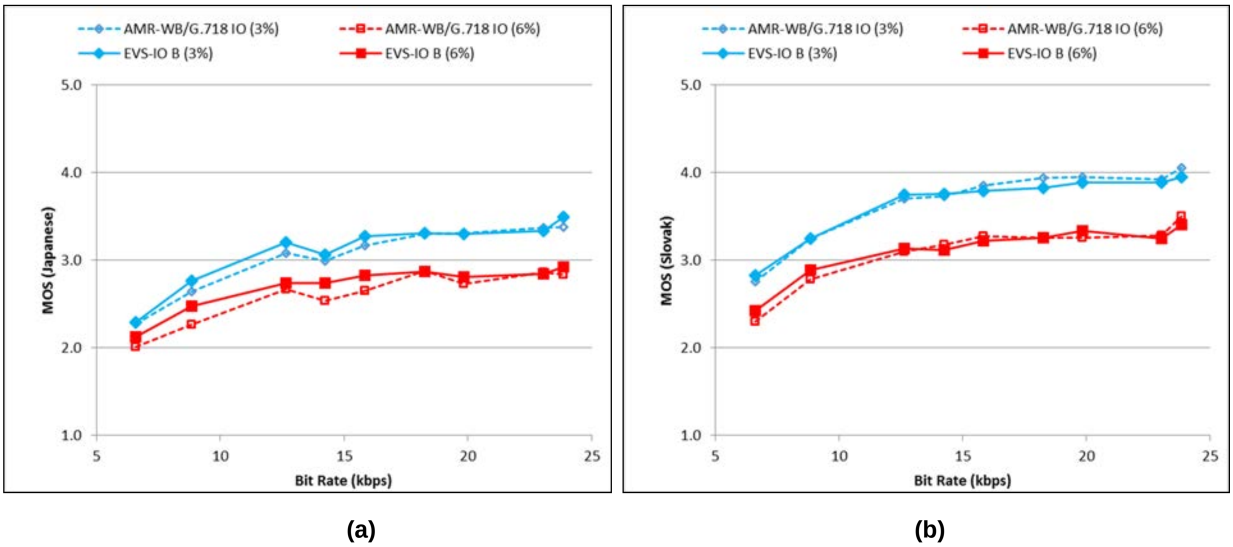


Figure 10.12: Experiment I2, testing EVS AMR-WB IO case B with clean speech under impaired channel conditions (a) with Japanese language and (b) with Slovak language

The test results of Experiment I2 in Figure 10.12 show statistically equal performance of EVS AMR-WB IO case B and AMR-WB / G.718IO.

10.1.10 Experiment I3

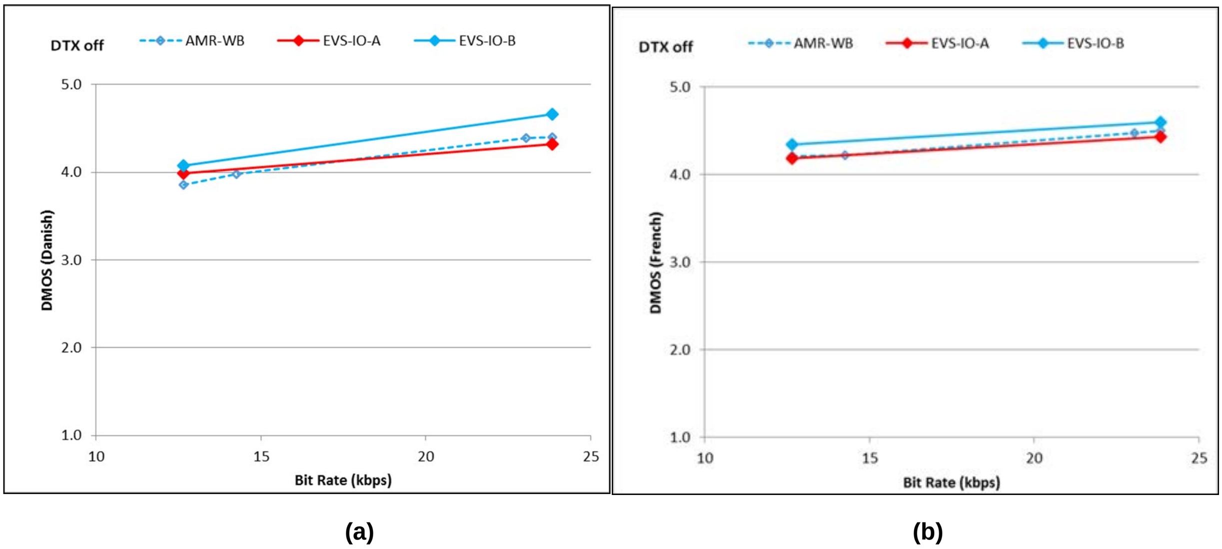
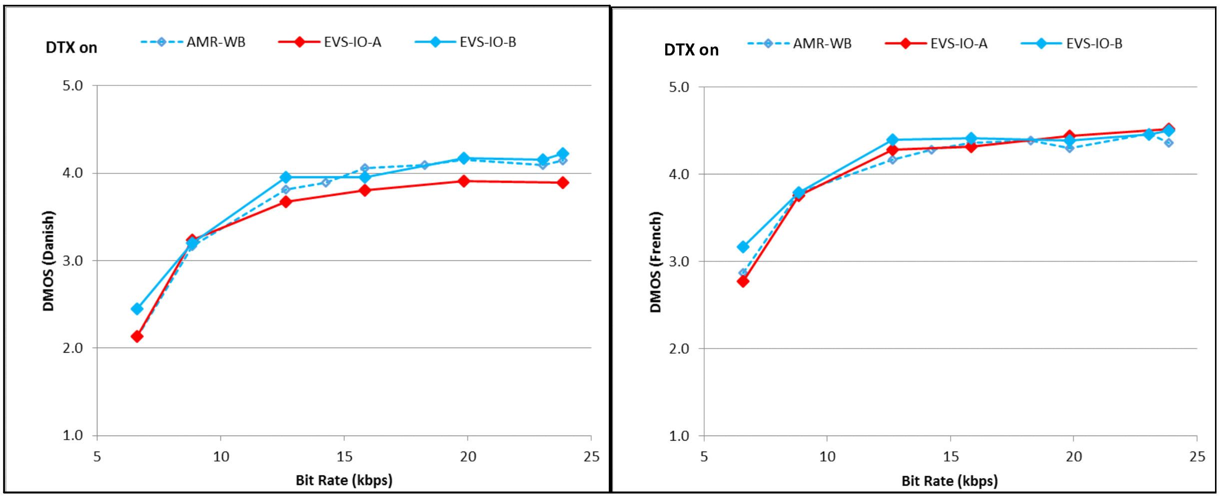


Figure 10.13: Experiment I3, testing EVS AMR-WB IO with noisy speech, DTX off, under clean channel condition (a) with Danish language and (b) with French language

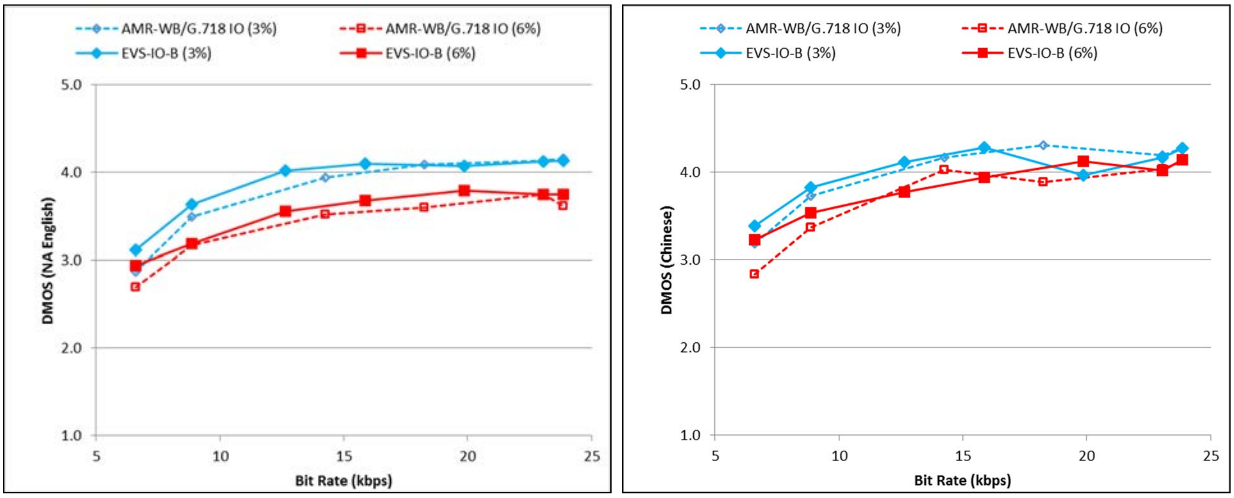


(a) (b)

Figure 10.14: Experiment I3, testing EVS AMR-WB IO with noisy speech, DTX on, under clean channel condition (a) with Danish language and (b) with French language

The test results of Experiment I3 in Figures 10.13 and 10.14 show similar performance of EVS AMR-WB IO case B and AMR-WB. For EVS AMR-WB IO case A, there is difference observed between the two languages.

10.1.11 Experiment I4

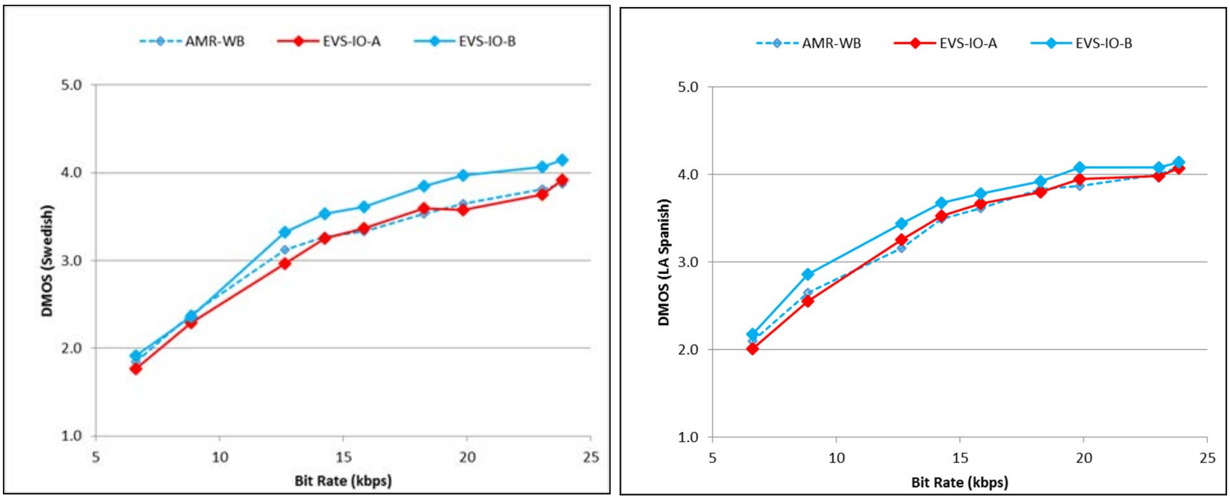


(a) (b)

Figure 10.15: Experiment I4, testing EVS AMR-WB IO with noisy speech under impaired channel conditions (a) with North American English language and (b) with Chinese language

The test results of Experiment I4 in Figures 10.15 show statistically equal performance of EVS AMR-WB IO and AMR-WB / G.718IO.

10.1.12 Experiment I5

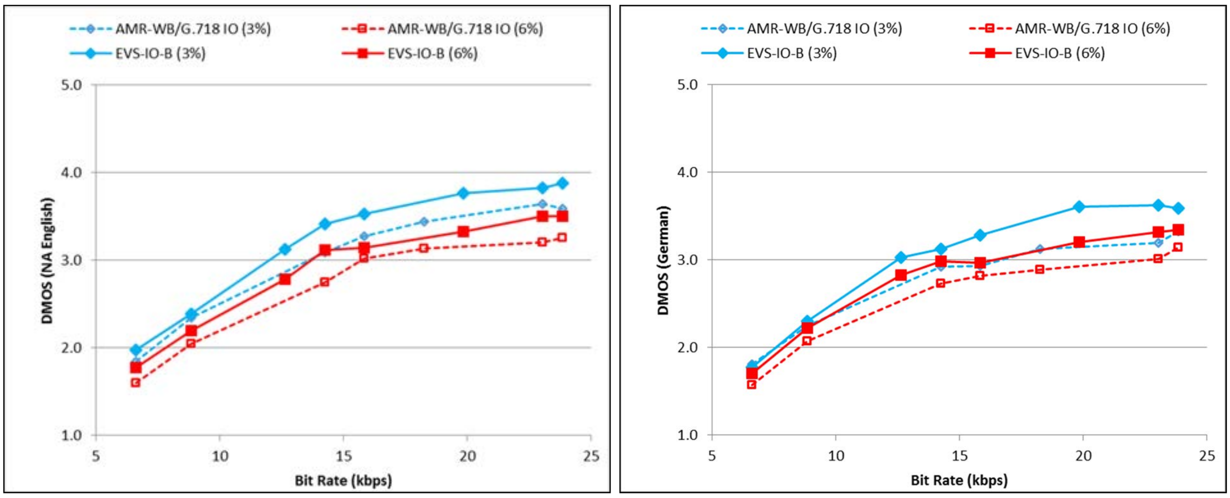


(a) (b)

Figure 10.16: Experiment I5, testing EVS AMR-WB IO case B with mixed contents and music under clean channel condition (a) with Swedish language and (b) with LA Spanish language

The test results of Experiment I5 in Figure 10.16 show equal performance of EVS AMR-WB IO and AMR-WB for Case A (EVS AMR-WB IO encoder and AMR-WB decoder). Case B (AMR-WB encoder and EVS AMR-WB IO decoder) shows overall an improvement over the AMR-WB decoder.

10.1.13 Experiment I6



(a) (b)

Figure 10.17: Experiment I6, testing EVS AMR-WB IO (case B) mixed contents and music under impaired channel conditions (a) with North American English language and (b) with German language

The test results of Experiment I6 in Figure 10.17 show equal or better performance of EVS AMR-WB IO over AMR-WB; specifically, the EVS AMR-WB IO mode (case B) shows improved performance compared to AMR-WB/G.718 IO mode for bit rates 14.25 kbps and 23.85 kbps with 3% FER. Also EVS AMR-WB IO mode (Case B) shows improved performance compared to AMR-WB/G.718 IO mode for bit rates 6.6, 8.85, 19.85, 23.05 and 23.85 kbps with 6 % FER.

10.2 WB Characterization Tests

10.2.0 List of experiments in the wideband frequency bandwidth

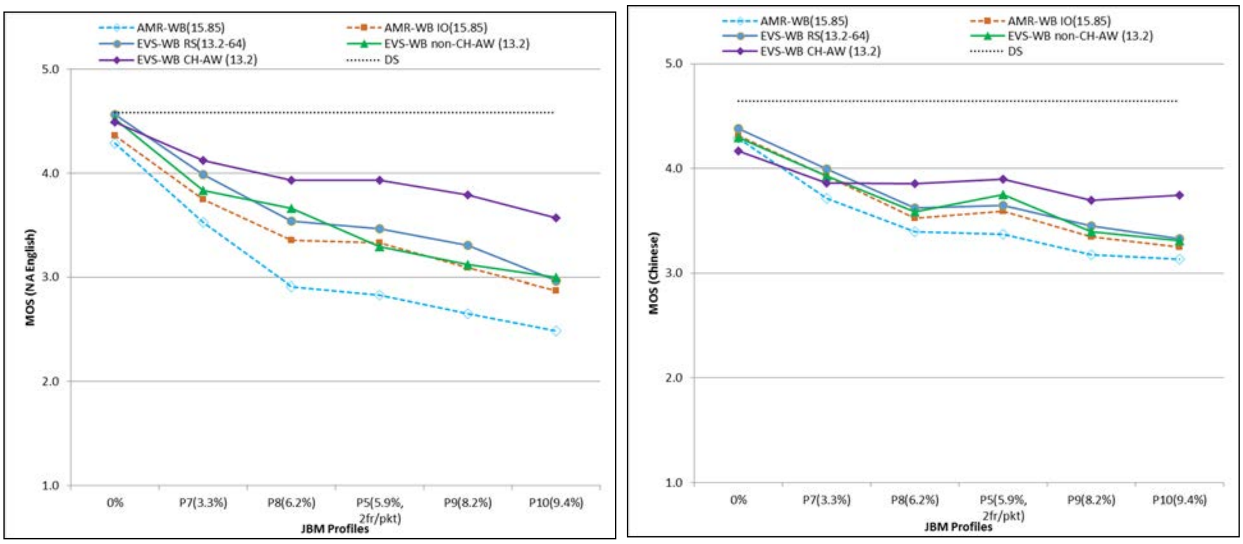
In characterization phase, five experiments, W1…W5 were designed to evaluate the performance of the EVS codec in wideband:

- Experiment W1 (ACR): clean speech in North American English and Chinese languages to evaluate rate switching and channel aware mode
- Experiment W2 (DCR): speech in Spanish and Slovak under background noise (office noise at 20 dB SNR) to evaluate rate switching
- Experiment W3 (ACR): clean speech in North American and Slovak languages to evaluate rate switching and EVS AMR-WB IO modes
- Experiment W4 (DCR): music and mixed content in North American English language to evaluate rate switching
- Experiment W5 (ACR): clean speech in Danish language to evaluate tandem and high FER conditions

In the characterization tests, with respect to impaired channels, AMR-WB was used without the G.718 IO mode concealment method to give a more complete picture of the performance.

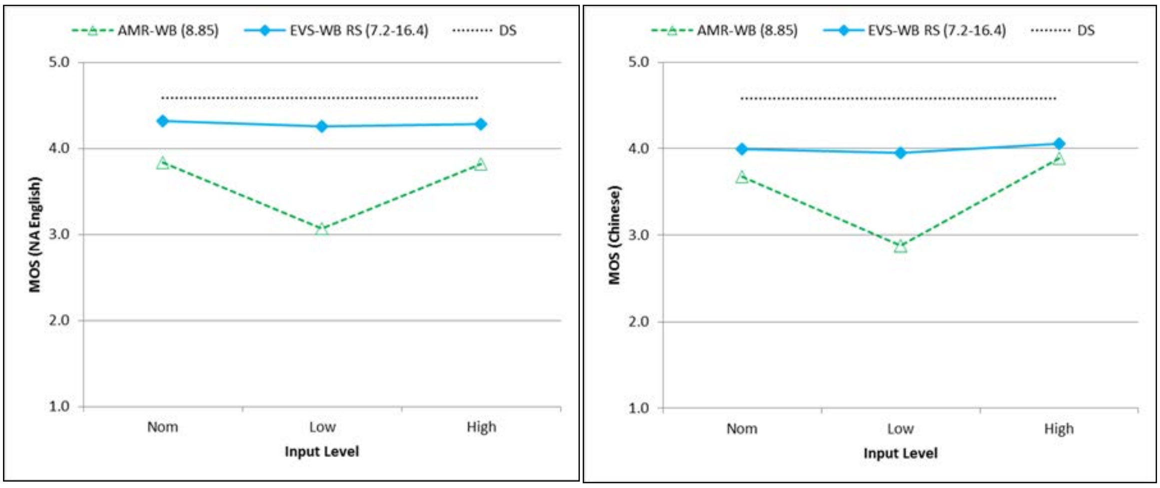
10.2.1 Experiment W1

Experiment W1 was conducted to evaluate the EVS codec WB clean speech performance under clean background conditions. Experiment W1 was conducted in two different listening labs: one in North American English, see Figure 10.18(a), 10.19(a) and another one in Chinese language, see Figure 10.18(b), 10.19(b). The codec performance was evaluated for rate switching conditions for low rates under three different input levels: nominal (-26 dBov), low (-36 dBov) and high (-16 dBov) levels as shown in Figure 10.19(a, b).



(a) **(b)**

Figure 10.18: Experiment W1, testing EVS-WB clean speech (a) with North American English language and (b) with Chinese language



(a) (b)

Figure 10.19: Experiment W1, testing EVS-WB clean speech (a) with North American English language and (b) with Chinese language

As shown in Figure 10.19 (a, b), EVS performs consistently across all input signal levels. Based on the languages tested, the W1 performance results were consistent across laboratories as well as languages tested.

Some observations from characterization experiment W1, Figures 10.18(a, b) and 10.19(a, b) include:

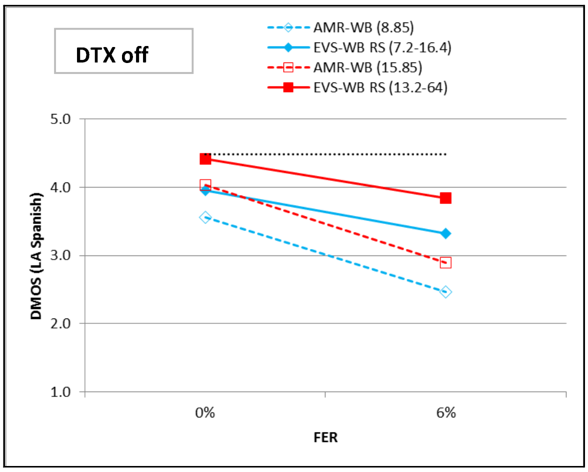
1) EVS-WB RS(13.2-64) performs comparable to 13.2 kbps in non-rate switching conditions.

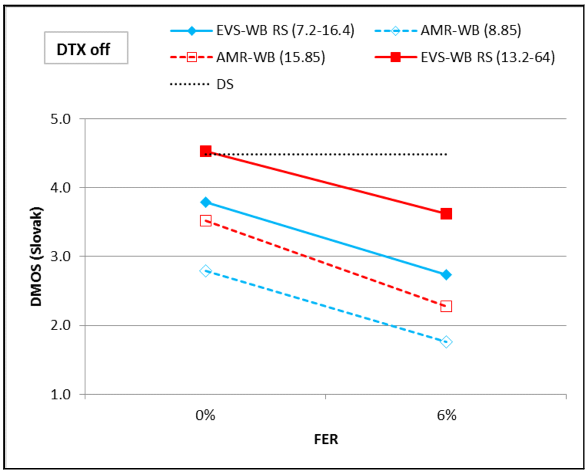
2) EVS-WB channel aware mode at 13.2 kbps demonstrates a significantly improved FER performance compared to AMR-WB and EVS AMR-WB IO at 15.85 kbps .

3) EVS-WB non-channel aware 13.2 kbps mode has shown improved performance compared to the AMR-WB 15.85 kbps mode.

10.2.2 Experiment W2

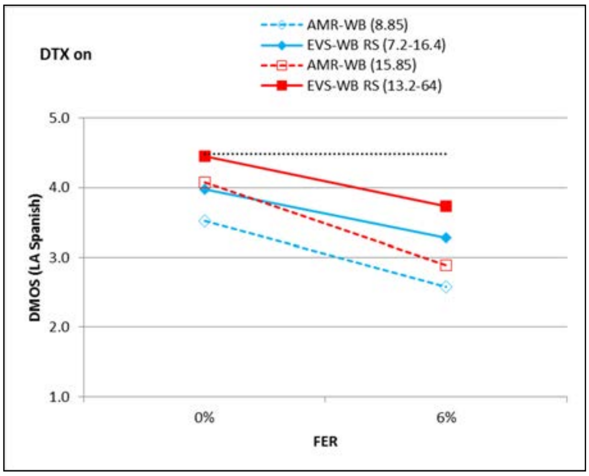
Experiment W2 was conducted to evaluate the EVS codec WB noisy speech performance under clean channel conditions to evaluate EVS codec under rate switching conditions. Experiment W2 was conducted in two different listening labs one in Latin American Spanish language, see Figure 10.20(a), 10.21(a), 10.22(a) and another in Slovak language, see Figure 10.20(b), 10.21(b), 10.22(b). The codec performance was evaluated for nominal level (-26 dBov) speech mixed with office noise with effective SNR of 20 dB.

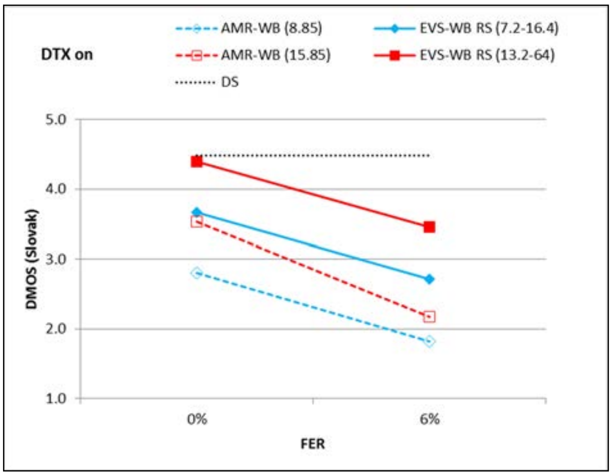




(a) (b)

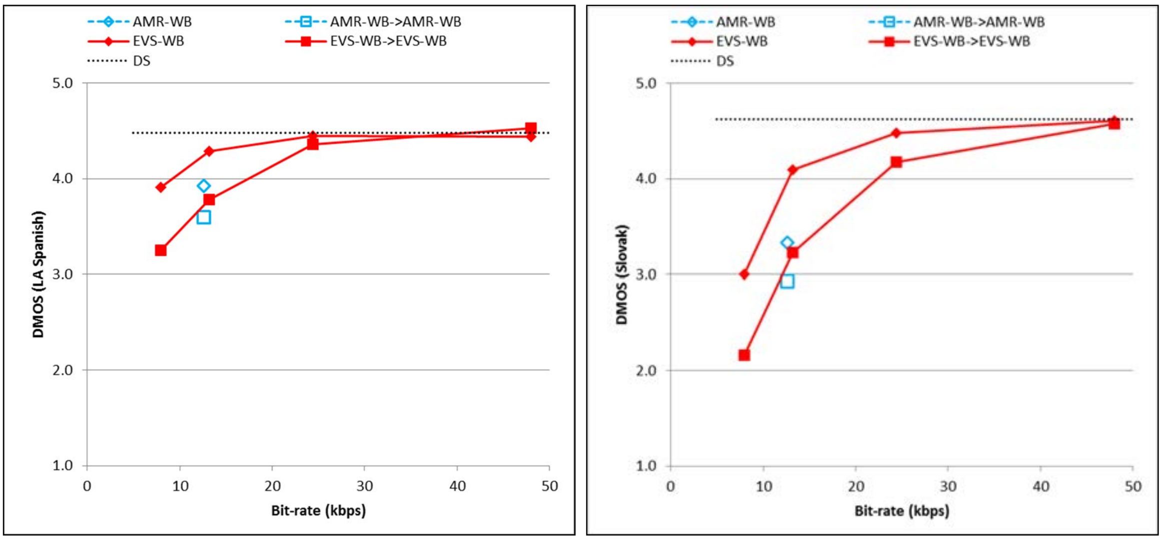
Figure 10.20: Experiment W2, testing EVS-WB noisy speech (office noise at 20 dB SNR), DTX off, (a) with LA Spanish language and (b) with Slovak language





(a) (b)

Figure 10.21: Experiment W2, testing EVS-WB noisy speech (office noise at 20 dB SNR), DTX on, (a) with LA Spanish language and (b) with Slovak language



(a) **(b)**

Figure 10.22: Experiment W2, testing EVS-WB noisy speech (office noise at 20 dB SNR) (a) with LA Spanish language and (b) with Slovak language

Based on the languages tested, the W2 performance results are consistent across laboratories as well as languages tested. Some observations from characterization experiment W2, Figures 10.20(a, b), 10.21(a, b) and 10.22(a, b) include:

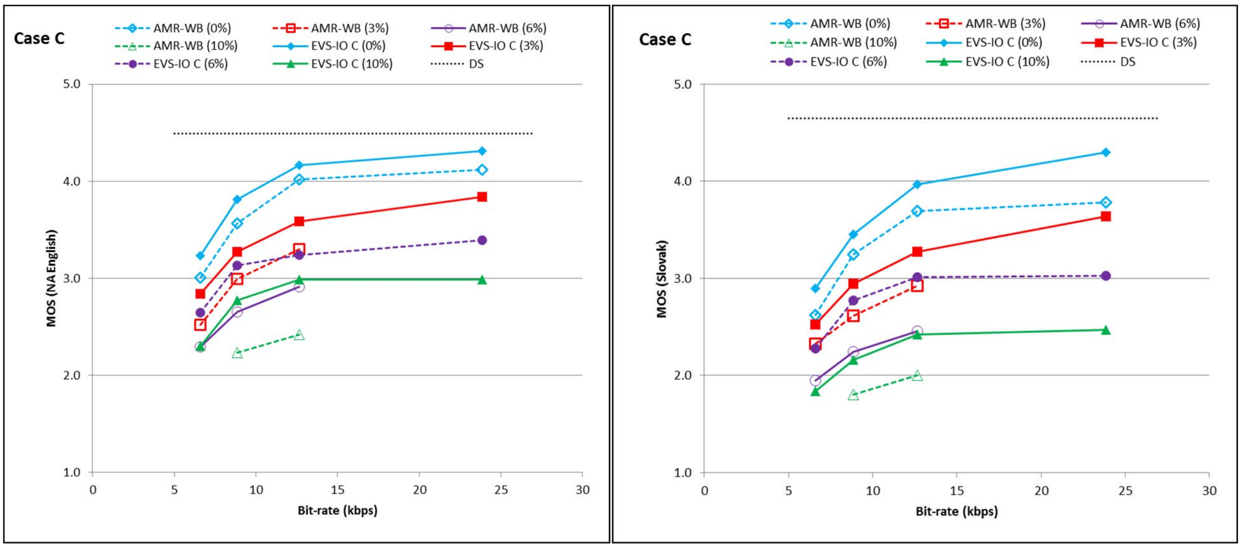
1) EVS-WB 7.2-16.4 RS has shown significant performance improvement compared to AMR-WB 8.85 kbps mode for both DTX on and DTX off conditions.

2) Similarly, EVS-WB 13.2-64 RS has shown significant performance improvement compared to AMR-WB 15.85 kbps mode for both DTX on and DTX off conditions in Figures 10.20(a,b) and 10.21(a,b).

3) EVS-WB 13.2 kbps self-tandeming condition has performed comparable to AMR-WB 12.65 kbps mode without tandeming as shown in Figures 10.22(a,b). The impact of self-tandeming to the EVS-WB codec performance is less pronounced at 24.4 kbps than at lower bit rates. At 48 kbps, the impact is negligible.

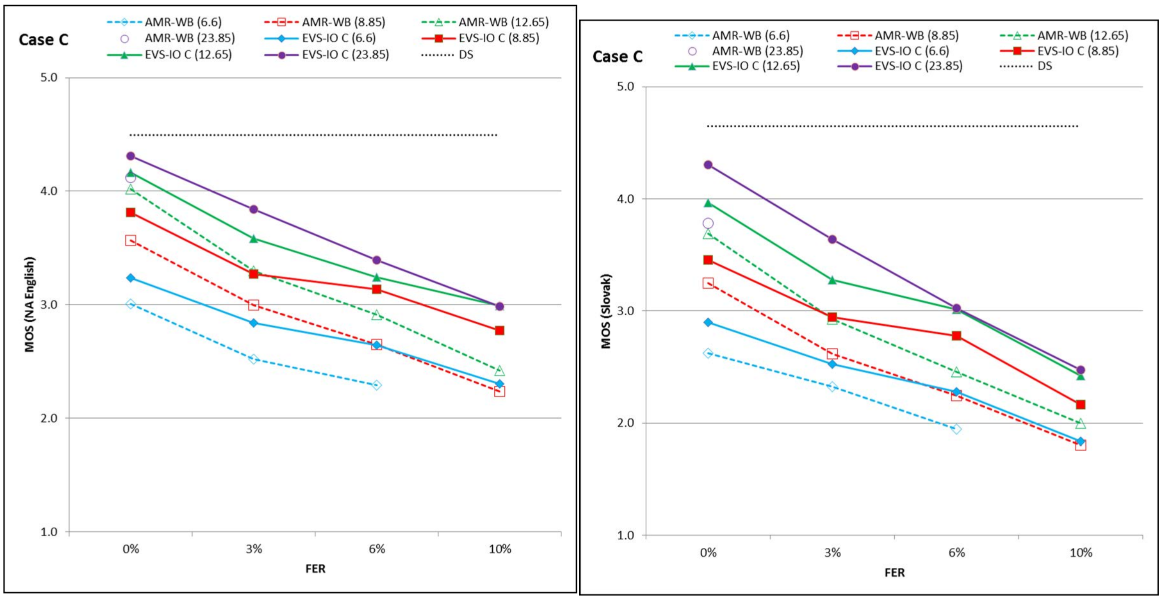
10.2.3 Experiment W3

Experiment W3 was conducted to evaluate EVS AMR-WB IO mode case C and case B clean speech performance under degraded channel conditions. Experiment W3 was conducted in two different listening labs: one in North American English language, see Figure 10.23(a), 10.24(a), 10.25(a), 10.25(a) and another one in Slovak language, see Figure 10.23(b), 10.24(b), 10.25(b), 10.25(b). The codec performance was evaluated for nominal level (-26 dBov) speech.



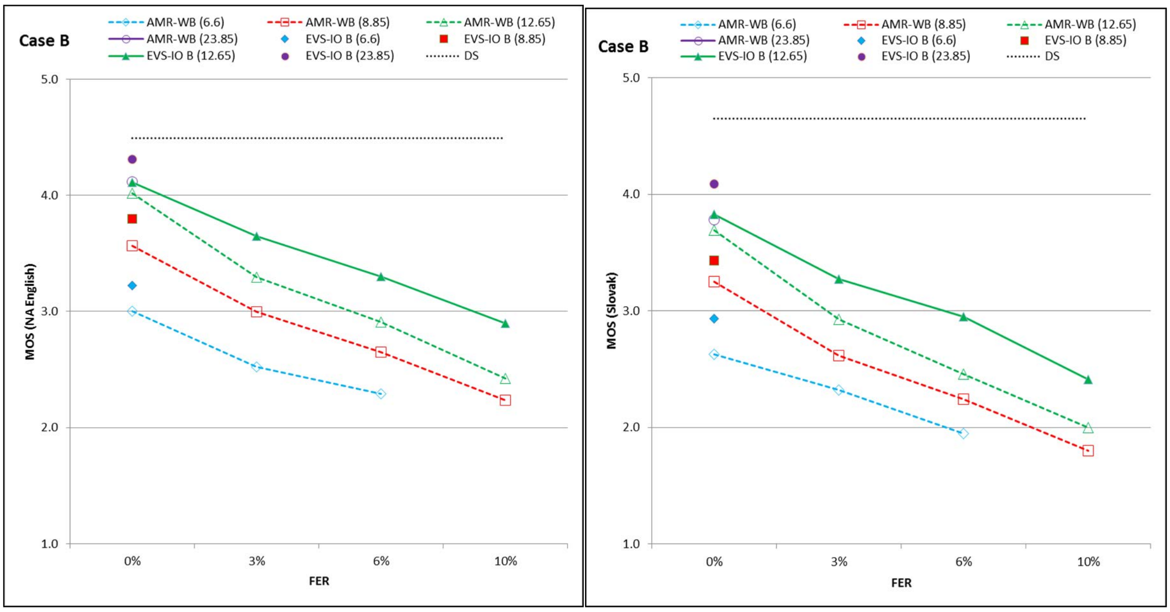
(a) (b)

Figure 10.23: Experiment W3, testing EVS-WB clean speech in EVS AMR-WB IO Case C (a) with North American English language and (b) with Slovak language



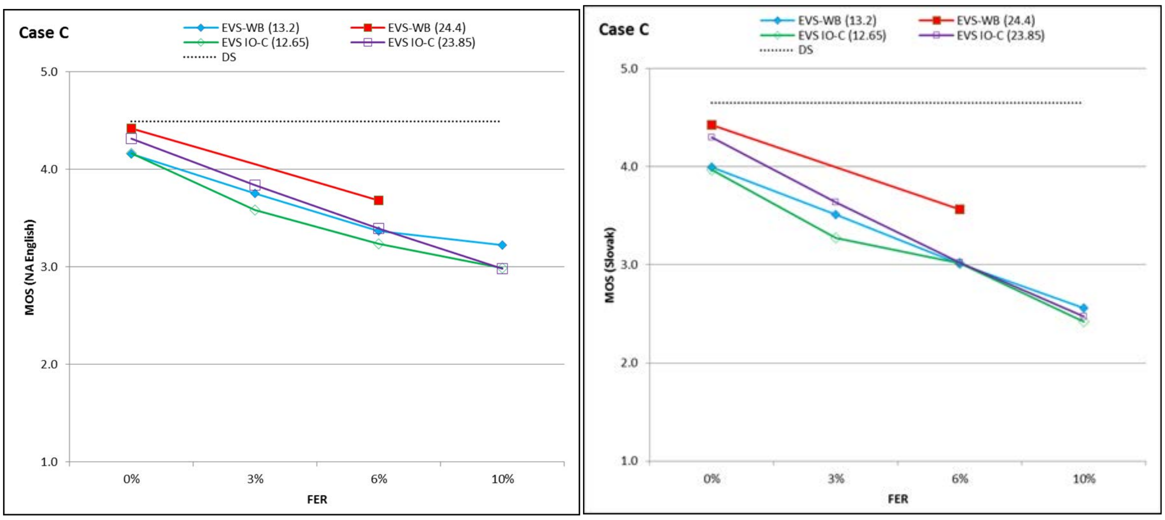
(a) (b)

Figure 10.24: Experiment W3, testing EVS-WB clean speech in EVS AMR-WB IO Case C (a) with North American English language and (b) with Slovak language



(a) (b)

Figure 10.25: Experiment W3, testing EVS-WB clean speech in EVS AMR-WB IO Case B (a) with North American English language and (b) with Slovak language



(a) (b)

Figure 10.26: Experiment W3, testing EVS-WB clean speech in EVS AMR-WB IO Case C (a) with North American English language and (b) with Slovak language

The test results of Experiment W3 in Figure 10.23, 10.24, 10.25 and 10.26 show that the EVS AMR-WB IO demonstrated improved performance compared to the AMR-WB codec at the same bit rate. Figure 10.26 shows EVS-WB has improved performance compared to EVS AMR-WB IO case C in FER conditions.

10.2.4 Experiment W4

Experiment W4 was conducted to evaluate the performance of EVS rate switching and EVS AMR-WB IO modes under clean channel conditions for music and mixed content material in North American English language. The results are shown in Figure 10.27. VBR mode is designed to achieve the average data rate (ADR) of 5.9 kbps for active speech. In order to further evaluate and confirm the performance of the VBR mode in music/mixed content, this experiment included the VBR condition. While achieving the ADR of 5.9 kbps for active speech, the VBR mode may result in a different ADR between 5.9 and 8 kbps for music/mixed content; the value of ADR was 7.52 kbps in this experiment. The codec performance was evaluated for nominal level (-26 dBov) test materials.

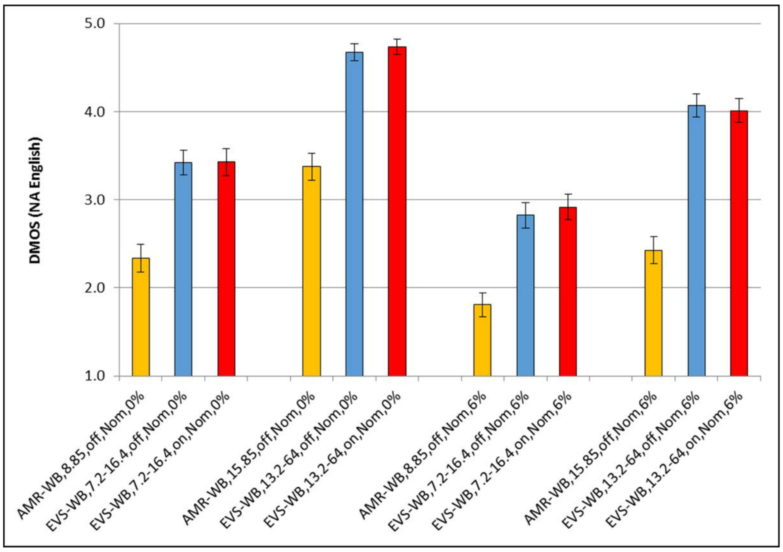
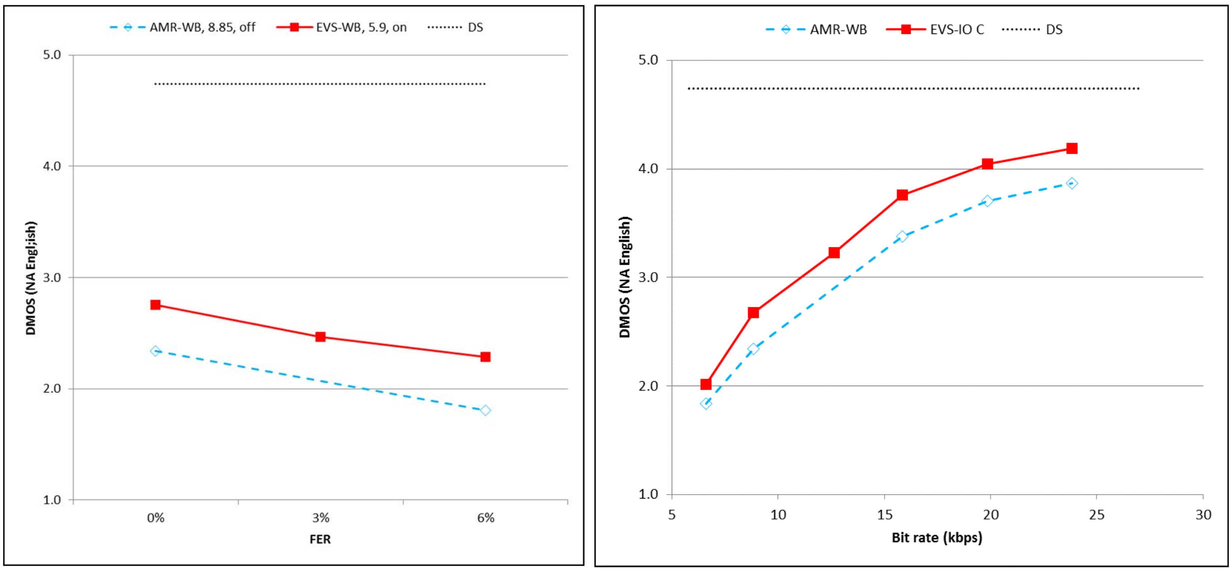


Figure 10.27: Experiment W4, testing EVS-WB music and mixed content with North American English language



(a) **(b)**

Figure 10.28: Experiment W4, testing EVS-WB music and mixed content with North American English language

The test results of Experiment W4 in Figure 10.27 and 10.28 show EVS AMR-WB IO mode demonstrated a consistent improvement over corresponding AMR-WB mode under clean, 3% and 6% FER channel error rates.

10.2.5 Experiment W5

Experiment W5 is conducted to evaluate the EVS codec performance under tandem, primary/IO-mode switching and high FER conditions. This clean speech test was conducted in Danish language, see Figure 10.28. Different tandem conditions such as AMR-WB->AMR-WB, EVS-WB→G.722, G.722→EVS-WB, G.711.1 R2b (u law) →EVS-WB, EVS-WB →G.711.1 R2b (µ-law), EVS-WB self-tandem, EVS-WB →AMR-WB and AMR-WB →EVS-WB were tested in this experiment. Furthermore, EVS primary/IO mode rate switching performance was evaluated for nominal (-26 dBov), high (-16 dBov) and low level (-36 dBov) speech signals.

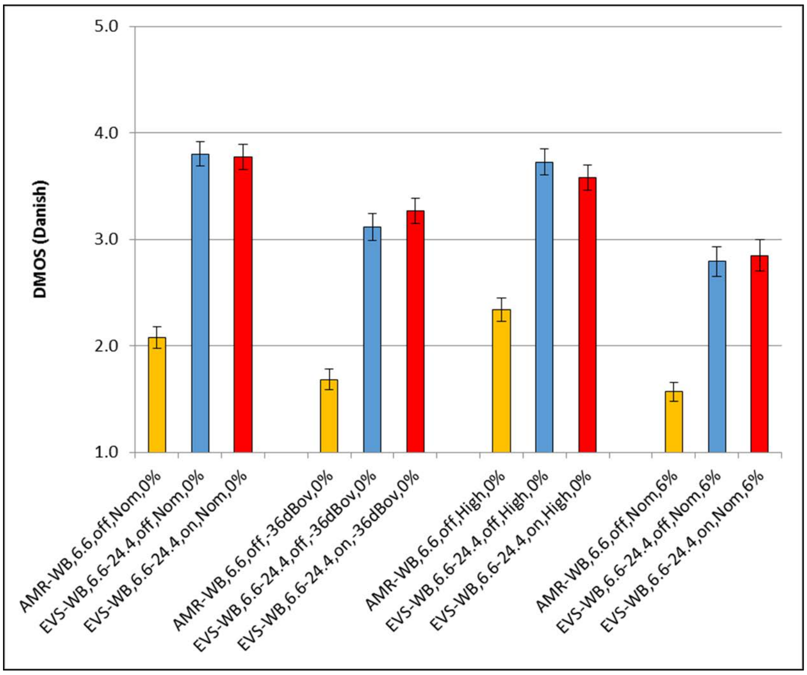


Figure 10.29: Experiment W5, testing EVS-WB clean speech with Danish language

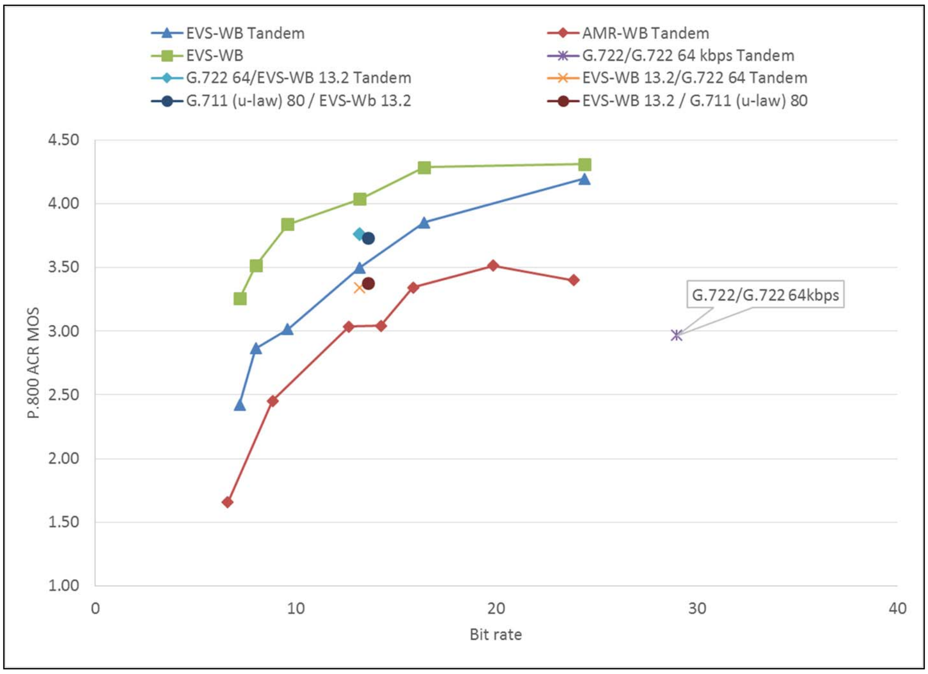


Figure 10.30: Experiment W5, testing EVS-WB clean speech with Danish language

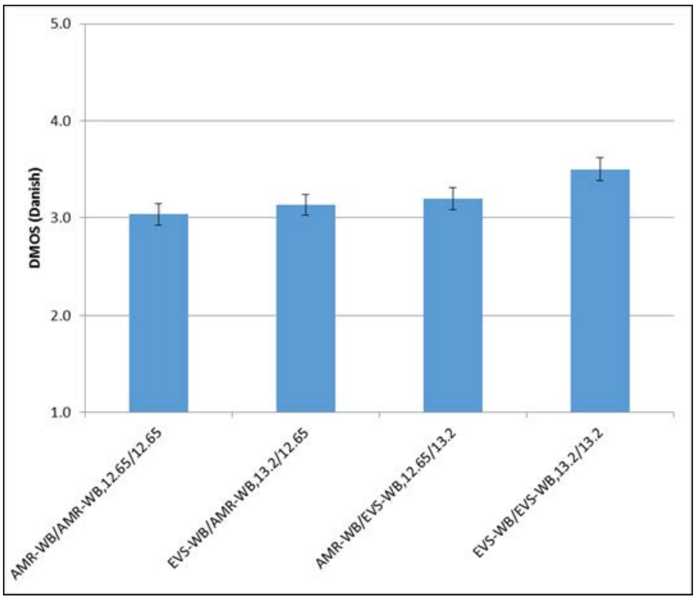


Figure 10.31: Experiment W5, testing EVS-WB clean speech with Danish language

Some observations from characterization experiment W5 (Figures 10.29, 10.30, 10.31) include:

1) When EVS primary modes are in tandem with non-EVS codecs (e.g. G.722, G.711.1), conditions with EVS –WB mode as the last codec, scores better than the cases when the last codec in the processing chain is not the EVS codec.

2) EVS-WB self-tandeming conditions perform comparable to the tandem conditions of AMR-WB codec operating at a considerably higher rate.

3) The EVS primary/IO rate switching conditions show a performance degradation for low level speech.

10.3 Conclusions on EVS Performance in Wideband

As discussed in the results shown in clauses 10.1 and 10.2, EVS codec in wideband mode demonstrates a significantly improved performance over prior codecs including AMR-WB, G.722 and G.718.

The improvement is evident across all bit rates, all rate conditions including rate switching, all tested languages, various input speech levels, different input content spanning speech / music / mixed content, as well as clean vs. noisy inputs. Furthermore, EVS demonstrates a high robust performance under channel error conditions compared to prior codecs including AMR-WB.

EVS-WB shows significant improvements over AMR-WB for all content types (clean speech, noisy speech, mixed content and music) and in clean channel as well as under impaired channel conditions. EVS is able to maintain good quality at bitrates much lower than AMR-WB, and scales to much higher qualities close to transparency for bitrates of 24.4 kbps and higher. EVS also performs much better with low level input signals.

The EVS AMR-WB-IO mode shows improvements over AMR-WB for mixed content and music and in impaired channel conditions especially when compared to AMR-WB without G.718 IO mode. In some cases, the EVS AMR-WB-IO mode also shows improvements for clean speech and for low input levels.

11 EVS Performance in Super-Wideband

11.1 SWB Selection Tests

In Selection phase, seven experiments, S1-S7, have been conducted to evaluate the performance of the EVS codec with super-wideband (SWB) content sampled at 32 kHz. While the experiments S1 through S5 used clean/noisy speech, the experiments S6 and S7 used mixed/music content for evaluating the SWB performance of the EVS codec. All the seven SWB experiments used the ITU-T P.800 DCR subjective test methodology.

- Experiment S1: SWB clean speech under clean channel condition including input level dependency
- Experiment S2: SWB clean speech under impaired channel conditions including delay/jitter profiles
- Experiment S3: SWB noisy speech (Street noise at 20 dB) under clean channel condition
- Experiment S4: SWB noisy speech (Office noise at 20 dB) under clean channel condition
- Experiment S5: SWB noisy speech (Car noise at 15 dB) under impaired channel condition
- Experiment S6: SWB mixed contents and music under clean channel condition
- Experiment S7: SWB mixed contents and music under impaired channel conditions including delay/jitter profiles

11.1.1 Experiment S1

Experiment S1 is conducted to evaluate the EVS codec SWB clean speech performance under clean channel conditions at three different active input levels, namely, 1) nominal level at -26 dBov, 2) low level at -36 dBov, and 3) high level at -16 dBov. Experiment S1 is conducted in two different listening labs in North American English language, see Figure 11.1(a) and in French language, see Figure 11.1(b).

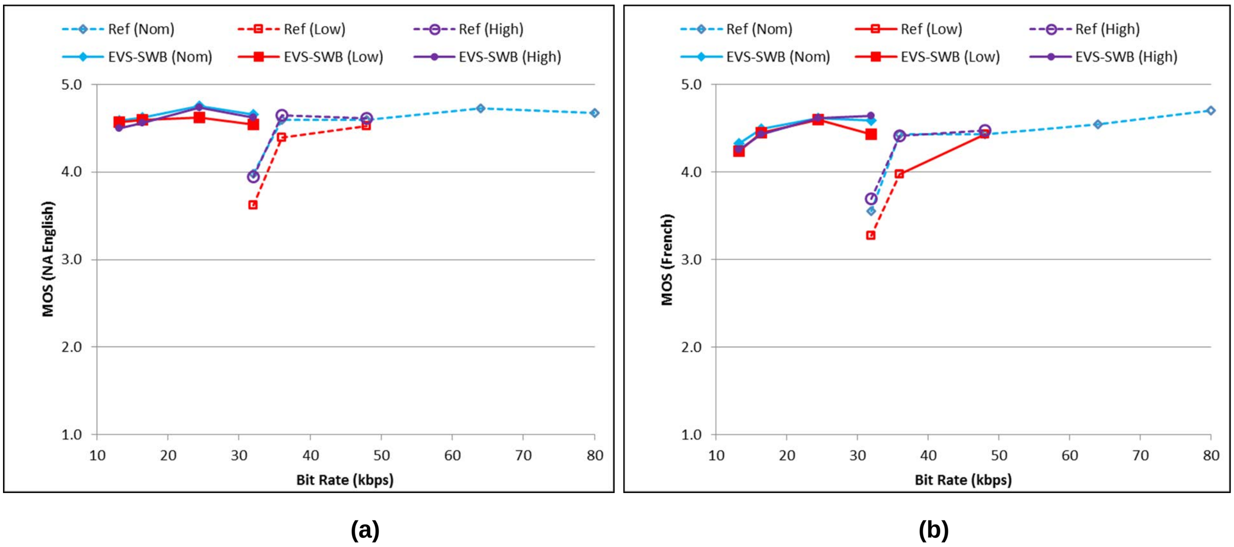


Figure 11.1: Experiment S1, testing EVS-SWB clean speech under clean channel condition at three different active input levels (nominal -26 dBov, low -36 dBov, and high -16 dBov). (a) North American English language, (b) French language.

Figure 11.1(a, b) shows the EVS codec SWB performance at bit rates of 13.2, 16.4, 24.4, and 32 kbps. In both the languages tested, the SWB performance results are consistent and furthermore comparable across the three different input signal levels (Nominal: blue, Low: red, and High: purple shown as markers with solid fill). In particular, it is observed that there is negligible performance degradation due to operating at a wider range of active input signal levels from -16 dBov to -36 dBov. The Reference SWB codecs shown in Figure 11.1(a,b) include G.722.1C at 32 kbps, G.718B at 36 kbps, and G.719 at 48, 64, and 80 kbps.

Some observations from Selection Experiment S1, Figure 11.1(a,b) include:

1) The EVS-SWB clean speech quality at 13.2 kbps is comparable to that of G.718B at 36 kbps and G.719 at 48 kbps.

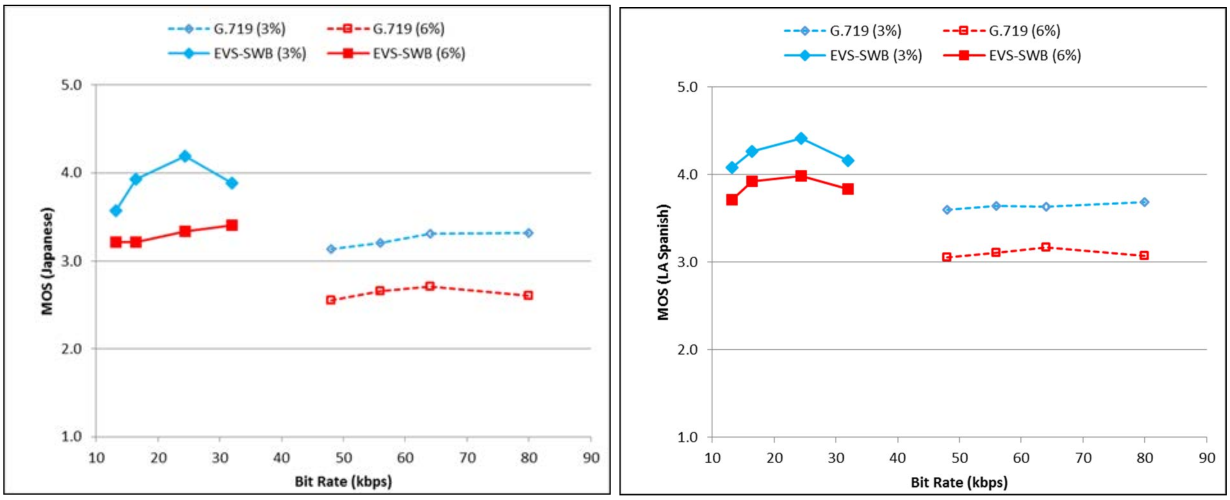
2) The SWB clean speech quality at 24.4 kbps already approaches that of the Direct Source quality based on the ITU-T P.800 DCR test methodology.

3) The EVS-SWB clean speech test shows that not only the performance requirements are met but it also shows that the codec exceeds the EVS selection performance objectives. Please refer to Annex A for ToR performance (Table A.1 and Table A.3) of SWB clean speech quality in EVS Selection Experiment S1.

11.1.2 Experiment S2

Experiment S2 is conducted to evaluate the EVS codec SWB clean speech performance under impaired channel conditions. In particular, the EVS-SWB codec performance at frame erasure rates (FERs) of 3% and 6% is evaluated in two different labs using the Japanese language (Figure 11.2(a)) and Spanish language (Figure 11.2(b)).

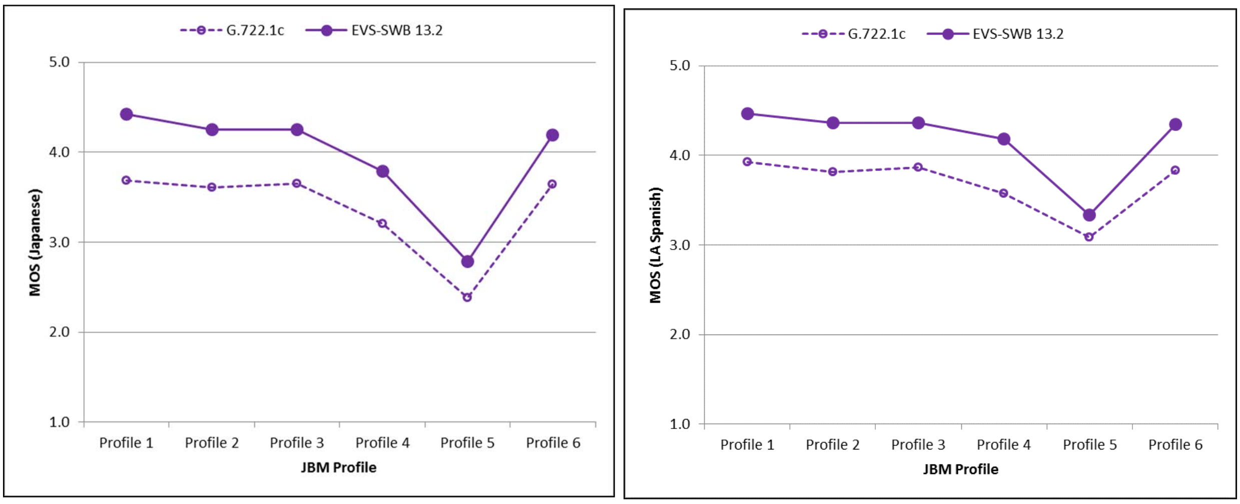
Figure 11.2(a, b) shows the EVS codec SWB performance at bit rates of 13.2, 16.4, 24.4, and 32 kbps. In both the languages, the EVS-SWB performance is significantly better than the Reference codecs used in the EVS Selection Experiment S2. The Reference SWB codecs shown in Figure 11.2(a,b) include G.719 at 48, 56, 64, and 80 kbps.



(a) (b)

Figure 11.2: Experiment S2, testing EVS-SWB clean speech under impaired channel condition with frame erasure rates of 3% and 6%. (a) Japanese language, (b) LA Spanish language.

Figure 11.3(a, b) shows the EVS codec SWB performance at 13.2 kbps under six different delay/loss profiles simulating impaired channel characteristic with varying delay and jitter as specified in the 26.114 specification. In both the languages, i.e., in Japanese Figure 11.3(a) and in Spanish (Figure 11.3(b)), the EVS-SWB performance is significantly better than the Reference codec. The Reference SWB codec shown in Figure 11.3(a, b) is G.722.1C at 32 kbps.



(a) (b)

Figure 11.3: Experiment S2, testing EVS-SWB clean speech under impaired channel conditions with frame erasures introduced through Delay/Loss Profiles 1-6. (a) Japanese language, (b) LA Spanish language.

Some observations from Selection Experiment S2, Figure 11.2(a, b) and Figure 11.3(a, b) include:

1) Under the frame erasure rate of 3%, the EVS-SWB clean speech performance at 13.2 kbps already is better than the G.719 at 80 kbps. Similar trend is seen at frame erasure rate of 6% where EVS-SWB clean speech performance at 13.2 kbps is significantly better than the G.719 at 80 kbps.

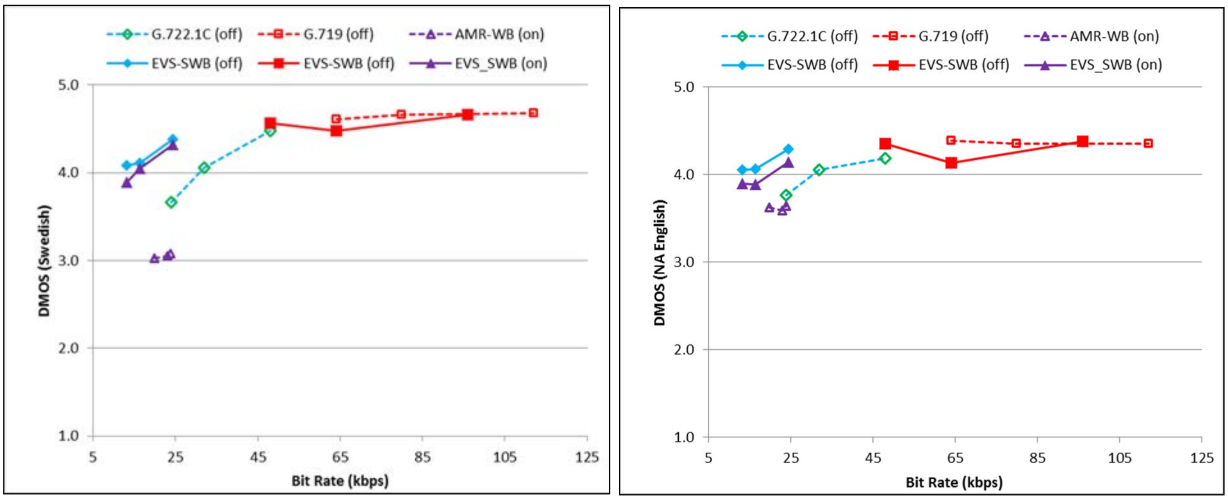
2) Under the frame erasure rate of 6%, the EVS-SWB clean speech performance is comparable or better than the performance of G.719 at 3% FER for all tested operating points.

3) The EVS-SWB clean speech test under impaired channel conditions (FER, delay-loss profiles) shows that not only the performance requirements are met but it also shows that the codec exceeds the EVS selection performance objectives. Please refer to Annex A for ToR performance (Table A.1 and Table A.3) of SWB clean speech quality under impaired channel in EVS Selection Experiment S2. There is a steady progression of quality increase from 13.2 kbps to 24.4 kbps in EVS. (In the standardized code a fixed-point implementation bug at 32 kbps was removed but it was still present in selection testing affecting the performance at 32 kbps).

11.1.3 Experiment S3

Experiment S3 is conducted to evaluate the EVS codec SWB noisy speech performance under Street noise that is mixed at 20 dB SNR. In particular, the EVS-SWB codec noisy speech performance under DTX on/off conditions is evaluated in two different labs using the Swedish language (Figure 11.4(a)) and North American English language (Figure 11.4(b)).

Figure 11.4(a, b) shows the EVS codec SWB noisy speech performance at bit rates of 13.2, 16.4, 24.4, 48, 64, and 96 kbps. In both the languages, the EVS-SWB performance is better than the Reference codecs. The three Reference codecs shown in Figure 11.4(a,b) include G.722.1C at 24, 32, and 48 kbps under DTX off; G.719 at 64, 80, and 112 kbps under DTX off; and AMR-WB at 19.85, 23.05, and 23.85 kbps under DTX on.



(a) (b)

Figure 11.4: Experiment S3, testing EVS-SWB noisy speech (Street Noise, at 20 dB) under clean channel condition. (a) Swedish language, (b) North American English language.

Some observations from Selection Experiment S3, Figure 11.4(a, b) include:

1) Under DTX off conditions (shown as "(off)" in the legend), the EVS-SWB noisy speech performance at 13.2 and 16.4 kbps is better than G.722.1C at 24 kbps and comparable to that of G.722.1C at 32 kbps. The EVS-SWB noisy speech performance at 24.4 kbps is comparable to that of G.722.1C at 48 kbps, and starts approaching the DMOS region of saturation and "Direct Source" quality.

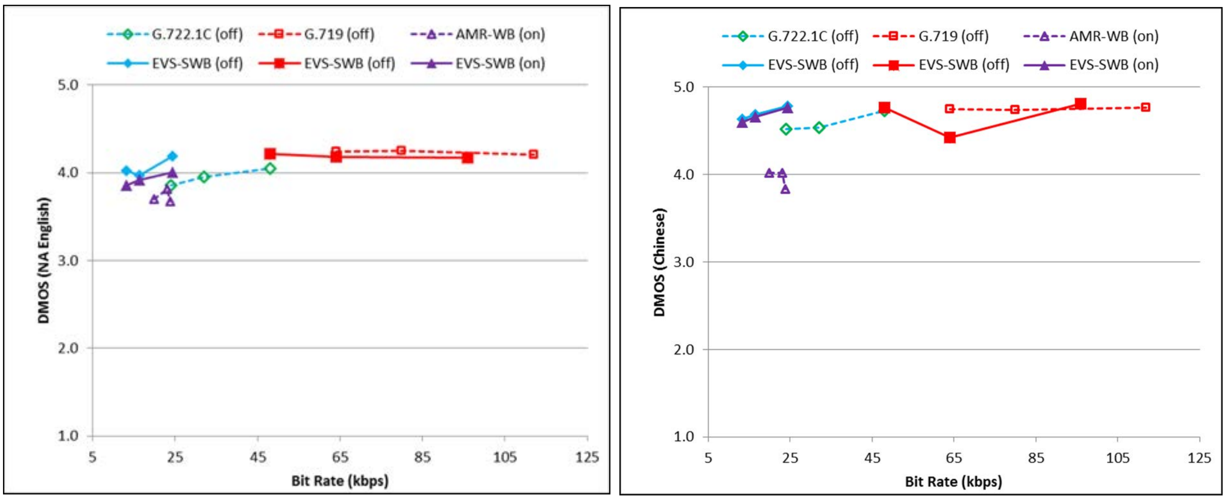
2) Under DTX off conditions (shown as "(off)" in the legend) at higher bit rates, i.e., 48 kbps, 64 kbps, and 96 kbps, the EVS-SWB codec performance is in the DMOS saturation region close to the "Direct Source." EVS codec data point at 64 kbps had a fixed point implementation bug that shows a significant quality drop (e.g. in North American English language) in selection testing that was subsequently corrected in characterization testing (see clause 12.1.2).

3) Under DTX on conditions (shown as "(on)" in the legend), the EVS-SWB codec noisy speech performance (at 13.2, 16.4, and 24.4 kbps) is benchmarked against the AMR-WB (19.85, 23.05, and 23.85 kbps). As can be seen from Figure 11.4(a, b) the EVS-SWB DTX performance is significantly better (more than about 0.5 DMOS) than the AMR-WB DTX performance.

4) There is a steady progression of quality increase from 13.2 kbps to 24.4 kbps in EVS-SWB noisy speech performance and tends to approach the region of transparency at higher bit rates.

11.1.4 Experiment S4

Experiment S4 is conducted to evaluate the EVS codec SWB noisy speech performance under office noise that is mixed at 20 dB SNR. In particular, the EVS-SWB codec noisy speech performance under DTX on/off conditions is evaluated in two different labs using the North American English language (Figure 11.5(a)) and Chinese language (Figure 11.5(b)).



(a) (b)

Figure 11.5: Experiment S4, testing EVS-SWB noisy speech (Office Noise, at 20 dB) under clean channel condition. (a) North American English language, (b) Chinese language.

Figure 11.5(a, b) shows the EVS codec SWB noisy speech performance at bit rates of 13.2, 16.4, 24.4, 48, 64, and 96 kbps. In both the languages, the EVS-SWB performance is better than the Reference codecs at lower bit rates. The three Reference codecs shown in Figure 11.4(a,b) include G.722.1C at 24, 32, and 48 kbps under DTX off; G.719 at 64, 80, and 112 kbps under DTX off; and AMR-WB at 19.85, 23.05, and 23.85 kbps under DTX on.

Similar observations from Selection Experiment S3 can be extended to Experiment S4 also:

1) Under DTX off conditions (shown as "(off)" in the legend in Figure 11.5(a, b)), the EVS-SWB noisy speech performance at 13.2 and 16.4 kbps is better than G.722.1C at 24 kbps and comparable to that of G.722.1C at 32 kbps. The EVS-SWB noisy speech performance at 24.4 kbps is comparable to that of G.722.1C at 48 kbps, and starts approaching the DMOS region of saturation and "Direct Source" quality.

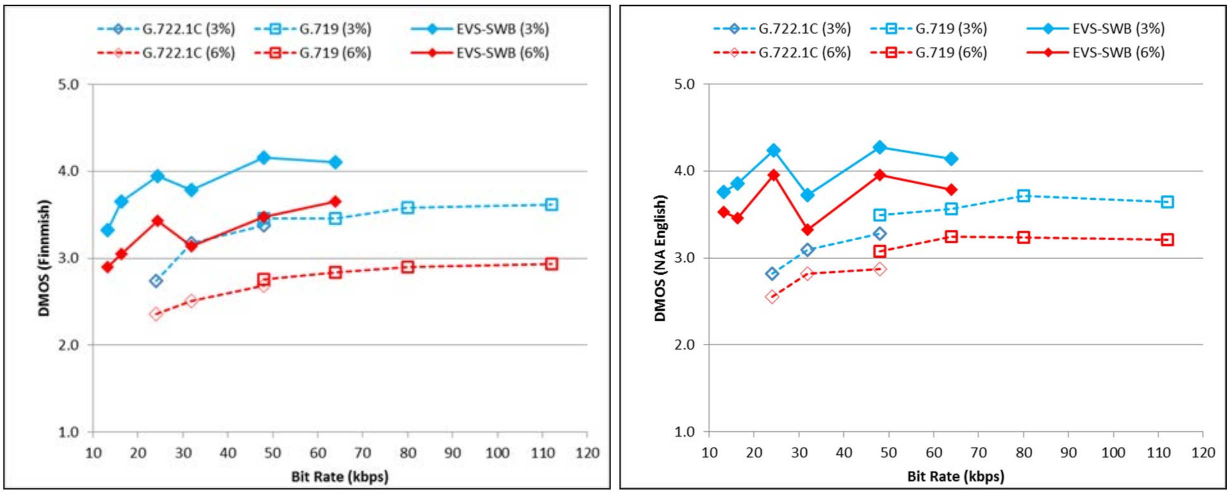
2) At higher bit rates, i.e., 48 kbps and 96 kbps, the EVS-SWB codec performance is in the DMOS saturation region close to the "Direct Source." The EVS-SWB at 64 kbps had a fixed-point implementation bug in selection testing which was eventually corrected in characterization testing (see clause 12.1.2).

3) Under DTX on conditions (shown as "(on)" in the legend in Figure 11.5(a, b)), the EVS-SWB codec noisy speech performance (at 13.2, 16.4, and 24.4 kbps) is benchmarked against the AMR-WB (19.85, 23.05, and 23.85 kbps). As can be seen from Figure 11.5(a, b) the EVS-SWB DTX performance is significantly better (more than about 0.5 DMOS) than the AMR-WB DTX performance.

11.1.5 Experiment S5

Experiment S5 is conducted to evaluate the EVS codec SWB noisy speech performance (with Car noise mixed at 15 dB SNR) under impaired channel conditions. In particular, the EVS-SWB codec performance at frame erasure rates (FERs) of 3% and 6% is evaluated in two different labs using the Finnish language (Figure 11.6(a)) and North American English language (Figure 11.6(b)).

Figure 11.6(a, b) shows the EVS codec SWB performance at bit rates of 13.2, 16.4, 24.4, 32, 48, and 64 kbps. In both the languages, the EVS-SWB performance is significantly better than the Reference codecs. The Reference SWB codecs shown in Figure 11.2(a,b) include G.722.1C at 24, 32, and 48 kbps; and G.719 at 48, 56, 64, 80, and 112 kbps.



(a) **(b)**

Figure 11.6: Experiment S5, testing EVS-SWB noisy speech (Car Noise, at 15 dB) under impaired channel condition. (a) Finnish language, (b) North American English language.

Some observations from Selection Experiment S5, Figure 11.6(a, b) include:

1) Under the frame erasure rate of 3 %, the EVS-SWB noisy speech performance at 13.2 kbps already is comparable to that of G.719 at 80 kbps. Similar trend is seen at frame erasure rate of 6 % where EVS-SWB noisy speech performance at 13.2 kbps is comparable to that of G.719 at 80 kbps.

2) There is a steady progression of quality increase from 13.2 kbps to 24.4 kbps in EVS, while meeting not only the performance requirements (PRs) but also exceeding the EVS Selection performance objectives in Experiment S5. Please refer to Annex A for ToR performance (Table A.1 and Table A.3) of SWB noisy speech quality under impaired channel in EVS Selection Experiment S5.

3) EVS codec data point at 32 kbps and 64 kbps had a fixed point implementation bug in Selection testing that was corrected in Characterization testing (See clause 11.2.2 and clause 12.1.2).

11.1.6 Experiment S6

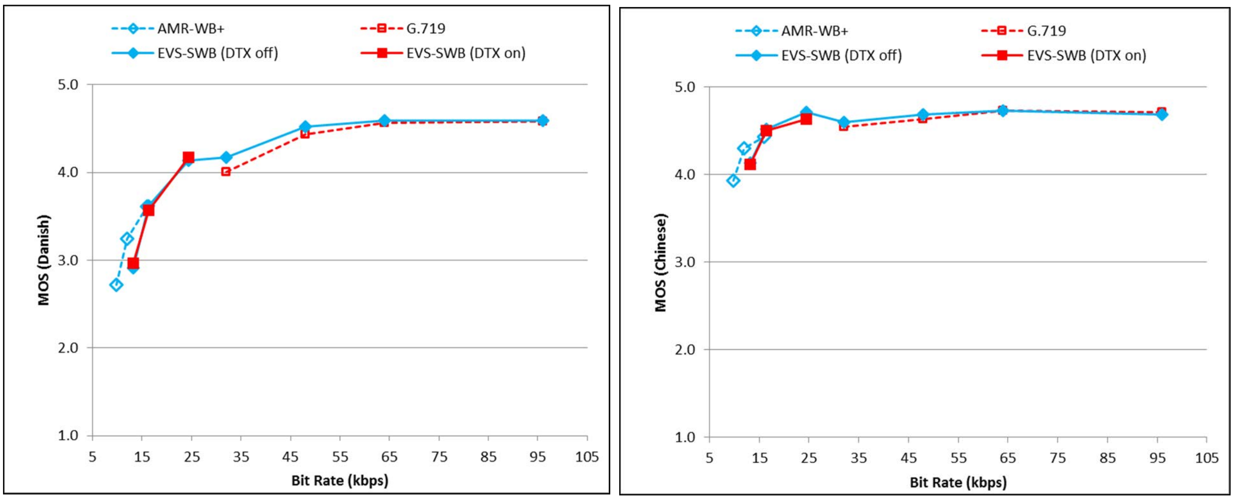
Experiment S6 is conducted to evaluate the EVS codec SWB mixed/music performance. In particular, the EVS-SWB codec performance is evaluated in two different labs using the Danish mixed/music content (Figure 11.7(a)) and Chinese mixed/music content (Figure 11.7(b)).

Figure 11.7(a, b) shows the EVS codec SWB performance at bit rates of 13.2-96 kbps. In both Danish and Chinese mixed/music content, the EVS-SWB performance is comparable to the Reference codecs. The Reference SWB codecs shown in Figure 11.7(a,b) include AMR-WB+ at 9.75, 12, and 16 kbps; and G.719 at 32, 48, 64, and 96 kbps.

Some observations from Selection Experiment S6, Figure 11.7(a,b) include:

1) The EVS-SWB mixed/music quality at lower bit rates (13.2, 16.4, and 24.4 kbps) is better than the high-delay codec AMR-WB+ (at 9.75, 12, and 16 kbps). At higher bit rates (32-96 kbps) the SWB mixed/music performance is comparable to that of the G.719 at the same bit rate (32-96 kbps).

2) There is negligible performance impact between DTX off and DTX on conditions at lower bit rates.



(a) (b)

Figure 11.7: Experiment S6, testing EVS-SWB mixed/music performance. (a) Danish mixed/music, (b) Chinese mixed/music

11.1.7 Experiment S7

Experiment S7 is conducted to evaluate the EVS codec SWB mixed/music performance under impaired channel conditions. In particular, the EVS-SWB codec performance at frame erasure rates (FERs) of 3% and 6% are evaluated in two different labs using the North American English mixed/music content (Figure 11.8(a)) and German mixed/music content (Figure 11.8(b)).

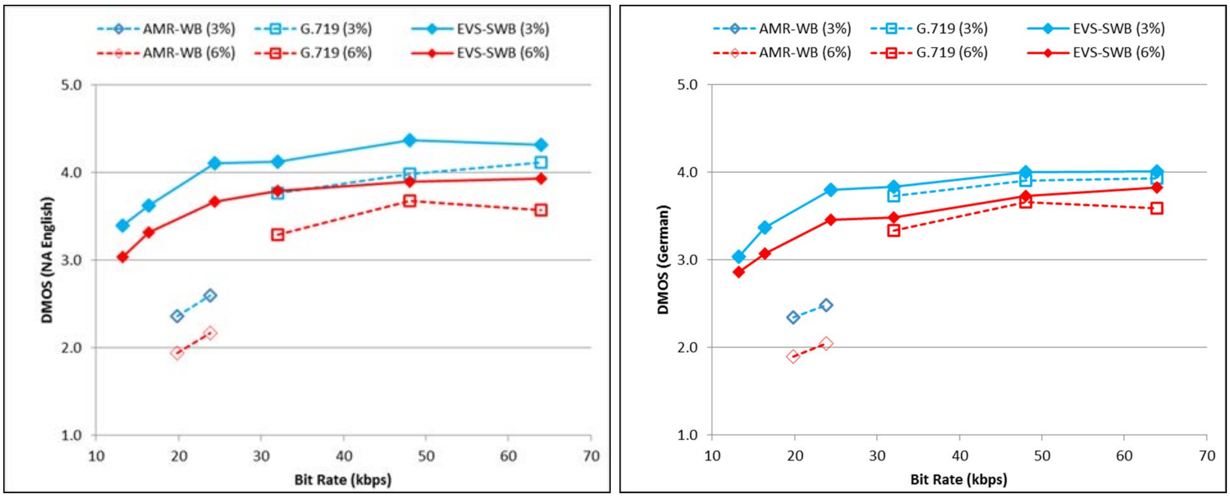
Figure 11.8(a, b) shows the EVS codec SWB performance at bit rates of 13.2-64 kbps. In both the NAE and German mixed/music content, the EVS-SWB performance is better than the Reference codecs. The Reference SWB codecs shown in Figure 11.8(a,b) include AMR-WB at 19.85, 23.85 kbps; and G.719 at 32, 48, and 64 kbps.

Figure 11.9(a, b) shows the EVS codec SWB performance at 24.4 kbps under six different delay/loss profiles simulating impaired channel characteristic with varying delay and jitter as specified in the 26.114 specification. In both the NAE and German mixed/music content, the EVS-SWB performance is better than the Reference codec. The Reference SWB codec shown in Figure 11.9(a,b) is G.722.1C at 24 kbps.

Some observations from Selection Experiment S7, Figure 11.8(a,b) and Figure 11.9(a,b) include:

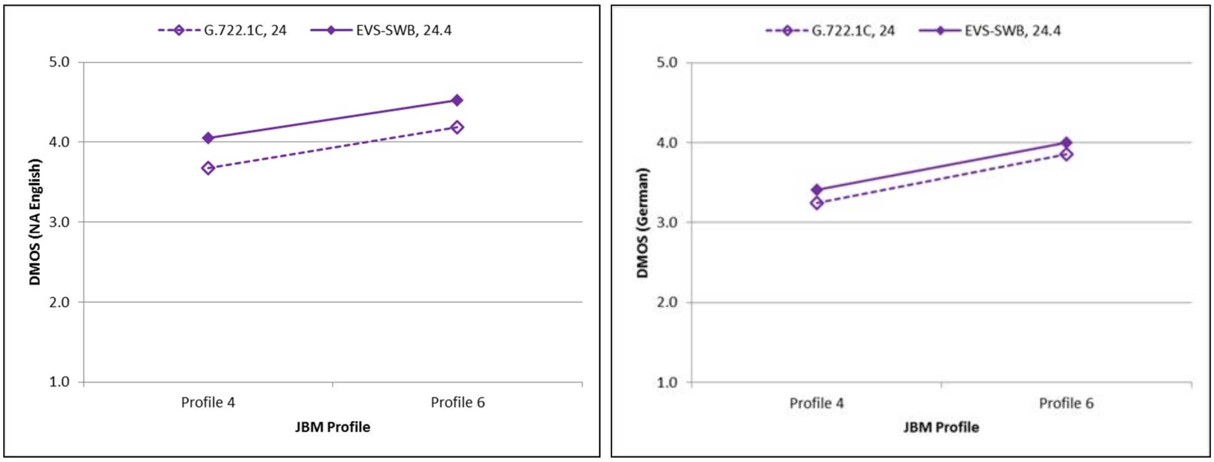
1) Under the frame erasure rate of 3%, the EVS-SWB mixed/music performance at lower bit rates is significantly better than the AMR-WB at 23.85 kbps. Similar trend is seen at frame erasure rate of 6%. At higher bit rates (32-64 kbps), the performance is better than G.719 at the same bit rate under impaired channel (both FER 3%, 6 %).

2) There is a steady progression of quality increase from 13.2 kbps to 64 kbps under mixed/music, while meeting not only the performance requirements (PRs) but also exceeding the EVS Selection performance objectives in Experiment S7. Please refer to Annex A for ToR performance (Table A.1 and Table A.3) of SWB mixed/music quality under impaired channel in EVS Selection Experiment S7.



(a) **(b)**

Figure 11.8: Experiment S7, testing EVS-SWB mixed/music performance under impaired channel condition with frame erasure rates of 3% and 6%. (a) North American English language, (b) German language



(a) **(b)**

Figure 11.9: Experiment S7, testing EVS-SWB mixed/music performance under impaired channel conditions with frame erasures introduced through Delay/Loss Profiles 4 and 6. (a) North American English language, (b) German language

11.2 SWB Characterization Tests

11.2.0 List of experiments in the super-wideband frequency bandwidth

In Characterization phase, four experiments, S1, S2, S3, and S1_Noisy have been conducted to evaluate the performance of the EVS codec with super-wideband (SWB) content sampled at 32 kHz. While the experiments S1, S2, and S1_Noisy used clean/noisy speech, the experiment S3 used mixed/music content for evaluating the SWB performance of the EVS codec. All the four SWB experiments used the ITU-T P.800 DCR subjective test methodology.

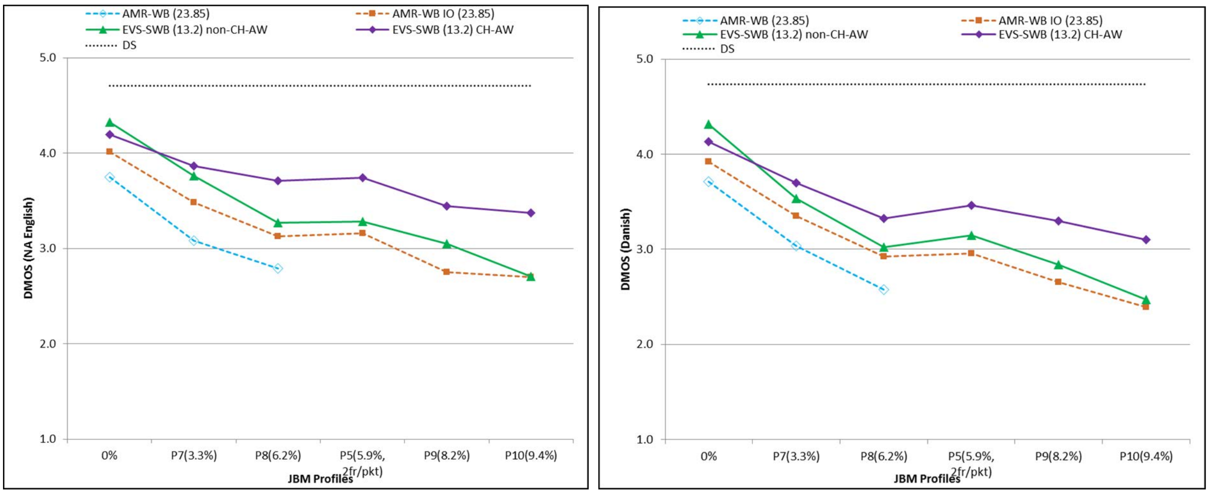
- Experiment S1: clean speech in North American English and Danish languages to evaluate rate switching, channel aware mode, and tandem conditions.

- Experiment S2: speech in Danish and French languages under background noise (car noise at 15 dB SNR) to evaluate rate switching and high FER conditions.
- Experiment S3: music and mixed content in Spanish language to evaluate rate switching, untested conditions in selection phase, and JBM.
- Experiment S1_Noisy: Noisy speech (car noise 15 dB) in North American English to evaluate rate switching, channel aware mode, and tandem conditions.

11.2.1 Experiment S1

Characterization Experiment S1 is conducted to evaluate the EVS codec SWB clean speech performance under clean channel and impaired channel conditions. Experiment S1 is conducted in two different listening labs in North American English language (Figure 11.10(a)) and in Danish language (Figure 11.10(b)).

Figure 11.10(a, b) shows the EVS-SWB channel aware mode performance at 13.2 kbps under clean channel as well as under five different delay/loss profiles (Profiles 5, 7, 8, 9, and 10) simulating impaired channel characteristic with varying delay and jitter. In both the languages, i.e., in North American English Figure 11.10(a) and in Danish (Figure 11.10(b)), the EVS-SWB performance under impaired channel with channel aware mode enabled is significantly better than the EVS-SWB without channel aware mode as well as the Reference codecs. The Reference codecs shown in Figure 11.10(a, b) include AMR-WB at 23.85 kbps, EVS AMR-WB-IO at 23.85 kbps, and the EVS-SWB non-channel aware mode at 13.2 kbps.



(a) **(b)**

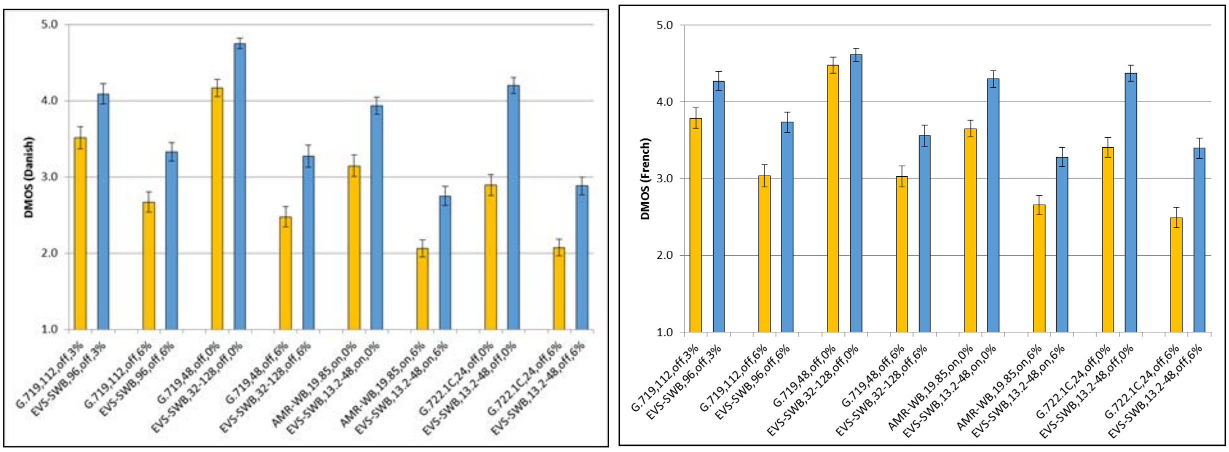
Figure 11.10: Experiment S1, testing EVS-SWB channel aware mode performance under clean and impaired channel conditions. (a) North American English language, (b) Danish language

Figure 11.10(a, b) indicate that compared against AMR-WB/EVS AMR-WB-IO modes, the subjective quality performance gap with EVS-SWB Channel Aware mode increases from about 0.3 DMOS to 0.75 DMOS when tested across lower to higher frame erasure rates (e.g. Profile 7 through Profile 10). In clean channel condition, the EVS-SWB channel aware mode at 13.2 kbps performs significantly better than AMR-WB / EVS AMR-WB IO at its highest bit rate of 23.85 kbps.

11.2.2 Experiment S2

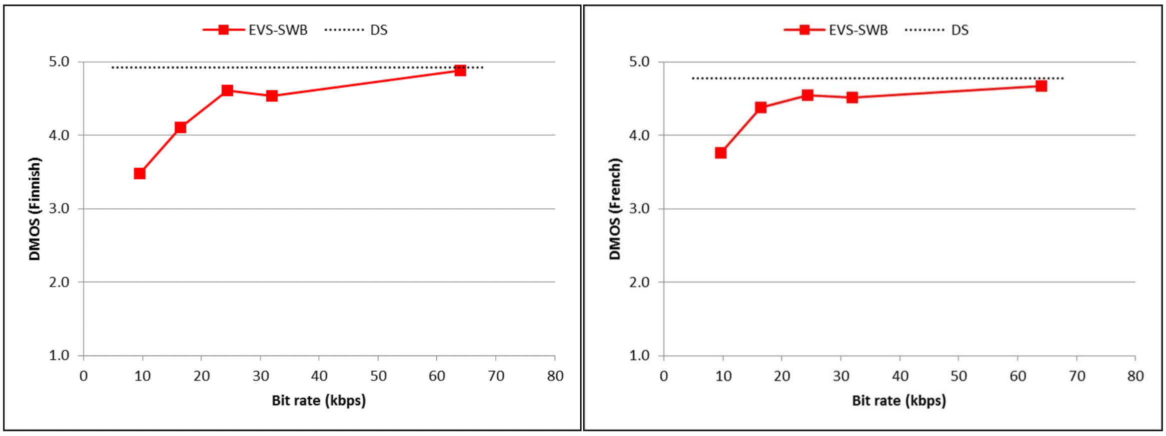
Characterization Experiment S2 is conducted to evaluate the EVS codec SWB noisy speech rate switching performance under clean channel and impaired channel conditions. Experiment S2 is conducted in two different listening labs in Danish language (Figure 11.11(a)) and in French language (Figure 11.11(b)).

From the bar charts shown in Figure 11.11(a, b), it is noted that the EVS-SWB performance under various rate switching conditions (e.g. rate switching conditions in the range of 13.2 to 48 kbps and 32 to 128 kbps both in DTX on and off) is significantly better than the corresponding Reference condition both under clean channel and under impaired channel.



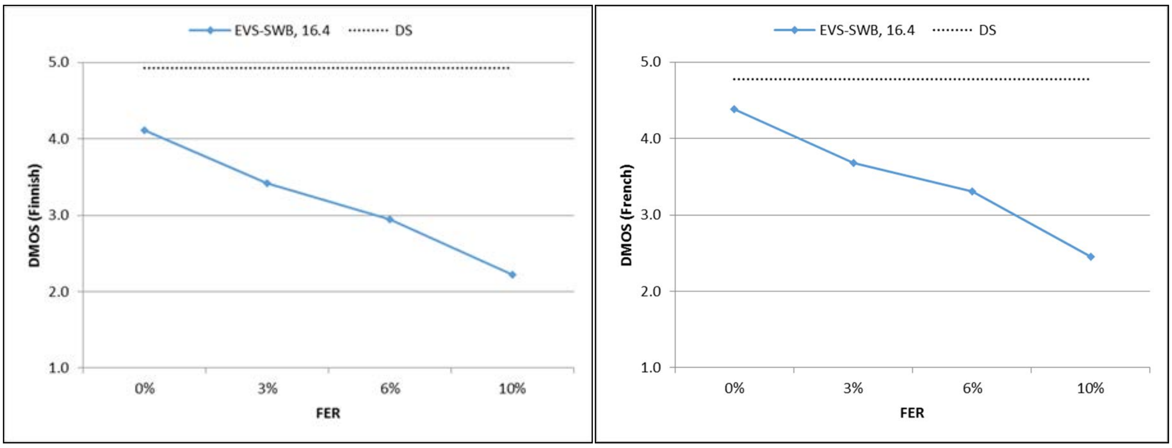
(a) **(b)**

Figure 11.11: Experiment S2, testing EVS-SWB noisy speech (Car noise 15 dB) to evaluate rate switching performance in DTX on and DTX off. (a) Danish language, (b) French language



(a) **(b)**

Figure 11.12 Experiment S2, testing EVS-SWB noisy speech (Car noise 15 dB) to evaluate rate switching performance in DTX on and DTX off. (a) Danish language, (b) French language



(a) **(b)**

Figure 11.13: Experiment S2, testing EVS-SWB noisy speech (Car noise 15 dB) to evaluate rate switching performance in DTX on and DTX off. (a) Danish language, (b) French language

11.2.3 Experiment S3

Characterization Experiment S3 is conducted to evaluate the EVS codec SWB mixed/music rate switching performance under clean channel and impaired channel conditions.

From the bar charts shown in Figure 11.14, it is noted that the EVS-SWB performance under various rate switching conditions is significantly better than the corresponding Reference condition both under clean channel and under impaired channel.

Figure 11.15 indicates the SWB results at 3% and 6% FER. Furthermore, Figure 11.16 shows the EVS-SWB performance at 24.4 kbps under six different delay/loss profiles (Profiles 1 through 6) simulating impaired channel characteristic. These profiles are explained and included in [13].

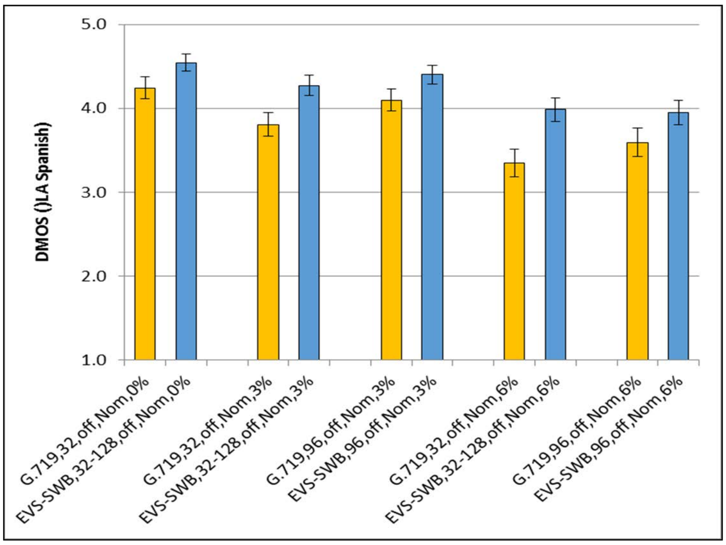
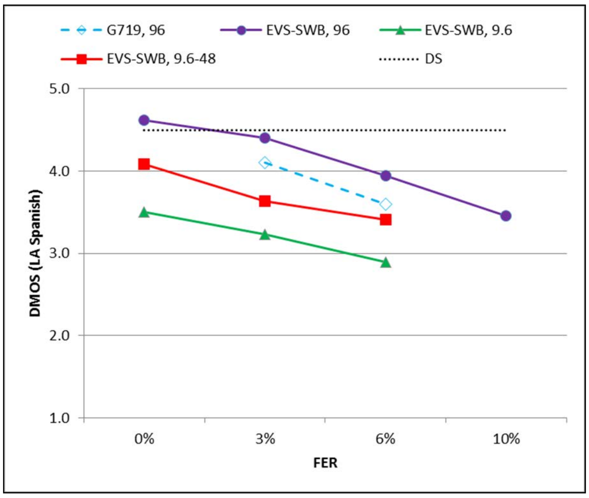
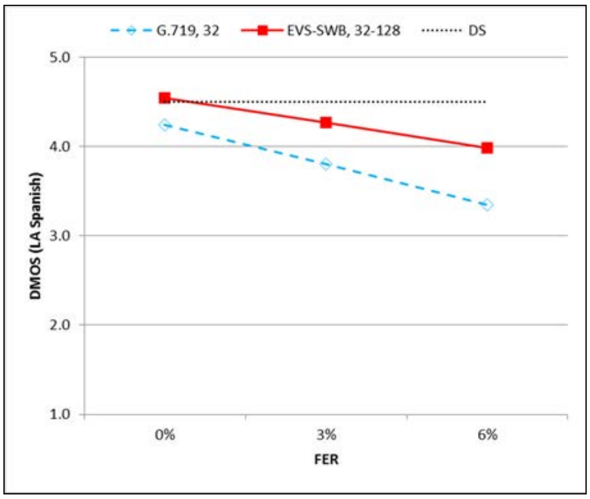


Figure 11.14: Experiment S3, testing EVS-SWB mixed/music content. LA Spanish language





(a) (b)

Figure 11.15: Experiment S3, testing EVS-SWB mixed/music content in LA Spanish language under impaired channel (3% and 6% FER)

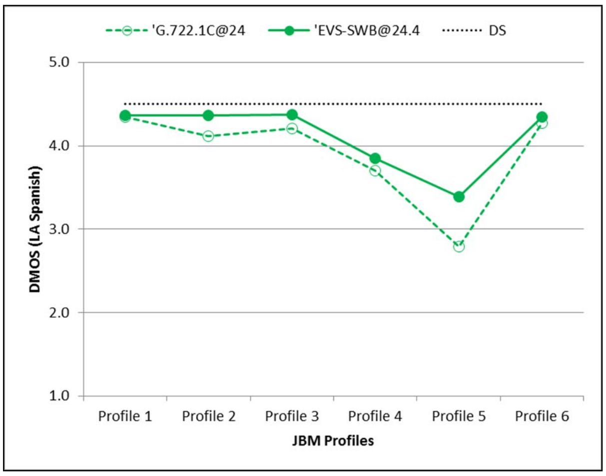


Figure 11.16: Experiment S3, testing EVS-SWB mixed/music content in LA Spanish language under impaired channel (MTSI profiles 1…6)

11.2.4 Experiment S1_Noisy

Characterization Experiment S1_Noisy was conducted to evaluate the EVS codec SWB noisy speech performance under clean channel and impaired channel conditions, in North American English language.

Figure 11.17 shows the EVS-SWB channel aware mode performance at 13.2 kbps under clean channel as well as under five different delay/loss profiles (Profiles 5, 7, 8, 9, and 10) simulating impaired channel characteristic with varying delay and jitter. The EVS-SWB performance under impaired channel with channel aware mode enabled is significantly better than the EVS-SWB without channel aware mode as well as the reference codecs. The reference codecs shown in Figure 11.10(a, b) include AMR-WB at 23.85 kbps, EVS AMR-WB-IO at 23.85 kbps, and EVS-SWB non-channel aware mode at 13.2 kbps.

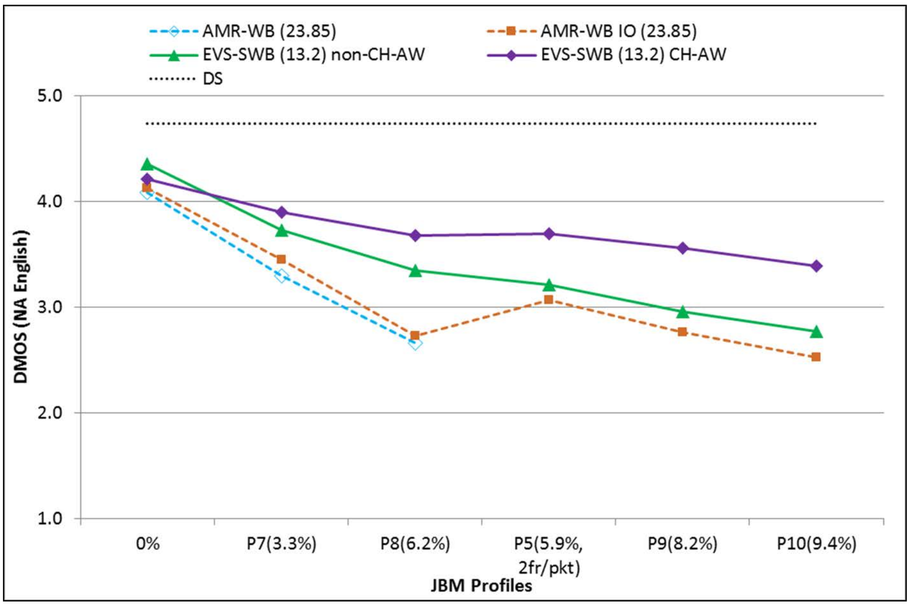


Figure 11.17: Experiment S1_Noisy, testing EVS-SWB channel aware mode performance in noisy speech (car noise at 15 dB) under clean and impaired channel conditions, with North American English language

Observations from characterization experiment S1 can be extended to S1_Noisy (Figure 11.17) test as follows:

When compared against AMR-WB/EVS AMR-WB-IO modes, the subjective quality performance advantage with EVS-SWB channel aware mode increases from about 0.3 DMOS to 0.75 DMOS when tested across lower to higher frame erasure rates (e.g. Profile 7 through Profile 10). In clean channel condition and noisy speech, the EVS-SWB channel aware mode at 13.2 kbps achieves subjective quality comparable to that of AMR WB and EVS AMR-WB-IO mode at its highest bit rate of 23.85 kbps.

11.3 Conclusions on EVS Performance in Super-Wideband

EVS-SWB shows significant improvements over existing Super-Wideband codecs, such as G.722.1C and G.719, especially for clean speech and speech with background noise, both in clean channel as well as under impaired channel conditions. EVS is able to achieve good quality at bitrates much lower than G.722.1C and G.719, and scales to much higher quality with increasing bit rate before saturation.

EVS-SWB also shows major improvements for mixed-content and music, performing equally or better than AMR-WB+, G.722.1C or G.719, at much lower algorithmic delay than those codecs.

When compared to AMR-WB in the same test, EVS-SWB modes outperform AMR-WB.

The channel aware coding mode of the 3GPP EVS codec offers a highly error resilient coding mode at 13.2 kbps. The channel aware mode quality degrades much more gracefully even out to the 10% FER of profile 10 (that may occur for example in best-effort channels), compared to the AMR-WB and EVS AMR-WB-IO conditions.

12 Mixed Bandwidth and Fullband Tests in Characterization

12.1 Mixed Bandwidth Tests

In characterization phase, three experiments, M1, M2, M3 have been conducted to evaluate the mixed bandwidth performance of the EVS codec. Across a wide range of bit rates, the mixed bandwidth tests may effectively characterize the codec, e.g. encoding up to SWB bandwidth relative to encoding up to only NB or WB. While the experiments M1 and M2 used clean/noisy speech, the experiment M3 used mixed/music content for evaluating extended bandwidth coding performance of the EVS codec beyond NB/WB. All the three mixed bandwidth experiments used the ITU-T P.800 DCR subjective test methodology.

- Experiment M1 (DCR): NB/WB/SWB clean speech in North American English language under clean channel condition
- Experiment M2 (DCR): NB/WB/SWB speech in Finnish language under background noise (car noise at 20 dB SNR) under clean channel condition
- Experiment M3 (DCR): NB/WB/SWB music and mixed content in Chinese language under clean channel condition

12.1.1 Experiment M1

Experiment M1 is conducted to evaluate the EVS codec multi-bandwidth NB/WB/SWB clean speech performance under clean channel conditions. Experiment M1 is conducted in North American English language.

Figure 12.1 and 12.2 shows the EVS codec NB/WB/SWB performance at bit rates ranging from 5.9 to 64 kbps. For benchmarking purposes, the previously standardized codecs AMR NB (7.4 to 12.2 kbps) and AMR-WB (8.85 to 23.85 kbps) are also tested in the same mixed bandwidth test.

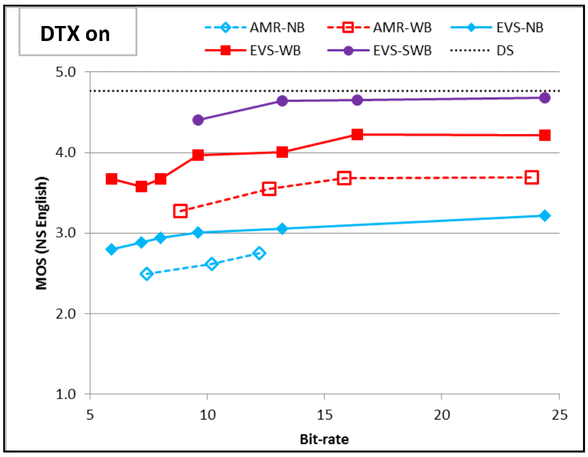


Figure 12.1: Experiment M1, testing EVS-NB, WB, SWB mixed-bandwidth test, DTX on, with clean speech in North American English language

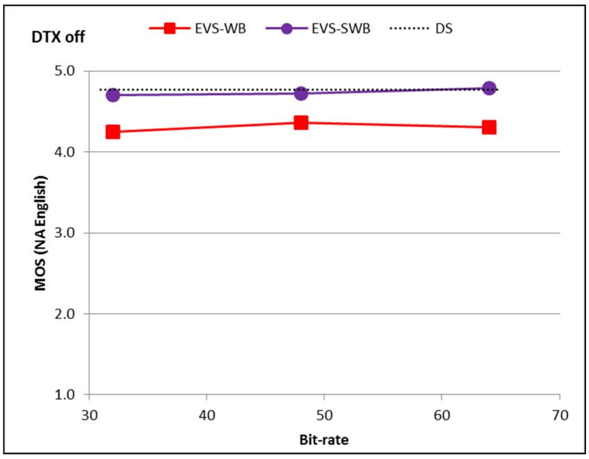


Figure 12.2: Experiment M1, testing EVS-NB, WB, SWB mixed bandwidth test, DTX off, with clean speech in North American English language

Observations from the characterization experiment M1 with clean speech, Figure 12.1 and 12.2, include:

1. The EVS-NB codec performance is significantly better than the AMR NB across a wide range of bit rates. In particular, the EVS-NB coding at its lowest bit rate VBR 5.9 kbps already achieves the subjective quality comparable to that of AMR NB at its highest bit rate of 12.2 kbps coding in clean speech. The EVS-NB coding at its highest bit rate 24.4 kbps starts converging into the subjective quality region of AMR-WB at its lowest bit rate 8.85 kbps.

2. The EVS-WB codec performance is significantly better than the previously standardized AMR-WB codec. In particular, the EVS-WB coding at 5.9 kbps achieves subjective quality better than the AMR-WB at 8.85 kbps and comparable to that of AMR-WB at 12.65 kbps in clean speech. The subjective quality of EVS-WB coding starting at 9.6 kbps is significantly better than the AMR-WB coding at its highest bit rate of 23.85 kbps. There is a steady progression of EVS-WB subjective quality from 5.9 kbps to 16.4 kbps (Figure 12.1) and tending towards the region of DMOS saturation for WB at higher bit rates (Figure 12.1, 12.2).

3. The EVS-SWB codec performance is significantly better than the previously standardized AMR-WB codec as well as the corresponding bit rates of EVS-WB. The subjective quality of EVS-SWB coding at 9.6 kbps is better than the AMR-WB at 23.85 kbps as well as EVS-WB at 24.4 kbps. There is a further significant quality increase starting from 13.2 kbps as compared to 9.6 kbps.

4. From the Mixed-Bandwidth experiment, M1, the EVS-SWB clean speech quality at 13.2 kbps already approaches that of the “Direct Source” quality based on the ITU-T P.800 DCR test methodology.

12.1.2 Experiment M2

Experiment M2 is conducted to evaluate the EVS codec mixed bandwidth NB/WB/SWB noisy speech performance (in car noisy at 20 dB SNR) under clean channel conditions, in Finnish language.

Figure 12.3 and 12.4 show the EVS codec NB/WB/SWB performance at bit rates ranging from 5.9 to 64 kbps. For benchmarking purposes, the previously standardized codecs AMR NB (7.4 to 12.2 kbps) and AMR-WB (8.85 to 23.85 kbps) are also tested in the same mixed bandwidth test.

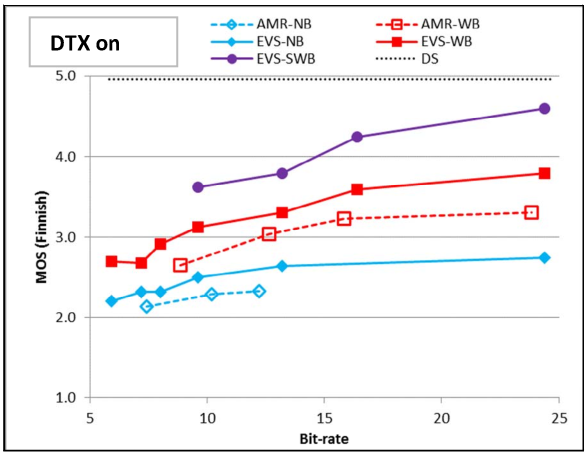


Figure 12.3: Experiment M2, testing EVS-NB, WB, SWB mixed bandwidth test, DTX on, with noisy speech (car noise at 20 dB SNR) in Finnish language

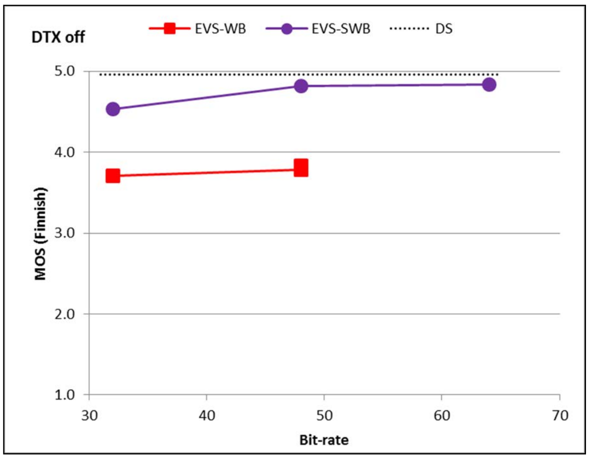


Figure 12.4: Experiment M2, testing EVS-NB, WB, SWB mixed bandwidth test, DTX off, with noisy speech (car noise at 20 dB SNR) in Finnish language

Observations from the characterization experiment M2 with noisy speech, Figure 12.3 and 12.4, include:

1) The EVS-NB codec performance is significantly better than the AMR NB across a wide range of bit rates. In particular, the subjective quality of EVS-NB coding at 7.2 kbps is comparable to that of AMR NB at its highest bit rate of 12.2 kbps coding. There is a steady increase in EVS-NB subjective quality performance from 5.9 kbps to 24.4 kbps.

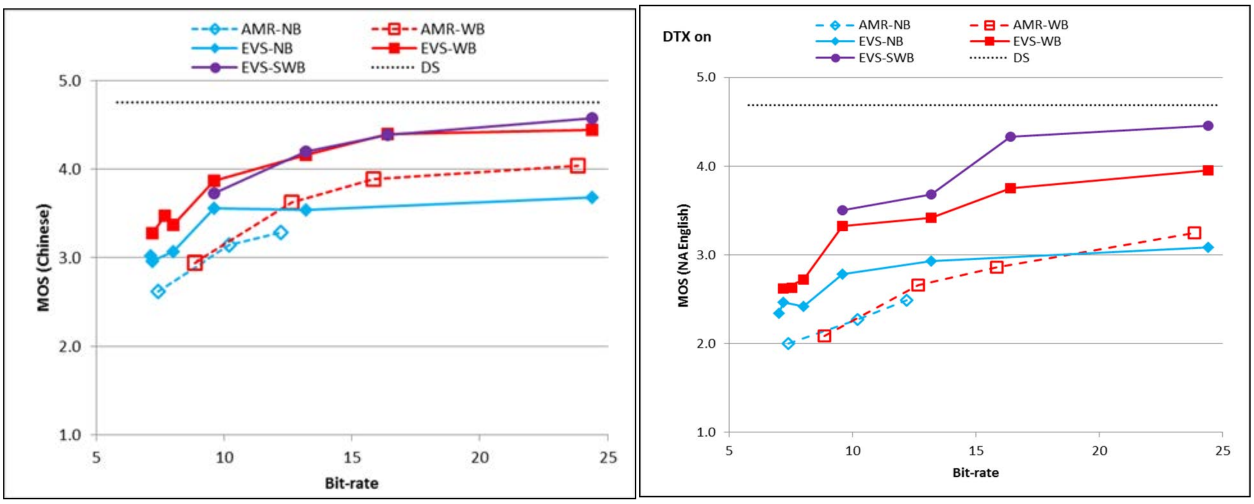
2) The EVS-WB codec performance is significantly better than the previously standardized AMR-WB codec. In particular, the EVS-WB coding at its lowest bit rate at 5.9 kbps achieves subjective quality comparable to that of AMR-WB at 8.85 kbps. The subjective quality of EVS-WB coding at 13.2 kbps is comparable to that of AMR-WB coding at its highest bit rate of 23.85 kbps in noisy speech. There is a consistent progression of EVS-WB subjective quality in DTX on from 5.9 kbps to 24.4 kbps (Figure 12.4) and tending towards the region of DMOS saturation for WB at higher bit rates tested in DTX off configuration (Figure 12.4).

3) The EVS-SWB codec performance is significantly better than the previously standardized AMR-WB codec as well as the corresponding bit rates of EVS-WB. The subjective quality of EVS-SWB coding at 13.2 kbps is significantly better than the AMR-WB at 23.85 kbps. There is a significant progression of EVS-SWB subjective quality in DTX on from 9.6 kbps to 24.4 kbps (Figure 12.4) and tending towards the region of DMOS saturation for SWB at higher bit rates tested in DTX off configuration (Figure 12.4).

12.1.3 Experiment M3

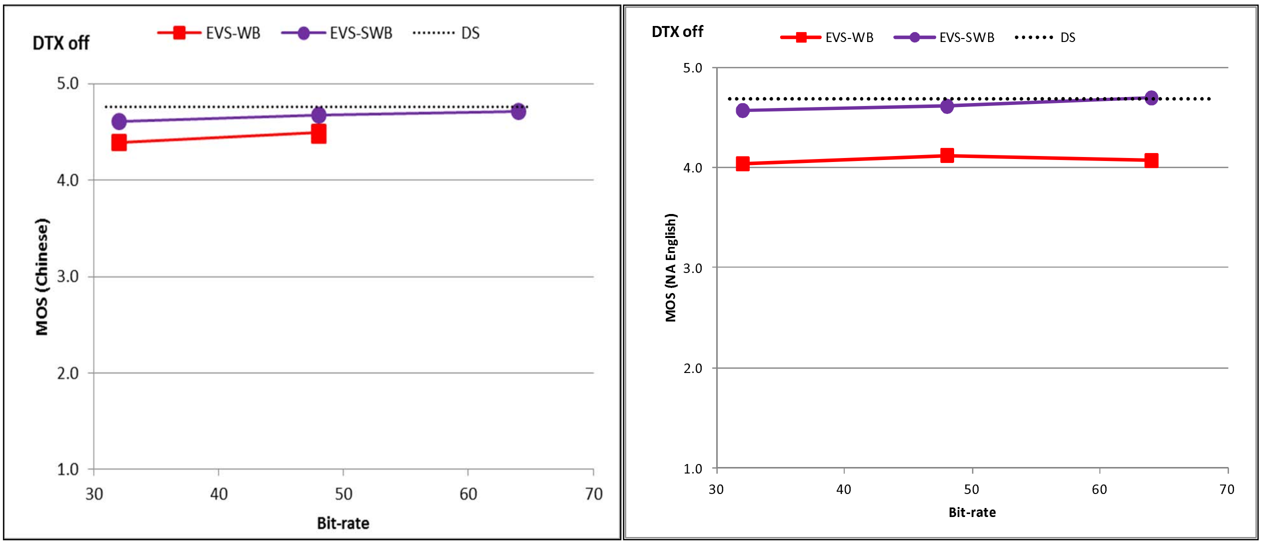
Experiment M3 is conducted to evaluate the EVS codec mixed bandwidth NB/WB/SWB mixed/music coding performance under clean channel conditions. Experiment M3 is conducted in Chinese and North American English mixed/music content.

Figures 12.5(a, b) and 12.6(a, b) show the EVS codec NB/WB/SWB performance at bit rates ranging from 5.9 to 64 kbps. VBR mode is designed to achieve the average data rate (ADR) of 5.9 kbps for active speech. In order to further evaluate and confirm the performance of the VBR mode in music/mixed content, this experiment included the VBR condition in NB and WB. While achieving the ADR of 5.9 kbps for active speech, the VBR mode may result in a different ADR between 5.9 and 8 kbps for music/mixed content; the ADR values were in this experiment M3 7.11 kbps for NB and 7.68 kbps for WB for Chinese music/mixed content, and 7.01 kbps for NB and 7.53 kbps for WB for North American English music/mixed content. For benchmarking purposes, the previously standardized codecs AMR (7.4 to 12.2 kbps) and AMR-WB (8.85 to 23.85 kbps) are also tested in the same mixed bandwidth test.



(a) (b)

Figure 12.5: Experiment M3, testing EVS-NB, WB, SWB music and mixed content in (a) Chinese language and (b) North American English language



(a) (b)

Figure 12.6: Experiment M3, testing EVS-NB, WB, SWB music and mixed content in (a) Chinese language and (b) North American English language

Observations from the characterization experiment M3 with mixed/music content, Figure 12.5(a, b) and 12.6(a, b), include:

1) The EVS-NB mixed/music codec performance is significantly better than the AMR across a wide range of bit rates. In particular, the subjective quality of EVS-NB coding at its lowest constant bit rate coding of 7.2 kbps is significantly better than that of AMR 7.4 kbps coding. Starting from 9.6 kbps, EVS-NB is significantly better than AMR at any bit rate. EVS-NB up to 13.2 kbps performs equal or better than AMR-WB at a similar bit rate.

2) The EVS-WB codec performance is significantly better than the previously standardized AMR-WB codec at any comparable bit rate. In particular, the EVS-WB coding starting at 13.2 kbps achieves subjective quality that is significantly better than that of AMR-WB at any bit rate; for North American English music and mixed content, EVS-WB coding starting already at 9.6 kbps achieves better quality than AMR-WB at any bit rate.

3) The EVS-SWB codec performance is significantly better than the previously standardized AMR-WB codec at any comparable bit rate. The subjective quality of EVS-SWB coding starting at 13.2 kbps is significantly better

than the AMR-WB at its highest bit rate 23.85 kbps. For North American English music and mixed content, EVS-SWB coding starting already at 9.6 kbps achieves better quality than AMR-WB at its highest bit rate. There is a steady progression of EVS-SWB subjective quality in DTX on from 9.6 kbps to 24.4 kbps (Figure 12.5) and tending towards the region of DMOS saturation for SWB at higher bit rates tested in DTX off configuration (Figure 12.6).

4) For North American English music and mixed content, EVS-SWB coding performs significantly better than EVS-WB at the same bit rate.

12.2 Fullband Tests

12.2.0 List of experiments in the fullband frequency bandwidth

In characterization phase, two experiments, F1 and F2 have been conducted to evaluate the EVS fullband coding performance. While the experiment F1 used clean speech, the experiment F2 used mixed/music content. Both the fullband experiments used the ITU-T P.800 DCR subjective test methodology.

- Experiment F1 (DCR): clean speech in German language to evaluate EVS-SWB and FB performance under clean channel condition.
- Experiment F2 (DCR): music and mixed content in Danish language to evaluate EVS-SWB and FB performance under clean and impaired channel conditions.

12.2.1 Experiment F1

Figure 12.7 and 12.8 show the EVS-FB (Red curve) performance relative to EVS-SWB (blue curve) coding at various bit rates. The fullband experiment F1 was conducted with clean speech in German language. The support for FB starts from 16.4 kbps and extends up to 128 kbps.

As shown in Figure 12.7 and 12.8, the subjective quality of EVS-FB coding (up to 20 kHz bandwidth) is quite comparable to that of EVS-SWB coding (up to 16 kHz bandwidth) in clean speech both in DTX on as well as DTX off conditions.

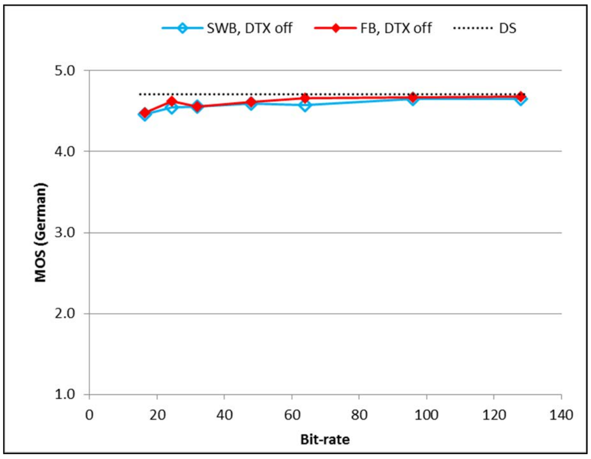


Figure 12.7: Experiment F1, testing EVS-SWB and EVS-FB clean speech in German language

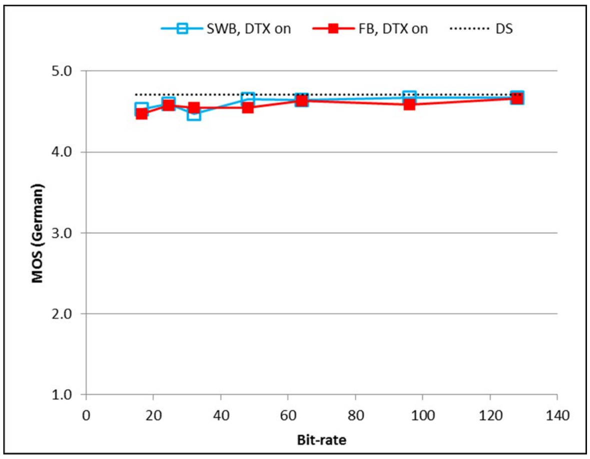


Figure 12.8: Experiment F1, testing EVS-SWB and EVS-FB clean speech in German language

12.2.2 Experiment F2

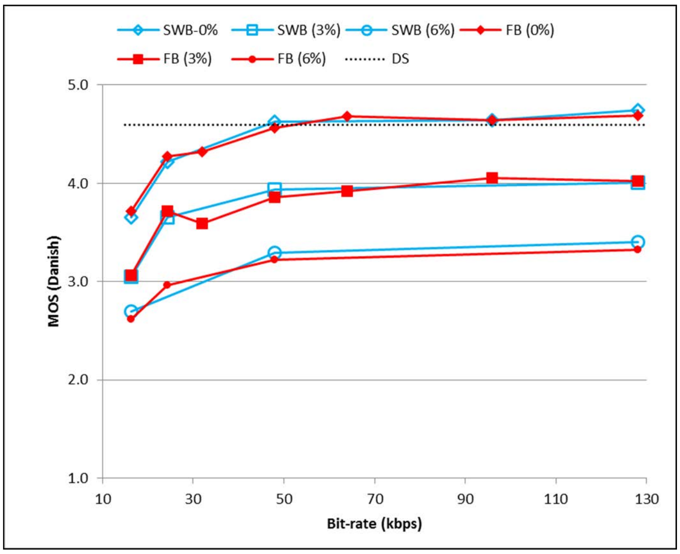


Figure 12.9: Experiment F2, testing EVS-SWB and EVS-FB music and mixed content with Danish language

Figure 12.9 shows the EVS-FB (Red curve) performance relative to EVS-SWB (blue curve) coding at various bit rates and testing under various impaired channels with frame erasure rates of 3 % and 6 %. The fullband experiment F2 was tested using Danish mixed/music content.

Observations from fullband tests F1 and F2, Figures 12.7, 12.8, and 12.9 include:

1) As shown in Figure 12.9, the subjective quality of EVS-FB coding (up to 20 kHz bandwidth) is quite comparable to that of EVS-SWB coding (up to 16 kHz bandwidth) in mixed/music coding.

2) There is a steady progression of music coding quality from lower rates of EVS-SWB/FB coding to higher rates of EVS-SWB/FB coding and tending towards the subjective quality of Direct Source.

13 Objective Evaluations

13.1 Selection Phase

13.1.1 Objective Measurements

The purpose of the objective measurement is to verify the performance of the EVS codec candidate algorithms using objective metrics. Those are applied to test the fulfillment of the design constraints defined in EVS-4 Permanent Document: Design Constraints [18] or to evaluate objective performance requirements defined in EVS-3 Permanent Document: Performance Requirements [17].

The objective metrics for selection consist of the following items:

1) Gain verification (design constraints) (**Gain**)

   - The EVS candidate codecs do not amplify the output signal relative to the input signal beyond limits.

2) JBM compliance to TS 26.114 (design constraints) (**JBM**)

   - A JBM solution conforming to the requirements in TS 26.114 [13], except for the functional requirement in sub-clause 8.2.2 of TS 26.114: "*Speech JBM used in MTSI shall support all the codecs as defined in clause 5.2.1*", will be provided with the candidate codecs.

3) Active Frame Ratio (AFR) (performance requirements) (**AFR**)

   - This part of the evaluation is based on a large database of speech and noisy speech of length (approximately 10 to 30 min) with an AFR of approximately 40 % (AFR is based on P.56 measured on clean speech). The requirements are set for clean speech, speech under background noise, music and mixed content, in narrowband, wideband, super-wideband, at all bit rates below 24.4 kbps.

4) Attenuation during inactive regions (performance requirements) (**Att.**)

   - For clean speech and speech under background noise, the attenuation of background noise level during inactive regions is constrained.

5) Average active speech bit rate of VBR and CBR (**BR**)

   - Verification of average active bit rate.

6) Complexity measurement (**Cmp**)

   - Processing details are described in EVS-7b [21].

Processing of objective performance requirements is described in EVS-7b Processing Plan, Annex A.

Table 13.1: List of databases used for objective evaluations

Database	Description	Gain	JBM	AFR	Att.	BR	Cmp
1	NB clean speech filtered by MSIN, in 8 kHz sampling, at -16, -26 and -36 dBov	pass	pass	n/a	pass	pass	pass
2	WB clean speech filtered by HP50, in 16 kHz sampling, at -16, -26 and -36 dBov	pass	pass	n/a	pass	pass	pass
3	SWB clean speech filtered by HP50, in 32 kHz sampling at -16, -26 and -36 dBov	pass	pass	n/a	pass	pass	pass
4	NB speech with car noise at [15]dB SNR filtered by MSIN, in 8 kHz sampling	pass	n/a	pass	pass	pass	pass
5	WB speech with car noise at [15] dB SNR filtered by HP50, in 16 kHz sampling	pass	n/a	pass	pass	pass	pass
6	SWB speech with car noise at [15] dB SNR filtered by HP50, in 32 kHz sampling	pass	n/a	pass	pass	pass	pass
7	NB speech with street noise at [20] dB SNR filtered by MSIN, in 8 kHz sampling	pass	n/a	pass	pass	pass	pass
8	WB speech with street noise at [20] dB SNR filtered by HP50, in 16 kHz sampling	pass	n/a	pass	pass	pass	pass
9	SWB speech with street noise at [20] dB SNR filtered by HP50, in 32 kHz sampling	pass	n/a	pass	pass	pass	pass
10	NB speech with office noise at [20] dB SNR filtered by MSIN, in 8 kHz sampling	pass	n/a	pass	pass	pass	pass
11	WB speech with office noise at [20] dB SNR filtered by HP50, in 16 kHz sampling	pass	n/a	pass	pass	pass	pass
12	SWB speech with office noise at [20] dB SNR filtered by HP50, in 32 kHz sampling	pass	n/a	pass	pass	pass	pass
13	NB mixed content and music filtered by MSIN, in 8 kHz sampling rate	pass	n/a	n/a	n/a	pass	pass
14	WB mixed content and music filtered by HP50, in 16 kHz sampling	pass	n/a	n/a	n/a	pass	pass
15	SWB mixed content and music filtered by HP50, in 32 kHz sampling	pass	n/a	n/a	n/a	pass	pass

For verification of the candidate solution several objective metrics will be evaluated by using the tools defined in EVS-7b [21]. PC should use the tools defined in EVS-7b to create JBM objective metrics, but may also use their own tool. Proponents reported how they have made the JBM objective metrics compliance assessment as part of the selection deliverables defined in EVS-6b [20].

13.1.2 Verification of Codec Performance with respect to Acoustic Test Cases based on the EVS Selection Phase Executable

13.1.2.1 Evaluation Setup

The first step in the evaluation is the scaling with regard to a certain overload point. Two overload points (3.0 vs 9.0 dBm0) for the conversion from the physical unit Volt to 16-bit scale were taken into account.

Since the overload point (OVLP) refers to a full-scale sine wave (with level T_{max} = -3.01 dBov according to [30]), for the scaling between dBov and dBm0 resp. dBV, the following notation can be made:

T_{max} = -3.01 dBov

$$x[\text{dBov}] = y[\text{dBm0}] - (OVLP\ [\text{dBm0}] - T_{max}[\text{dBov}])$$

$$\leftrightarrow x[\text{dBov}] = z[\text{dBV}] + 2.21dB - (OVLP\ [\text{dBm0}] - T_{max}[\text{dBov}])$$

Example: For an OVLP of 3.0 dBm0, the scaling between dBV and dBov is defined as:

$$x[\text{dBov}] = z[\text{dBV}] + 2.21\ \text{dB} - (3.0\ \text{dBm} + 3.01\ \text{dBov}))$$

$$\leftrightarrow x[ov] = \frac{z[V]}{10^{3.81/_{20}}} = \frac{z[V]}{1.5488}$$

The scaling back to the physical unit Volt was applied in the corresponding inverse way.

After scaling, the next step included the encoding and decoding of the audio data. This was conducted with the provided command line executable. The source code was not recompiled to a new binary.

For the evaluation of narrowband, wideband, and super-wideband mode, all bit rates which are available in each bandwidth mode according to Table 1 of [2] were used.

13.1.2.2 General

Several tests according to 3GPP TS 26.132 [29] were performed in order to evaluate the performance according to [14] of the EVS codec. Only frequency response results are reported in this subclause. Detailed results can be found in Annex D.

TS 26.132 is originally intended for acoustic testing of terminals. Since the EVS codec is regarded as the "device under test", only electrical insertions are reasonable for testing and thus only measurements in (acoustic) receiving direction are taken into account.

In narrowband the test signal bandlimitation as defined in 3GPP TS 26.132 [29] was used. For superwideband the ITU-T P.501 [30] test signals were downsampled to 32 kHz, cut-off frequency 14.4 kHz, >80 dB/oct. For fullband the original speech signals from ITU-T P.501 [30] were used.

With this approach, the EVS codec can be evaluated with typical test scenarios, which will occur in real-life applications with mobile phones.

The following graphs include multiple curves representing the different bit rates within each bandwidth mode. For the sake of clarity, the corresponding legends are not repeated in each graph, Table 13.2 shows the legends used in the following clauses.

Table 13.2: Legends for different bit rates

5.9 kbit/s 7.2 kbit/s 8.0 kbit/s 9.6 kbit/s 13.2 kbit/s 16.4 kbit/s 24.4 kbit/s	5.9 kbit/s 7.2 kbit/s 8.0 kbit/s 9.6 kbit/s 13.2 kbit/s 16.4 kbit/s 24.4 kbit/s 32.0 kbit/s 48.0 kbit/s 64.0 kbit/s 96.0 kbit/s 128.0 kbit/s
NB mode	**WB mode**
9.6 kbit/s 13.2 kbit/s 16.4 kbit/s 24.4 kbit/s 32.0 kbit/s 48.0 kbit/s 64.0 kbit/s 96.0 kbit/s 128.0 kbit/s	16.4 kbit/s 24.4 kbit/s 32.0 kbit/s 48.0 kbit/s 64.0 kbit/s 96.0 kbit/s 128.0 kbit/s
SWB mode	**FB mode**

13.1.2.3 EVS-Mode: Narrowband (NB) -- Frequency Response with Real Speech

In narrowband mode, a sampling rate of 8 kHz and all bit rates (5.9, 7.2, 8.0, 9.6, 13.2, 16.4 and 24.4 kbit/s) according to Table 1 of [2] were used. The two possible target overload points 3.0 and 9.0 dBm0 were used by default for all analyses.

It should be noted, that even though Table 1 of [2] states that a narrowband signal can be encoded with 16.4 and 24.4 kbit/s, the provided command line executable produces a bitstream with 13.2 kbit/s. Therefore, the magenta, the blue, and the grey curves are identical in the plots of this section and the last three rows of Table 2 and Table 3 state the same values.

The following results are produced by applying the measurement instructions according to clause 7.4.2 of [29]. To simulate also the impact of level variations, additional overload points of 21.0 and 39.0 dBm0 were also simulated. These overload points do not represent a realistic conversion, they are only used for checking the linearity of the codec and can be regarded as attenuations of 18.0 resp. 36.0 dB compared to the overload point of 3.0 dBm0.

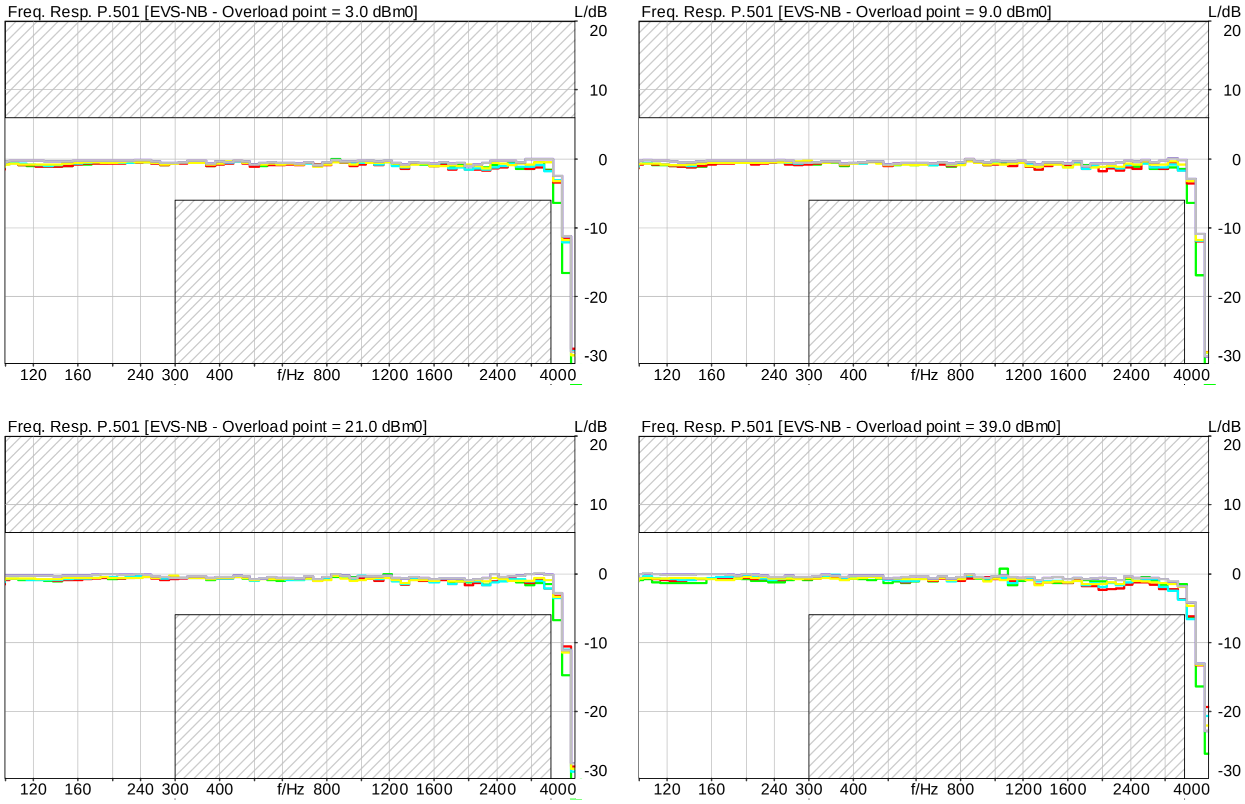


Figure 13.1: Frequency response for NB for different overload points

The results of this analysis are shown in Figure 13.1. For the default and extra overload points, the codec meets the given tolerance scheme according to [14] for all bit rates.

13.1.2.4 EVS-Mode: Wideband (WB) – Frequency Response with Real Speech

In wideband mode, a sampling rate of 16 kHz and all bit rates (5.9, 7.2, 8.0, 9.6, 13.2, 16.4, 24.4, 32.0, 48.0, 64.0, 96.0 and 128.0 kbit/s) according to Table 1 of [2] were used. The two possible target overload points 3.0 and 9.0 dBm0 were used by default for all analyses.

The following results are produced by applying the measurement instructions according to clause 8.4.2 of [29]. To simulate also the impact of level variations, additional overload points of 21.0 and 39.0 dBm0 were also simulated. These overload points do not represent a realistic conversion, they are only used for checking the linearity of the codec and can be regarded as attenuations of 18.0 resp. 36.0 dB compared to the overload point 3.0 dBm0.

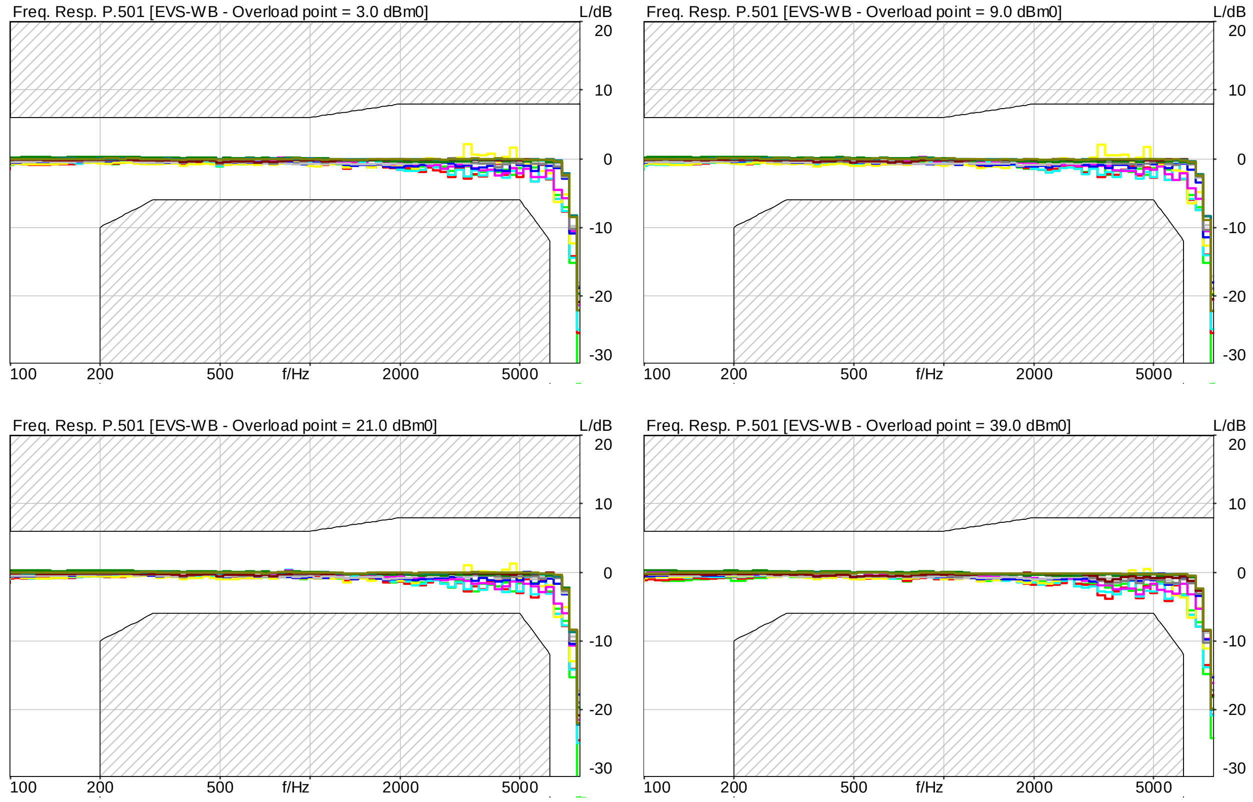


Figure 13.2: Frequency response for WB for different overload points

The results of this analysis are shown in Figure 13.2. For the default and extra overload points, the codec meets the given tolerance scheme according to [14] for all bit rates.

13.1.2.5 EVS-Mode: Super-Wideband (SWB) – Frequency Response with Real Speech

In super-wideband mode, a sampling rate of 32 kHz and all bit rates (9.6, 13.2, 16.4, 24.4, 32.0, 48.0, 64.0, 96.0 and 128.0 kbit/s) according to Table 1 of [2] were used. The two possible target overload points 3.0 and 9.0 dBm0 were used by default for all analyses.

It should be noted, that even though Table 1 of [2] states that a signal can be encoded in super-wideband mode with 9.6 kbit/s, this mode was not tested during Selection Phase of EVS development. Unfortunately a software bug was present in the version used for Selection and the provided command line executable encodes effectively only the wideband bandwidth up to 8 kHz. Therefore, the green curve violates the tolerance schema in Figure 13.3. The software bug was corrected in v.12.1.0 – See clause 13.4.1

The following results are produced by applying measurement instructions similar to clause 8.4.2 of [29] which are adapted to super-wideband by replacing the source signal with a fullband version of the same file and extending the tolerance schema to 14 kHz. To simulate also the impact of level variations, additional overload points of 21.0 and 39.0 dBm0 were also simulated. These overload points do not represent a realistic conversion; they are only used for checking the linearity of the codec and can be regarded as attenuations of 18.0 resp. 36.0 dB compared to the overload point 3.0 dBm0.

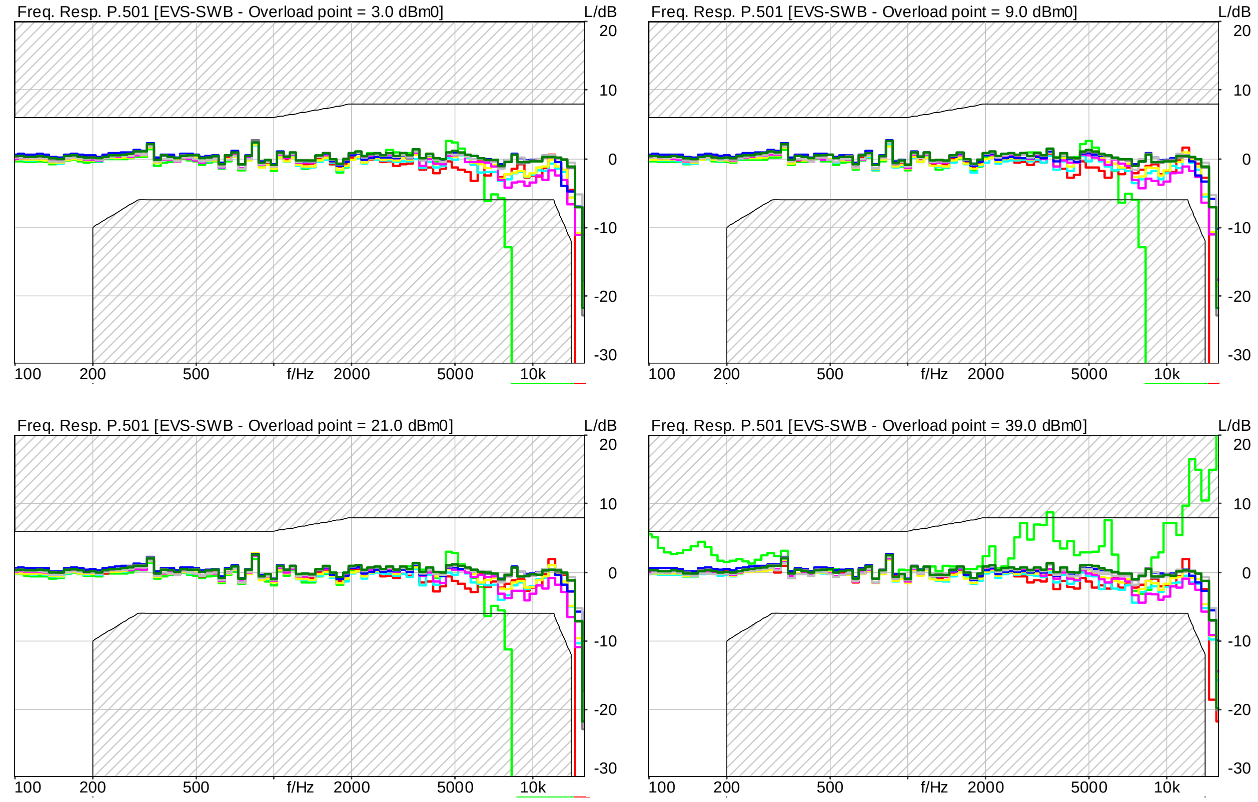


Figure 13.3: Frequency response for SWB for different overload points

The results of this analysis are shown in Figure 13.3. For the default and extra overload points, the codec meets the given tolerance scheme according to [14] for all bit rates but 9.6 kbit/s as explained above.

13.1.2.6 Conclusions

In narrowband mode, all bit rates pass all requirements of [14].

Across all bitrates and for various operating bandwidths NB/WB/SWB, the EVS codec version tested during the Selection Phase meets the frequency response tolerance requirements over an extended range of overload points when tested using real speech.

Note that more detailed results can be found in Annex D.

13.2 Complexity and Delay Analysis

With all features supported and measured according to EVS-8b, the worst case complexity of the coder is 87.97 WMOPS which splits up to 56.25 WMOPS for encoder (24.4 kbit/s SWB with DTX on) and 31.72 WMOPS for decoder (9.6 kbit/s SWB with DTX on, FER=30%). The coder uses 149 kW of RAM (with no JBM included), 147 kW of ROM, and 114500 program instructions.

The JBM solution was measured to consume 18 WMOPS and 49 kW RAM.

The computational complexity and program ROM (PROM) were measured with ITU-T STL2009 [27]. RAM and ROM are reported using 16 bit word-size.

Typical, or mean, and maximum complexity figures for the various sample rates and coded bandwidths of operation have been further analysed and are provided in Tables 13.2a and 13.2b. In these tables, as in the above analysis, the complexity has been analysed on a per-audio-frame basis and the means and maxima collected. The source file comprised 8.5minutes of mixed speech and music.

NOTE: EVS AMR-WB IO is considered to be included in the calculated WB 16 kHz and 32 kHz figures.

The figures in Table 13.2a are calculated using the mean of the per-audio-frame complexity over the whole file. They represent the highest values of the mean complexity when comparing all of the available bit rates for the given bandwidth and sample rate. In the case of the balanced combined figures, the bit rate is identical for the encoder and decoder. In the worst/worst combined figures the complexity of the most complex encoder bit rate is added to that of the most complex decoder bit rate, even though this combination might not occur in actual use.

In both Tables 13.2a and 13.2b, the bold figures represent the complexities for the native (i.e. minimum appropriate) sample rates for each coded bandwidth.

Table 13.2a: Highest values of the mean complexity for different sample rates and coded bandwidths

Coded Bandwidth	Sample Rate (kHz)	Encoder Complexity (WMOPS)	Decoder Complexity (WMOPS)	Balanced Combined Complexity (WMOPS)	Worst/Worst Combined Complexity (WMOPS)
NB	**8**	**29.98**	**15.40**	**43.97**	**45.38**
	16	31.52	16.38	47.35	47.90
	32	37.34	19.37	56.71	56.71
WB	**16**	**38.25**	**18.29**	**51.38**	**56.54**
	32	43.32	22.88	60.92	66.20
SWB	**32**	**45.12**	**22.67**	**65.56**	**67.80**

Worst-case complexity figures for the various sample rates and coded bandwidths of operation are provided in Table 13.2b. In this table the per-audio-frame maxima have been used.

Table 13.2b: Highest values of the worst-case complexity for different sample rates and coded bandwidths

Coded Bandwidth	Sample Rate (kHz)	Encoder Complexity (WMOPS)	Decoder Complexity (WMOPS)	Balanced Combined Complexity (WMOPS)	Worst/Worst Combined Complexity (WMOPS)
NB	**8**	**47.95**	**25.59**	**73.54**	**73.54**
	16	49.48	25.81	75.30	75.30
	32	52.06	28.33	79.49	80.39
WB	**16**	**50.63**	**27.69**	**76.71**	**78.32**
	32	55.27	30.63	84.80	85.90
SWB	**32**	**56.25**	**31.72**	**87.13**	**87.97**

It can be seen from Tables 13.2a and 13.2b that the complexity of the EVS codec broadly scales with coded bandwidth and sample rate of operation and that the mean computational load is significantly lower than the worst-case figures.

Table 13.2c provides the highest values of the mean complexity for AMR-WB calculated using ITU-T STL2009 [27], the same methodology and testfile described above and Table 13.2d provides the relative complexity increase for the EVS codec compared to AMR-WB. Such figures clearly reflect the additional battery drain due to the codec for an EVS service over one based upon AMR-WB.

Table 13.2c: Highest values of the mean complexity for AMR-WB

Coded Bandwidth	Sample Rate (kHz)	Encoder Complexity (WMOPS)	Decoder Complexity (WMOPS)	Balanced Combined Complexity (WMOPS)	Worst/Worst Combined Complexity (WMOPS)
WB	**16**	**32.30**	**7.80**	**38.73**	**40.11**

Table 13.2d: Incremental mean complexity for EVS relative to AMR-WB

Coded Bandwidth	Sample Rate (kHz)	Encoder Complexity (WMOPS)	Decoder Complexity (WMOPS)	Balanced Combined Complexity (WMOPS)	Worst/Worst Combined Complexity (WMOPS)
NB	**8**	**-7.18%**	**+97.4%**	**+13.5%**	**+13.1%**
	16	-2.41%	+110%	+22.3%	+19.4%
	32	+15.6%	+148%	+46.4%	+41.4%
WB	**16**	**+18.4%**	**+134%**	**+32.7%**	**+41.0%**
	32	+34.1%	+193%	+57.3%	+65.0%
SWB	**32**	**+39.8%**	**+191%**	**+69.3%**	**+69.0%**

From Table 13.2d it is clear that there is a significant complexity increase in the EVS decoder compared to that of AMR-WB which is closely linked to enhancements in post-processing and frame erasure concealment leading to the significant performance improvements of EVS in error-free and impaired channels. It should also be noted that the EVS codec complexities also include sample rate conversion to and from the different sampling rates. Overall though, the worst/worst combined complexity increase in Table 13.2d, which represents a doubling of the audio bandwidth of AMR-WB (SWB at 32kHz = +69.0%), is well within the complexity goal for EVS which was to be less than twice the complexity of AMR-WB.

The coder operates on 20 msec frames and the algorithmic delay (frame size plus look-ahead) is less than or equal to 32 msec.

13.3 EVS JBM Objective Performance Evaluation in Channel Aware Mode

The objective performance evaluation results on EVS channel aware mode presented below are pertaining to the modes that were tested in EVS characterization testing. In particular, the delay and JICO (jitter induced concealment operations) objective performance conformance are reported as defined by 3GPP TS 26.114 [13] for delay error profiles 1-6 (MTSI profiles) and 7-10 (VoLTE profiles). The four VoLTE profiles were defined in Annex D of the EVS-7c [23] document and used for EVS Characterization Testing. The evaluation was done by setting the frame offset for partial redundancy to 3 (o=3) and frame erasure rate to HI (p = HI).

The test conditions used and the JBM performance evaluation results are given in Table 13.3.

Table 13.3: Conditions tested and conformance results of JBM objective performance evaluation for EVS channel aware mode

Label	Condition	Codec	Bit rate	DTX	Level	FER/Profile	RF	Obj Requirements Pass/Fail
1	channel aware mode clean channel (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 1	on	Pass
2	channel aware mode frame erasures (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 2	on	Pass
3	channel aware mode frame erasures (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 3	on	Pass
4	channel aware mode frame erasures (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 4	on	Pass
5	channel aware mode MTSI bundled frame erasures (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 5	on	Pass
6	channel aware mode clean channel (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 6	on	Pass
7	channel aware mode clean channel (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 7	on	Pass
8	channel aware mode frame erasures (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 8	on	Pass
9	channel aware mode frame erasures (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 9	on	Pass
10	channel aware mode frame erasures (p=HI, o=3)	EVS-WB	13.2 kbps	on	-26 dBov	Profile 10	on	Pass
11	channel aware mode clean channel (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 1	on	Pass
12	channel aware mode frame erasures (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 2	on	Pass
13	channel aware mode frame erasures (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 3	on	Pass
14	channel aware mode frame erasures (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 4	on	Pass
15	channel aware mode MTSI bundled frame erasures (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 5	on	Pass
16	channel aware mode clean channel (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 6	on	Pass
17	channel aware mode clean channel (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 7	on	Pass
18	channel aware mode frame erasures (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 8	on	Pass
19	channel aware mode frame erasures (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 9	on	Pass
20	channel aware mode frame erasures (p=HI, o=3)	EVS-SWB	13.2 kbps	on	-26 dBov	Profile 10	on	Pass

The measurements are performed using scripts derived from the selection processing with necessary updates to run the channel aware mode.

Item 2 of TS 26.114, clause 8.2.3.1 (design guidelines for JBM minimum performance requirements) underscores the need for handling conditions that cause higher JICO even at the expense of increased buffering times.

"2. If the limit of jitter induced concealment operations cannot be met, it is always preferred to increase the buffering time in order to avoid growing jitter induced concealment operations going beyond the stated limit above. This guideline applies even if that means that end-to-end delay requirement given in 3GPP TS 22.105 [26] can no longer be met;".

Increased error burst lengths are common with network conditions that result in higher JICO. In such cases of larger error bursts, the EVS channel aware mode can be configured to utilize increased buffering times via higher FEC offsets to maximize the availability of partial copies. In particular, EVS channel aware mode is designed to use FEC offsets 5 and 7 to address the above mentioned scenario.

The results in Table 13.3 show that the EVS JBM in channel aware mode conforms to the delay and JICO objective requirements defined by 3GPP TS 26.114 for all configurations with FEC offsets 2 and 3.

13.4 Frequency Response

13.4.1 Evaluation of Codec Performance with respect to Acoustic Test Cases based on EVS v.12.1.0 [7]

The measurement results presented in this clause are based on the release codec version 12.1.0. The measurement results presented include the fullband version which is available in the release version. It shows that the error found in the pre-release version in wideband 9.6 kb/s mode is corrected. For all tests except 5.9 kb/s bitrate DTX was deactivated.

13.4.1.1 Evaluation Setup

The first step in the evaluation is the scaling with regard to a certain overload point. Two overload points (3.0 vs 9.0 dBm0) for the conversion from the physical unit Volt to 16-bit scale were taken into account.

Since the overload point (OVLP) refers to a full-scale sine wave (with level T_{max} = -3.01 dBov according to [29]), for the scaling between dBov and dBm0 resp. dBV, the following notation can be made:

T_{max} = -3.01 dBov

$$x[\mathrm{dBov}] = y[\mathrm{dBm0}] - (OVLP\ [\mathrm{dBm0}] - T_{max}[\mathrm{dBov}])$$

$$\leftrightarrow x[\mathrm{dBov}] = z[\mathrm{dBV}] + 2.21dB - (OVLP\ [\mathrm{dBm0}] - T_{max}[\mathrm{dBov}])$$

Example: For an OVLP of 3.0 dBm0, the scaling between dBV and dBov is defined as:

$$x[\mathrm{dBov}] = z[\mathrm{dBV}] + 2.21\ \mathrm{dB} - (3.0\ \mathrm{dBm} + 3.01\ \mathrm{dBov}))$$

$$\leftrightarrow x[ov] = \frac{z[V]}{10^{3.81}/_{20}} = \frac{z[V]}{1.5488}$$

The scaling back to the physical unit Volt was applied in the corresponding inverse way.

After scaling, the next step included the encoding and decoding of the audio data. This was conducted with the provided command line executable. The source code was not recompiled to a new binary.

For the evaluation of narrowband, wideband, and super-wideband mode, all bit rates which are available in each bandwidth mode according to Table 1 of [2] were used.

13.4.1.2 General

Several tests according to 3GPP TS 26.132 [29] were performed in order to evaluate the performance according to [14] of the EVS codec. Only frequency response results are reported in this subclause. Further results, e.g. distortion measurement results, can be found in Annex D.

TS 26.132 is originally intended for acoustic testing of terminals. Since the EVS codec is regarded as the "device under test", only electrical insertions are reasonable for testing and thus only measurements in (acoustic) receiving direction are taken into account.

In narrowband the test signal bandlimitation as defined in 3GPP TS 26.132 [29] was used. For superwideband the ITU-T P.501 [30] test signals were downsampled to 32 kHz. For fullband the original speech signals from ITU-T P.501 [30] were used.

With this approach, the EVS codec can be evaluated with typical test scenarios, which will occur in real-life applications with mobile phones.

The following graphs include multiple curves representing the different bit rates within each bandwidth mode. For the sake of clarity, the corresponding legends are not repeated in each graph, Table 13.4 shows the legends used in the following sections.

Table 13.4: Legends for different bit rates

5.9 kbit/s 7.2 kbit/s 8.0 kbit/s 9.6 kbit/s 13.2 kbit/s 16.4 kbit/s 24.4 kbit/s	5.9 kbit/s 7.2 kbit/s 8.0 kbit/s 9.6 kbit/s 13.2 kbit/s 16.4 kbit/s 24.4 kbit/s 32.0 kbit/s 48.0 kbit/s 64.0 kbit/s 96.0 kbit/s 128.0 kbit/s
NB mode	**WB mode**
9.6 kbit/s 13.2 kbit/s 16.4 kbit/s 24.4 kbit/s 32.0 kbit/s 48.0 kbit/s 64.0 kbit/s 96.0 kbit/s 128.0 kbit/s	16.4 kbit/s 24.4 kbit/s 32.0 kbit/s 48.0 kbit/s 64.0 kbit/s 96.0 kbit/s 128.0 kbit/s
SWB mode	**FB mode**

13.4.1.3 EVS-Mode: Narrowband (NB) – Frequency Response with Real Speech

In narrowband mode, a sampling rate of 8 kHz and all bit rates (5.9, 7.2, 8.0, 9.6, 13.2, 16.4 and 24.4 kbit/s) according to Table 1 of [2] were used. The two possible target overload points 3.0 and 9.0 dBm0 were used by default for all analyses.

The following results are produced by applying the measurement instructions according to clause 7.4.2 of [29]. To simulate also the impact of level variations, additional overload points of 21.0 and 39.0 dBm0 were also simulated. These overload points do not represent a realistic conversion, they are only used for checking the linearity of the codec and can be regarded as attenuations of 18.0 resp. 36.0 dB compared to the overload point of 3.0 dBm0.

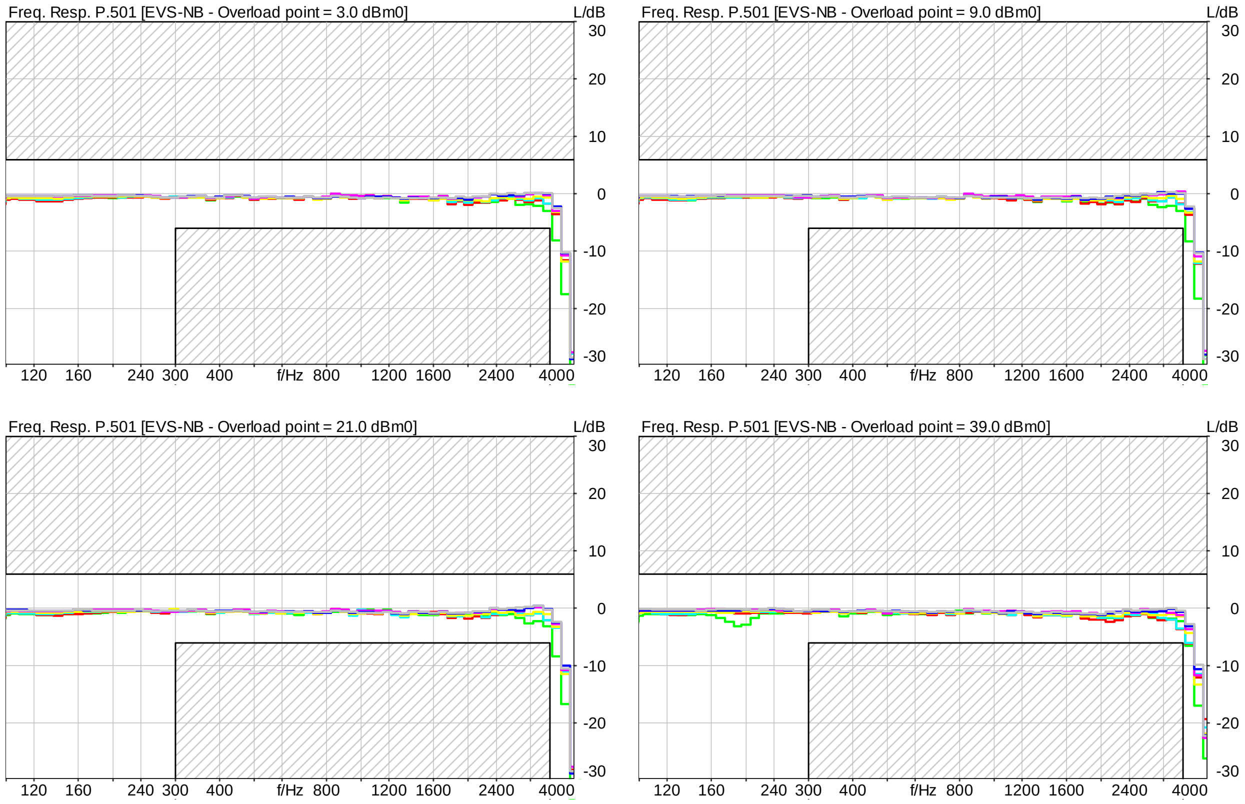


Figure 13.4: Frequency response for NB for different overload points

The results of this analysis are shown in Figure 13.4. For the default and extra overload points, the codec does not violate the given tolerance scheme according to [14] for all bit rates.

The results when using 1/3rd octave instead of 1/12th octave analysis are in Figure 13.5.

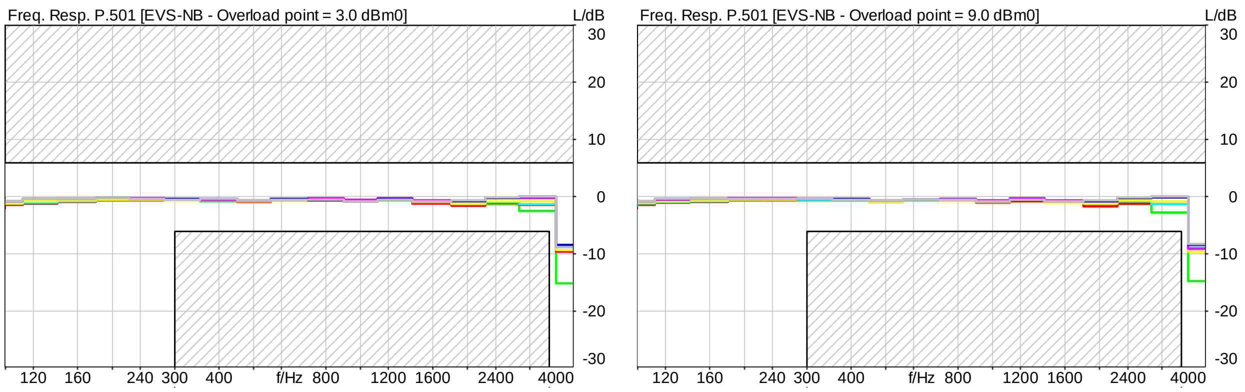


Figure 13.5: Frequency response EVS NB in 1/3rd Oct. for different overload points with P.501 speech signals

13.4.1.4 EVS-Mode: Wideband (WB) – Frequency Response with Real Speech

In wideband mode, a sampling rate of 16 kHz and all bit rates (5.9, 7.2, 8.0, 9.6, 13.2, 16.4, 24.4, 32.0, 48.0, 64.0, 96.0 and 128.0 kbit/s) according to Table 1 of [2] were used. The two possible target overload points 3.0 and 9.0 dBm0 were used by default for all analyses.

The following results are produced by applying the measurement instructions according to clause 8.4.2 of [28]. To simulate also the impact of level variations, additional overload points of 21.0 and 39.0 dBm0 were also simulated. These overload points do not represent a realistic conversion, they are only used for checking the linearity of the codec and can be regarded as attenuations of 18.0 resp. 36.0 dB compared to the overload point 3.0 dBm0.

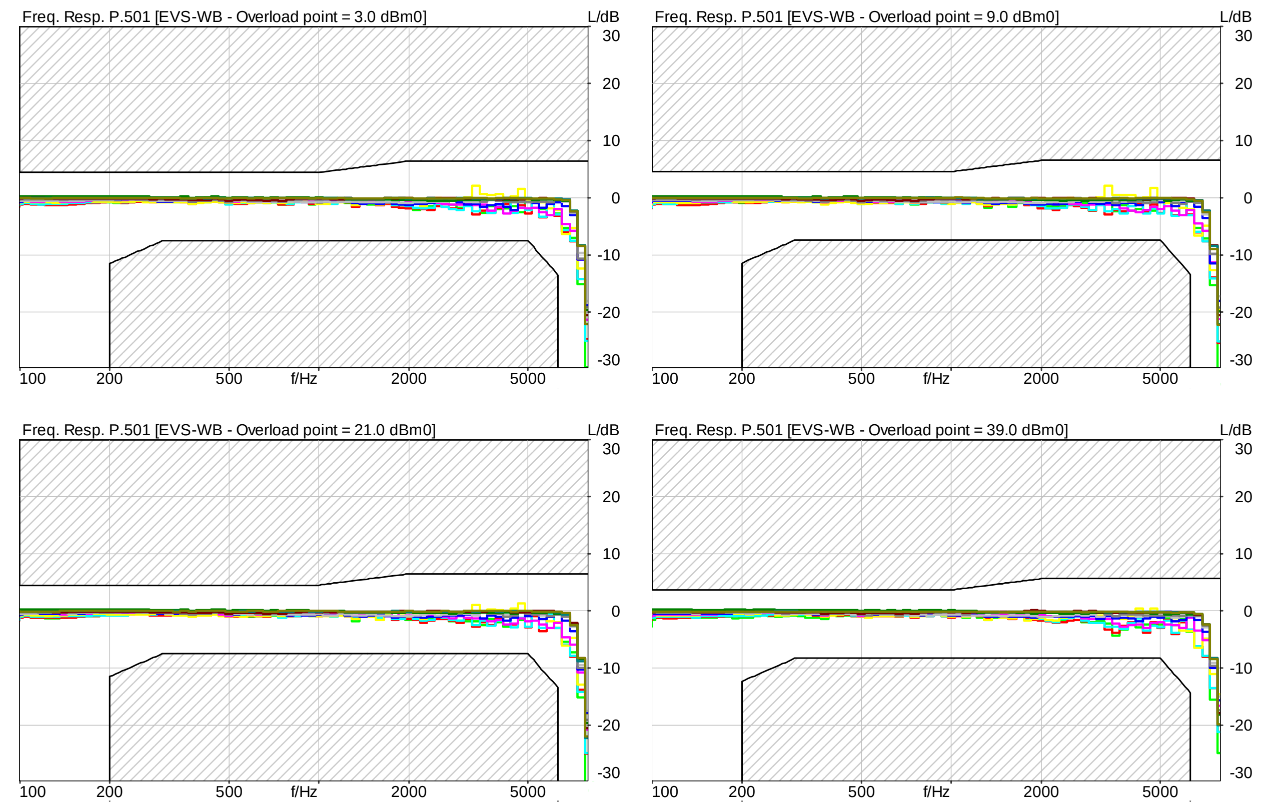


Figure 13.6: Frequency response for WB for different overload points

The results of this analysis are shown in Figure 13.6. For the default and extra overload points, the codec does not violate the given tolerance scheme according to [14] for all bit rates.

Instead of 1/12th octave analysis was used. The results using 1/3rd octave analysis are shown in Figure 13.7.

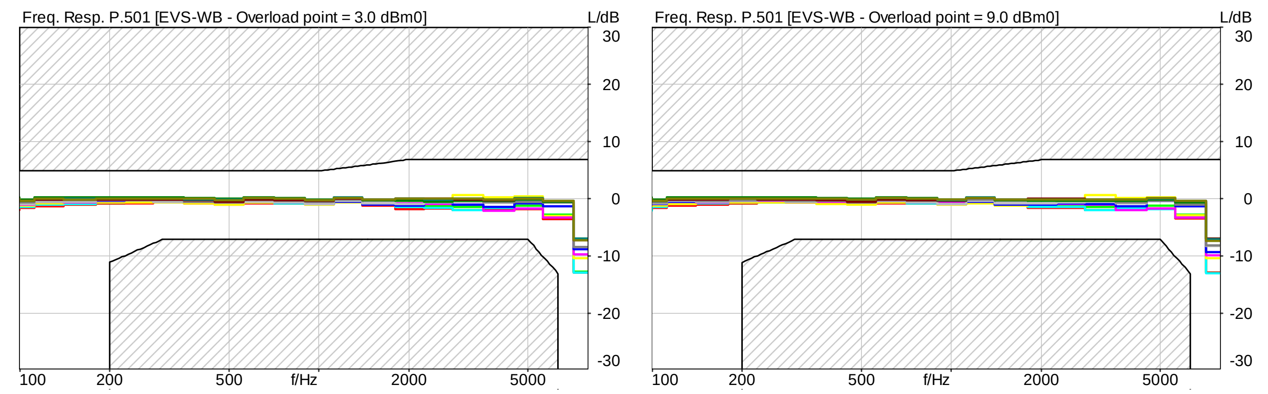


Figure13.7: Frequency response EVS WB in 1/3rd Oct. for different overload points with P.501 speech signals

13.4.1.5 EVS-Mode: Super-Wideband (SWB) – Frequency Response with Real Speech

In super-wideband mode, a sampling rate of 32 kHz and all bit rates (9.6, 13.2, 16.4, 24.4, 32.0, 48.0, 64.0, 96.0 and 128.0 kbit/s) according to Table 1 of [2] were used. The two possible target overload points 3.0 and 9.0 dBm0 were used by default for all analyses.

The following results are produced by applying measurement instructions similar to clause 8.4.2 of [28] which are adapted to super-wideband by replacing the source signal with a fullband version of the same file. To simulate also the impact of level variations, additional overload points of 21.0 and 39.0 dBm0 were also simulated. These overload points do not represent a realistic conversion; they are only used for checking the linearity of the codec and can be regarded as attenuations of 18.0 resp. 36.0 dB compared to the overload point 3.0 dBm0.

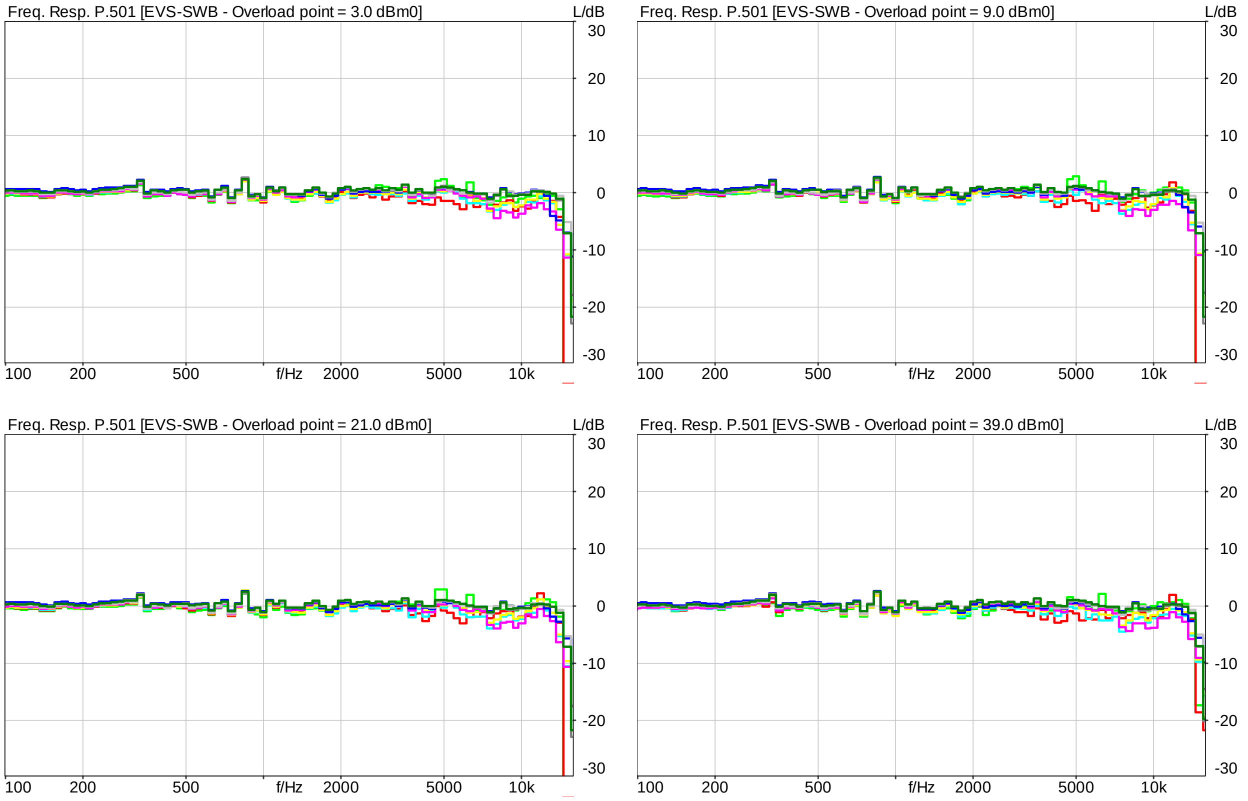


Figure 13.8: Frequency response for SWB for different overload points

The results of this analysis are shown in Figure 13.8. For the default and extra overload points, the codec provides an accurate transmission behavior for all bit rates.

The results for the 1/3rd octave analysis are shown in Figure 13.9.

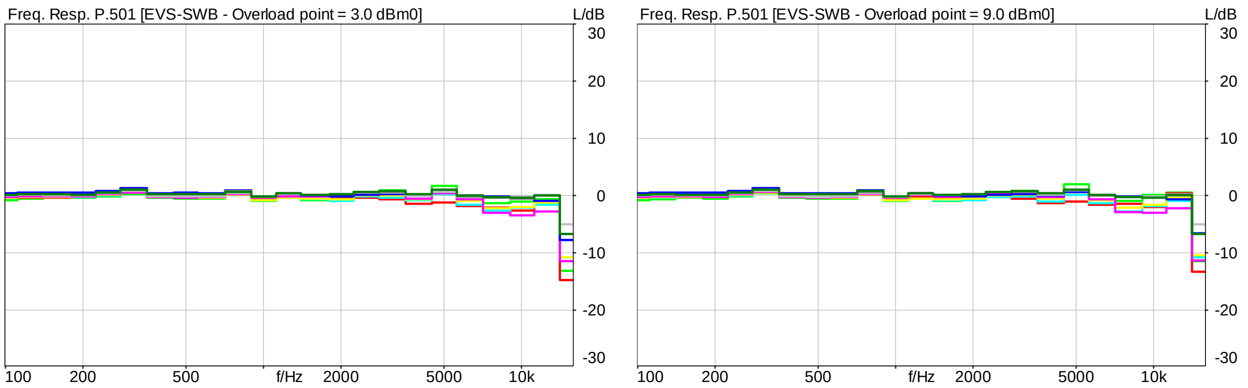


Figure13.9: Frequency response EVS SWB in 1/3rd Oct. for different overload points with P.501 speech signals

13.4.1.6 EVS-Mode: Fullband (FB) – Frequency Response with Real Speech

In fullband mode, a sampling rate of 48 kHz and all possible bit rates (16.4, 24.4, 32.0, 48.0, 64.0, 96.0 and 128.0 kbit/s) according to Table 1 of [2] were used. The two possible target overload points 3.0 and 9.0 dBm0 were used by default for all analyses.

The following results are produced by applying measurement instructions similar to clause 8.4.2 of [28] which are adapted to fullband by replacing the source signal with a fullband version of the same. To simulate also the impact of level variations, additional overload points of 21.0 and 39.0 dBm0 were also simulated. These overload points do not represent a realistic conversion; they are only used for checking the linearity of the codec and can be regarded as attenuations of 18.0 resp. 36.0 dB compared to the overload point 3.0 dBm0. The results of this analysis are shown in Figure 13.10.

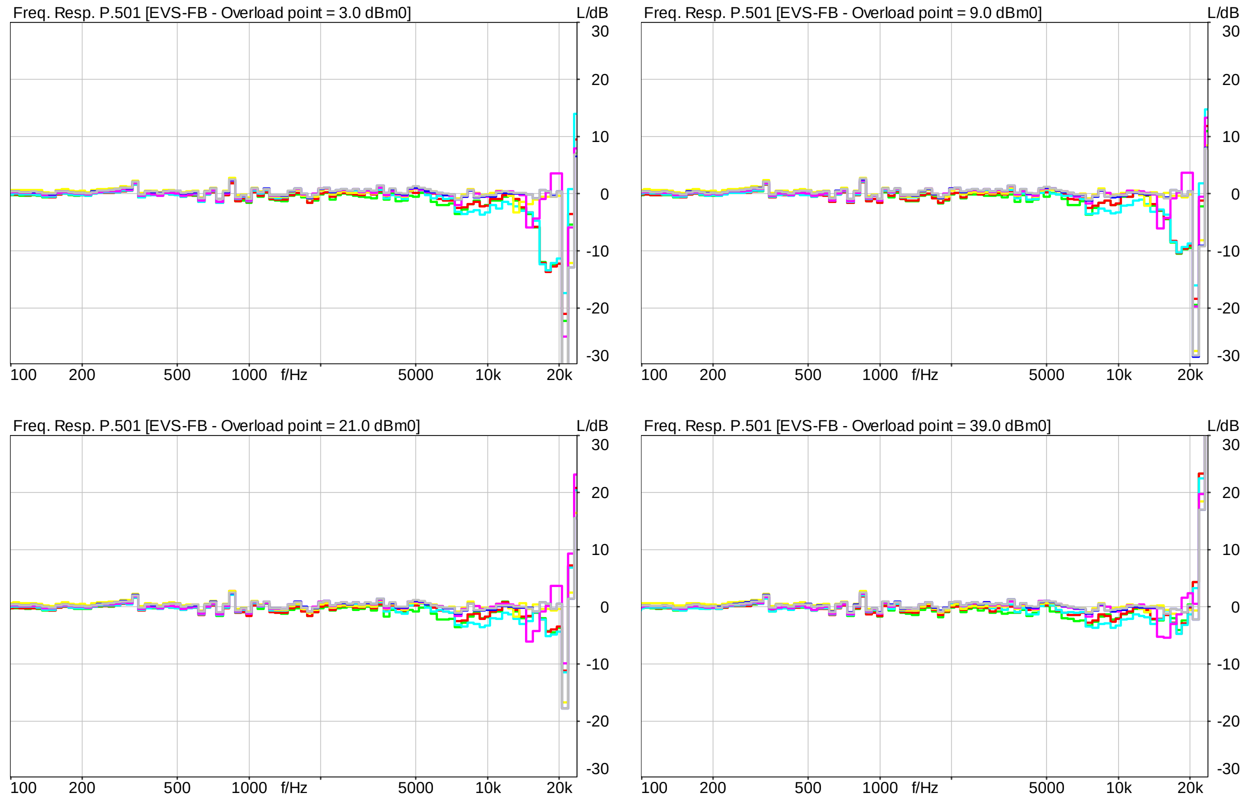


Figure 13.10: Frequency response for FB for different overload points

For lower codec rates (16.4 and 24.4 kbit/s, red/green curves), the transmission characteristics show some slight degradations for frequencies between 15 and 20 kHz.

13.4.1.7 Conclusions

In narrowband mode, all bit rates pass all requirements of [14].

Across all bitrates and for various operating bandwidths NB/WB/SWB, the EVS codec version 12.1.0 exceeds the frequency response tolerance requirements over an extended range of overload points when tested using both real speech and composite source signal.

The fullband mode was introduced with the release version 12.1.0. Since no requirements are available for this mode, results can only be reported. However, besides some slight degradation in the frequency response evaluations, the codec accurately performs on all tests.

Note that more detailed results can be found in Annex D.

13.5 Further Evaluations

Further evaluation results, e.g. distortion measurement results, can be found in Annex D.

13.6 Conclusions on Objective Evaluations

The objective evaluation results on the single joint EVS candidate demonstrated a "pass" at every condition. With this, the EVS coder fully meets all objective performance requirements for each database and also the acoustic requirements.

Annex A: ToR Tests in Selection Phase

A.1 ToR Tests for Requirements

Table A.1 summarizes the results for the Requirements ToR tests over the 24 Experiments. Each row of the table shows results of ToR tests for a single Experiment -- results for Test#1 on the left and for Test#2 on the right. For each Experiment, the table shows the #Requirement ToR's followed by the Test#1 label, #ToRs passed, #ToRs failed, Test#2 label, #ToRs passed, #ToRs failed and finally the Percent of ToRs passed across both Tests within the Experiment.

On the far right side of the table, "Percent ToRs Passed" values are shown for each of the four Groups of Experiments: 100% for NB, 99.58% for WB, 90.38% for EVS AMR-WB IO, and 97.93% for SWB. Finally, at the bottom of the table, "Percent ToRs Passed" for the entire Selection Phase is reported - 25 of 778 ToRs were failed for a Percent Passed value of 96.79%. It is noted that 20 (i.e., 80%) of those 25 failures occurred in the EVS AMR-WB IO Experiments.

Over all of the DGTT comparisons where the test was "CuT Not Worse Than REF", 61% of those comparisons showed that the CuT was significantly "Better Than" the REF.

Table A.1: Summary ToR Test Results for Requirements

#Req	Test#1			Test#2			% Passed	Group
	Label	#PASS	#FAIL	Lavel	#PASS	#FAIL		
23	bn1	23	0	cn1	23	0	100.0%	
18	an2	18	0	bn2	18	0	100.0%	
15	an3	15	0	cn3	15	0	100.0%	NB
21	an4	21	0	bn4	21	0	100.0%	100.00%
24	bw1	24	0	cw1	24	0	100.0%	
24	bw2	24	0	cw2	24	0	100.0%	
12	aw3	11	1	bw3	12	0	95.8%	
16	aw4	16	0	bw4	16	0	100.0%	
16	bw5	16	0	cw5	16	0	100.0%	
16	aw6	16	0	cw6	16	0	100.0%	WB
12	aw7	12	0	cw7	12	0	100.0%	99.58%
18	ai1	17	1	bi1	17	1	94.4%	
18	ai2	18	0	ci2	11	7	80.6%	
17	ai3	12	5	ci3	17	0	85.3%	
16	bi4	16	0	ci4	14	2	93.8%	
19	ai5	16	3	bi5	18	1	89.5%	AMR-IO
16	bi6	16	0	ci6	16	0	100.0%	90.38%
16	bs1	16	0	cs1	16	0	100.0%	
14	as2	14	0	bs2	14	0	100.0%	
9	as3	7	2	bs3	8	1	83.3%	
9	bs4	9	0	cs4	8	1	94.4%	
14	as5	14	0	bs5	14	0	100.0%	
10	as6	10	0	cs6	10	0	100.0%	SWB
16	bs7	16	0	cs7	16	0	100.0%	97.73%
389							96.79%	
		# REQ ToRs			778			
		# REQ Failures			25			
753 of 778 Requirements passed (96.79%)								

Table A.2 shows a list of the 25 failures for Requirement ToRs. Note that, consistent with the attached spreadsheet, the Experiment column is color-coded for the LL that conducted the test

Delta	Dynastat	Mesaqin

Table A.2: Requirement ToR Failures

#	Exp	Test	Mean	Ref	Mean	Diff	SEmd	T	Set	REQ	Test vs Ref	ToR Result
1	aw3	c19	3.229	c08	3.365	0.135	0.063	2.157	1A	NWT	WT	FAIL
2	ai1	c42	3.510	c01	4.203	0.693	0.069	9.974	5A	NWT OR	WT	FAIL
		c42	3.510	c23	3.667	0.156	0.064	2.432		NWT	WT	
3	bi1	c39	4.229	c01	4.500	0.271	0.066	4.101	5A	NWT OR	WT	FAIL
		c39	4.229	c17	4.339	0.109	0.066	1.663		(NWT AND	WT	
		c39	4.229	c18	4.198	-0.031	0.072	-0.432		NWT)	NWT	
4	ci2	c29	3.792	c01	4.755	0.964	0.070	13.725	5A	NWT OR	WT	FAIL
		c29	3.792	c12	3.938	0.146	0.060	2.413		NWT	WT	
5	ci2	c30	3.823	c01	4.755	0.932	0.061	15.191	5A	NWT OR	WT	FAIL
		c30	3.823	c13	3.953	0.130	0.065	1.998		NWT	WT	
6	ci2	c32	3.885	c01	4.755	0.870	0.064	13.522	5A	NWT OR	WT	FAIL
		c32	3.885	c14	3.927	0.042	0.070	0.595		(NWT AND	NWT	
		c32	3.885	c15	4.057	0.172	0.064	2.693		NWT)	WT	
7	ci2	c33	3.948	c01	4.755	0.807	0.063	12.800	5A	NWT OR	WT	FAIL
		c33	3.948	c14	3.927	-0.021	0.067	-0.312		(NWT AND	NWT	
		c33	3.948	c15	4.057	0.109	0.062	1.753		NWT)	WT	
8	ci2	c37	3.120	c20	3.276	0.156	0.064	2.432	5A	NWT	WT	FAIL
9	ci2	c41	3.245	c23	3.286	0.042	0.054	0.776	5A	NWT AND	NWT	FAIL
		c41	3.245	c24	3.505	0.260	0.067	3.859		NWT	WT	
10	ci2	c42	3.406	c23	3.286	-0.120	0.067	-1.801	5A	NWT AND	BT	FAIL
		c42	3.406	c24	3.505	0.099	0.059	1.681		NWT	WT	
11	ai3	c26	3.672	c13	3.818	0.146	0.065	2.237	5A	NWT	WT	FAIL
12	ai3	c27	3.802	c15	4.057	0.255	0.064	3.984	5A	NWT	WT	FAIL
13	ai3	c28	3.906	c17	4.151	0.245	0.073	3.370	5A	NWT	WT	FAIL
14	ai3	c29	3.891	c19	4.146	0.255	0.074	3.469	5A	NWT	WT	FAIL
15	ai3	c33	3.958	c16	4.094	0.135	0.071	1.899	5A	NWT	WT	FAIL
16	ci4	c26	3.964	c01	4.734	0.771	0.067	11.530	5A	NWT OR	WT	FAIL
		c26	3.964	c12	4.193	0.229	0.067	3.428		NWT	WT	
17	ci4	c32	3.776	c17	4.026	0.250	0.071	3.529	5A	NWT	WT	FAIL
18	ai5	c20	2.964	c09	3.125	0.161	0.076	2.124	5A	NWT	WT	FAIL
19	ai5	c27	1.917	c07	1.844	-0.073	0.072	-1.010	5A	BT OR NWT	NWT	FAIL
		c27	1.917	c08	2.391	0.474	0.081	5.885			WT	
20	ai5	c28	2.359	c08	2.391	0.031	0.077	0.405	5A	BT OR NWT	NWT	FAIL
		c28	2.359	c09	3.125	0.766	0.077	9.920			WT	
21	bi5	c27	2.177	c07	2.099	-0.078	0.069	-1.135	5A	BT OR NWT	NWT	FAIL
		c27	2.177	c08	2.656	0.479	0.078	6.132			WT	
22	as3	c18	4.380	c09	4.479	0.099	0.057	1.736	2A	NWT	WT	FAIL
23	as3	c20	4.479	c01	4.651	0.172	0.061	2.809	2A	NWT OR	WT	FAIL
		c20	4.479	c11	4.667	0.188	0.049	3.844		NWT	WT	
24	bs3	c20	4.130	c01	4.417	0.286	0.059	4.825	2A	NWT OR	WT	FAIL
		c20	4.130	c11	4.354	0.224	0.062	3.633		NWT	WT	
25	cs4	c20	4.417	c01	4.776	0.359	0.064	5.653	2A	NWT OR	WT	FAIL
		c20	4.417	c11	4.740	0.323	0.070	4.635		NWT	WT	

There were two instances where a Requirement ToR was failed in both of the LLs that conducted the Test. Those two "Systematic Failures" are listed below.

- Failures #19 and #21, Test *ai5* and *bi5*, CuT condition c27 vs. REF conditions c07 and c08 for the combination ToR "c27 BT c07 OR c27 NWT c08"
- Failures #23 and #24, Test *as3* and *bs3*, CuT condition c20 vs. REF conditions c01 and c11 for the combination ToR "c20 NWT c01 OR c20 NWT c11"

Annex A contains a complete description of the CuT and REF conditions involved in the 25 Requirement ToRs that were failed.

In Table A.2, the systematic noisy speech failure at 64 kbps in Experiment S3 is due to a rather serious fixed point implementation bug that affected the conditions #23, #24, and #25. In particular, the quantization steps used in the encoder at 64 kbps were twice as large as in decoder. This bug has been in time resolved for Characterization testing and is shown to address the issue from Experiments S2 and M2 in Characterization testing. Furthermore, another systematic bug that is related to EVS AMR-WB IO in mixed/music testing in Experiment I5 conditions #19 and #21 was noted at lower bit rates (i.e., at 6.6 kbps). From the Selection Experiment I5, while the subjective quality from the EVS AMR-WB IO Case B (at 6.6 kbps) shows minor improvement to AMR-WB at same bit rate, the performance requirement of better than AMR-WB 6.6 kbps could not be met.

A.2 ToR Tests for Objectives

Table A.3 summarizes the results for the Objectives ToR tests over the 24 Experiments. The table is organized in the same manner as the Summary results for Requirement ToRs shown in Table A.1.

On the far right side of the table, "Percent ToRs Passed" values are shown for each of the four Groups of Experiments: 98.15% for NB, 92.90% for WB, 55.42% for EVS AMR-WB IO, and 90.58% for SWB. Finally, at the bottom of the table, "Percent ToRs Passed" for all of the Objectives in the Selection Phase is reported - 110 of 590 ToRs were not-passed for a Percent Passed value of 81.36%. It is noted that 67% of those 110 ToRs that were "not-passed" occurred in the EVS AMR-WB IO Experiments.

Table A.3: Summary ToR Test Results for Objectives

ToRs for OBJECTIVES								
#Obj	Test#1			Test#2			% Passed	Group
	Label	#PASS	#Not-passed	Label	#PASS	#Not-passed		
23	bn1	22	1	cn1	23	0	97.8%	
12	an2	11	1	bn2	12	0	95.8%	
12	an3	12	0	cn3	10	2	91.7%	NB
9	an4	7	2	bn4	9	0	88.9%	98.15%
20	bw1	20	0	cw1	12	8	80.0%	
14	bw2	14	0	cw2	13	1	96.4%	
12	aw3	9	3	bw3	10	2	79.2%	
4	aw4	4	0	bw4	4	0	100.0%	
15	bw5	15	0	cw5	15	0	100.0%	
10	aw6	8	2	cw6	9	1	85.0%	WB
12	aw7	12	0	cw7	12	0	100.0%	92.90%
10	ai1	8	2	bi1	7	3	75.0%	
18	ai2	9	9	ci2	6	12	41.7%	
16	ai3	5	11	ci3	10	6	46.9%	
8	bi4	2	6	ci4	2	6	25.0%	
19	ai5	13	6	bi5	11	8	63.2%	AMR-IO
12	bi6	9	3	ci6	10	2	79.2%	55.42%
14	bs1	14	0	cs1	14	0	100.0%	
8	as2	8	0	bs2	8	0	100.0%	
6	as3	3	3	bs3	4	2	58.3%	
6	bs4	6	0	cs4	6	0	100.0%	
14	as5	14	0	bs5	14	0	100.0%	
9	as6	6	3	cs6	7	2	72.2%	SWB
12	bs7	12	0	cs7	9	3	87.5%	90.58%
295							81.36%	
		# OBJ ToRs			590			
		# OBJ Not-passed			110			
480 of 590 Objectives passed (81.36%)								

A.3 ToR Tests by Sets

Table A.4 shows ToR results by Sets, where Sets were defined by the EVS sub-working group. The Total number of ToRs and the number of ToRs failed for Requirements are shown in the left-hand side of the table. For Objectives, the Total number of ToRs and the number of ToRs not-passed are shown in the right-hand side of the table.

Table A.4: ToR Test Results for Requirements and Objectives by Set

Requirement ToRs				Objective ToRs		
Set	#ToRs	#FAIL		Set	#ToRs	#Not Passed
1A	88	1		1A	86	15
1B	40	0		1B	32	1
2A	68	4		2A	52	5
3A	42	0		3A	36	1
3B	20	0		3B	18	5
4A	92	0		4A	52	2
4B	44	0		4B	24	0
4C	56	0		4C	44	0
4D	120	0		4D	80	7
5A	208	20		5A	166	74
Total	778	25		Total	590	110

The ToRs in Set 5A account for most of the ToRs failed for Requirements (80%) and also for most of the ToRs "Not-passed" for Objectives (67%).

A.4 Comparison of Listening Labs

There was a discrepancy among the LLs in the number of Requirement ToRs failed. The LL running the most Tests (Lab-b with 18) and therefore the most ToR conditions showed the fewest (3) Requirement ToR failures. A logical hypothesis for this discrepancy is that the LL with the lowest failure rate might have lower sensitivity to quality differences and lower resolving-power in the T-tests. Table 9 shows results of analyses designed to test that hypothesis. The table shows that Lab-b had the lowest ToR failure rate (1%) but also had virtually the same sensitivity to differences as the other two LLs. In fact, the Minimum Significant Differences for all three LLs were remarkably similar, 0.118 for Lab-a, 0.116 for Lab-b, and 0.114 for Lab-c. Note that these values have been adjusted to take into account the differences in the Average Range of the Rating Scale used by the listeners tested in the three individual LLs.

Table A.5: Comparison of LLs for the Sensitivity and Precision of the Requirements ToRs

Listening Lab	# Tests Conducted	# ToRs Tested	# ToRs Failed	Failure Rate	Avg. Range	Avg. SE_{MD}	Adjusted SE_{MD}	Minimum Signif. Diff.
Delta (a)	15	229	12	5.2%	3.45	0.0728	0.0715	0.118
Dynastat (b)	18	301	3	1.0%	3.23	0.0667	0.0699	0.116
Mesaqin (c)	15	248	10	4.0%	3.48	0.0711	0.0692	0.114

Table A.6 Shows Means and Standard Deviations across conditions for each of the two Tests conducted within each of the 24 Experiments involved in the Selection Phase. In the last column on the right side of the table is the correlation of the condition Mean scores between the two Tests/LLs. Annex B contains two plots for each Experiment. The first plot shows MOS/DMOS for the MNRU Reference conditions for the two Tests within the Experiment. The second plot shows a scatter-plot of MOS/DMOS for the two tests within the Experiment.

Table A.6: Comparison of scores for the two Tests/LLs within each Experiment

Experiment			Test#1			Test#2			Correlation
Label	Method	# cond	LL	Mean	Stdev	LL	Mean	Stdev	
n1	ACR	42	Dynastat	3.933	0.625	Mesaqin	3.688	0.643	0.983
n2	ACR	36	Delta	2.990	0.552	Dynastat	3.750	0.611	0.958
n3	DCR	36	Delta	3.388	0.644	Mesaqin	3.718	0.708	0.982
n4	ACR	48	Delta	2.751	0.502	Dynastat	3.484	0.466	0.946
w1	ACR	48	Dynastat	3.840	0.742	Mesaqin	3.894	0.812	0.983
w2	ACR	48	Dynastat	3.474	0.680	Mesaqin	3.065	0.575	0.966
w3	DCR	30	Delta	3.498	1.008	Dynastat	3.963	0.871	0.920
w4	DCR	36	Delta	3.430	0.775	Dynastat	3.582	0.675	0.984
w5	DCR	30	Dynastat	3.708	0.967	Mesaqin	3.745	1.003	0.990
w6	DCR	36	Delta	3.033	0.830	Mesaqin	3.116	0.717	0.981
w7	DCR	24	Delta	3.340	0.942	Mesaqin	3.831	0.842	0.942
i1	ACR	48	Delta	3.029	0.934	Dynastat	3.794	0.733	0.901
i2	ACR	42	Delta	2.797	0.622	Mesaqin	3.294	0.738	0.989
i3	DCR	36	Delta	3.638	0.921	Mesaqin	3.991	0.845	0.970
i4	DCR	36	Dynastat	3.567	0.707	Mesaqin	3.805	0.726	0.967
i5	DCR	36	Delta	3.194	0.823	Dynastat	3.449	0.782	0.977
i6	DCR	36	Dynastat	3.062	0.835	Mesaqin	2.868	0.719	0.983
s1	DCR	36	Dynastat	4.317	0.729	Mesaqin	4.175	0.760	0.985
s2	DCR	36	Delta	3.389	0.831	Dynastat	3.721	0.681	0.972
s3	DCR	24	Delta	3.979	0.866	Dynastat	3.914	0.694	0.956
s4	DCR	24	Dynastat	3.832	0.690	Mesaqin	4.389	0.789	0.960
s5	DCR	36	Delta	3.264	0.812	Dynastat	3.465	0.742	0.959
s6	DCR	24	Delta	3.785	0.854	Mesaqin	4.361	0.716	0.888
s7	DCR	36	Dynastat	3.593	0.792	Mesaqin	3.385	0.697	0.977

Annex B: Overall Characterization of the EVS Codec

EVS is the next generation codec in 3GPP which provides an advantage over existing 3GPP coders in terms of:

- Extended audio bandwidth (super-wideband, fullband)
- Improved performance for narrowband and wideband speech
- Improved robustness against transmission errors
- Lower average bit rate through discontinuous transmission and through source-controlled variable bit rate operation for active speech
- Better performance for music and mixed content in all bandwidths
- Backward interoperability to AMR-WB by inclusion of EVS AMR-WB IO modes

The fixed-point EVS codec was rigorously tested using the ITU-T P.800 [25] methodology with naïve listeners, demonstrating fulfillment of all testable EVS WID objectives. The extensive Selection and Characterization testing required a budget exceeding 1Million €. The first EVS WID objective was to provide improvements for NB and WB services, and the NB improvement over AMR at all comparable bit rates was demonstrated. Similarly, the WB improvements over AMR-WB that EVS offers are apparent, including equivalence to Direct at 24.4 kbps.

To fulfill the SWB EVS WID objective, the EVS codec provides state-of-the-art SWB performance, both in benign conditions and in more realistic conditions of background noise and frame erasures. The WID objective for the robustness of EVS was demonstrated in SWB and also confirmed in NB and WB testing. Addressing an objective for improved performance in mixed content and music, the EVS codec provides a significant improvement over legacy codecs.

3GPP's rigorous and transparent standardization process involved the definition of demanding terms of reference (ToR's). The EVS codec was tested against these ToR's in three test phases and with extensive independent evaluations using an unprecedented budget. The test campaign included 70 subjective tests performed in 10 languages, several input signal categories, and using independent test labs.

The standardization successfully delivered the EVS codec standard with greatly enhanced performance as compared current codec standards from 3GPP, ITU-T and the IETF.

EVS is currently the best available codec for all mobile and VoIP applications.

The performance of the EVS codec excels, especially at low bit rates of up to 24.4 kbps, a feature of utmost importance for the deployment of cost-effective mobile services, a cornerstone of mobile operator businesses.

Annex C: EVS Permanent Documents in 3GPP FTP-site

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

Table C.1: EVS Permanent Project Documents

SA4 TDoc number	P-doc	Title
S4-140756	EVS-1	EVS codec development overview
S4-141016	EVS-2	EVS Project plan
S4-130522	EVS-3	EVS performance requirements
S4-130778	EVS-4	EVS design constraints
S4-140208	EVS-5b	Selection Rules
S4-140630	EVS-6b	Selection Deliverables
S4-141026	EVS-7b	Processing functions for selection phase
S4-141126	EVS-7c	Processing functions for characterization phase
S4-141036	EVS-8b	EVS Permanent Document EVS-8b: Test plan for selection phase
S4-141131	EVS-8c	EVS Permanent Document EVS-8c: Test plan for characterization phase
S4-100547	EVS-10	List of potential reference codecs
S4-140983	EVS-11	EVS Verification Items
S4-140918	EVS-12	Incorporating EVS into TS 26.114

The latest version of these documents can be found in the following link:

http://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/EVS_Permanent_Documents/

Annex D: Attachments

Attachements to the present document include:

1) Excel sheet containing the EVS Selection Phase Test Results
2) Excel sheet containing the EVS Characterization Phase Test Results
3) Contribution S4-150559: HEAD acoustics, EVS – Objective Codec Evaluation - including results of codec release version 12.1.0, 3GPP SA4#83 Meeting, 13-17 April 2015

Annex E: Change History

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2014-12	66	SP-140730			Version 1.0.0 presented for approval at SA66		1.0.0
2014-12	66					1.0.0	12.0.0
2015-03	67	SP-150088	0001	2	Correction of values and figures	12.0.0	12.1.0
2015-06	68	SP-150205	0002	2	Inclusion of verification results	12.1.0	12.2.0
2015-09	69	SP-150434	0003	1	Corrections to the Complexity and Delay Analysis	12.2.0	12.3.0
2015-12	70				Version for Release 13	12.3.0	13.0.0
2016-03	71	SP-160064	0005	1	Corrections	13.0.0	13.1.0

History

Document history		
V13.0.0	January 2016	Publication
V13.1.0	April 2016	Publication