International Telecommunication Union

ITU-T

Technical Paper

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(06 November 2009)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS
Digital sections and digital line system – Access networks

GSTP-GVBR
Performance of ITU-T G.718

ITU-T

Summary

This Technical Paper compiles performance assessment of Recommendation ITU-T G.718. In particular, it presents all the subjective results from both steps of the Characterization Tests. Also, the verification results for the floating-point implementation of G.718 are provided. Finally, information on G.718 complexity, memory requirements and algorithmic delay are also presented.

Change Log

This document contains Version 1 of the ITU-T Technical Paper on "*Performance of ITU-T G.718*" approved at the ITU-T Study Group 16 meeting held in Geneva, 26 October – 6 November 2009.

Editor:	Milan Jelinek VoiceAge Corporation Canada	Tel: Fax: Email:	+1 819 821 8000 / 63893 +1 819 821 7937 milan.jelinek@usherbrooke.ca

Contents

Page

List of Tables

List of Figures

ITU-T Technical Paper GSTP-GVBR

Performance of ITU-T G.718

Summary

This Technical Paper compiles performance assessment of Recommendation ITU-T G.718 "Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s". In particular, it presents all the subjective results from both steps of the Characterization Tests. Also, the verification results for the floating-point implementation of G.718 are provided. Finally, information on G.718 complexity, memory requirements and algorithmic delay are also presented.

1 Scope

This Technical Paper compiles performance assessment of Recommendation ITU-T G.718. It presents the subjective results of the Characterization tests,. the verification results for the floating-point implementation, the complexity, the memory requirements and the algorithmic delay of G.718.

2 References

[1] Processing plan version 1.03 of the G.718 Characterization step 1 test, 8 February 2008.

[2] Listening test plan version 2.1of the G.718 Characterization step 1 test, 8 February 2008.

[3] TD 289 R1 (WP3/16), "Report of Question 9/16 Variable Bit Rate Coding of Speech Signals", Source: Rapporteur Q9/16, ITU-T SG16 meeting, June-July 2007.

[4] AH-08-58, "Summary of results for G.EV-VBR", Source: Q.7/12 Rapporteurs, Q7/12 meeting, April 2008.

[5] AH-08-25, "Listening Laboratory Report for the G.EV-VBR Characterization/Optimization Phase 1 Quality Assessment Test – Motorola Experiments 2a, 3a, and 4a", Source: Dynastat, Q7/12 meeting, April 2008.

[6] AH-08-26, "Listening Laboratory Report for the G.EV-VBR Characterization/Optimization Phase 1 Quality Assessment Test – Texas Instruments Experiments 3c and 4c", Source: Dynastat, Q7/12 meeting, April 2008.

[7] AH-08-27, "Listening Laboratory Report for the G.EV-VBR Characterization/Optimization Phase 1 Quality Assessment Test – Qualcomm Experiments 2c, 3b, and 4b", Source: Dynastat, Q7/12 meeting, April 2008.

[8] AH-08-29, "Results of EV-VBR Quality Assessment Optimization/Characterization Phase I Tests, Experiment 3b (NB-DCR, Interfering Talker, Music), Swedish", Source: L.M. Ericsson, Q7/12 meeting, April 2008.

[9] AH-08-32, "VoiceAge listening laboratory report on the results of Experiments 1a, 2b, and 4a of the EV-VBR Characterization, Phase I", Source: VoiceAge Corporation, Q7/12 meeting, April 2008.

[10] AH-08-32, "G.EV-VBR listening test report from Huawei Listening", Source: Huawei, Q7/12 meeting, April 2008.

[11] AH-08-34, "EV-VBR listening laboratory report (Experiment 3a, 4b, and 4e in Japanese)", Source: Panasonic (Q.7/12 Rapporteurs), Q7/12 meeting, April 2008.

[12] AH-08-51, "Results of EV-VBR Quality Assessment Optimization/Characterization Phase I Tests, Experiment 2c (WB-ACR Music), Experiment 4c (WB-DCR, Car Noise), Chinese", Source: Beijing Institute of Technology, L.M. Ericsson, Q7/12 meeting, April 2008.

[13] AH-08-52, "Nokia Listening Test Laboratory Report for EV-VBR Characterization", Source: Nokia, Q7/12 meeting, April 2008.

[14] AH-08-53, "EV-VBR: Listening laboratory Report", Source: France Telecom, Q7/12 meeting, April 2008.

[15] Processing plan version 1.7 of the G.718 Characterization Step 2 test, 20 October 2008.

[16] Listening test plan version 0.6 of the G.718 Characterization Step 2 test., 3 February 2009.

[17] TD 27 (WP 1/12), "Summary of the characterisation step2 of G.718", Source: Rapporteurs for Question 7/12, ITU-T SG12 meeting, March 2009.

[18] COM 12 – C 9 – E, "VoiceAge listening laboratory report on the results for Experiments 1c & 4b of the G.718 Characterization, Phase II", Source: VoiceAge, ITU-T SG12 meeting, March 2009.

[19] COM 12 – C 10 – E, "Listening Laboratory Report for the G.718 Characterization Phase Step 2 Quality Assessment Test – Qualcomm Experiments 1b and 2", Source: Dynastat, ITU-T SG12 meeting, March 2009.

[20] COM 12 – C 11 – E, "Listening Laboratory Report for the G.718 Characterization Phase Step 2 Quality Assessment Test – Motorola Experiments 1a and 1c ", Source: Dynastat, ITU-T SG12 meeting, March 2009.

[21] COM 12 – C 19 – E "Japanese listening laboratory report on the results of experiments 1a and 3a of the G.718 characterization phase step2", Source: Panasonic corporation, ITU-T SG12 meeting, March 2009.

[22] COM 12 – C 34 – E, "Report for Exp.3b and Exp4a for the G.718 Characterization Phase II Test in Chinese Language ", Source: Huawei Technologies, ITU-T SG12 meeting, March 2009.

[23] COM 12 – C 35 – E, "Results of G.718 Quality Assessment Optimization/Characterization Phase II Tests, Experiment 2 (WB-ACR Music), Experiment 4a (WB-DCR, Noisy speech car noise), Swedish ", Source: L.M. Ericsson, ITU-T SG12 meeting, March 2009.

[24] COM 12 – C 37 – E, "Results of characterization phase step2 of G.718 in French language ", Source: France Telecom, ITU-T SG12 meeting, March 2009.

[25] "Nokia Listening Test Laboratory Report for G.718 Characterization, Phase II", Source: Nokia, ITU-T SG12 meeting, March 2009.

[26] COM 16 – C 477 – E, "Proposed revision of the EV-VBR fixed-point code", Source: Nokia, ITU-T SG16 meeting, April 2008.

[27] AC-0809-Q09-04-R1, "Proposed maintenance of the fixed point simulation of the G.718 standard", Source: VoiceAge, L.M. Ericsson, ITU-T WP3/16 meeting, September 2008.

[28] COM 16 – C 49 – E, "Maintenance of G.718", Source: VoiceAge, ITU-T SG16 meeting, January-February 2009.

[29] AC-0907-Q09-02, "G.718 maintenance", Source: VoiceAge, ITU-T WP3/16 meeting, July 2009.

[30] COM 16 – C 48 – E, "Objective evaluation and submission of G.718 floating point Annex", Source: VoiceAge, ITU-T SG16 meeting, January-February 2009.

[31] ITU-T Rec. P.862 (2001), *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*

[32] ITU-T Rec.P.862.2 (2007), *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*

3 Abbreviations and acronyms

The following is the list of abbreviations and acronyms used in this document.

ACR Absolute category rating

BFER Bursty frame erasure rate

BT Better than *(in statistical analysis)*

CNG Comfort noise generator

DCR Degradation category rating

DMOS Degradation mean opinion score

DTX Discontinuous transmission

EV-VBR Embedded variable bit-rate

FER Frame erasure rate

MOS Mean opinion score

NB Narrowband

NR Noise reduction

NWT Not worse than *(in statistical analysis)*

PESQ Perceptual evaluation of speech quality

SD Standard deviation

SNR Signal-to-noise ratio

TC Transition Coding

VAF Voice activity factor

WB Wideband

WB-PESQ Wideband extension to PESQ

WMOPS Weighted million operations per second

4 History of G.718 "Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s"

ITU-T started the standardization of embedded variable bit rate coding in 1999. The codec development was pursued in Question 9 of Study Group 16, under the working name of EV-VBR (embedded variable bit-rate) codec. For EV-VBR, wideband rendering was mandated for all embedded layers while narrowband rendering was mandatory only for the lowest layers.

An initial phase of the codec evaluation was scheduled for March 2007 to select the baseline for further optimization. Four candidate codecs were evaluated in the selection phase. A solution jointly developed by Ericsson, Motorola, Nokia, Texas Instruments, and VoiceAge won the competition and was selected as the baseline codec.

Unlike in previous ITU-T standardization efforts, the following optimization phase was opened to all companies interested in improving the performance of the baseline codec. Nine other companies declared an intention to participate in this effort. Out of these nine companies, Matsushita, Huawei, France Telecom and Qualcomm contributed to the final solution. The optimization phase was completed in April 2008 by finalizing Step 1 of a comprehensive subjective evaluation of the codec called Characterization test. The fixed point description of the codec was adopted as Recommendation ITU-T G.718 in June 2008. The corresponding floating point description was approved in March 2009.

In the Step 2 of the Characterization Tests, the codec was subjectively tested in conditions not covered during the Step 1. In particular, this test characterized the G.718 codec in extremely difficult channel environments. The Step 2 was completed in March 2009 and this concluded the evaluation and assessment of the G.718 performance.

5 Scope of the Codec

In the following, the applications foreseen for G.718 are listed. These applications are partitioned into two groups: a primary group and a secondary group. The primary group comprises those applications that should benefit from an embedded scheme while having a great potential use i.e. applications that are most likely to employ G.718 early and in large numbers. As a result, primary applications are expected to "drive" the development of the standard, at least as regards schedule. The secondary group comprises applications likely to benefit from the availability of G.718 standard, but which are either unlikely to employ large numbers of G.718 audio coding devices or, at least on an interim basis, can also utilise some other audio coding standards without adversely impacting the economics of their application.

The following applications were targeted as primary applications:

– Packetized voice (VoIP, VoATM, IP phone, private networks)
– High quality audio/video conferencing
– Applications that benefit from congestion control
– Applications that benefit from differentiated QoS
– Applications that benefit from 3G and future wireless (e.g., 4G, WiFi) systems (packet switched conversational multimedia, multimedia content distribution)
– Multimedia streaming (e.g. video + audio involving bit-rate tradeoff)
– Multiple access home gateway

The following applications were targeted as secondary applications:

– Multicast content distribution (offline/online)
– Message retrieval systems
– CME/Trunking equipment
– Applications that require music on hold
– Network-based speech recognition using speech codec

Some general guidelines have driven the drafting of the preliminary ToR, as follows:

– Speech was a signal of primary interest but in high quality audio conferencing, background signals were considered, not as the noise anymore, but as a part of the signals that convey information
– To cope with heterogeneous accesses and terminals, it was important to consider not only bit-rate scalability but also bandwidth scalability and complexity scalability
– Narrowband/wideband signal capability with hi-fi bandwidth was a requirement and stereo/multi-channel capability was an objective (up to 20 kHz)
– Smoothen the bandwidth switching effects

– The bit range should cover low bit rate (around 8 kbit/s) to higher bit rate (approximately 32 kbit/s); for mobile users, it was highly desirable to introduce bitrates compatible with mobile links
– Fine-grain bit-rate scalability was a highly desirable feature to allow trade-off between speech and audio quality and the quality of other services (e.g., video)
– It was necessary to maintain the overall delay as low as possible to maintain a good quality of services requiring interactivity (however, delay requirement tend to have less importance in applications involving packetized voice, possibly combined with other media and/or in heterogeneous network environment); a trade-off was found between low delays and flexibility (scalability, ability to operate in various conditions with many types of signals etc.)

The G.718 codec operates on 20 ms frames and comprises five fixed-rate layers (Table 1), referred to as L1 (core layer) through L5 (the highest extension layer). It can accept wideband or narrowband signals sampled at either 16 or 8 kHz, respectively. The decoder can also provide output sampled at 8 or 16 kHz, which may be different from the sampling rate of the input. The wideband rendering is supported for all layers. The narrowband rendering is supported only for L1 and L2, meaning that if the encoder is presented with a narrowband input, only the first two layers are encoded. Similarly, if the narrowband option is invoked at the decoder, the highest synthesized layer is limited to L2.

6 Algorithm overview

G.718 is a narrowband (NB) and wideband (WB) embedded variable bit-rate coding algorithm for speech and audio operating in the range from 8 to 32 kbit/s. G.718 is designed to be highly robust to frame erasures, thereby enhancing the speech quality when used in IP transport applications on fixed, wireless and mobile networks. Despite its embedded nature, the codec also performs well with both NB and WB generic audio signals.

The G.718 bitstream may be truncated at the decoder side or by any component of the communication system to instantaneously adjust the bit rate to the desired value without the need for out-of-band signalling. The encoder produces an embedded bitstream structured in five layers corresponding to the five available bit rates: 8, 12, 16, 24 and 32 kbit/s.

The G.718 encoder can accept WB sampled signals at 16 kHz, or NB signals sampled at either 16 or 8 kHz. Similarly, the decoder output can be 16 kHz WB, in addition to 16 or 8 kHz NB. Input signals sampled at 16 kHz, but with bandwidth limited to NB, are detected by the encoder. The output of the G.718 codec is capable of operating with a bandwidth of 300-3400 Hz at 8 and 12 kbit/s and 50-7000 Hz from 8 to 32 kbit/s.

The codec operates on 20 ms frames and has a maximum algorithmic delay of 42.875 ms for wideband input and wideband output signals. The maximum algorithmic delay for narrowband input and narrowband output signals is 43.875 ms. The codec may also be employed in a low-delay mode when the encoder and decoder maximum bit rates are set to 12 kbit/s. In this case the maximum algorithmic delay is reduced by 10 ms.

The codec also incorporates an alternate coding mode, with a minimum bit rate of 12.65 kbit/s, which is bitstream interoperable with ITU-T Recommendation G.722.2, 3GPP AMR-WB and 3GPP2 VMR-WB mobile WB speech coding standards. This option replaces Layer 1 and Layer 2, and the layers 3-5 are similar to the default option with the exception that in Layer 3 fewer bits are used to compensate for the extra bits of the 12.65 kbit/s core. The decoder is further able to decode all other G.722.2 operating modes. G.718 also includes discontinuous transmission mode (DTX) and comfort noise generation (CNG) algorithms that enable bandwidth savings during inactive periods. An integrated noise reduction algorithm can be used provided that the communication session is limited to 12 kbit/s.

The underlying algorithm is based on a two-stage coding structure: the lower two layers are based on Code-Excited Linear Prediction (CELP) coding of the band (50-6400 Hz) where the core layer takes advantage of signal-classification to use optimized coding modes for each frame. The higher layers encode the weighted error signal from the lower layers using overlap-add MDCT transform coding. Several technologies are used to encode the MDCT coefficients to maximize performance for both speech and music.

7 Codec complexity and memory

Table 1 and Table 2 show the G.718 complexity in terms of weighted million operations per second (WMOPS) and its memory requirements, respectively.

Table 1: G.718 complexity in WMOPS

Option	Bitrate	Encoder	Decoder*	Total
Default, WB	8 kbit/s	31.2	11.6	42.8
	12 kbit/s	36.0	11.9	48.0
	16 kbit/s	39.4	13.3	52.8
	24 kbit/s	42.8	12.1	54.9
	32 kbit/s	43.4	12.5	55.9
	Overall	*43.4*	*13.3*	*56.7*
Default, NB	8 kbit/s	30.9	13.0	43.9
	12 kbit/s	37.0	13.4	50.4
	Overall	*37.0*	*13.4*	*50.4*
G.722.2-interoperable	12.65 kbit/s	30.0	12.1	42.1
	16 kbit/s	35.4	12.0	47.4
	24 kbit/s	36.9	10.1	47.0
	32 kbit/s	37.8	11.2	49.0
	Overall	*37.8*	*12.1*	*49.9*

* Note: Estimated with 5% frame erasure rate.

Table 2: G.718 memory consumption in kWords

Memory	Type	Encoder	Decoder	Common	Total
Table ROM	–	2.0	0.3	30.3	32.6
Program ROM*	–	–	–	–	18.1
RAM	Static	6.0	5.7	–	11.8
	Dynamic	8.4	5.8	–	14.3
	Overall	*14.4*	*11.6*	–	*26.1*

* Note: Evaluated as the number of basic operators used.

8 Codec algorithmic delay

The codec algorithmic delay depends on the sampling rate of the input and the output signal, and on the number of decoded layers. At higher layers (L3-L5), a 10 ms decoder delay is required for overlap-add operation of the transform coding. At lower layers (L1 and L2), the transform coding is not used and this delay can be either saved, or it can be used to improve the frame erasure concealment and the narrowband music quality at 8 or 12 kbit/s. The low-delay option can be used only if the decoder is prevented from switching to layer 3 and above during the call. The algorithmic delay for different configurations is summarized in Table 3.

Table 3: G.718 algorithmic delay

Input sampling rate	Output sampling rate	Algorithmic delay	
		Normal decoding	Low-delay decoding
16 kHz	16 kHz	42.875 ms	32.875 ms*
8 kHz	8 kHz	43.875 ms	33.875 ms
16 kHz	8 kHz	42.8125 ms	32.8125 ms
8 kHz	16 kHz	43.9375 ms	33.9375 ms*

* Note: Low-delay decoding is applicable only up to L2 decoding.

9 Characterization Tests, Step 1

Recommendation ITU-T G.718 was formally evaluated through extensive ITU-T characterization tests. The first step (Step 1) was completed in April 2008. The subjective quality tests in this step evaluated the codec for narrowband and wideband speech and music signals for different input levels, background noises, channel impairment characteristics, and in tandem with other speech and audio coding standards. Overall, 9 listening laboratories participated in Step 1 of the Characterization Tests and each condition was evaluated by two laboratories using different languages. The latest versions of the G.718 Characterization Step 1 Processing Plan and Listening Test Plan can be found in [1] and [2].

During Step 1, the codec was evaluated for 80 reference conditions (the Terms of Reference (ToR) can be found in Annex A of the June/July 2007 meeting Report of Question 9/16 [3]). For 78 conditions, the codec met the requirements in both testing laboratories and for 2 conditions the codec met the requirements in at least one laboratory. No requirement was failed in both testing laboratories. The test showed that the most significant progress with respect to state-of-the-art references had been made in the low bit rate wideband speech conditions in the presence of channel errors. While not primarily designed for narrowband inputs, very good quality has also been achieved for narrowband clean speech inputs where R1 at 8 kbit/s performed not worse than G.729 Annex E at 11.8 kbit/s. Finally, the codec performed very well in noisy conditions for both narrowband and wideband inputs. The Step 1 test results are summarized in the following subsections.

9.1 Organization of the Characterization Tests, Step 1

Table 4 shows a summary of the parameters tested in each experiment of the Characterization Step 1 testing phase. In Table 4, WB means wideband input and wideband output, and NB means narrowband input and narrowband output. The “Background noise” column specifies the type and the level of background noise signal applied to the input (measured as SNR with respect to the level of the input signal). The test method is either the Absolute Category Rating (ACR) or Degradation Category Rating (DCR). The “Errors” column shows whether frame erasures were tested and if so, the percentage of Frame Erasure Rate (FER). The “Rates” column specifies the tested rates of the G.718 codec, i.e. R1-R5 corresponding to the bitrates of 8, 12, 16, 24 and 32kbit/s. The following acronyms are used in Table 4, and throughout the following sections: LD stands for Low Delay decoding mode, DTX for discontinuous transmission and INT corresponds to conditions where first two G.718 default layers are replaced with the mode interoperable with G.722.2 at 12.65kbit/s.

Table 5 indicates the testing laboratories in the Characterization step 1 testing phase and the languages testing in each of them.

Table 4: Organization of Characterization Tests, Step 1

Exp	Input	NB/ WB	Background noise	Test method	Errors	Bit rates	Remarks
1a	Speech	NB	-	ACR	3% FER	R1,R2	Input level: -16 dBov and -36 dBov WB input → NB output LD, DTX
1b	Music	NB	-	ACR	-	R1, R2	
2a	Speech	WB	-	ACR	3-8% FER	R1, R2, R3	Input levels: -16 dBov and -36 dBov LD, DTX, INT, switching
2b	Speech	WB	-	ACR	3-5% FER 0-12% FER	R4, R5	Input level: -16 dBov and -36 dBov INT, 0-12% correlated FER on R1-5
2c	Music	WB	-	ACR	-	R3, R4, R5	INT
3a	Speech	NB	Car @ 15dB SNR	DCR	-	R1, R2	DTX
3b	Speech	NB	Interfering talker @ 15dB SNR Music @ 25dB SNR	DCR	-	R1, R2	-
3c	Speech	NB	Babble @ 25dB SNR Office @ 20dB SNR	DCR	-	R1, R2	DTX
4a	Speech	WB	Interfering talker @ 15dB SNR	DCR	-	R1, R2, R3, R4, R5	INT
4b	Speech	WB	Music @ 25dB SNR	DCR	-	R1, R2, R3, R4, R5	INT
4c	Speech	WB	Car @ 15dB SNR	DCR	-	R1, R2, R3, R4, R5	INT
4d	Speech	WB	Office @ 20dB SNR	DCR	-	R1, R2, R3, R4, R5	DTX, INT
4e	Speech	WB	Babble @ 25dB SNR Street @ 20dB SNR	DCR	-	R1, R2, R3, R4, R5	-

Table 5: Testing laboratories and associated languages in Characterization Tests, Step 1

Exp	Lab A	Language – Lab A	Lab B	Language – Lab B
1a	Huawei	Chinese	VoiceAge	Canadian French
1b	Nokia	Finnish	France Telecom	French
2a	Dynastat (Motorola)	American English	Huawei	Chinese
2b	Huawei	Chinese	VoiceAge	Canadian French
2c	BIT (Ericsson)	Chinese	Dynastat (Qualcomm)	American English
3a	Matsushita	Japanese	Dynastat (Motorola)	American English
3b	Dynastat (Qualcomm)	American English	Ericsson	Swedish
3c	Nokia	Finnish	Dynastat (Texas Instruments)	American English
4a	Dynastat (Motorola)	American English	VoiceAge	Canadian French
4b	Matsushita	Japanese	Dynastat (Qualcomm)	American English
4c	BIT (Ericsson)	Chinese	Dynastat (Texas Instruments)	American English
4d	Nokia	Finnish	France Telecom	French
4e	Matsushita	Japanese	France Telecom	French

9.2 Test results

The summary of the test results of Characterization Tests, Step 1 can be found in [4]. Individual listening laboratory reports can be found in [5], [6], [7], [8], [9], [10], [11], [12], [13] and [14]. Note that all "better than" (BT) criteria are systematically supplemented with "or not worse than (NWT) direct" criteria, to account for test saturation. The NWT direct criterion is not made explicit in the tables on “Verification against terms of reference”, but is always assumed to be part of the BT criteria. In all tests, if not mentioned explicitly, the default level of input signal is -26 dBov. The number of votes per condition is 192. The test results are divided into three categories: requirements, objectives and informative. In the figures, different test cases are logically grouped and colours are used to identify the test conditions and the reference conditions. Please note, that a particular reference condition may be used for several test conditions. The colour code in the “Verification against terms of reference” tables visually groups test conditions addressing requirements, objectives and informational test items; and pass / fail in bold typeface indicate the requirements.

9.2.1 Experiment 1a: Narrowband clean speech

Experiment 1a has been run twice, once in Chinese (Lab A) and once in Canadian French (Lab B). The purpose of this experiment was to evaluate the quality for narrowband speech, both in clean channel and frame erasure conditions. Different input levels have been used. Performance in DTX operation and low delay mode was also evaluated. Finally, the experiment also assessed the codec performance in case of wideband input, narrowband output scenario. The test method used was the ACR method. If not explicitly mentioned, the input and output signals in all test conditions of this experiment were narrowband.

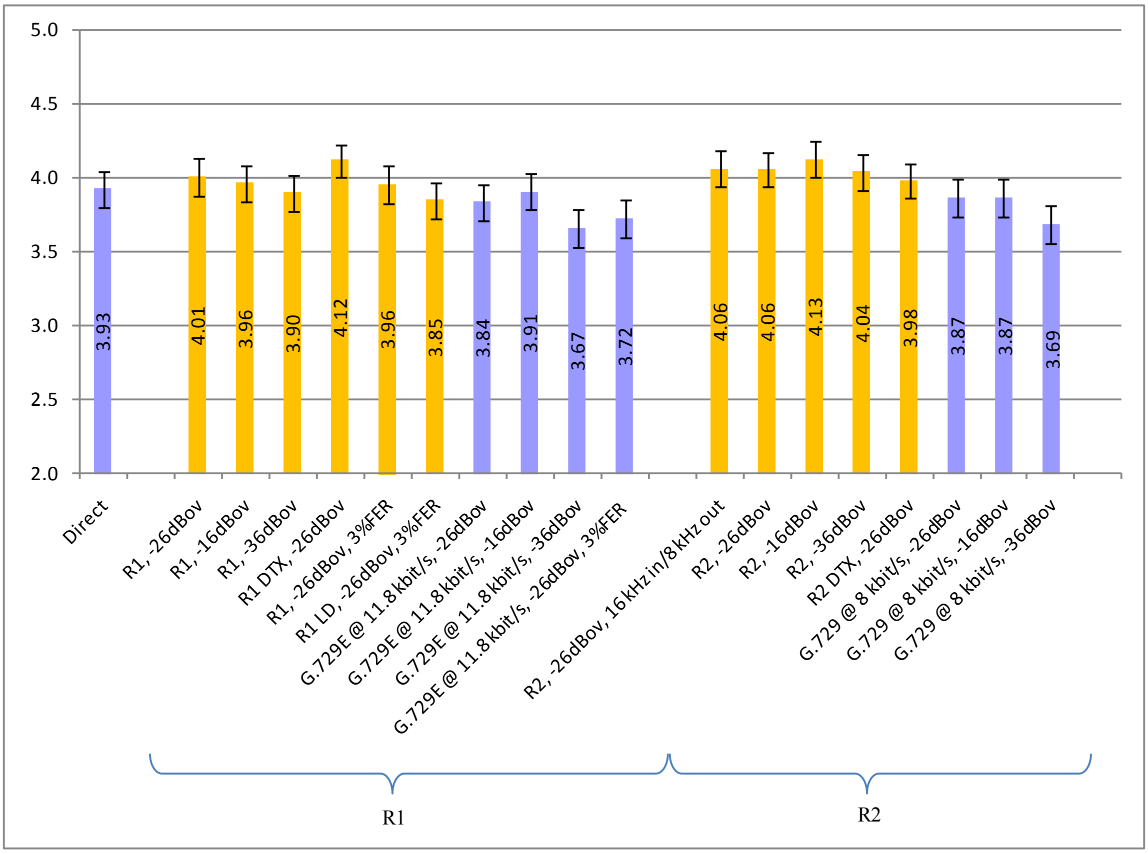

Figure 1: Experiment 1a results for Lab A, Chinese

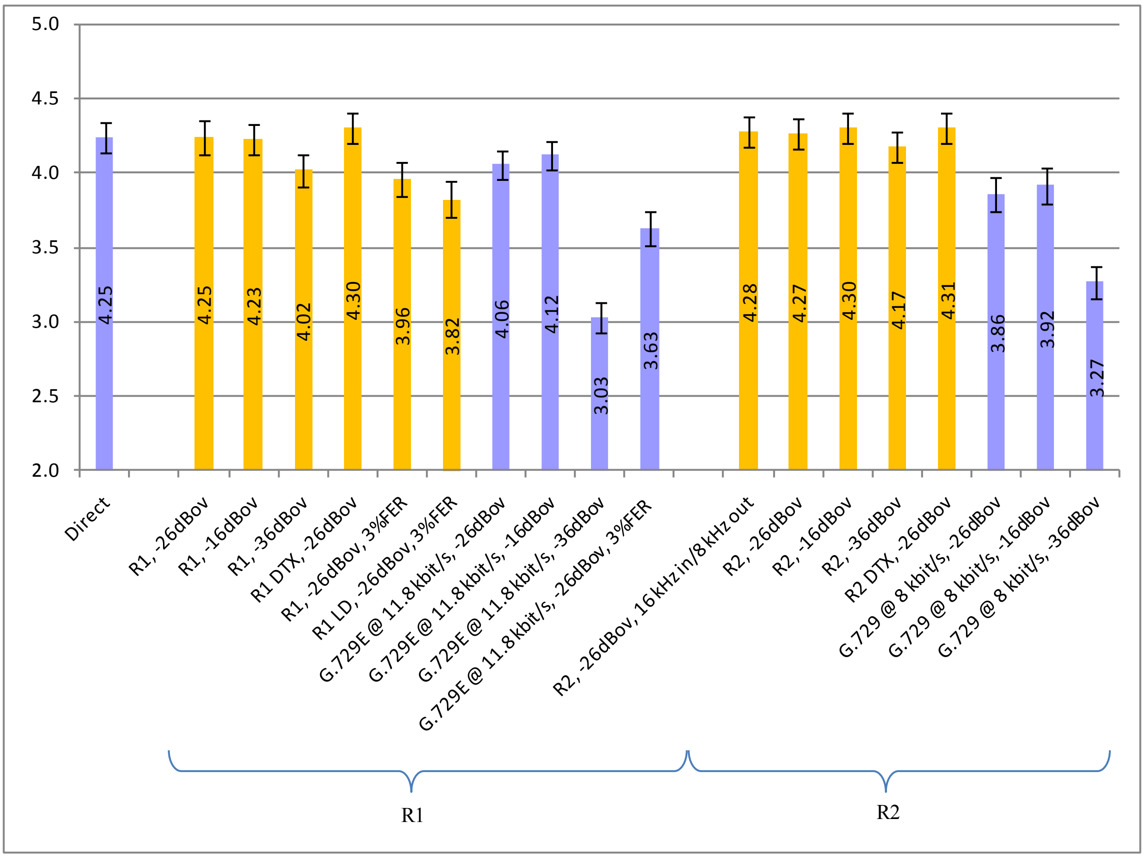

Figure 2: Experiment 1a results for Lab B, Canadian French

Table 6: Experiment 1a results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	3.93	0.87	4.25	0.71
R1, -26 dBov	4.01	0.89	4.25	0.82
R1, -16 dBov	3.96	0.90	4.23	0.72
R1, -36 dBov	3.90	0.86	4.02	0.77
R1 DTX, -26 dBov	4.12	0.78	4.30	0.70
R1, -26 dBov, 3%FER	3.96	0.92	3.96	0.84
R1 LD, -26 dBov, 3%FER	3.85	0.89	3.82	0.86
R2, -26 dBov, 16kHz in/8kHz out	4.06	0.85	4.28	0.70
R2, -26 dBov	4.06	0.84	4.27	0.72
R2, -16 dBov	4.13	0.85	4.30	0.70
R2, -36 dBov	4.04	0.84	4.17	0.73
R2 DTX, -26 dBov	3.98	0.81	4.31	0.71
G.729 @ 8 kbit/s, -26 dBov	3.84	0.90	4.06	0.77
G.729E @ 11.8 kbit/s, -26 dBov	3.91	0.85	4.12	0.68
G.729E @ 11.8 kbit/s, -16 dBov	3.67	0.85	3.03	0.68
G.729E @ 11.8 kbit/s, -36 dBov	3.72	0.92	3.63	0.74
G.729E @ 11.8 kbit/s, -26 dBov, 3%FER	3.87	0.93	3.92	0.83
G.729 @ 8 kbit/s, -16 dBov	3.69	0.91	3.27	0.85
G.729 @ 8 kbit/s, -36 dBov	3.93	0.93	4.25	0.74

Table 7: Verification against terms of reference - Experiment 1a

	Test Condition	Reference Condition	Criterion	Result LabA	Result LabB
Requirements	R1, -26 dBov	G.729E @ 11.8 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1, -16 dBov	G.729E @ 11.8 kbit/s, -16 dBov	NWT	**PASS**	**PASS**
	R1, -36 dBov	G.729E @ 11.8 kbit/s, -36 dBov	NWT	**PASS**	**PASS**
	R1 DTX, -26 dBov	G.729E @ 11.8 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1, -26 dBov, 3%FER	G.729E @ 11.8 kbit/s, -26 dBov 3%FER	NWT	**PASS**	**PASS**
	R1 LD, -26 dBov, 3%FER	G.729E @ 11.8 kbit/s, -26 dBov 3%FER	NWT	**PASS**	**PASS**
	R2, -26 dBov, 16kHz in/8kHz out	G.729 @ 8 kbit/s, -26 dBov	BT	**PASS**	**PASS**
	R2, -26 dBov	G.729 @ 8 kbit/s, -26 dBov	BT	**PASS**	**PASS**
	R2, -16 dBov	G.729 @ 8 kbit/s, -16 dBov	BT	**PASS**	**PASS**
	R2, -36 dBov	G.729 @ 8 kbit/s, -36 dBov	BT	**PASS**	**PASS**
	R2 DTX, -26 dBov	G.729 @ 8 kbit/s, -26 dBov	BT	**PASS**	**PASS**

9.2.2 Experiment 1b: Narrowband music

Experiment 1b has been run in two laboratories. The purpose of this experiment was to evaluate the quality for narrowband music. The test method used was ACR.

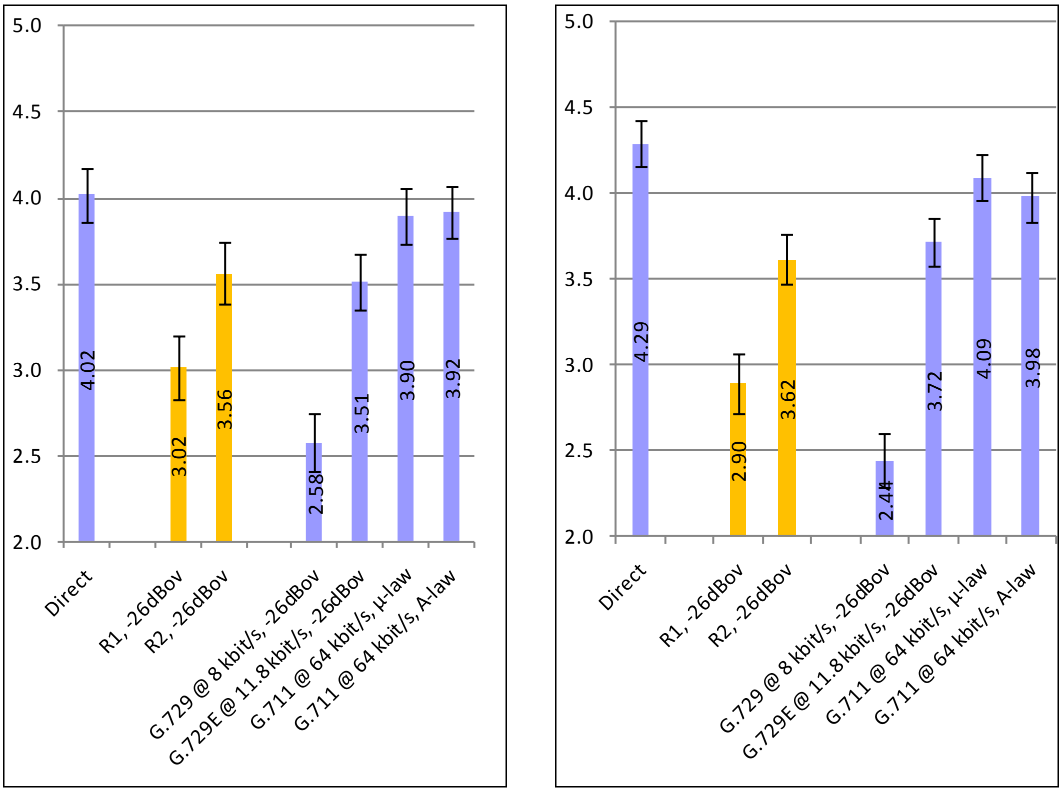

Figure 3: Experiment 1b results for Lab A and Lab B

Table 8: Experiment 1b results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.02	0.93	4.29	0.77
R1, -26 dBov	3.02	1.07	2.90	1.00
R2, -26 dBov	3.56	1.03	3.62	0.83
G.729 @ 8 kbit/s, -26 dBov	2.58	0.96	2.44	0.92
G.729E @ 11.8 kbit/s, -26 dBov	3.51	0.93	3.72	0.81
G.711 @ 64 kbit/s, µ-law	3.90	0.92	4.09	0.79
G.711 @ 64 kbit/s, A-law	3.92	0.88	3.98	0.82

Table 9: Verification against terms of reference - Experiment 1b

	Test Condition	Reference Condition	Criterion	Result LabA	Result LabB
Req.	R1, -26 dBov	G.729 @ 8 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R2, -26 dBov	G.729E @ 11.8 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
Obj.	R1, -26 dBov	G.729E @ 11.8 kbit/s, -26 dBov	NWT	FAIL	FAIL
	R2, -26 dBov	G.711 @ 64 kbit/s, µ-law	NWT	FAIL	FAIL
	R2, -26 dBov	G.711 @ 64 kbit/s, A-law	NWT	FAIL	FAIL
Inf.	R1, -26 dBov	G.729 @ 8 kbit/s, -26 dBov	BT	-	PASS
	R2, -26 dBov	G.729 @ 8 kbit/s, -26 dBov	BT	-	PASS

9.2.3 Experiment 2a: Wideband clean speech (lower rates)

Experiment 2a has been run twice, once in American English (Lab A) and once in Chinese (Lab B). The purpose of this experiment was to evaluate the performance of the codec for wideband clean speech (free of background noise) at lower layers (R1 at 8 kbit/s, R2 at 12 kbit/s, and R3 at 16 kbit/s) for different input levels. The performance of the interoperable modes R2 INT, R3 INT and R2-R5 INT was also evaluated as well as low delay mode and the DTX operation. The codec was evaluated with different percentage of random frame erasures. Also, this experiment evaluated the performance under slow 1Hz and fast 5Hz switching between different rates. The test method used was ACR.

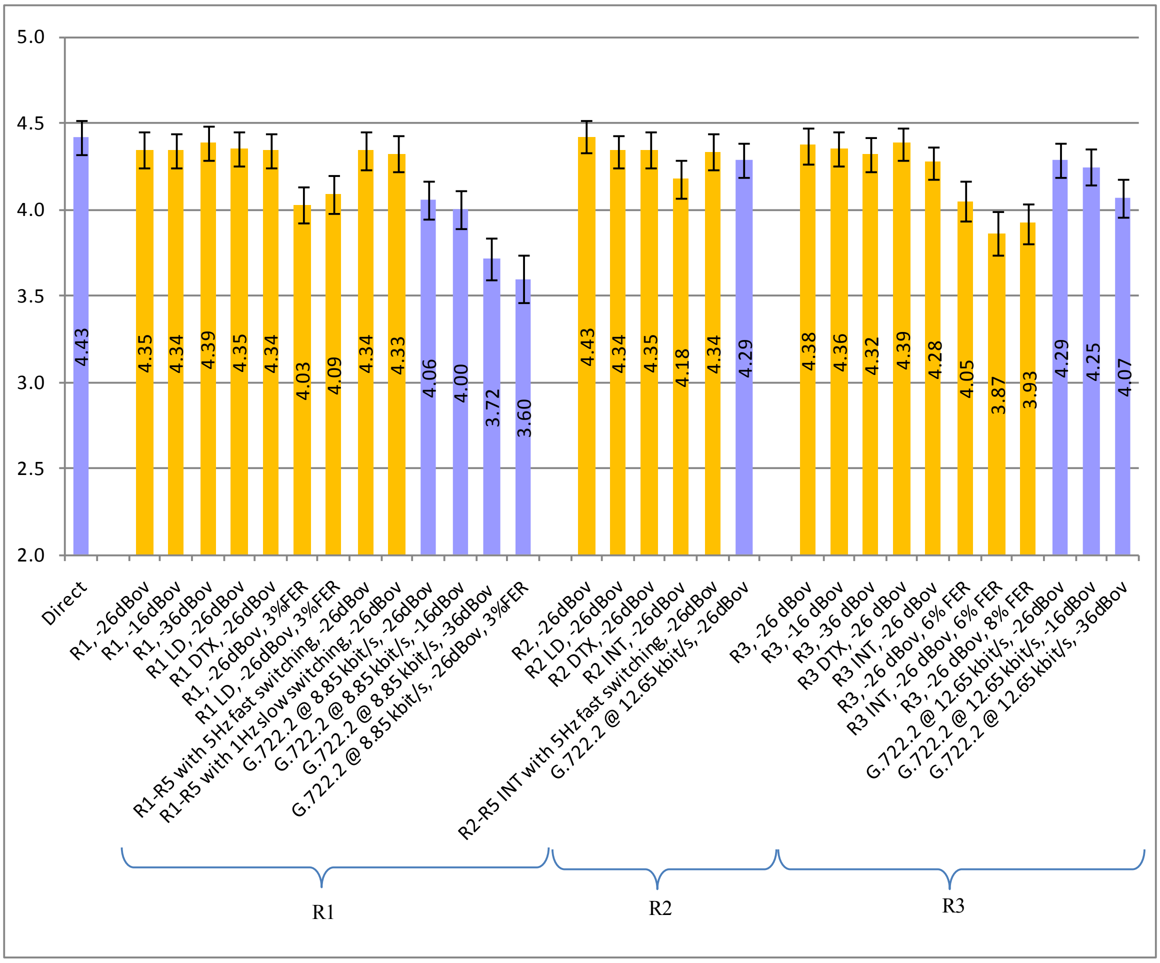

Figure 4: Experiment 2a results for Lab A, American English

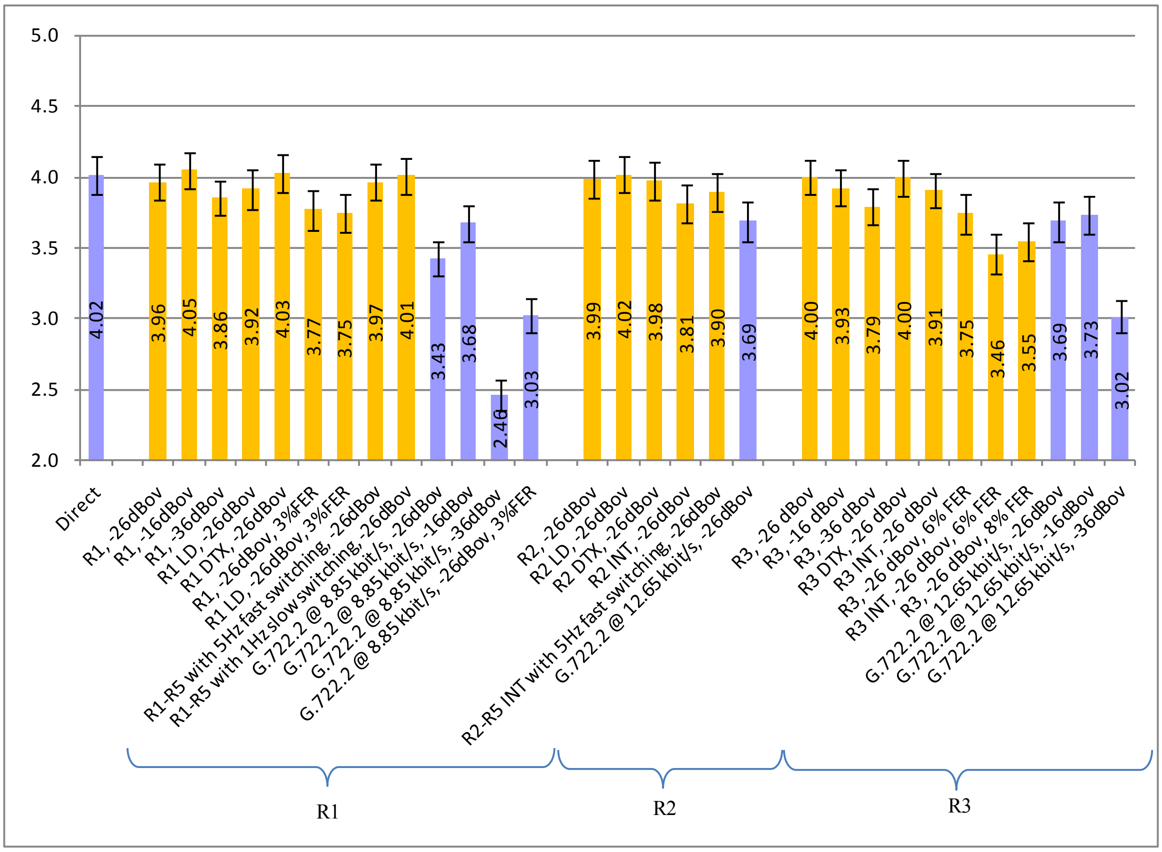

Figure 5: Experiment 2a results for Lab B, Chinese

Table 10: Experiment 2a results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.43	0.71	4.02	0.92
R1, -26 dBov	4.35	0.72	3.96	0.89
R1, -16 dBov	4.34	0.71	4.05	0.91
R1, -36 dBov	4.39	0.69	3.86	0.87
R1 LD, -26 dBov	4.35	0.72	3.92	0.98
R1 DTX, -26 dBov	4.34	0.70	4.03	0.95
R1, -26 dBov, 3%FER	4.03	0.77	3.77	0.98
R1 LD, -26 dBov, 3%FER	4.09	0.79	3.75	0.93
R2, -26 dBov	4.43	0.67	3.99	0.93
R2 LD, -26 dBov	4.34	0.65	4.02	0.90
R2 DTX, -26 dBov	4.35	0.74	3.98	0.94
R2 INT, -26 dBov	4.18	0.76	3.81	0.97
R1-R5 with 5Hz fast switching, -26 dBov	4.34	0.75	3.97	0.91
R1-R5 with 1Hz slow switching, -26 dBov	4.33	0.74	4.01	0.87
R2-R5 INT with 5Hz fast switching, -26 dBov	4.34	0.72	3.90	0.95
R3, -26 dBov	4.38	0.72	4.00	0.87
R3, -16 dBov	4.36	0.71	3.93	0.89
R3, -36 dBov	4.32	0.69	3.79	0.92
R3 DTX, -26 dBov	4.39	0.68	4.00	0.90
R3 INT, -26 dBov	4.28	0.68	3.91	0.83
R3, -26 dBov, 6% FER	4.05	0.83	3.75	0.99
R3 INT, -26 dBov, 6% FER	3.87	0.91	3.46	0.98
R3, -26 dBov, 8% FER	3.93	0.82	3.55	0.93
G.722.2 @ 8.85 kbit/s, -26 dBov	4.06	0.76	3.43	0.83
G.722.2 @ 8.85 kbit/s, -16 dBov	4.00	0.78	3.68	0.89
G.722.2 @ 8.85 kbit/s, -36 dBov	3.72	0.88	2.46	0.77
G.722.2 @ 8.85 kbit/s, -26 dBov, 3%FER	3.60	0.97	3.03	0.85
G.722.2 @ 12.65 kbit/s, -26 dBov	4.29	0.69	3.69	0.98
G.722.2 @ 12.65 kbit/s, -16 dBov	4.25	0.74	3.73	0.91
G.722.2 @ 12.65 kbit/s, -36 dBov	4.07	0.77	3.02	0.81

Table 11: Verification against terms of reference - Experiment 2a

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Requirements	R1, -26 dBov	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1, -16 dBov	G.722.2 @ 8.85 kbit/s, -16 dBov	NWT	**PASS**	**PASS**
	R1, -36 dBov	G.722.2 @ 8.85 kbit/s, -36 dBov	NWT	**PASS**	**PASS**
	R1 LD, -26 dBov	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1 DTX, -26 dBov	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1, -26 dBov, 3%FER	G.722.2 @ 8.85 kbit/s, -26 dBov, 3%FER	NWT	**PASS**	**PASS**
	R1 LD, -26 dBov, 3%FER	G.722.2 @ 8.85 kbit/s, -26 dBov, 3%FER	NWT	**PASS**	**PASS**
	R1-R5 with 5Hz fast switching, -26 dBov	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1-R5 with 1Hz slow switching, -26 dBov	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R3, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R3, -16 dBov	G.722.2 @ 12.65 kbit/s, -16 dBov	NWT	**PASS**	**PASS**
	R3, -36 dBov	G.722.2 @ 12.65 kbit/s, -36 dBov	NWT	**PASS**	**PASS**
	R3 DTX, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R3, -26 dBov, 6% FER	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
Objectives	R2, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R2 LD, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R2 DTX, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R2 INT, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	FAIL	PASS
	R2-R5 INT with 5Hz fast switching, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R3, -26 dBov, 8% FER	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	FAIL	PASS
	R3 INT, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R3 INT, -26 dBov, 6% FER	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	FAIL	PASS

9.2.4 Experiment 2b: Wideband clean speech (higher rates)

Experiment 2b has been run twice, once in Chinese (Lab A) and once in Canadian French (Lab B). The purpose of this experiment was to evaluate the performances of the codec for wideband clean speech (free of background noise) at higher rates (R4 at 24 kbit/s, R5 at 32 kbit/s) for different input levels. The performance of the interoperable modes R4 INT and R5 INT was also evaluated. The codec was evaluated with different percentage of random frame erasures, and also for conditions where higher layers have higher probability of frame erasures than lower layers. The test method used was the ACR.

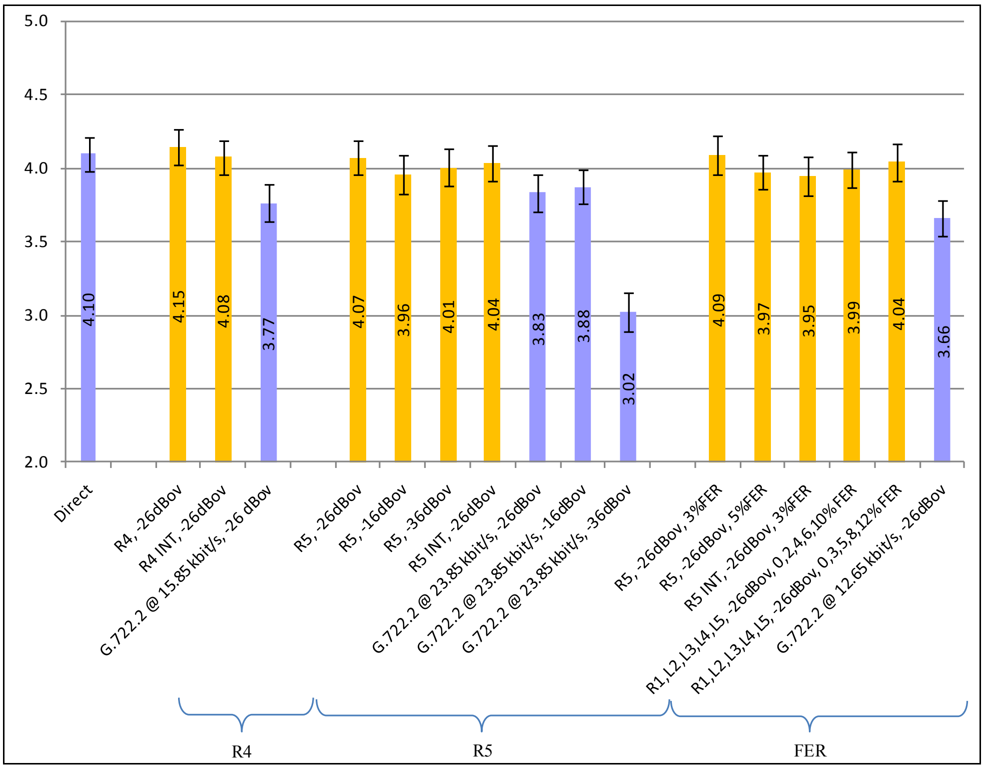

Figure 6: Experiment 2b results for Lab A, Chinese

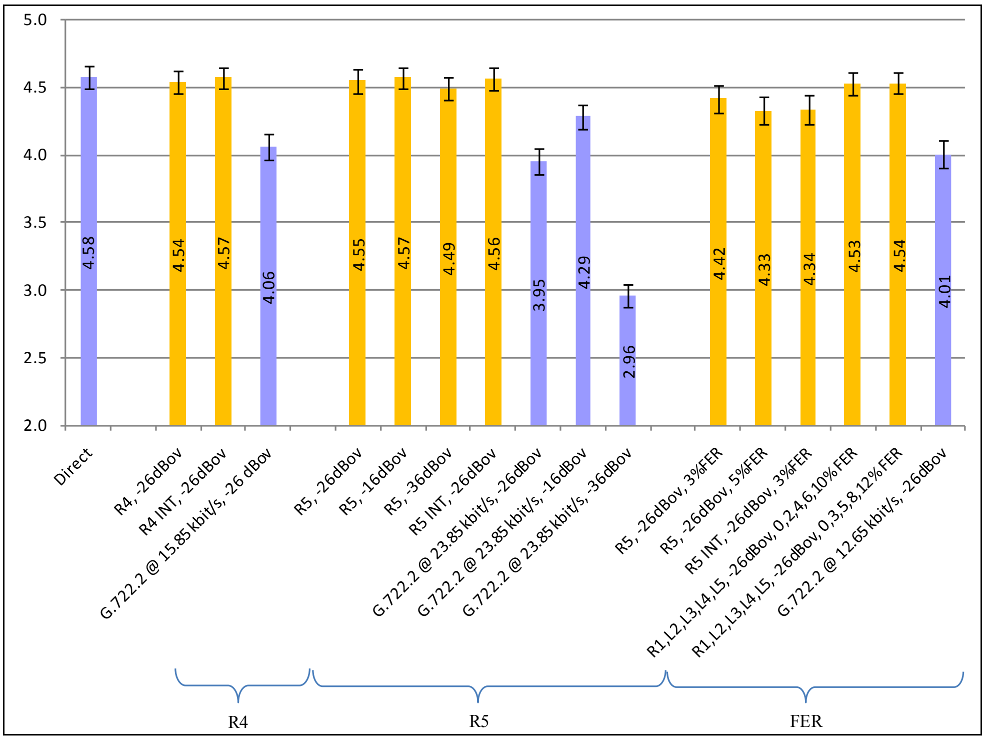

Figure 7: Experiment 2b results for Lab B, Canadian French

Table 12: Experiment 2b results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.10	0.83	4.58	0.60
R5, -26 dBov	4.07	0.82	4.55	0.63
R5, -16 dBov	3.96	0.92	4.57	0.57
R5, -36 dBov	4.01	0.88	4.49	0.60
R5, -26 dBov, 3%FER	4.09	0.92	4.42	0.74
R5, -26 dBov, 5%FER	3.97	0.82	4.33	0.73
R1,L2,L3,L4,L5, -26 dBov, 0, 2, 4, 6, 10% FER	3.99	0.87	4.53	0.62
R1,L2,L3,L4,L5, -26 dBov, 0, 3, 5, 8, 12% FER	4.04	0.89	4.54	0.59
R5 INT, -26 dBov	4.04	0.86	4.56	0.58
R5 INT, -26 dBov, 3%FER	3.95	0.91	4.34	0.76
R4, -26 dBov	4.15	0.86	4.54	0.60
R4 INT, -26 dBov	4.08	0.83	4.57	0.56
G.722.2 @ 23.85 kbit/s, -26 dBov	3.83	0.87	3.95	0.70
G.722.2 @ 23.85 kbit/s, -16 dBov	3.88	0.83	4.29	0.67
G.722.2 @ 23.85 kbit/s, -36 dBov	3.02	0.93	2.96	0.58
G.722.2 @ 12.65 kbit/s, -26 dBov	3.66	0.85	4.01	0.71
G.722.2 @ 15.85 kbit/s, -26 dBov	3.77	0.92	4.06	0.67

Table 13: Verification against terms of reference - Experiment 2b

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Requirements	R5, -26 dBov	G.722.2 @ 23.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R5, -16 dBov	G.722.2 @ 23.85 kbit/s, -16 dBov	NWT	**PASS**	**PASS**
	R5, -36 dBov	G.722.2 @ 23.85 kbit/s, -36 dBov	NWT	**PASS**	**PASS**
	R5, -26 dBov, 3%FER	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1,L2,L3,L4,L5, -26 dBov, 0, 2, 4, 6, 10% FER	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R4, -26 dBov	G.722.2 @ 15.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
Objectives	R5, -26 dBov, 5% FER	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R1,L2,L3,L4,L5, -26 dBov, 0, 3, 5, 8, 12% FER	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R5 INT, -26 dBov	G.722.2 @ 23.85 kbit/s, -26 dBov	NWT	PASS	PASS
	R5 INT, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R4 INT, -26 dBov	G.722.2 @ 15.85 kbit/s, -26 dBov	NWT	PASS	PASS

9.2.5 Experiment 2c: Wideband music

Experiment 2c has been run in two laboratories. The purpose of this experiment was to evaluate the performance of the codec for wideband music at higher rates (R3 at 16 kbit/s, R4 at 24 kbit/s, R5 at 32 kbit/s). The performance of the interoperable modes R4 INT and R5 INT was also evaluated. The test method used was ACR.

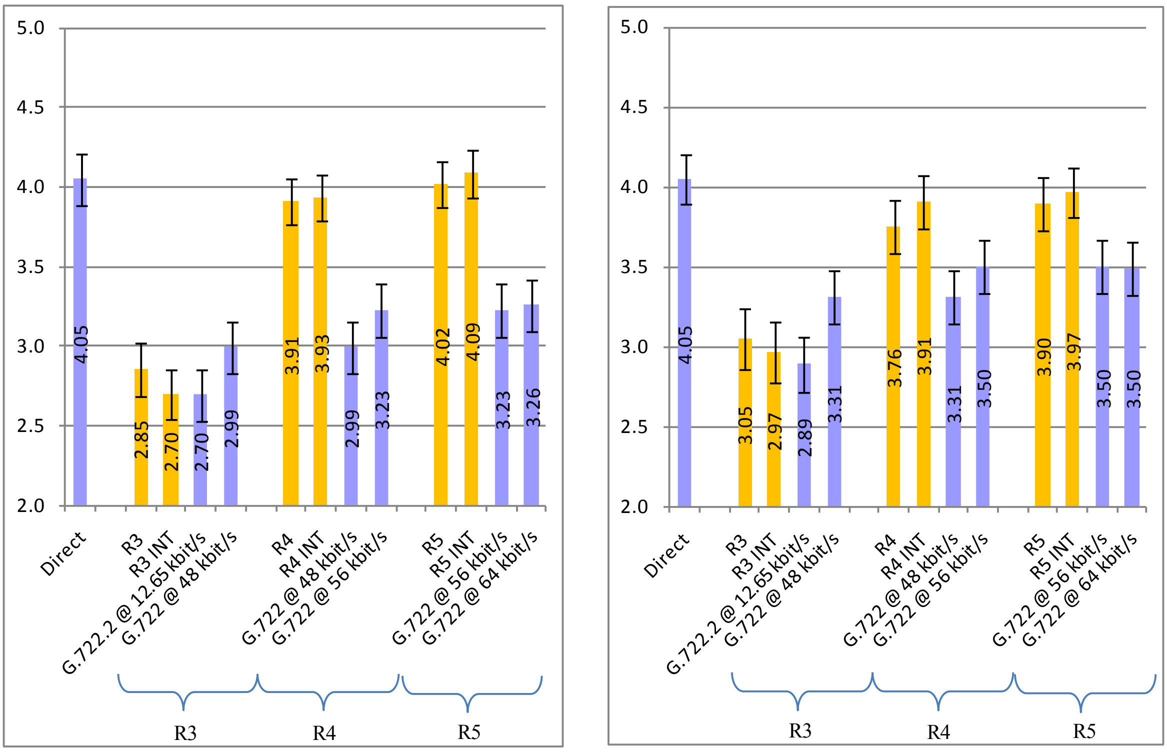

Figure 8: Experiment 2c results for Lab A and Lab B

Table 14: Experiment 2c results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.05	0.93	4.05	0.90
R3, -26 dBov	2.85	0.97	3.05	1.11
R3 INT, -26 dBov	2.70	0.92	2.97	1.08
R4 INT, -26 dBov	3.93	0.83	3.91	0.96
R4, -26 dBov	3.91	0.83	3.76	0.96
R5 INT, -26 dBov	4.09	0.84	3.97	0.90
R5, -26 dBov	4.02	0.85	3.90	0.95
G.722.2 @ 12.65 kbit/s, -26 dBov	2.70	0.93	2.89	1.01
G.722 @ 48 kbit/s, -26 dBov	2.99	0.93	3.31	0.96
G.722 @ 56 kbit/s, -26 dBov	3.23	0.99	3.50	0.99
G.722 @ 64 kbit/s, -26 dBov	3.26	0.92	3.50	0.97

Table 15: Verification against terms of reference - Experiment 2c

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R3, -26 dBov	G.722.2 @ 12.65kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R4, -26 dBov	G.722 @ 48 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R5, -26 dBov	G.722 @ 56 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
Objectives	R3, -26 dBov	G.722 @ 48 kbit/s, -26 dBov	NWT	FAIL	FAIL
	R4, -26 dBov	G.722 @ 56 kbit/s, -26 dBov	NWT	PASS	PASS
	R5, -26 dBov	G.722 @ 64 kbit/s, -26 dBov	NWT	PASS	PASS
	R3 INT, -26 dBov	G.722.2 @ 12.65kbit/s, -26 dBov	NWT	PASS	PASS
	R4 INT, -26 dBov	G.722 @ 48 kbit/s, -26 dBov	NWT	PASS	PASS
	R5 INT, -26 dBov	G.722 @ 56 kbit/s, -26 dBov	NWT	PASS	PASS

9.2.6 Experiment 3a: Narrowband noisy speech (car, street)

Experiment 3a has been run twice, once in Japanese (Lab A) and once in American English (Lab B). The purpose of this experiment was to evaluate the quality of the codec for narrowband noisy speech in clean channel conditions. Two different background noise types were tested: car noise at 15 dB SNR, and street noise at 20 dB SNR. The codec was evaluated at two different rates: R1 at 8 kbit/s and R2 at 12 kbit/s. Performance of the DTX operation was also evaluated. The test method used was DCR.

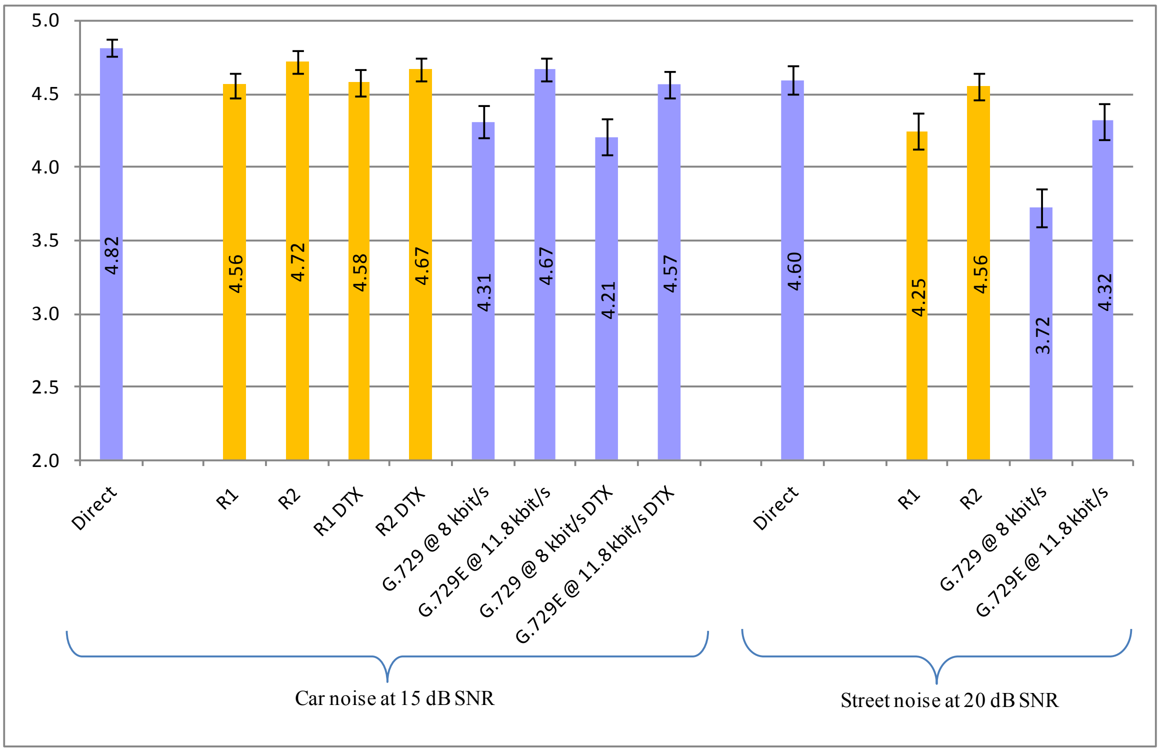

Figure 9: Experiment 3a results for Lab A, Japanese

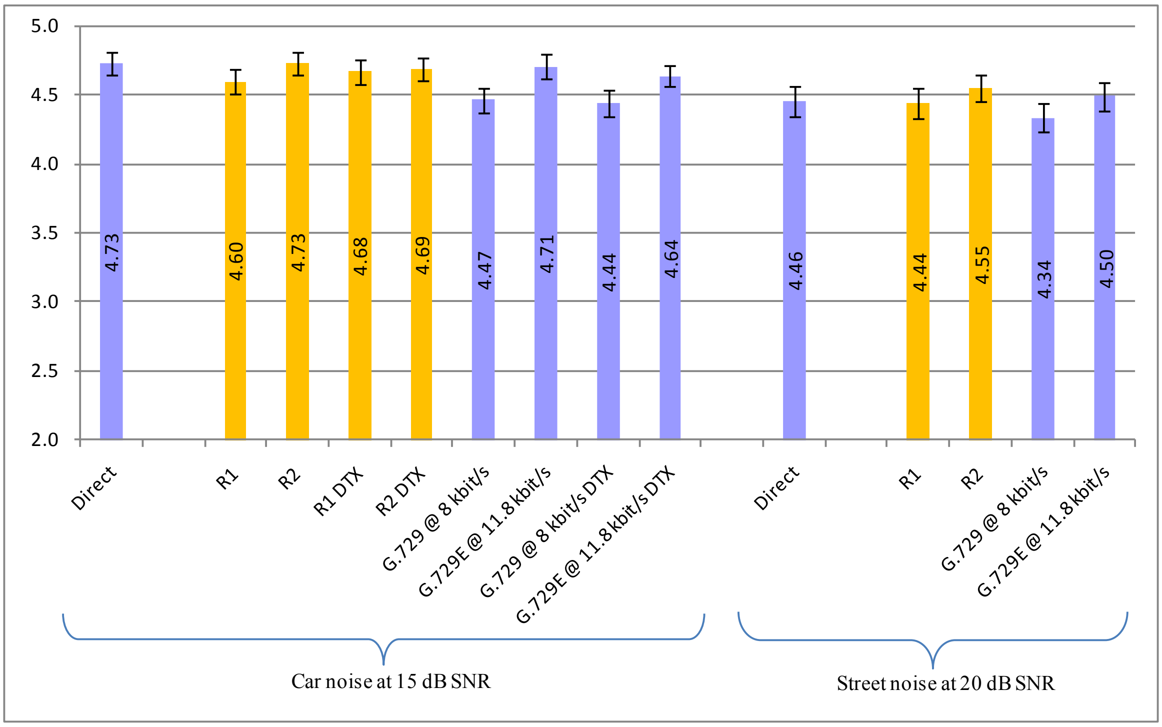

Figure 10: Experiment 3a results for Lab B, American English

Table 16: Experiment 3a results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct, 15 dB car noise	4.82	0.44	4.73	0.59
R1, 15 dB car noise	4.56	0.62	4.60	0.64
R2, 15 dB car noise	4.72	0.57	4.73	0.56
R1 DTX, 15 dB car noise	4.58	0.67	4.68	0.64
R2 DTX, 15 dB car noise	4.67	0.56	4.69	0.59
G.729 @ 8 kbit/s, 15 dB car noise	4.31	0.76	4.47	0.65
G.729E @ 11.8 kbit/s, 15 dB car noise	4.67	0.56	4.71	0.62
G.729 @ 8 kbit/s DTX, 15 dB car noise	4.21	0.87	4.44	0.71
G.729E @ 11.8 kbit/s DTX, 15 dB car noise	4.57	0.62	4.64	0.56
Direct, 20 dB street noise	4.60	0.68	4.46	0.78
R1, 20 dB street noise	4.25	0.87	4.44	0.74
R2, 20 dB street noise	4.56	0.66	4.55	0.67
G.729 @ 8 kbit/s, 20 dB street noise	3.72	0.92	4.34	0.73
G.729E @ 11.8 kbit/s, 20 dB street noise	4.32	0.87	4.50	0.75

Table 17: Verification against terms of reference - Experiment 3a

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Requirements	R1, 15dB car noise	G.729 @ 8 kbit/s, 15dB car noise	NWT	**PASS**	**PASS**
	R2, 15dB car noise	G.729 @ 8 kbit/s DTX, 15dB car noise	NWT	**PASS**	**PASS**
	R1 DTX, 15dB car noise	G.729E @ 11.8 kbit/s DTX, 15dB car noise	NWT	**PASS**	**PASS**
	R2 DTX, 15dB car noise	G.729E @ 11.8 kbit/s DTX, 15dB car noise	NWT	**PASS**	**PASS**
	R1, 20dB street noise	G.729 @ 8 kbit/s, 20dB street noise	NWT	**PASS**	**PASS**
	R2, 20dB street noise	G.729E @ 11.8 kbit/s, 20dB street noise	NWT	**PASS**	**PASS**

9.2.7 Experiment 3b: Narrowband noisy speech (background music, int. talker)

Experiment 3b has been run twice, once in American English (Lab A) and once in Swedish (Lab B). The purpose of this experiment was to evaluate the quality for narrowband noisy speech in clean channel conditions. Two different background noise types were tested: background music at 25 dB SNR, and interfering talker at 15 dB SNR. The codec was evaluated at two different rates: R1 at 8 kbit/s and R2 at 12 kbit/s. The test method used was DCR.

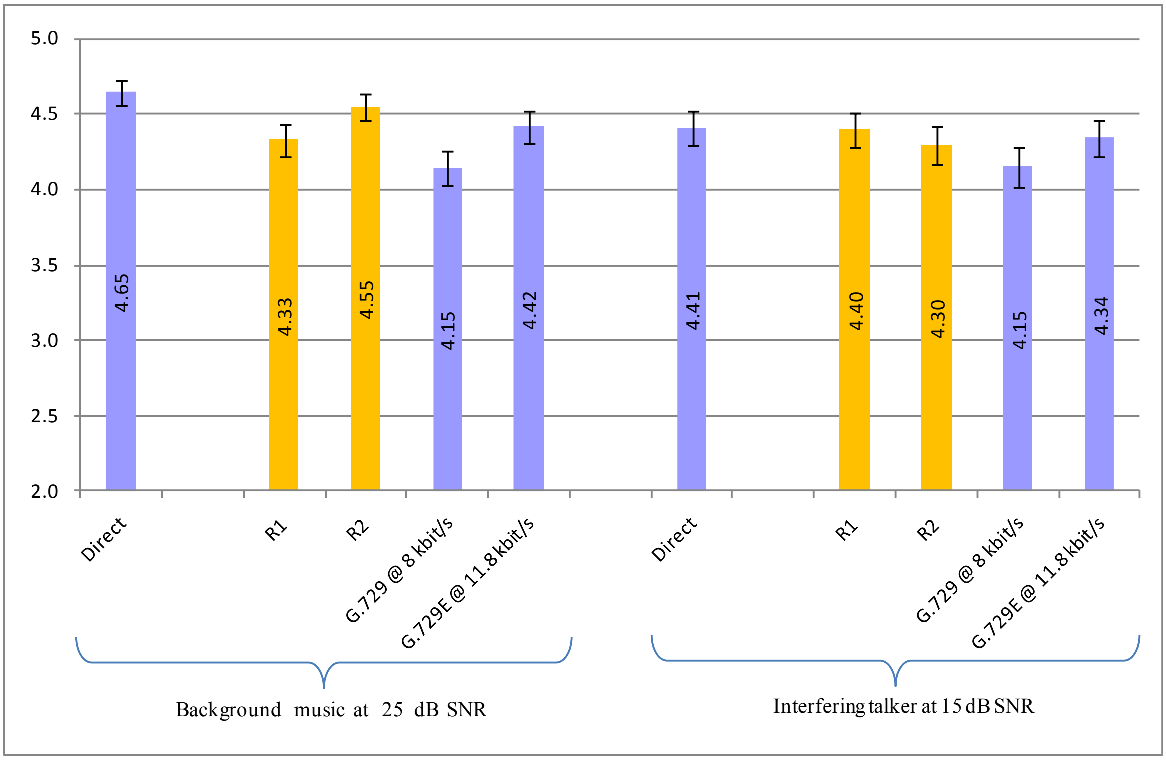

Figure 11: Experiment 3b results for Lab A, American English

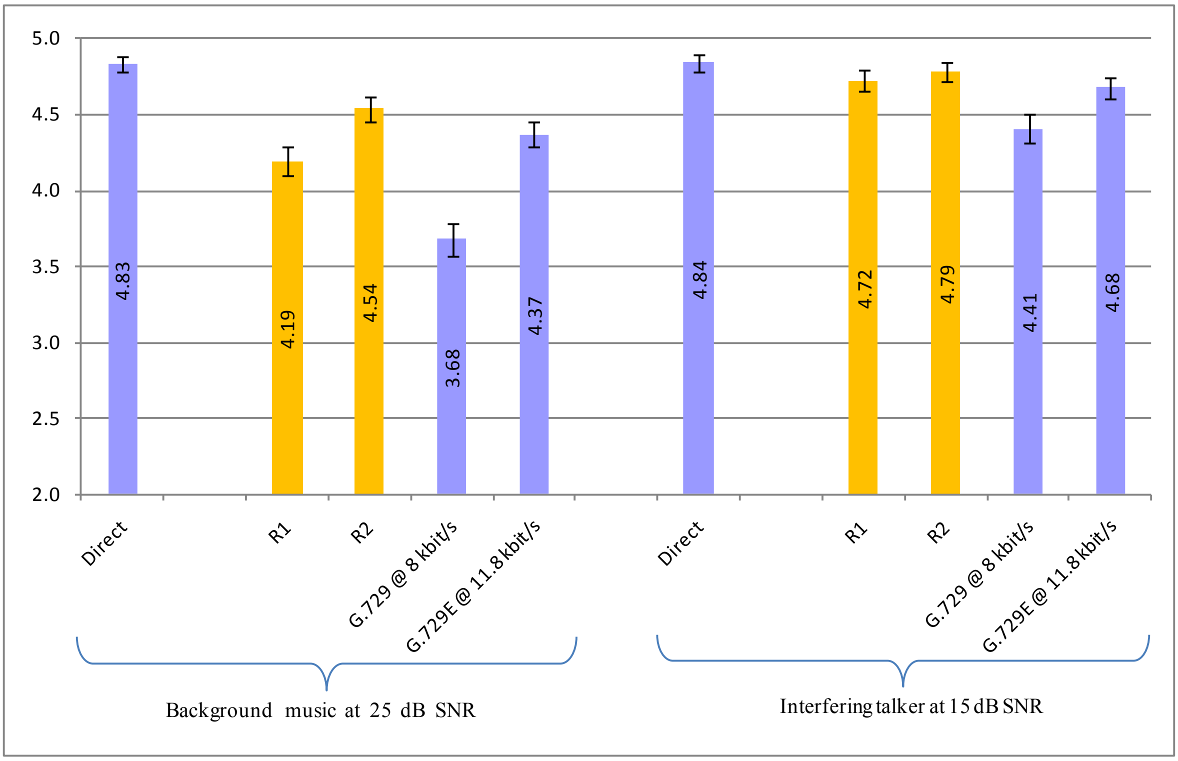

Figure 12: Experiment 3b results for Lab B, Swedish

Table 18: Experiment 3b results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct, 25 dB music	4.65	0.57	4.83	0.37
R1, 25 dB music	4.33	0.73	4.19	0.69
R2, 25 dB music	4.55	0.63	4.54	0.60
G.729 @ 8 kbit/s, 25 dB music	4.15	0.81	3.68	0.76
G.729E @ 11.8 kbit/s, 25 dB music	4.42	0.74	4.37	0.59
Direct, 15 dB int. talker	4.41	0.81	4.84	0.38
R1, 15 dB int. talker	4.40	0.80	4.72	0.48
R2, 15 dB int. talker	4.30	0.92	4.79	0.45
G.729 @ 8 kbit/s, 15 dB int. talker	4.15	0.93	4.41	0.66
G.729E @ 11.8 kbit/s, 15 dB int. talker	4.34	0.85	4.68	0.49

Table 19: Verification against terms of reference - Experiment 3b

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R1, 25dB music	G.729 @ 8 kbit/s, 25dB music	NWT	**PASS**	**PASS**
	R2, 25dB music	G.729E @ 11.8 kbit/s, 25dB music	NWT	**PASS**	**PASS**
	R1, 15dB int. talker	G.729 @ 8 kbit/s, 15dB int. talker	NWT	**PASS**	**PASS**
	R2, 15dB int. talker	G.729E @ 11.8 kbit/s, 15dB int. talker	NWT	**PASS**	**PASS**

9.2.8 Experiment 3c: Narrowband noisy speech (babble, office)

Experiment 3c has been run twice, once in Finnish (Lab A) and once in American English (Lab B). The purpose of this experiment was to evaluate the quality for narrowband noisy speech in clean channel conditions. Two different background noise types are tested: babble noise at 25 dB SNR, and office noise at 20 dB SNR. The codec was evaluated at two different rates: R1 at 8 kbit/s and R2 at 12 kbit/s. Performance in DTX operation was also evaluated. The procedure which was used is the DCR method.

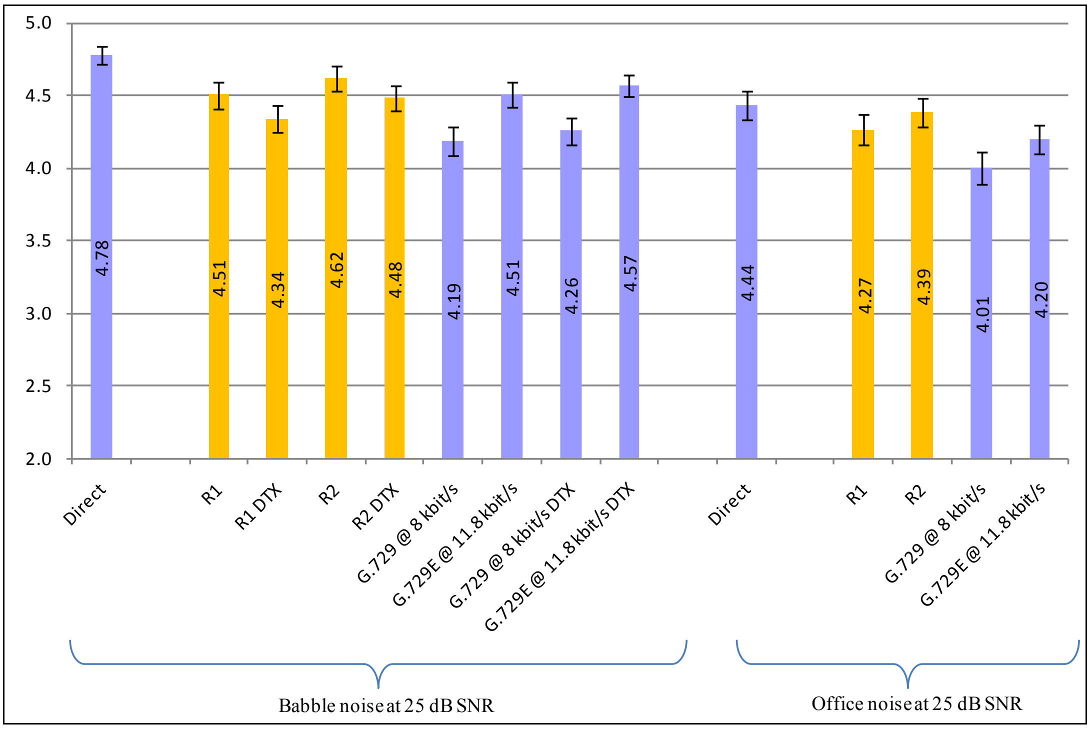

Figure 13: Experiment 3c results for Lab A, Finnish

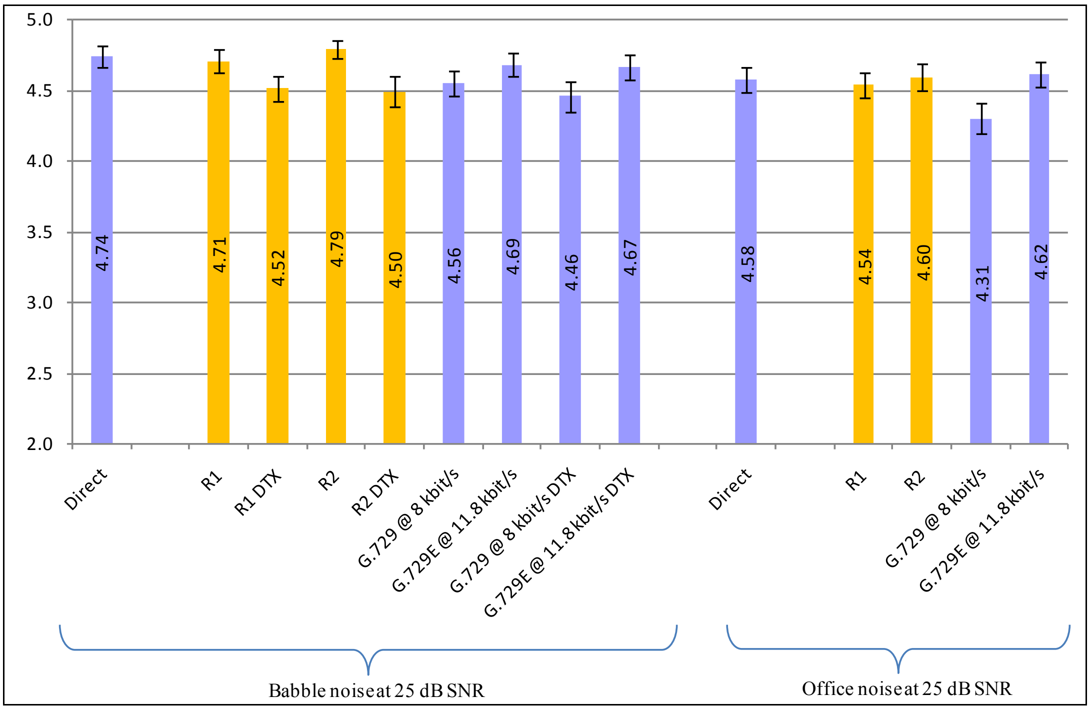

Figure 14: Experiment 3c results for Lab B, American English

Table 20: Experiment 3c results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct, 25 dB babble	4.78	0.46	4.74	0.54
R1, 25 dB babble	4.51	0.66	4.71	0.61
R1 DTX, 25 dB babble	4.34	0.64	4.52	0.62
R2, 25 dB babble	4.62	0.59	4.79	0.44
R2 DTX, 25 dB babble	4.48	0.63	4.50	0.75
G.729 @ 8 kbit/s, 25 dB babble	4.19	0.71	4.56	0.63
G.729E @ 11.8 kbit/s, 25 dB babble	4.51	0.60	4.69	0.58
G.729 @ 8 kbit/s DTX, 25 dB babble	4.26	0.68	4.46	0.78
G.729E @ 11.8 kbit/s DTX, 25 dB babble	4.57	0.57	4.67	0.62
Direct, 25 dB office	4.44	0.68	4.58	0.66
R1, 25 dB office	4.27	0.74	4.54	0.63
R2, 25 dB office	4.39	0.68	4.60	0.66
G.729 @ 8 kbit/s, 25 dB office	4.01	0.78	4.31	0.77
G.729E @ 11.8 kbit/s, 25 dB office	4.20	0.73	4.62	0.62

Table 21: Verification against terms of reference - Experiment 3c

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Requirements	R1, 25 dB babble	G.729 @ 8 kbit/s, 25 dB babble	NWT	**PASS**	**PASS**
	R1 DTX, 25 dB babble	G.729 @ 8 kbit/s DTX, 25 dB babble	NWT	**PASS**	**PASS**
	R2, 25 dB babble	G.729E @ 11.8 kbit/s, 25 dB babble	NWT	**PASS**	**PASS**
	R2 DTX, 25 dB babble	G.729E @ 11.8 kbit/s DTX, 25 dB babble	NWT	**PASS**	**FAIL**
	R1, 25 dB office	G.729 @ 8 kbit/s, 25 dB office	NWT	**PASS**	**PASS**
	R2, 25 dB office	G.729E @ 11.8 kbit/s, 25 dB office	NWT	**PASS**	**PASS**

9.2.9 Experiment 4a: Wideband noisy speech (interfering talker)

Experiment 4a has been run twice, once in American English (Lab A) and once in Canadian French (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband noisy speech in clean channel conditions. One type of background noise was considered: interfering talker at 15 dB SNR. The codec was evaluated at different rates including interoperable modes R2-R5 INT. The test method was used was DCR.

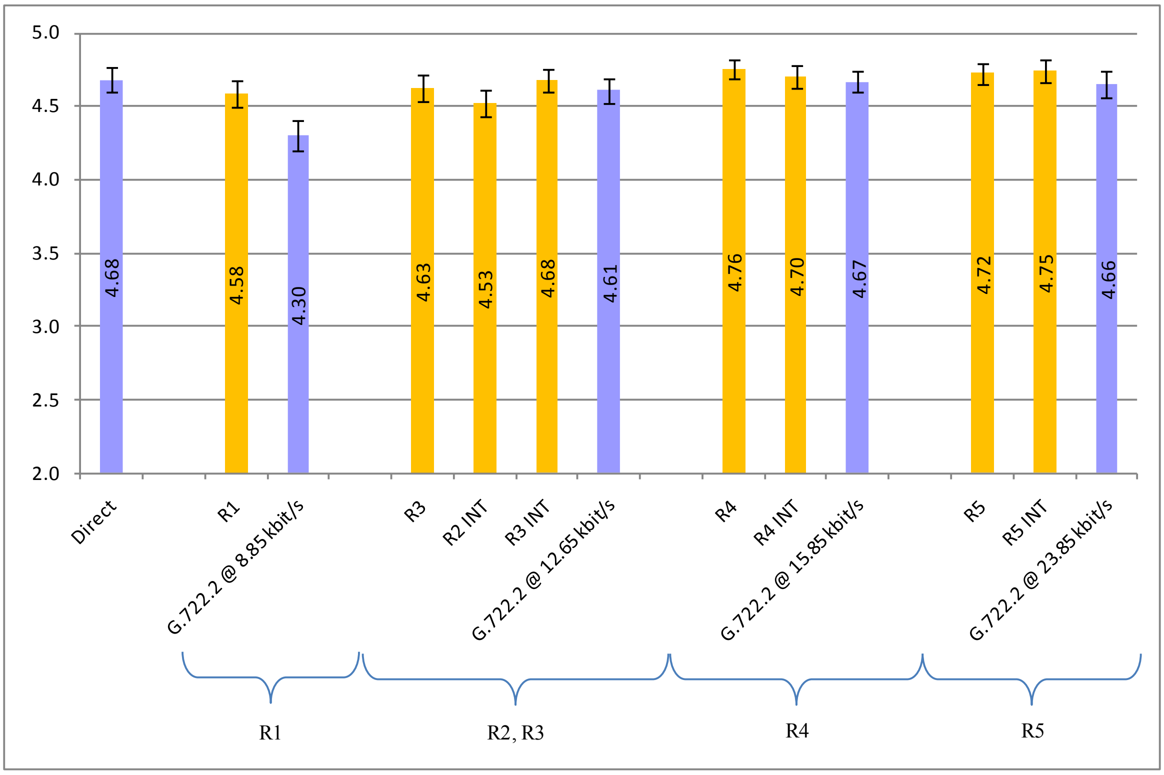

Figure 15: Experiment 4a results for Lab A, American English

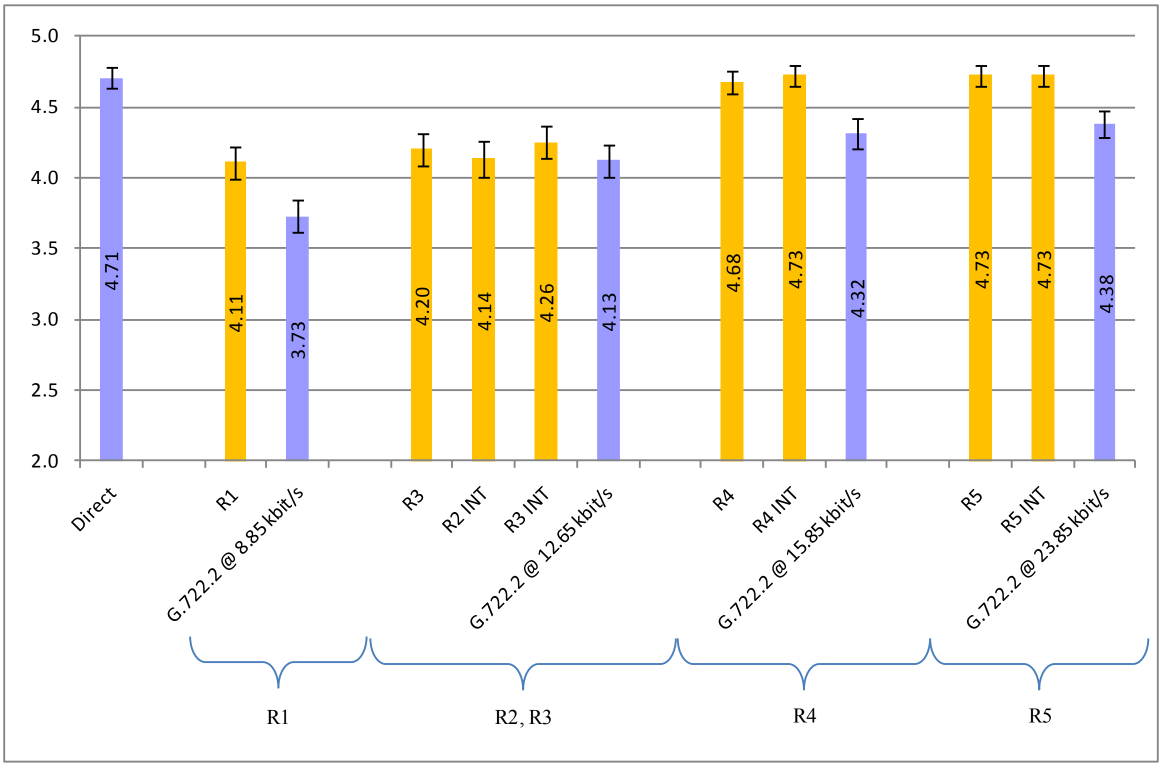

Figure 16: Experiment 4a results for Lab B, Canadian French

Table 22: Experiment 4a results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.68	0.59	4.71	0.50
R1	4.58	0.64	4.11	0.84
R3	4.63	0.61	4.20	0.80
R2 INT	4.53	0.64	4.14	0.86
R3 INT	4.68	0.55	4.26	0.81
R4	4.76	0.48	4.68	0.59
R4 INT	4.70	0.54	4.73	0.51
R5	4.72	0.52	4.73	0.53
R5 INT	4.75	0.53	4.73	0.53
G.722.2 @ 8.85 kbit/s	4.30	0.76	3.73	0.79
G.722.2 @ 12.65 kbit/s	4.61	0.62	4.13	0.81
G.722.2 @ 15.85 kbit/s	4.67	0.52	4.32	0.75
G.722.2 @ 23.85 kbit/s	4.66	0.64	4.38	0.69

Table 23: Verification against terms of reference - Experiment 4a

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R1	G.722.2 @ 8.85 kbit/s	NWT	**PASS**	**PASS**
	R3	G.722.2 @ 12.65 kbit/s	NWT	**PASS**	**PASS**
	R4	G.722.2 @ 15.85 kbit/s	NWT	**PASS**	**PASS**
	R5	G.722.2 @ 23.85 kbit/s	NWT	**PASS**	**PASS**
Obj.	R2 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R3 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R4 INT	G.722.2 @ 15.85 kbit/s	NWT	PASS	PASS
	R5 INT	G.722.2 @ 23.85 kbit/s	NWT	PASS	PASS

9.2.10 Experiment 4b: Wideband noisy speech (background music)

Experiment 4b has been run twice, once in Japanese (Lab A) and once in American English (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband noisy speech in clean channel conditions. One type of background noise was considered: background music at 25 dB SNR. The codec was evaluated at different rates including interoperable modes R2-R5 INT. The test method used was DCR.

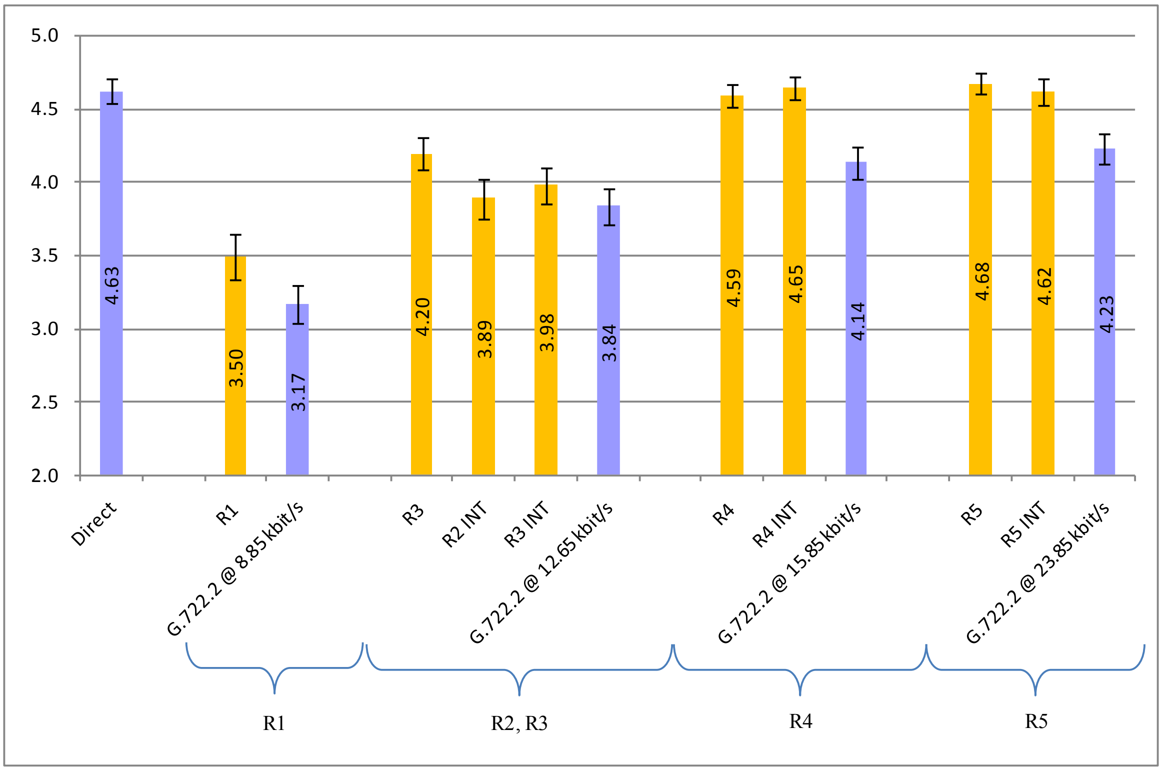

Figure 17: Experiment 4b results for Lab A, Japanese

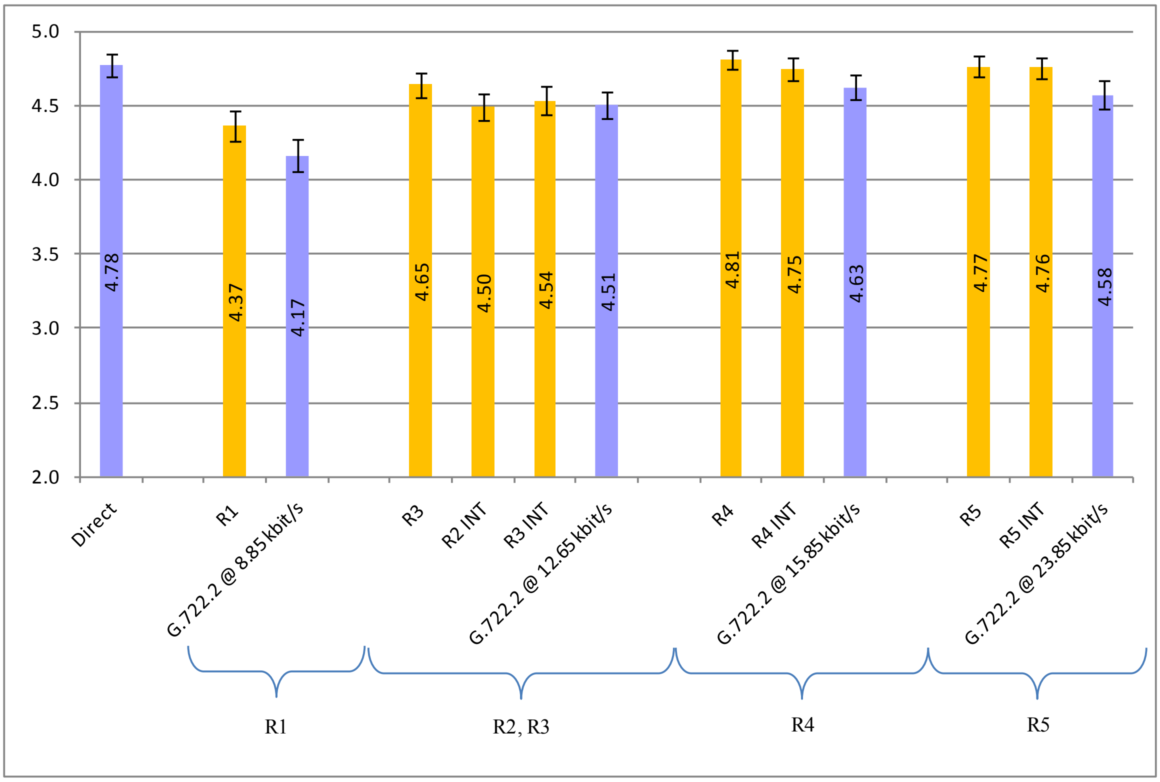

Figure 18: Experiment 4b results for Lab B, American English

Table 24: Experiment 4b results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.63	0.62	4.78	0.52
R1	3.50	1.08	4.37	0.72
R3	4.20	0.75	4.65	0.60
R2 INT	3.89	0.95	4.50	0.62
R3 INT	3.98	0.88	4.54	0.69
R4	4.59	0.56	4.81	0.44
R4 INT	4.65	0.56	4.75	0.52
R5	4.68	0.52	4.77	0.50
R5 INT	4.62	0.64	4.76	0.50
G.722.2 @ 8.85 kbit/s	3.17	0.93	4.17	0.75
G.722.2 @ 12.65 kbit/s	3.84	0.87	4.51	0.61
G.722.2 @15.85 kbit/s	4.14	0.75	4.63	0.60
G.722.2 @ 23.85 kbit/s	4.23	0.73	4.58	0.66

Table 25: Verification against terms of reference - Experiment 4b

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R1	G.722.2 @ 8.85 kbit/s	NWT	**PASS**	**PASS**
	R3	G.722.2 @ 12.65 kbit/s	NWT	**PASS**	**PASS**
	R4	G.722.2 @ 15.85 kbit/s	NWT	**PASS**	**PASS**
	R5	G.722.2 @ 23.85 kbit/s	NWT	**PASS**	**PASS**
Obj.	R2 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R3 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R4 INT	G.722.2 @ 15.85 kbit/s	NWT	PASS	PASS
	R5 INT	G.722.2 @ 23.85 kbit/s	NWT	PASS	PASS

9.2.11 Experiment 4c: Wideband noisy speech (car)

Experiment 4c has been run twice, once in Chinese (Lab A) and once in American English (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband noisy speech in clean channel conditions. One type of background noise was considered: car noise at 15 dB SNR. The codec was evaluated at different rates including interoperable modes R2-R5 INT. The test method used was DCR.

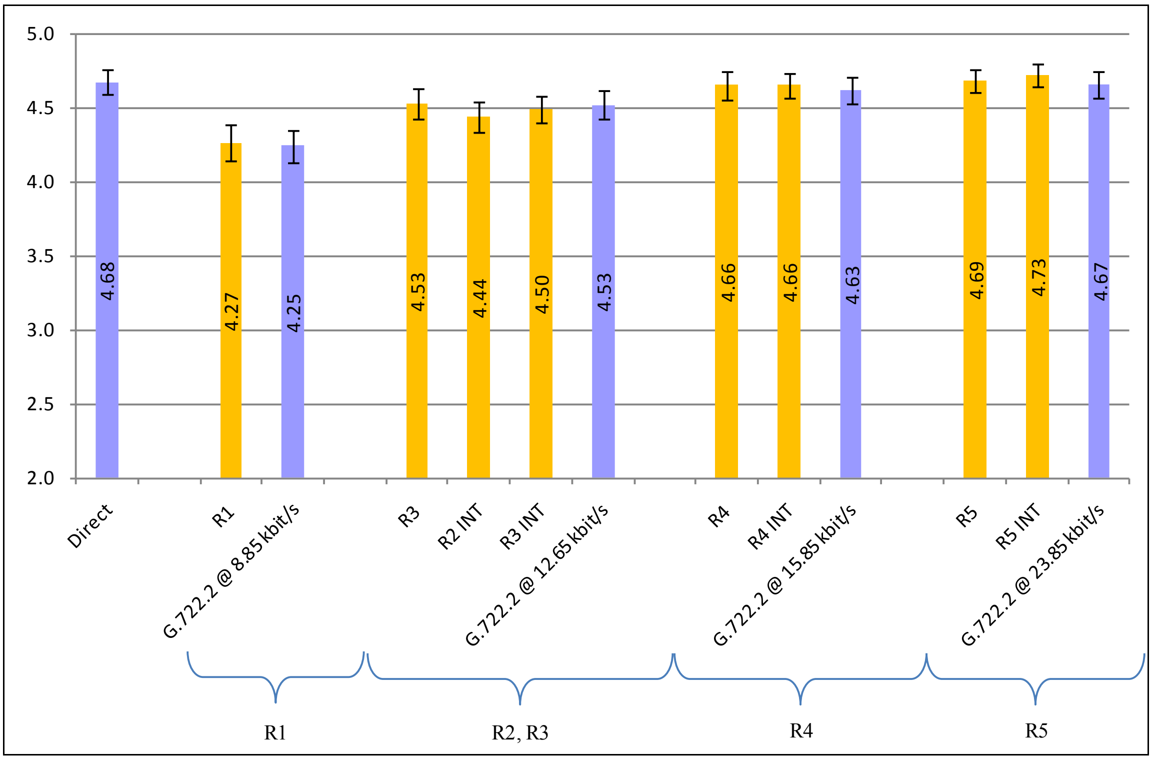

Figure 19: Experiment 4c results for Lab A, Chinese

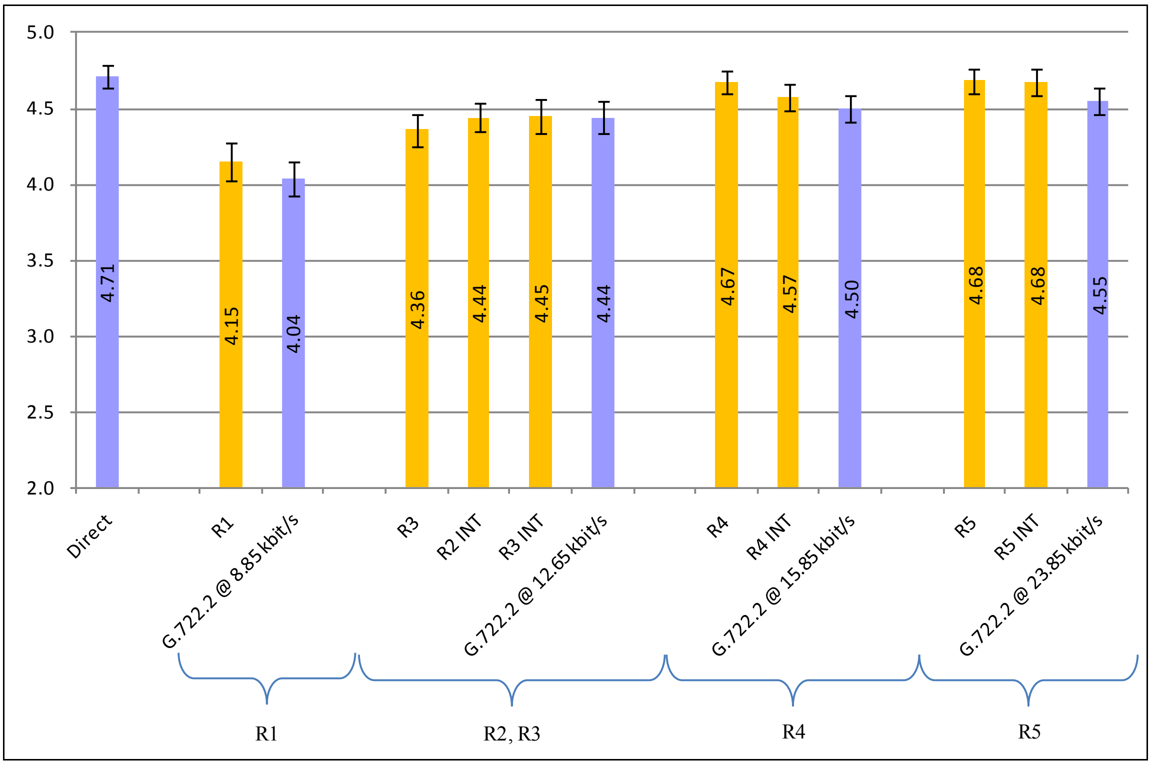

Figure 20: Experiment 4c results for Lab B, American English

Table 26: Experiment 4c results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.68	0.60	4.71	0.53
R1	4.27	0.85	4.15	0.87
R3	4.53	0.72	4.36	0.73
R2 INT	4.44	0.76	4.44	0.68
R3 INT	4.50	0.63	4.45	0.79
R4	4.66	0.64	4.67	0.52
R4 INT	4.66	0.60	4.57	0.60
R5	4.69	0.56	4.68	0.59
R5 INT	4.73	0.54	4.68	0.59
G.722.2 @ 8.85 kbit/s	4.25	0.79	4.04	0.81
G.722.2 @ 12.65 kbit/s	4.53	0.70	4.44	0.75
G.722.2 @ 15.85 kbit/s	4.63	0.62	4.50	0.62
G.722.2 @ 23.85 kbit/s	4.67	0.63	4.55	0.64

Table 27: Verification against terms of reference - Experiment 4c

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R1	G.722.2 @ 8.85 kbit/s	NWT	**PASS**	**PASS**
	R3	G.722.2 @ 12.65 kbit/s	NWT	**PASS**	**PASS**
	R4	G.722.2 @ 15.85 kbit/s	NWT	**PASS**	**PASS**
	R5	G.722.2 @ 23.85 kbit/s	NWT	**PASS**	**PASS**
Obj.	R2 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R3 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R4 INT	G.722.2 @ 15.85 kbit/s	NWT	PASS	PASS
	R5 INT	G.722.2 @ 23.85 kbit/s	NWT	PASS	PASS

9.2.12 Experiment 4d: Wideband noisy speech (office)

Experiment 4d has been run twice, once in Finnish (Lab A) and once in French (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband noisy speech in clean channel conditions. One type of background noise was considered: office noise at 20 dB SNR. The codec was evaluated at different rates including interoperable modes R2-R5 INT. Also, the codec was tested for interoperability with G.722.2/AMR-WB where encoder of one codec and decoder of another codec were employed. Performance in DTX operation was also evaluated. The test method used was DCR.

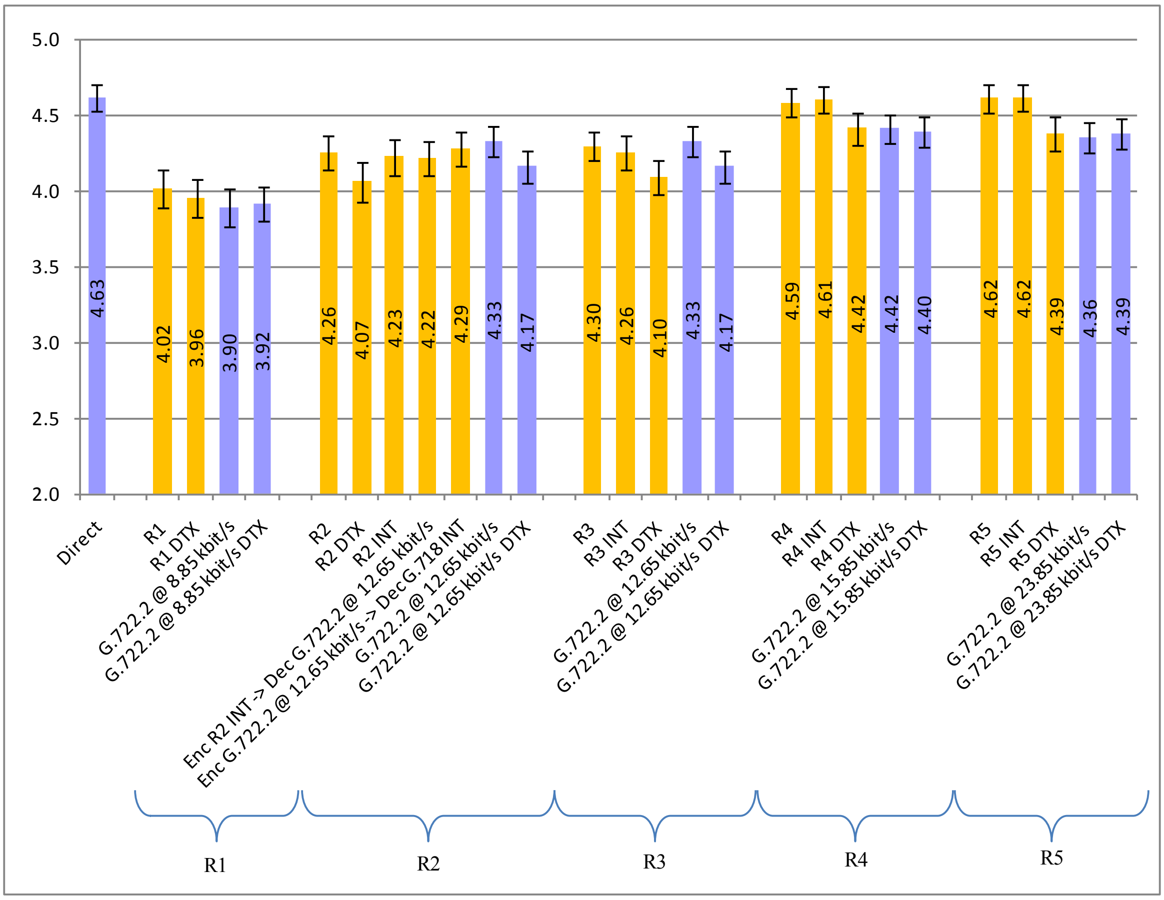

Figure 21: Experiment 4d results for Lab A, Finnish

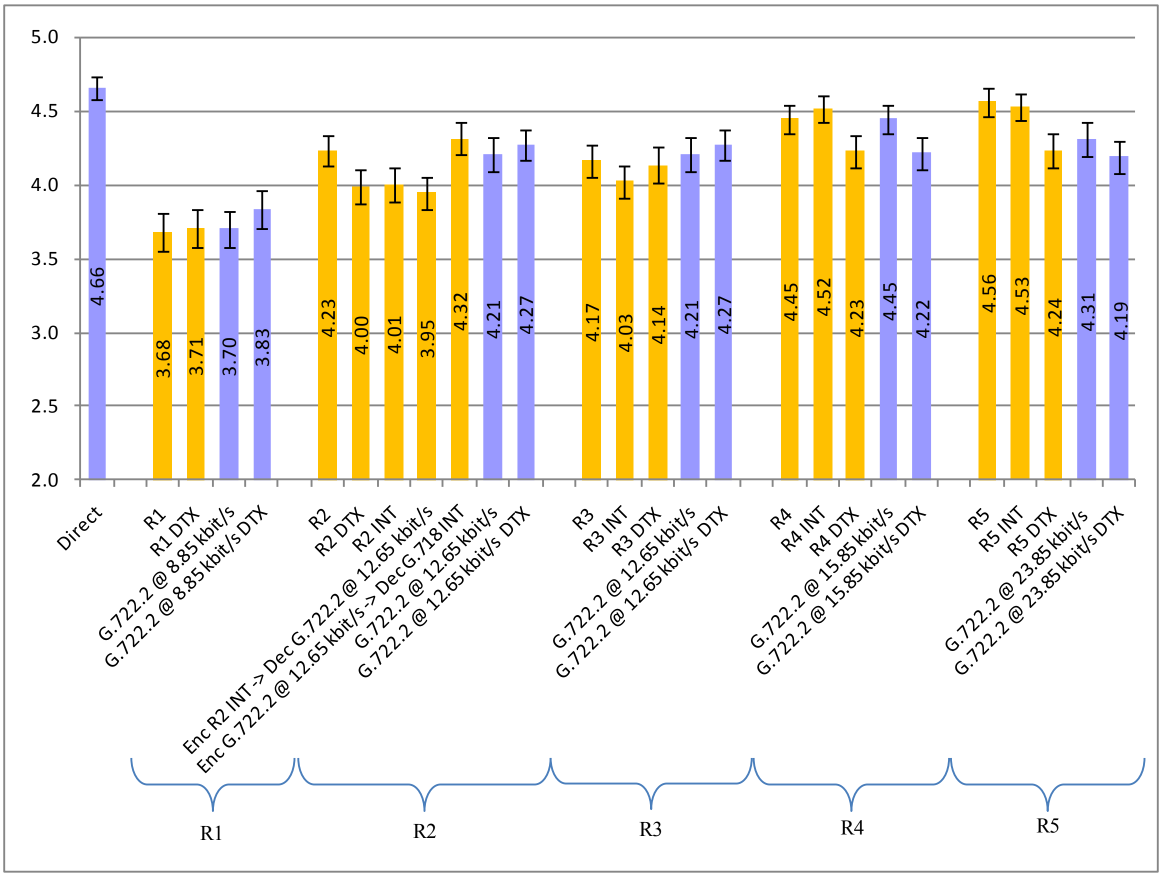

Figure 22: Experiment 4d results for Lab B, French

Table 28: Experiment 4d results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.63	0.64	4.66	0.53
R1	4.02	0.87	3.68	0.90
R1 DTX	3.96	0.88	3.71	0.92
R2	4.26	0.82	4.23	0.75
R2 DTX	4.07	0.94	4.00	0.83
R2 INT	4.23	0.81	4.01	0.78
Enc R2 INT → Dec G.722.2 @ 12.65 kbit/s	4.22	0.81	3.95	0.79
Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT	4.29	0.84	4.32	0.74
R3	4.30	0.69	4.17	0.78
R3 INT	4.26	0.81	4.03	0.76
R3 DTX	4.10	0.82	4.14	0.85
R4	4.59	0.67	4.45	0.69
R4 INT	4.61	0.60	4.52	0.62
R4 DTX	4.42	0.78	4.23	0.79
R5	4.62	0.67	4.56	0.64
R5 INT	4.62	0.63	4.53	0.62
R5 DTX	4.39	0.82	4.24	0.82
G.722.2 @ 8.85 kbit/s	3.90	0.90	3.70	0.85
G.722.2 @ 12.65 kbit/s	4.33	0.72	4.21	0.82
G.722.2 @ 15.85 kbit/s	4.42	0.70	4.45	0.69
G.722.2 @ 23.85 kbit/s	4.36	0.71	4.31	0.81
G.722.2 @ 8.85 kbit/s DTX	3.92	0.82	3.83	0.92
G.722.2 @ 12.65 kbit/s DTX	4.17	0.75	4.27	0.74
G.722.2 @ 15.85 kbit/s DTX	4.40	0.69	4.22	0.76
G.722.2 @ 23.85 kbit/s DTX	4.39	0.74	4.19	0.79

Table 29: Verification against terms of reference - Experiment 4d

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Requirements	R1	G.722.2 @ 8.85 kbit/s	NWT	**PASS**	**PASS**
	R1 DTX	G.722.2 @ 8.85 kbit/s DTX	NWT	**PASS**	**PASS**
	R3	G.722.2 @ 12.65 kbit/s	NWT	**PASS**	**PASS**
	R3 DTX	G.722.2 @ 12.65 kbit/s DTX	NWT	**PASS**	**FAIL**
	R4	G.722.2 @ 15.85 kbit/s	NWT	**PASS**	**PASS**
	R4 DTX	G.722.2 @ 15.85 kbit/s DTX	NWT	**PASS**	**PASS**
	R5	G.722.2 @ 23.85 kbit/s	NWT	**PASS**	**PASS**
	R5 DTX	G.722.2 @ 23.85 kbit/s DTX	NWT	**PASS**	**PASS**
Objectives	R2	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R2 DTX	G.722.2 @ 12.65 kbit/s DTX	NWT	PASS	FAIL
	R2 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	FAIL
	R2 INT → G.722.2 @ 12.65 kbit/s	G.722.2 @ 12.65 kbit/s	NWT	FAIL	FAIL
	G.722.2 @ 12.65 kbit/s → R2 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R3 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	FAIL
	R4 INT	G.722.2 @ 15.85 kbit/s	NWT	PASS	PASS
	R5 INT	G.722.2 @ 23.85 kbit/s	NWT	PASS	PASS
Inf.	R3 DTX	G.722.2 @ 12.65 kbit/s	NWT	-	PASS

9.2.13 Experiment 4e: Wideband noisy speech (babble and street)

Experiment 4e has been run twice, once in Japanese (Lab A) and once in French (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband noisy speech in clean channel conditions. Two types of background noise were considered: babble noise at 25 dB SNR, and street noise at 20 dB SNR. The codec was evaluated at different rates. The test method used was DCR.

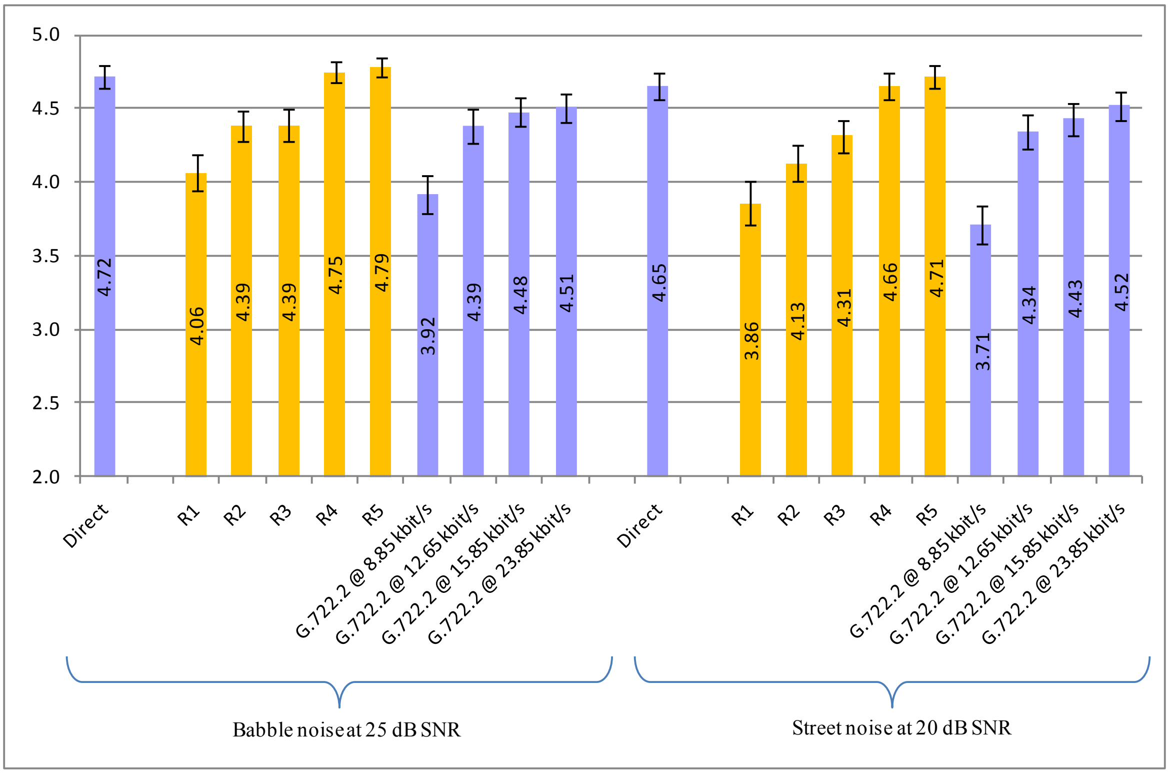

Figure 23: Experiment 4e results for Lab A, Japanese

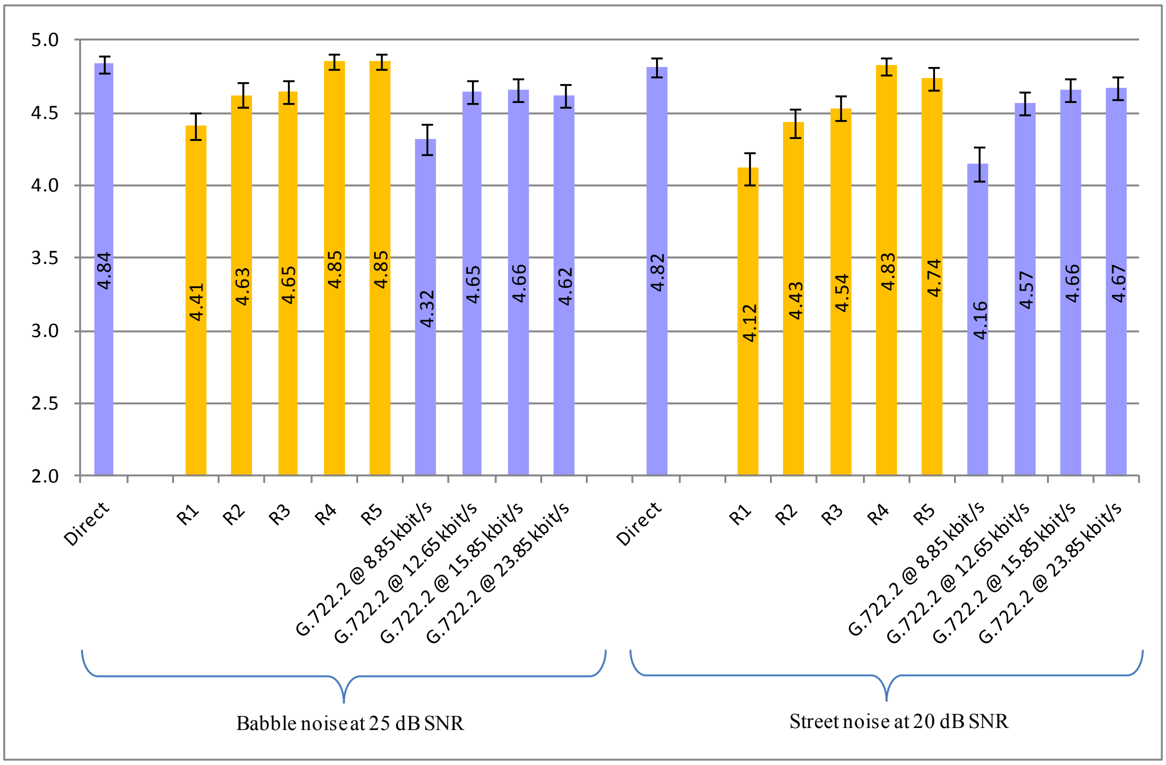

Figure 24: Experiment 4e results for Lab B, French

Table 30: Experiment 4e results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct, 25 dB babble	4.72	0.54	4.84	0.40
R1, 25 dB babble	4.06	0.87	4.41	0.63
R2, 25 dB babble	4.39	0.74	4.63	0.58
R3, 25 dB babble	4.39	0.76	4.65	0.57
R4, 25 dB babble	4.75	0.52	4.85	0.38
R5, 25 dB babble	4.79	0.47	4.85	0.38
G.722.2 @ 8.85 kbit/s, 25 dB babble	3.92	0.89	4.32	0.72
G.722.2 @ 12.65 kbit/s, 25 dB babble	4.39	0.80	4.65	0.55
G.722.2 @ 15.85 kbit/s, 25 dB babble	4.48	0.72	4.66	0.55
G.722.2 @ 23.85 kbit/s, 25 dB babble	4.51	0.68	4.62	0.58
Direct, 20dB street	4.65	0.65	4.82	0.43
R1, 20 dB street	3.86	1.03	4.12	0.79
R2, 20 dB street	4.13	0.86	4.43	0.65
R3, 20 dB street	4.31	0.78	4.54	0.60
R4, 20 dB street	4.66	0.61	4.83	0.41
R5, 20 dB street	4.71	0.56	4.74	0.56
G.722.2 @ 8.85 kbit/s, 20dB street	3.71	0.94	4.16	0.82
G.722.2 @ 12.65 kbit/s, 20dB street	4.34	0.80	4.57	0.57
G.722.2 @ 15.85 kbit/s, 20dB street	4.43	0.76	4.66	0.55
G.722.2 @ 23.85 kbit/s, 20dB street	4.52	0.66	4.67	0.55

Table 31: Verification against terms of reference - Experiment 4e

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Requirements	R1, 25 dB babble	G.722.2 @ 8.85 kbit/s, 25 dB babble	NWT	**PASS**	**PASS**
	R3, 25 dB babble	G.722.2 @ 12.65 kbit/s, 25 dB babble	NWT	**PASS**	**PASS**
	R4, 25 dB babble	G.722.2 @ 15.85 kbit/s, 25 dB babble	NWT	**PASS**	**PASS**
	R5, 25 dB babble	G.722.2 @ 23.85 kbit/s, 25 dB babble	NWT	**PASS**	**PASS**
	R1, 20 dB street	G.722.2 @ 8.85 kbit/s, 20 dB street	NWT	**PASS**	**PASS**
	R3, 20 dB street	G.722.2 @ 12.65 kbit/s, 20 dB street	NWT	**PASS**	**PASS**
	R4, 20 dB street	G.722.2 @ 15.85 kbit/s, 20 dB street	NWT	**PASS**	**PASS**
	R5, 20 dB street	G.722.2 @ 23.85 kbit/s, 20 dB street	NWT	**PASS**	**PASS**
Obj.	R2, 25 dB babble	G.722.2 @ 12.65 kbit/s, 25 dB babble	NWT	PASS	PASS
	R2, 20 dB street	G.722.2 @ 12.65 kbit/s, 20 dB street	NWT	FAIL	FAIL

9.3 Summary of non bit-exact corrections to the source code after Step 1

This clause summarizes all non bit-exact corrections of the source code implemented after the Characterization Tests, Step 1. The intention is to provide all information necessary for accurate interpretation of the test results.

9.3.1 Wrong erasure concealment attenuation factor

Description: In case of frame erasures, synthesis signal is attenuated. The attenuation factor was not correctly translated into the fixed point for some cases of unvoiced speech concealment. For more information, see [26].

Performance Impact: This problem manifested itself only for a special case of concealment of non-stationary unvoiced signal, when at least two consecutive frames were erased. The attenuation was then faster than intended. Given a very low probability of occurrence of this situation, the perceptual impact of this problem was negligible.

9.3.2 Wrong subframe length in concealment resynchronization

Description: This problem affected the resynchronization module of the frame erasure concealment and only in the case when the last pitch lag (the pitch value of the last correctly received frame) was lower than a half-frame. The half-frame length taken into account in that decision was not correct. For more information, see [26].

Performance Impact: The bug affected the efficiency of the resynchronization module in the specific case described above. The perceptual impact was therefore very small.

9.3.3 Missing initialization in the algebraic codebook search in the G.722.2 interoperable mode

Description: An uninitialized variable was found in the fixed point source code for the 36-bit fixed codebook in a mode interoperable with G.722.2/AMR-WB at 12.65 kbit/s. This missing initialization resulted in a sub-optimal performance of the search mechanism for the 36-bit codebook. For more information, see [26].

Performance Impact: The omitted initialization affected the perceptual quality of the synthesized signal in the G.722.2/AMR-WB interoperable mode at 12.65 kbit/s. It was found that this problem was the reason why, in the Characterization Tests, Step 1, the R2 INT and R3 INT test conditions performed lower than R2 and R3, respectively and also lower than G.722.2 at 12.65 kbit/s.

The positive impact of the correction was verified through an objective evaluation. Below is a summary of the SNR values before and after the correction of the problem for clean and noisy speech, evaluated on the whole database. The values in Table 32 are segmental SNR values computed on a weighted signal at the output of the codebook search.

The quality improvement was also verified using the WB-PESQ algorithm in Rec. ITU-T P.862.2 [32] for the configuration when the codec is interoperated with G.722.2 at 12.65 kbit/s. The results are shown in Table 33.

Table 32: Improvement after initialization of variable for 36-bit codebook search

Stage	Clean speech	Office noise
Before correction (Characterization Tests, Step 1)	SNR = 9.514 dB	SNR = 6.357 dB
After correction	SNR = 10.105 dB	SNR = 6.647 dB

Table 33: P.862.2 scores for G.718 interoperating with G.722.2 at 12.65 kbit/s

Encoder	Decoder	Correction	P.862.2 results
G.722.2	G.722.2	–	3.385
G.718	G.722.2	Before	3.312
G.718	G.722.2	After	3.398
G.722.2	G.718	Before	3.264
G.722.2	G.718	After	3.336

9.3.4 Problem in pitch extrapolation in case of frame erasures

Description: In case of erasures of voiced speech frames in low delay mode, the pitch evolution in the lost frame is extrapolated based on the mean of past pitch values. However, it was found that a sum of the differences of the past pitch values has been used instead of their mean. For more information, see [27].

Performance Impact: The general impact of this error was rather limited, and no perceptual impact was observed on the test results of the Characterization Tests, Step 1. This problem affected only the case of frame erasures following frames classified as voiced speech, and only in the case of low delay mode. After fixing this error, a perceptual improvement could be observed in these very rare occasions.

9.3.5 Problem in ISF interpolation in case of frame erasures

Description: In case of frame erasures preceding a Transition Coding (TC) frame and when in the low delay mode, a mismatch between the interpolated synthesis filter and the excitation could occasionally cause a strong artefact. For more information, see [27].

Performance Impact: The problem manifested itself on very rare occasions (TC frames are used in about 6% of active speech). When happened, however, its perceptual impact was sometimes important. In a simulation of the whole Characterization Tests, Step 1, only one perceptually audible artefact has been removed by the correction of this problem.

9.3.6 Wrong initialization of a normalization in FPC

Description: In the fixed point transcription of the source code, two digits have been swapped causing the constant 19418 being erroneously coded as 19148 in the initialization of the normalization in the Factorial Pulse Coding (FPC) of MDCT coefficients. For more information, see [27].

Performance Impact: None.

9.3.7 Non-optimal use of the dual low frequency (bass) post filter in G.722.2 interoperable modes

Description: During the decoding of G.722.2 frames at the following bitrates: 12.65 14.25 15.85 18.25 19.85 23.05 23.85 kbit/s, and also during the decoding of R2-R5 INT frames, the dual low frequency (bass) post filter was not used properly. In these cases, the pitch approximation was less accurate than in the default mode. For more information, see [27].

Performance Impact: Very small perceptual impact when decoding in the G.722.2 interoperable mode.

9.3.8 Out-of-memory access in extrapolation of ISFs

Description: Out-of-memory access was identified during the extrapolation of ISF parameters from 12.8 kHz domain to 16 kHz domain in the fixed point implementation. For more information, see [27].

Performance Impact: The problem occurred on very rare occasions. It was observed that it happened only in one case of music encoding/decoding when it caused a clearly perceptual artefact. It is highly unlikely that it affected the results of the Characterization Tests, Step 1.

9.3.9 Insufficient resolution in the computation of the total encoder excitation of Layer 2 in fixed point

Description: In rare cases of music signal encoding/decoding, the fixed point resolution was not sufficient for the gain representation of the innovative part of the excitation signal of Layer 2, resulting in a zero innovation vector. For more information, see [27].

Performance Impact: This problem did not affect speech inputs. For music inputs, it manifested itself in highly exceptional circumstances when unvoiced coding mode was used for music inputs. The problem was observed only once when it caused a clear perceptual artefact. It is highly unlikely that it affected the results of the Characterization Tests, Step 1.

9.3.10 Limited precision in the fixed point implementation of the innovation gain interpolation in case of frame erasures

Description: The problem was due to a limited precision of a fixed point variable. It happened only in the G.722.2 interoperable mode and during frame erasure concealment, when the excitation gains were interpolated. Due to the insufficient precision of the fixed point variable used to store the innovation gain, the random part of the excitation was zeroed instead of having its energy interpolated in the lost frame. For more information, see [28].

Performance Impact: The random part of the excitation was zeroed instead of having its energy interpolated in the lost frame, yielding a drop in the energy, at some places clearly annoying.

10 Characterization Tests, Step 2

The second step (Step 2) of the Characterization Tests was completed in March 2009. Step 2 extended the codec evaluation to conditions not covered in Step 1. One of the goals was to evaluate the codec in extreme transmission conditions, covering frame erasure rates up to 12% (random FER) and also in situations where long sequences of consecutive frames are lost (bursty FER). Also, in Step 2, the codec was evaluated in tandem conditions with other codecs. The latest versions of the G.718 Characterization Step 2 Processing Plan and Listening Test Plan can be found in [15] and [16].

10.1 Organization of the Characterization Tests, Step 2

Table 34shows a summary of the parameters tested in each experiment of the Characterization step 2 testing phase.

In Table 34, WB means wideband input and wideband output, and NB means narrowband input and narrowband output. The "Background noise" column specifies the type and the level of background noise signal applied to the input (measured as SNR with respect to the level of the input signal). The test method is ACR, DCR or ITU-T P.835. The "Errors" column shows whether frame erasures were tested and if so, the percentage of random Frame Erasure Rate (FER) or Bursty Frame Erasure Rate (BFER). The "Rates" column specifies the tested rates of the G.718 codec, i.e. R1-R5 corresponding to the bitrates of 8, 12, 16, 24 and 32kbit/s. The following acronyms are used in the table above, and throughout the following sections: LD stands for low delay decoding mode, DTX for discontinuous transmission, NR for noise reduction and INT corresponds to conditions where first two default layers are replaced with layers interoperable with G.722.2 at 12.65kbit/s. For experiments 3a, 3b, 4a and 4b voice activity factors (VAF) have been calculated for certain DTX conditions. They appear in the right-most column of tables with test results of each laboratory (tables with MOS and SD values). The VAF values indicate the percentage of all active frames, i.e. those that were not classified as "No data" or "Silence descriptor".

The experiments are assigned to different listening laboratories. Every experiment was run in two different languages. In Experiment 2, music items for all types but modern type are different across laboratories. The laboratories and languages are indicated in Table 35.

Table 34: Organization of Characterization Tests, Step 2

Exp	Input	NB/ WB	Background noise	Test method	Errors	Rates	Remarks
1a	Speech	NB	-	ACR	3, 6% FER, BFER	R1,R2	Tandem conditions switching
1b	Speech	WB	-	ACR	3, 6, 8% FER, BFER	R1, R2	Tandem conditions INT, switching
1c	Speech	WB	-	ACR	3, 6, 8, 12% FER, BFER	R3, R5	Tandem conditions INT
2	Music	WB	-	ACR	-	R1, R2, R3, R4, R5	INT
3a	Speech	NB	Car @ 15 dB SNR	DCR	3% FER	R1, R2	Input level: -16 dBov and -36 dBov; DTX
3b	Speech	NB	Office @ 20 dB SNR	DCR	3% FER	R1, R2	Input level: -16 dBov and -36 dBov; DTX
4a	Speech	WB	Car @ 15 dB SNR	DCR	3% FER	R1, R2, R3, R5	Input level: -16 dBov and -36 dBov INT, DTX, switching
4b	Speech	WB	Street @ 20 dB SNR	P.835	-	R1, R2, R3, R4	INT, DTX, NR

Table 35: Testing laboratories and associated languages in Characterization Tests, Step 2

Exp	Lab A	Language – Lab A	Lab B	Language – Lab B
1a	Panasonic	Japanese	Dynastat (Motorola)	American English
1b	Dynastat (Qualcomm)	American English	France Telecom	French
1c	Dynastat (Motorola)	American English	VoiceAge	Canadian French
2	Dynastat (Qualcomm)	Music	Ericsson	Music
3a	Panasonic	Japanese	Nokia	Finnish
3b	Huawei	Chinese	Nokia	Finnish
4a	Huawei	Chinese	Ericsson	Swedish
4b	France Telecom	French	VoiceAge	Canadian English

10.2 Test results

The summary of the test results of Characterization Tests, Step 2 can be found in [17]. Individual listening laboratory reports can be found in [18], [19], [20], [20], [22], [23], [24] and [25]. Note that all "better than" (BT) criteria are systematically supplemented with "or not worse than (NWT) direct" criteria, to account for test saturation. The NWT direct criterion is not made explicit in the tables on “Verification against terms of reference”, but is always assumed to be part of the BT criteria. In all tests, if not mentioned explicitly, the default level of input signal is -26 dBov. The number of votes per condition is 192. The test results are divided into three categories: requirements, objectives and informative. In the graphical representation, different test cases are logically grouped and colours are used to identify the test conditions and the reference conditions. Please note, that a particular reference condition may be used for several test conditions. In the subsections below, BFER refers to bursty frame erasures, whereas FER indicates random frame erasures. The colour code in the “Verification against terms of reference” tables visually groups test conditions addressing requirements, objectives and informational test items; and pass / fail in bold typeface indicate the requirements.

10.2.1 Experiment 1a: Narrowband clean speech

Experiment 1a has been run twice, once in Japanese (Lab A) and once in American English (Lab B). The purpose of this experiment was to evaluate the quality of the codec for narrowband speech both in clean channel and in frame erasures conditions (random and bursty). The codec was tested on two rates: R1 (8 kbit/s) and R2 (12 kbit/s). This experiment also evaluated the performance of the codec in tandem with other narrowband speech coding standards. The test method used was ACR.

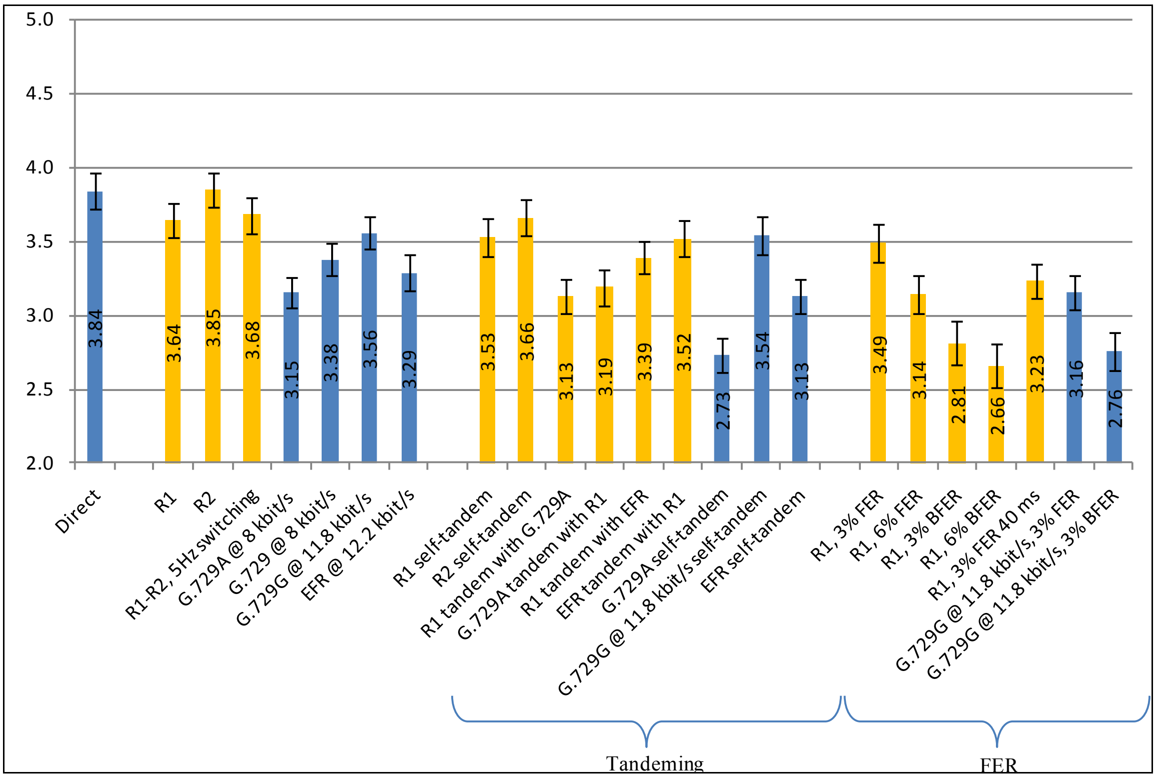

Figure 25: Experiment 1a results for Lab A, Japanese

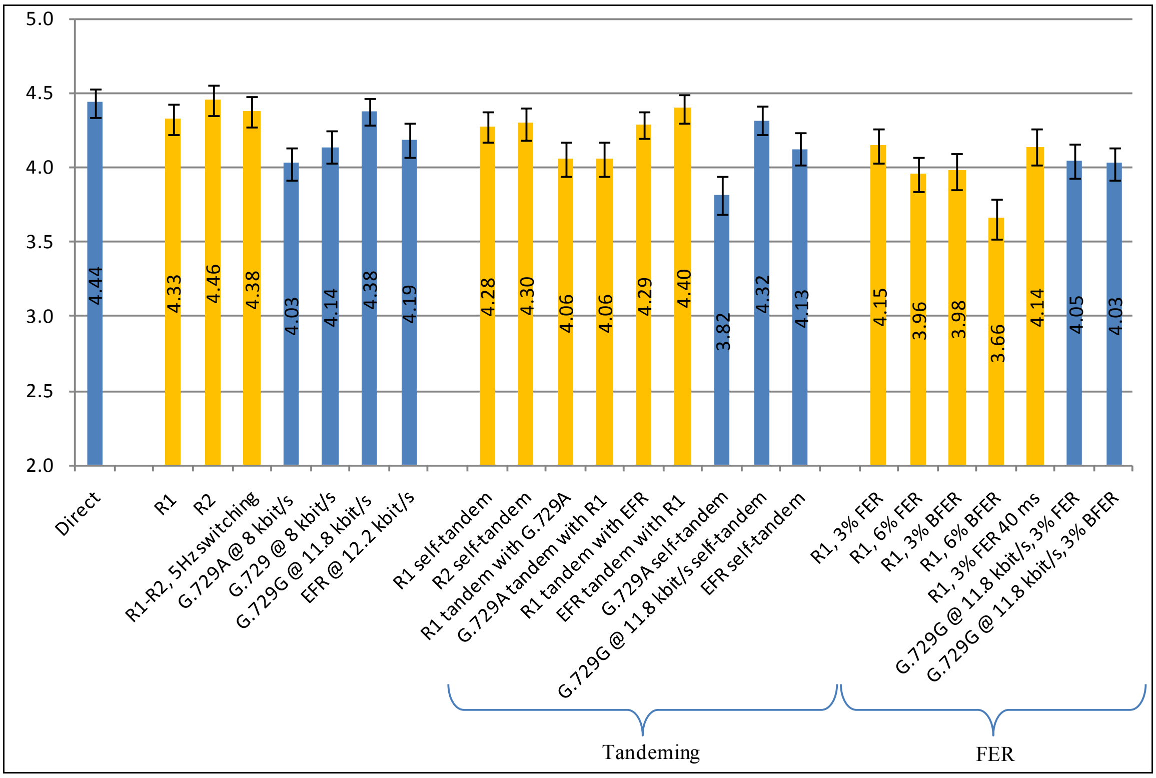

Figure 26: Experiment 1a results for Lab B, American English

Table 36: Experiment 1a results for both testing laboratories

Condition	Japanese		American English	
	MOS	SD	MOS	SD
Direct	3.84	0.86	4.44	0.69
R1	3.64	0.83	4.33	0.71
R2	3.85	0.81	4.46	0.72
R1-R2, 5Hz switching	3.68	0.86	4.38	0.68
G.729A @ 8 kbit/s	3.15	0.73	4.03	0.76
G.729 @ 8 kbit/s	3.38	0.81	4.14	0.76
G.729G @ 11.8 kbit/s	3.56	0.81	4.38	0.67
EFR @ 12.2 kbit/s	3.29	0.90	4.19	0.81
R1 self tandem	3.53	0.93	4.28	0.75
R2 self tandem	3.66	0.88	4.30	0.78
R1 tandem with G.729A	3.13	0.80	4.06	0.83
G.729A tandem with R1	3.19	0.89	4.06	0.83
R1 tandem with EFR	3.39	0.79	4.29	0.66
EFR tandem with R1	3.52	0.88	4.40	0.69
G.729A self tandem	2.73	0.81	3.82	0.93
G.729G @ 11.8 kbit/s self tandem	3.54	0.92	4.32	0.68
EFR self tandem	3.13	0.84	4.13	0.74
R1, 3% FER	3.49	0.88	4.15	0.81
R1, 6% FER	3.14	0.92	3.96	0.79
R1, 3% BFER	2.81	1.04	3.98	0.87
R1, 6% BFER	2.66	1.02	3.66	0.97
R1, 3% FER 40 ms	3.23	0.81	4.14	0.84
G.729G @ 11.8 kbit/s, 3% FER	3.16	0.81	4.05	0.81
G.729G @ 11.8 kbit/s, 3% BFER	2.76	0.89	4.03	0.80

Table 37: Verification against terms of reference - Experiment 1a

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Requirements	R1	G.729G @ 11.8 kbit/s	NWT	**PASS**	**PASS**
	R2	G.729G @ 11.8 kbit/s	BT	**PASS**	**PASS**
	R1, 3% FER	G.729G @ 11.8 kbit/s, 3% FER	NWT	**PASS**	**PASS**
	R1, 3% BFER	G.729G @ 11.8 kbit/s, 3% BFER	NWT	**PASS**	**PASS**
	R1-R2, 5Hz switching	R1	NWT	**PASS**	**PASS**
Informative	R1 self tandem	G.729A self tandem	NWT	PASS	PASS
	R2 self tandem	G.729G @ 11.8 kbit/s self tandem	NWT	PASS	PASS
	R1 tandem with G.729A	G.729A @ 8 kbit/s	NWT	PASS	PASS
	G.729A tandem with R1	G.729A @ 8 kbit/s	NWT	PASS	PASS
	R1, 3% FER 40 ms	R1, 3% FER	NWT	FAIL	PASS
	R1 tandem with EFR	EFR self tandem	NWT	PASS	PASS
	EFR tandem with R1	EFR self tandem	NWT	PASS	PASS
	R1, 6% FER	R1, 3% FER	NWT	FAIL	FAIL
	R1, 6% BFER	R1, 3% BFER	NWT	PASS	FAIL

10.2.2 Experiment 1b: Wideband clean speech (lower rates)

Experiment 1b has been run twice, once in American English (Lab A) and once in French (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband speech at lower rates, i.e. R1 at 8 kbit/s, R2 at 12 kbit/s and R2 INT at 12.65 kbit/s (interoperable with G.722.2 at 12.65 kbit/s). The codec was evaluated both, in clean channel and frame erasure conditions (random and bursty). Also, this experiment evaluated the performance of the codec in tandem with other wideband speech coding standards and its interoperability with G.722.2/AMR-WB. The test method used was ACR.

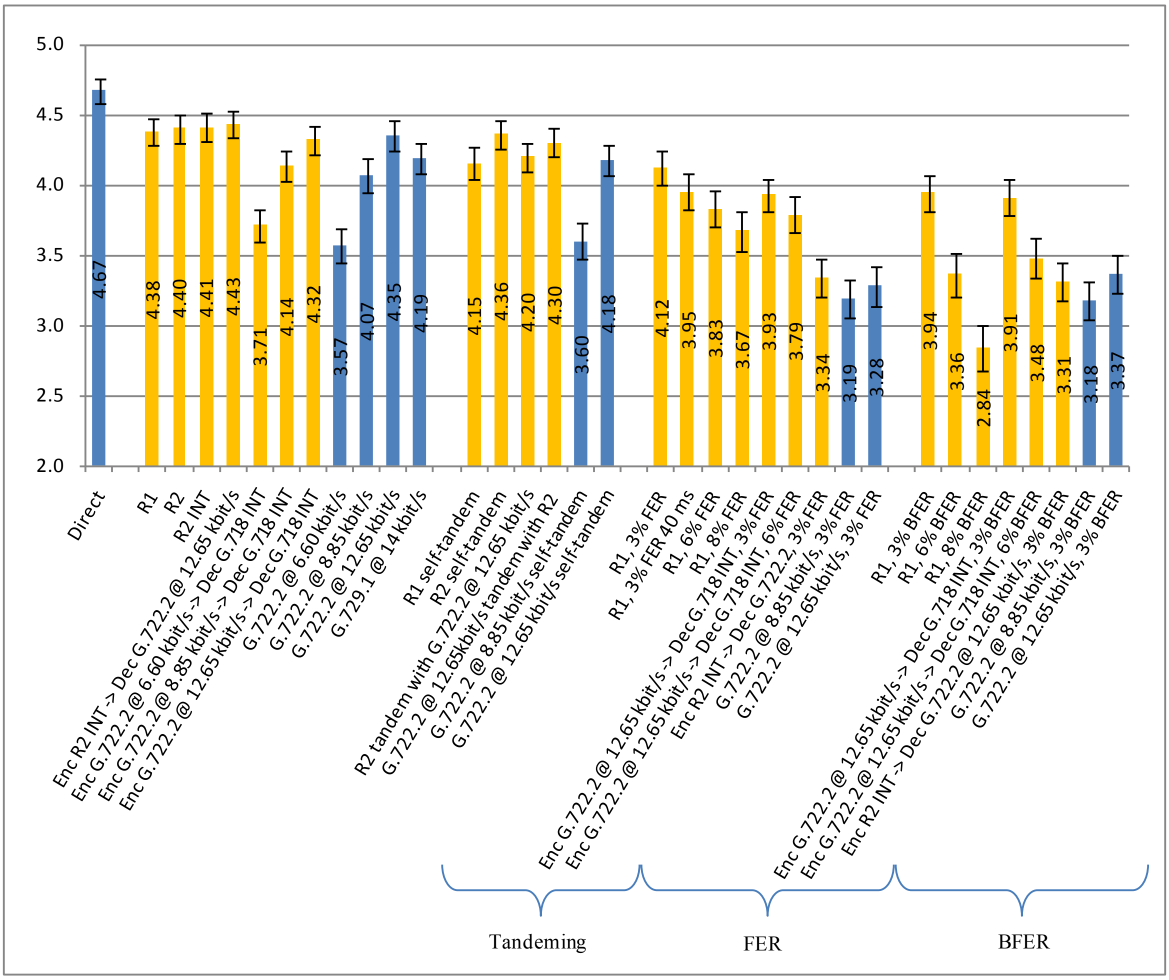

Figure 27: Experiment 1b results for Lab A, American English

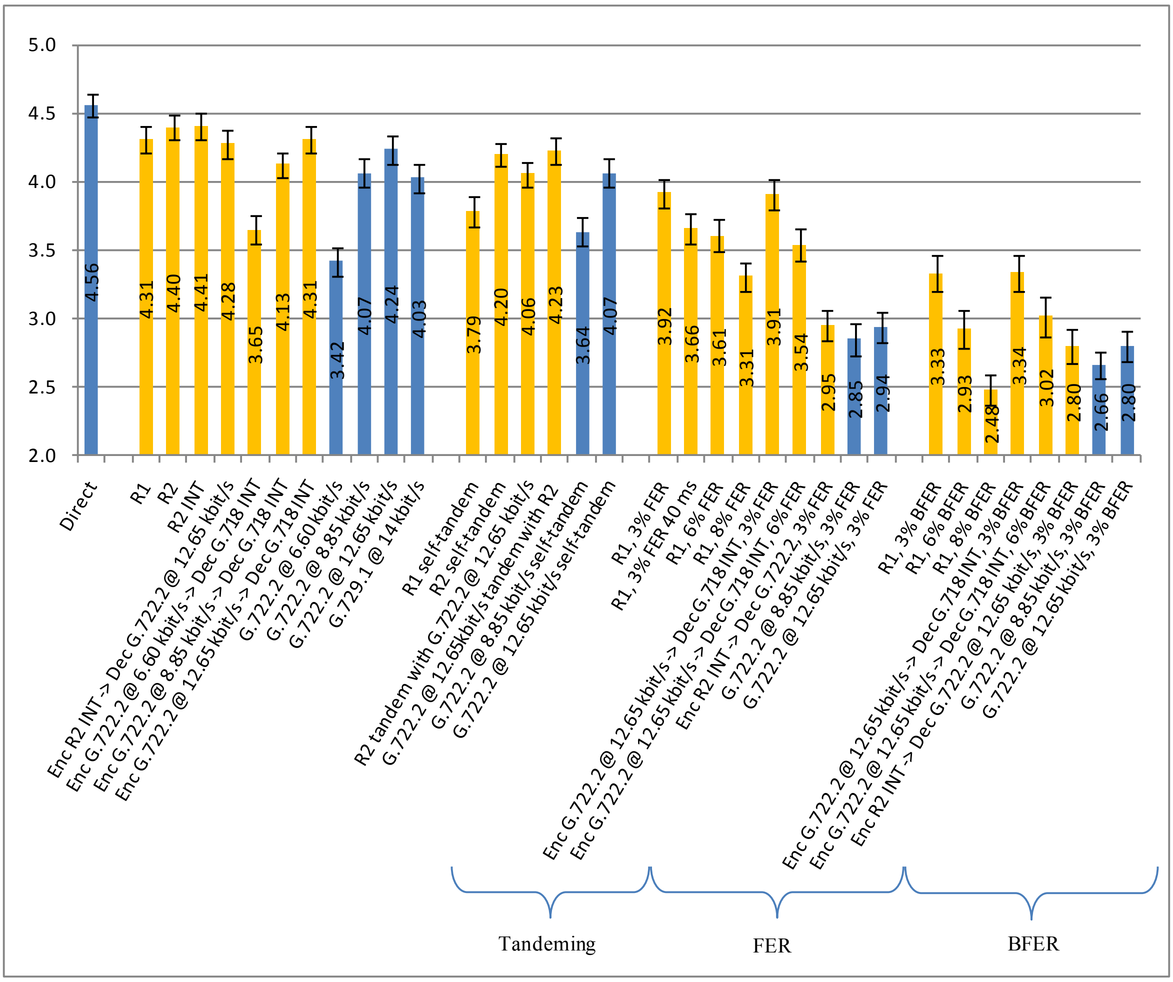

Figure 28: Experiment 1b results for Lab B, French

Table 38: Experiment 1b results for both testing laboratories

Condition	American English		French	
	MOS	SD	MOS	SD
Direct	4.67	0.58	4.56	0.57
R1	4.38	0.70	4.31	0.68
R2	4.40	0.72	4.40	0.66
R2 INT @ 12.65 kbit/s	4.41	0.70	4.41	0.65
Enc R2 INT → Dec G.722.2 @ 12.65 kbit/s	4.43	0.68	4.28	0.70
Enc G.722.2 @ 6.60 kbit/s → Dec G.718 INT	3.71	0.85	3.65	0.72
Enc G.722.2 @ 8.85 kbit/s → Dec G.718 INT	4.14	0.75	4.13	0.64
Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT	4.32	0.70	4.31	0.67
G.722.2 @ 6.60 kbit/s	3.57	0.88	3.42	0.75
G.722.2 @ 8.85 kbit/s	4.07	0.84	4.07	0.72
G.722.2 @ 12.65 kbit/s	4.35	0.73	4.24	0.71
G.729.1 @ 14 kbit/s	4.19	0.75	4.03	0.71
R1 self tandem	4.15	0.81	3.79	0.77
R2 self tandem	4.36	0.70	4.20	0.61
R2 tandem with G.722.2 @ 12.65 kbit/s	4.20	0.72	4.06	0.66
G.722.2 @ 12.65kbit/s tandem with R2	4.30	0.73	4.23	0.72
G.722.2 @ 8.85 kbit/s self tandem	3.60	0.90	3.64	0.77
G.722.2 @ 12.65 kbit/s self tandem	4.18	0.77	4.07	0.71
R1, 3% FER	4.12	0.84	3.92	0.71
R1, 3% FER 40 ms	3.95	0.89	3.66	0.78
R1, 6% FER	3.83	0.90	3.61	0.81
R1, 8% FER	3.67	0.97	3.31	0.76
Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 3% FER	3.93	0.83	3.91	0.78
Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 6% FER	3.79	0.91	3.54	0.84
Enc R2 INT → Dec G722.2 @ 12.65 kbit/s, 3% FER	3.34	0.96	2.95	0.80
G.722.2 @ 8.85 kbit/s, 3% FER	3.19	0.96	2.85	0.85
G.722.2 @ 12.65 kbit/s, 3% FER	3.28	1.03	2.94	0.78
R1, 3% BFER	3.94	0.87	3.33	0.92
R1, 6% BFER	3.36	1.11	2.93	0.99
R1, 8% BFER	2.84	1.11	2.48	0.81
Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 3% BFER	3.91	0.89	3.34	0.94
Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 6% BFER	3.48	1.02	3.02	1.04
Enc R2 INT → Dec G.722.2 @ 12.65 kbit/s, 3% BFER	3.31	0.94	2.80	0.85
G.722.2 @ 8.85 kbit/s, 3% BFER	3.18	0.94	2.66	0.70
G.722.2 @ 12.65 kbit/s, 3% BFER	3.37	0.94	2.80	0.75

Table 39: Verification against terms of reference - Experiment 1b

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R1	G.722.2 @ 8.85 kbit/s	NWT	**PASS**	**PASS**
Req.	R1, 3% FER	G.722.2 @ 8.85 kbit/s, 3% FER	NWT	**PASS**	**PASS**
Req.	R1, 3% BFER	G.722.2 @ 8.85 kbit/s, 3% BFER	NWT	**PASS**	**PASS**
Objectives	R2	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
Objectives	R2 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
Objectives	Enc G.722.2 @12.65 kbit/s → Dec G.718-INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
Objectives	Enc R2 INT → Dec G.722.2 @ 12.65 kbit/s	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
Informative	R1 self tandem	G.722.2 @ 8.85 kbit/s self tandem	NWT	PASS	PASS
Informative	R2 self tandem	G.722.2 @ 12.65 kbit/s self tandem	NWT	PASS	PASS
Informative	R2 tandem with G.722.2 @ 12.65 kbit/s	G.722.2 @ 12.65 kbit/s self tandem	NWT	PASS	PASS
Informative	G.722.2 @ 12.65 kbit/s tandem with R2	G.722.2 @ 12.65 kbit/s self tandem	NWT	PASS	PASS
Informative	R1, 6% FER	R1, 3% FER	NWT	FAIL	FAIL
Informative	R1, 8% FER	R1, 6% FER	NWT	FAIL	FAIL
Informative	R1, 6% BFER	R1, 3% BFER	NWT	FAIL	FAIL
Informative	R1, 8% BFER	R1, 6% BFER	NWT	FAIL	FAIL
Informative	R1, 3% FER 40 ms	R1, 3% BFER	NWT	FAIL	FAIL
Informative	Enc G.722.2 @ 6.60 kbit/s → Dec G.718 INT	G.722.2 @ 6.60 kbit/s	NWT	PASS	PASS
Informative	Enc G.722.2 @ 8.85 kbit/s → Dec G.718 INT	G.722.2 @ 8.85 kbit/s	NWT	PASS	PASS
Informative	Enc R2 INT → Dec G.722.2, 3% FER	G.722.2 @ 12.65 kbit/s, 3% FER	NWT	PASS	PASS
Informative	Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 3% FER	G.722.2 @ 12.65 kbit/s, 3% FER	NWT	PASS	PASS
Informative	Enc R2 INT → Dec G.722.2 @ 12.65 kbit/s, 3% BFER	G.722.2 @ 12.65 kbit/s, 3% BFER	NWT	PASS	PASS
Informative	Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 3% BFER	G.722.2 @ 12.65 kbit/s, 3% BFER	NWT	PASS	PASS
Informative	Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 6% FER	Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 3% FER	NWT	FAIL	FAIL
Informative	Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 6% BFER	Enc G.722.2 @ 12.65 kbit/s → Dec G.718 INT, 3% BFER	NWT	FAIL	FAIL

10.2.3 Experiment 1c: Wideband clean speech (higher rates)

Experiment 1c has been run twice, once in American English (Lab A) and once in Canadian French (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband speech at higher layers, i.e. R3 at 16 kbit/s and R5 at 32 kbit/s. The codec was evaluated both, in clean channel and frame erasure conditions (random and bursty). Also, this experiment evaluated the performance of the codec in tandem with other wideband speech coding standards and its interoperability with G.722.2/AMR-WB. The test method used was ACR.

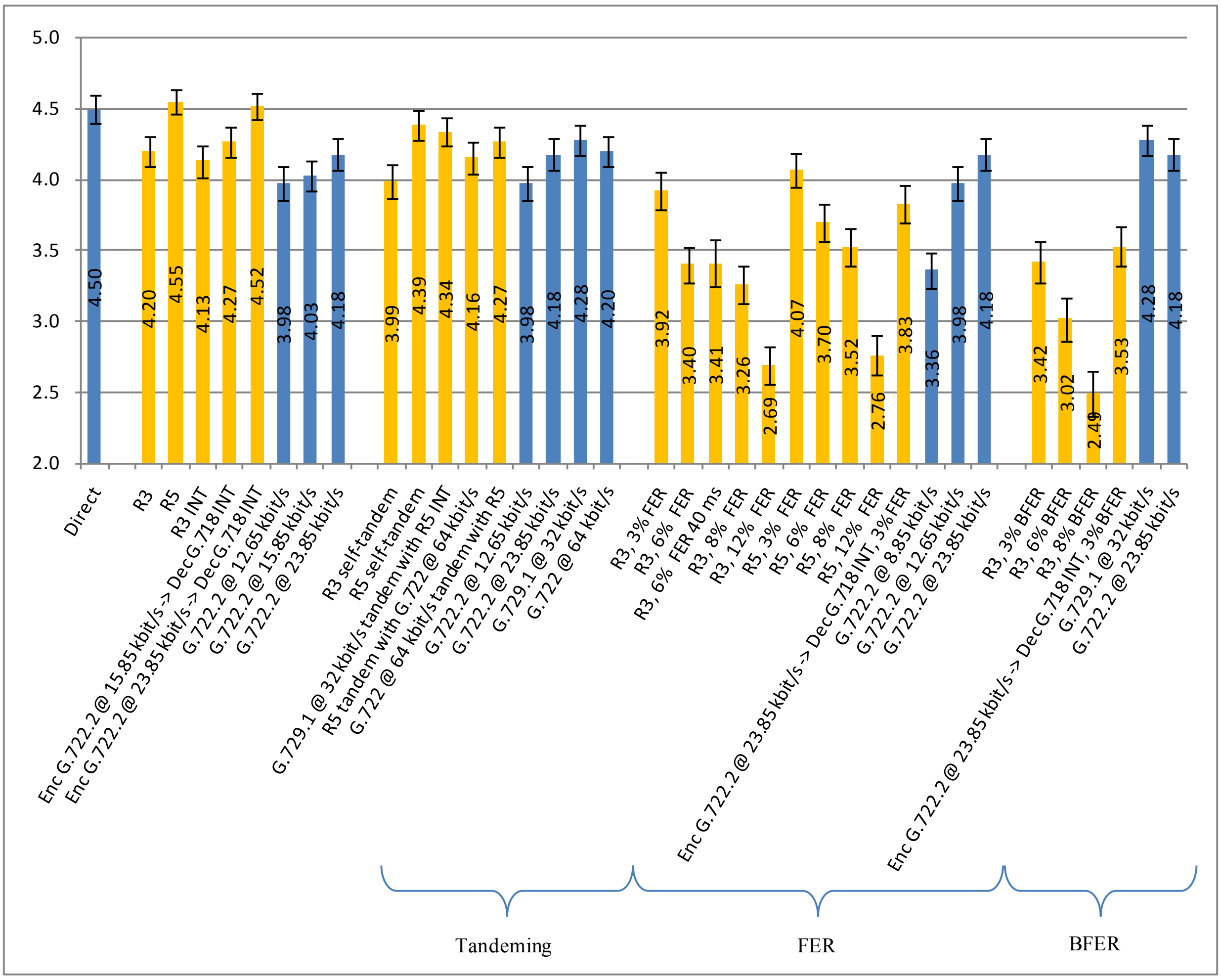

Figure 29: Experiment 1c results for Lab A, American English

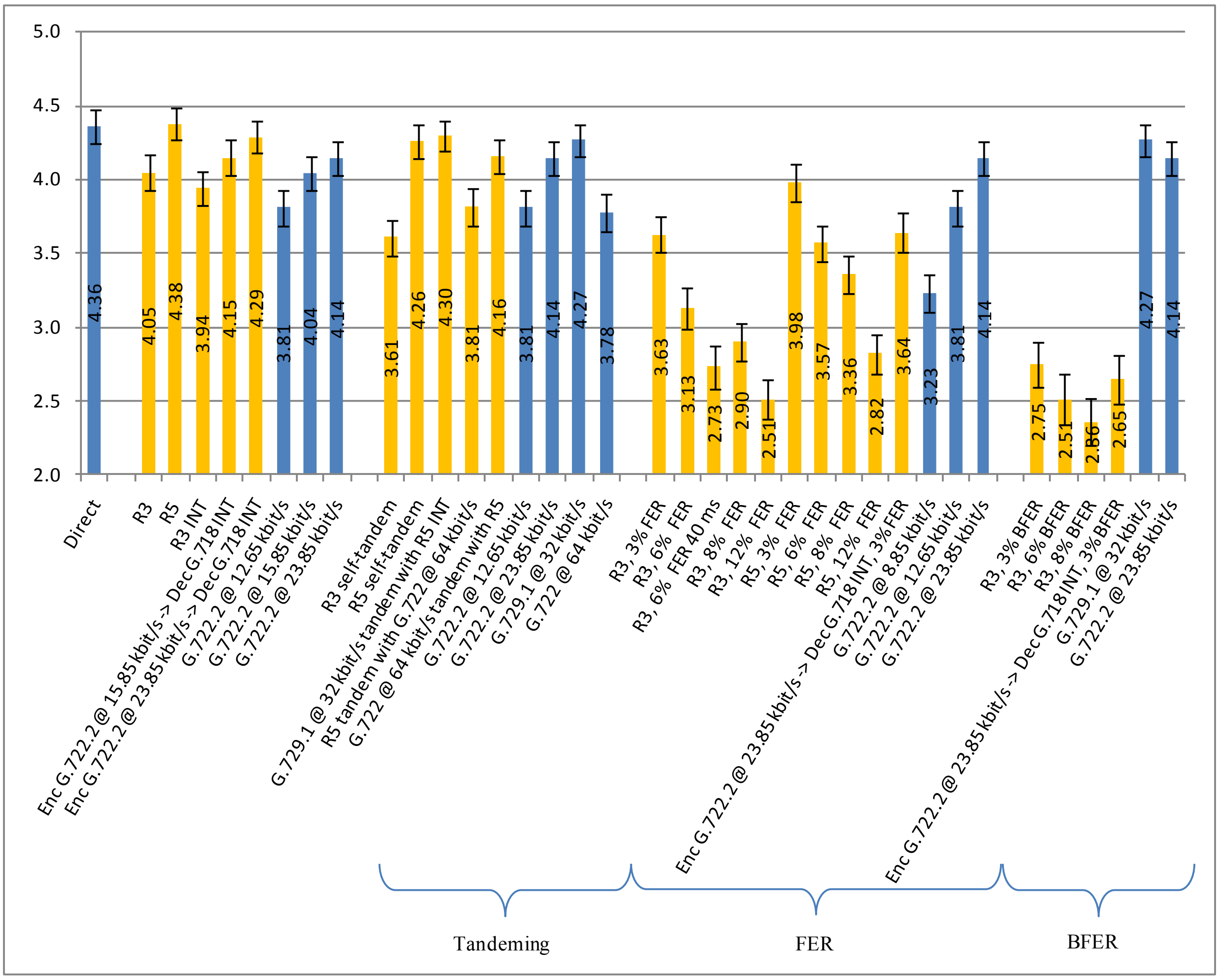

Figure 30: Experiment 1c results for Lab B, Canadian French

Table 40: Experiment 1c results for both testing laboratories

Condition	American English		Canadian French	
	MOS	SD	MOS	SD
Direct	4.50	0.69	4.36	0.78
R3	4.20	0.74	4.05	0.81
R5	4.55	0.61	4.38	0.77
R3 INT	4.13	0.76	3.94	0.80
Enc G.722.2 @ 15.85 kbit/s → Dec G.718 INT	4.27	0.77	4.15	0.82
Enc G.722.2 @ 23.85 kbit/s → Dec G.718 INT	4.52	0.64	4.29	0.74
G.722.2 @ 8.85 kbit/s	3.36	0.91	3.23	0.88
G.722.2 @ 12.65 kbit/s	3.98	0.84	3.81	0.84
G.722.2 @ 15.85 kbit/s	4.03	0.77	4.04	0.82
G.722.2 @ 23.85 kbit/s	4.18	0.77	4.14	0.80
G.729.1 @ 24 kbit/s	4.10	0.76	4.12	0.79
G.729.1 @ 32 kbit/s	4.28	0.74	4.27	0.76
G.722 @ 64 kbit/s	4.20	0.73	3.78	0.92
R3 self tandem	3.99	0.89	3.61	0.85
R5 self tandem	4.39	0.72	4.26	0.78
G729.1 @ 32 kbit/s tandem with R5 INT	4.34	0.74	4.3	0.74
R5 tandem with G.722 @ 64 kbit/s	4.16	0.81	3.81	0.89
G.722 @ 64 kbit/s tandem with R5	4.27	0.76	4.16	0.78
R3, 3% FER	3.92	0.93	3.63	0.88
R3, 6% FER	3.4	0.92	3.13	0.95
R3, 6% FER 40 ms	3.41	1.16	2.73	1.03
R3, 8% FER	3.26	0.97	2.9	0.87
R3, 12% FER	2.69	0.95	2.51	0.93
R5, 3% FER	4.07	0.84	3.98	0.87
R5, 6% FER	3.7	0.93	3.57	0.87
R5, 8% FER	3.52	0.94	3.36	0.88
R5, 12% FER	2.76	0.98	2.82	0.92
Enc G.722.2 @ 23.85 kbit/s → Dec G.718 INT, 3% FER	3.83	0.94	3.64	0.95
R3, 3% BFER	3.42	1.03	2.75	1.06
R3, 6% BFER	3.02	1.09	2.51	1.22
R3, 8% BFER	2.49	1.09	2.36	1.15
Enc G.722.2 @ 23.85 kbit/s → Dec G.718 INT, 3% BFER	3.53	0.97	2.65	1.19

Table 41: Verification against terms of reference - Experiment 1c

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R3	G.722.2 @ 12.65 kbit/s	NWT	**PASS**	**PASS**
	R5	G.722.2 @ 23.85 kbit/s	NWT	**PASS**	**PASS**
	R3, 6% FER	G.722.2 @ 8.85 kbit/s	NWT	**PASS**	**PASS**
Obj.	R3 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R3, 8% FER	G.722.2 @ 8.85 kbit/s	NWT	PASS	FAIL
	R5, 6% FER	G.722.2 @ 12.65 kbit/s	NWT	FAIL	FAIL
Informative	R3 self tandem	G.722.2 @ 12.65 kbit/s	NWT	FAIL	FAIL
	R5 self tandem	G.722.2 @ 23.85 kbit/s	NWT	PASS	PASS
	R5 tandem with G.722 @ 64 kbit/s	G.722 @ 64 kbit/s	NWT	PASS	PASS
	G.722 @ 64 kbit/s tandem with R5	G.722 @ 64 kbit/s	NWT	PASS	PASS
	G.729.1 @ 32 kbit/s tandem with R5 INT	G.729.1 @ 32 kbit/s	NWT	PASS	PASS
	Enc G.722.2 @ 23.85 kbit/s → Dec G.718 INT	G.722.2 @ 23.85 kbit/s	NWT	PASS	PASS
	Enc G.722.2 @ 15.85 kbit/s → Dec G.718 INT	G.722.2 @ 15.85 kbit/s	NWT	PASS	PASS
	Enc G.722.2 @ 23.85 kbit/s → Dec G.718 INT, 3% FER	G.722.2 @ 23.85 kbit/s	NWT	FAIL	FAIL
	Enc G.722.2 @ 23.85 kbit/s → Dec G.718 INT, 3% BFER	G.722.2 @ 23.85 kbit/s	NWT	FAIL	FAIL
	R3, 3% FER	R3	NWT	FAIL	FAIL
	R3, 3% BFER	R3	NWT	FAIL	FAIL
	R3, 6% FER 40 ms	R3, 6% FER	NWT	PASS	FAIL
	R3, 6% FER 40 ms	G.722.2 @ 8.85 kbit/s	NWT	PASS	FAIL
	R3, 6% BFER	G.729.1 @ 32 kbit/s	NWT	FAIL	FAIL
	R3, 8% FER	R3, 6% FER	NWT	FAIL	FAIL
	R3, 8% BFER	R3, 6% BFER	NWT	FAIL	PASS
	R3, 8% BFER	G.729.1 @ 32 kbit/s	NWT	FAIL	FAIL
	R3, 12% FER	R3, 8% FER	NWT	FAIL	FAIL
	R3, 12% FER	G.722.2 @ 8.85 kbit/s	NWT	FAIL	FAIL
	R5, 3% FER	R5	NWT	FAIL	FAIL
	R5, 8% FER	R5, 6% FER	NWT	FAIL	FAIL
	R5, 8% FER	G.722.2 @ 12.65 kbit/s	NWT	FAIL	FAIL
	R5, 12% FER	R5, 8% FER	NWT	FAIL	FAIL
	R5, 12% FER	G.722.2 @ 12.65 kbit/s	NWT	FAIL	FAIL

10.2.4 Experiment 2: Wideband music

Experiment 2 has been run in two laboratories. The purpose of this experiment was to evaluate the performance of the codec for wideband music at all rates. The codec was evaluated in clean channel conditions. Also, this experiment evaluated the interoperability of the codec with G.722.2/AMR-WB. The test method used was ACR.

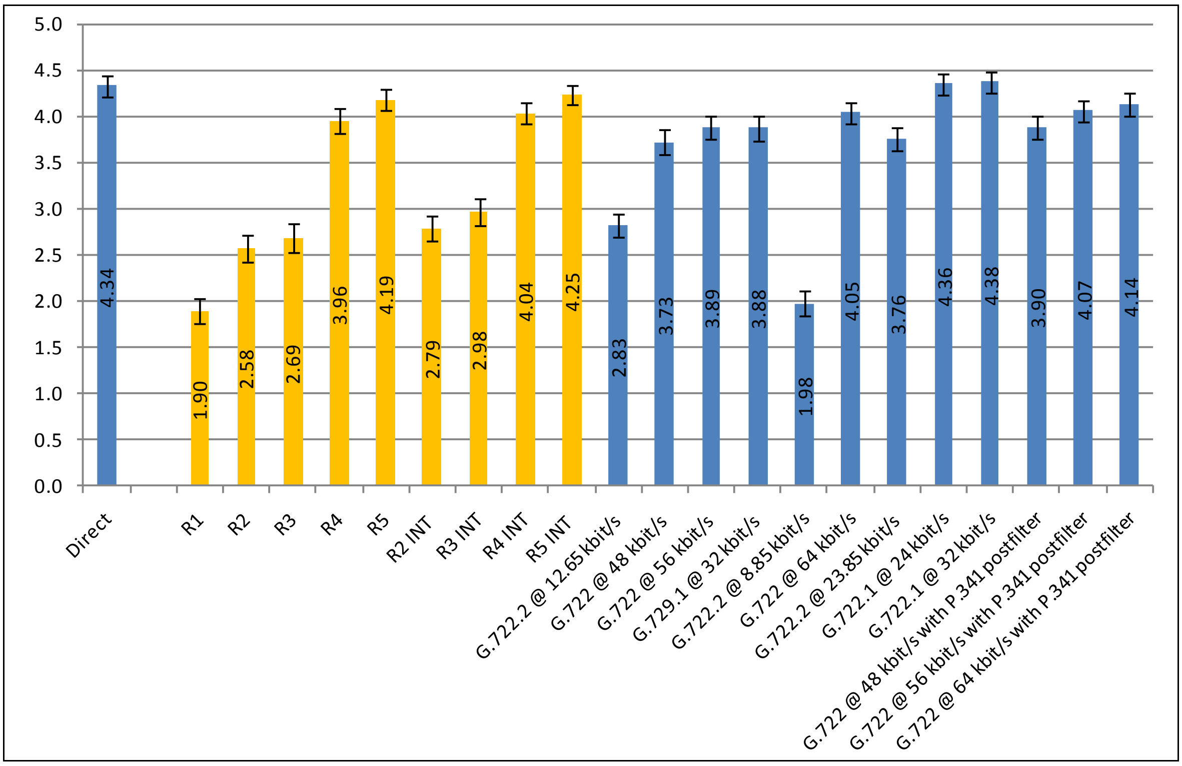

Figure 31: Experiment 2 results for Lab A

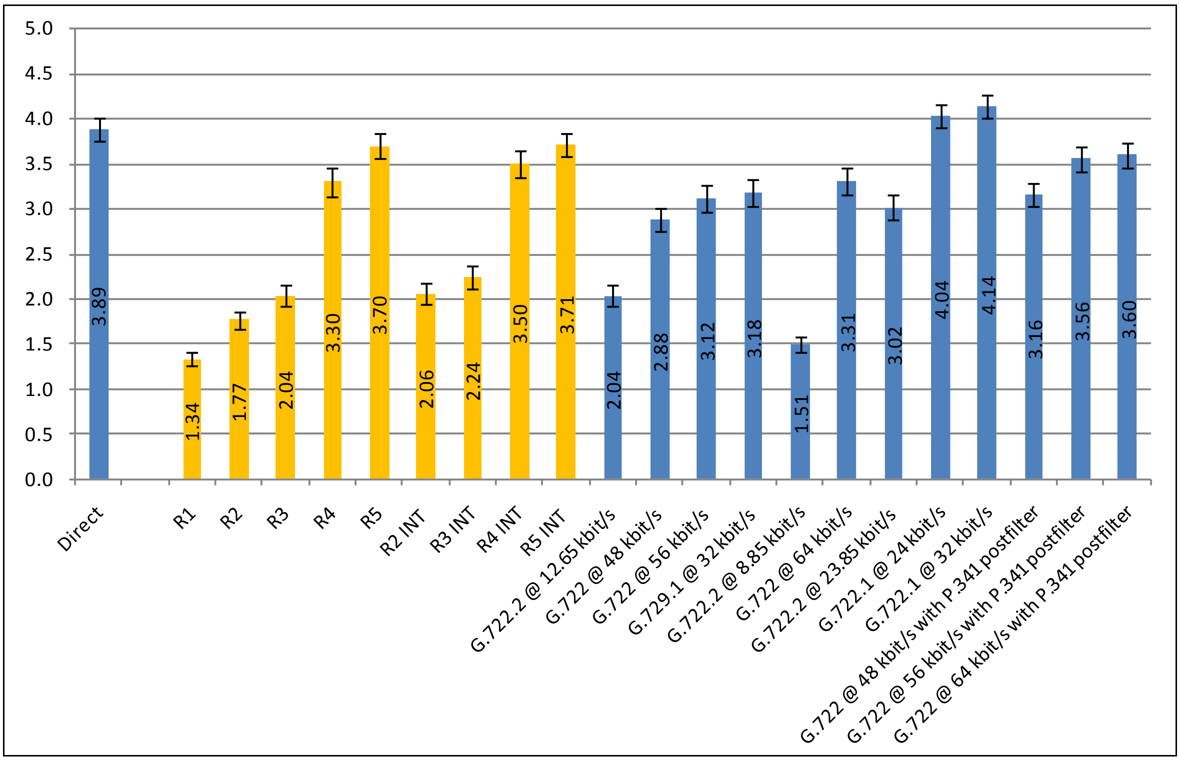

Figure 32: Experiment 2 results for Lab B

Table 42: Experiment 2results for both testing laboratories

Condition	Lab A		Lab B	
	MOS	SD	MOS	SD
Direct	4.34	0.77	3.89	0.93
R1	1.90	0.95	1.34	0.57
R2	2.58	1.01	1.77	0.71
R3	2.69	1.06	2.04	0.81
R4	3.96	0.93	3.30	1.07
R5	4.19	0.84	3.70	0.98
R2 INT	2.79	0.98	2.06	0.83
R3 INT	2.98	1.01	2.24	0.87
R4 INT	4.04	0.83	3.50	1.01
R5 INT	4.25	0.76	3.71	0.89
G.722.2 @ 12.65 kbit/s	2.83	0.94	2.04	0.83
G.722 @ 48 kbit/s	3.73	0.97	2.88	0.93
G.722 @ 56 kbit/s	3.89	0.91	3.12	1.05
G.729.1 @ 32 kbit/s	3.88	0.92	3.18	1.01
G.722.2 @ 8.85 kbit/s	1.98	0.98	1.51	0.60
G.722 @ 64 kbit/s	4.05	0.85	3.31	1.10
G.722.2 @ 23.85 kbit/s	3.76	0.90	3.02	1.00
G.722.1 @ 24 kbit/s	4.36	0.75	4.04	0.89
G.722.1 @ 32 kbit/s	4.38	0.78	4.14	0.86
G.722 @ 48 kbit/s with P.341 postfilter	3.90	0.88	3.16	0.95
G.722 @ 56 kbit/s with P.341 postfilter	4.07	0.86	3.56	0.99
G.722 @ 64 kbit/s with P.341 postfilter	4.14	0.85	3.60	1.01

Table 43: Verification against terms of reference - Experiment 2

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Requirements	R3	G.722.2 @ 12.65 kbit/s	NWT	**FAIL**	**PASS**
	R4	G.722 @ 48 kbit/s	NWT	**PASS**	**PASS**
	R4	G.722 @ 48 kbit/s with P.341 postfilter	NWT	**PASS**	**PASS**
	R5	G.722 @ 56 kbit/s	NWT	**PASS**	**PASS**
	R5	G.722 @ 56 kbit/s with P.341 postfilter	NWT	**PASS**	**PASS**
Informative	R1	G.722.2 @ 8.85 kbit/s	NWT	PASS	FAIL
	R2	G.722.2 @ 12.65 kbit/s	NWT	FAIL	FAIL
	R2 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R3 INT	G.722.2 @ 12.65 kbit/s	NWT	PASS	PASS
	R4 INT	G.722 @ 48 kbit/s	NWT	PASS	PASS
	R4 INT	G.722 @ 48 kbit/s with P.341 postfilter	NWT	PASS	PASS
	R5 INT	G.722 @ 56 kbit/s	NWT	PASS	PASS
	R5 INT	G.722 @ 56 kbit/s with P.341 postfilter	NWT	PASS	PASS

10.2.5 Experiment 3a: Narrowband noisy speech (car)

Experiment 3a has been run twice, once in Japanese (Lab A) and once in Finnish (Lab B). The purpose of this experiment was to evaluate the quality of the codec for narrowband noisy speech. One type of background noise was considered, namely car noise at 15 dB SNR. The codec was tested on two rates: R1 at 8 kbit/s and R2 at 12 kbit/s both, in clean channel and frame erasure conditions. Also, this experiment evaluated the performance of the codec in DTX operation. The test method used was DCR. The measured voice activity factors (VAF) are also reported.

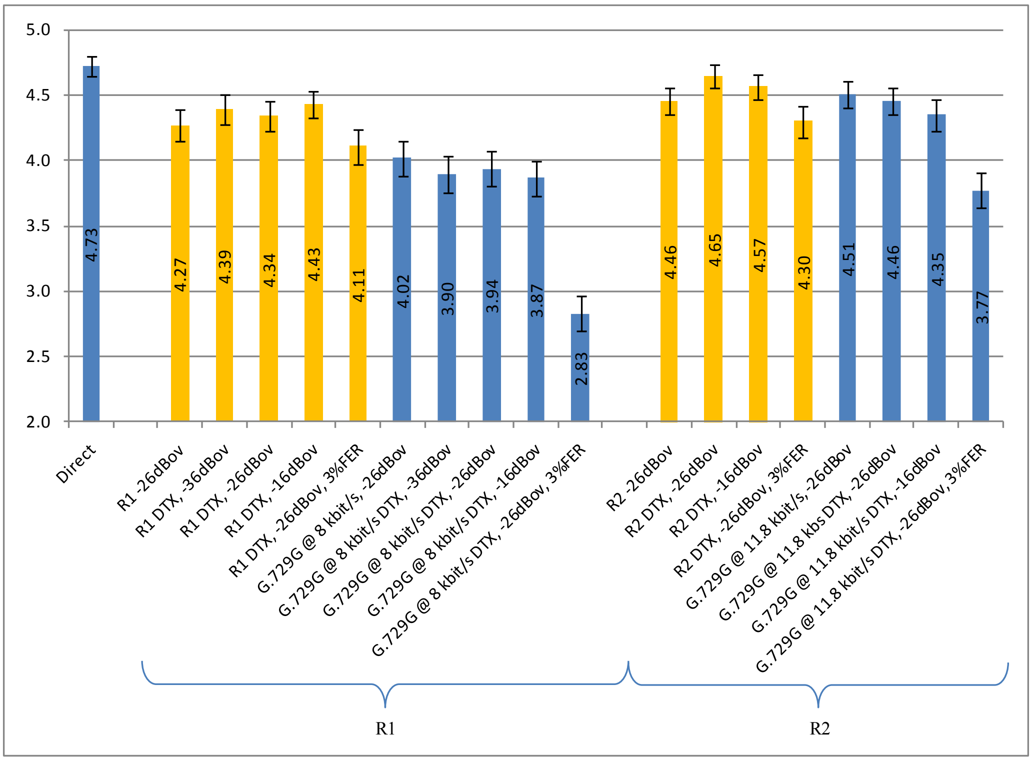

Figure 33: Experiment 3a results for Lab A, Japanese

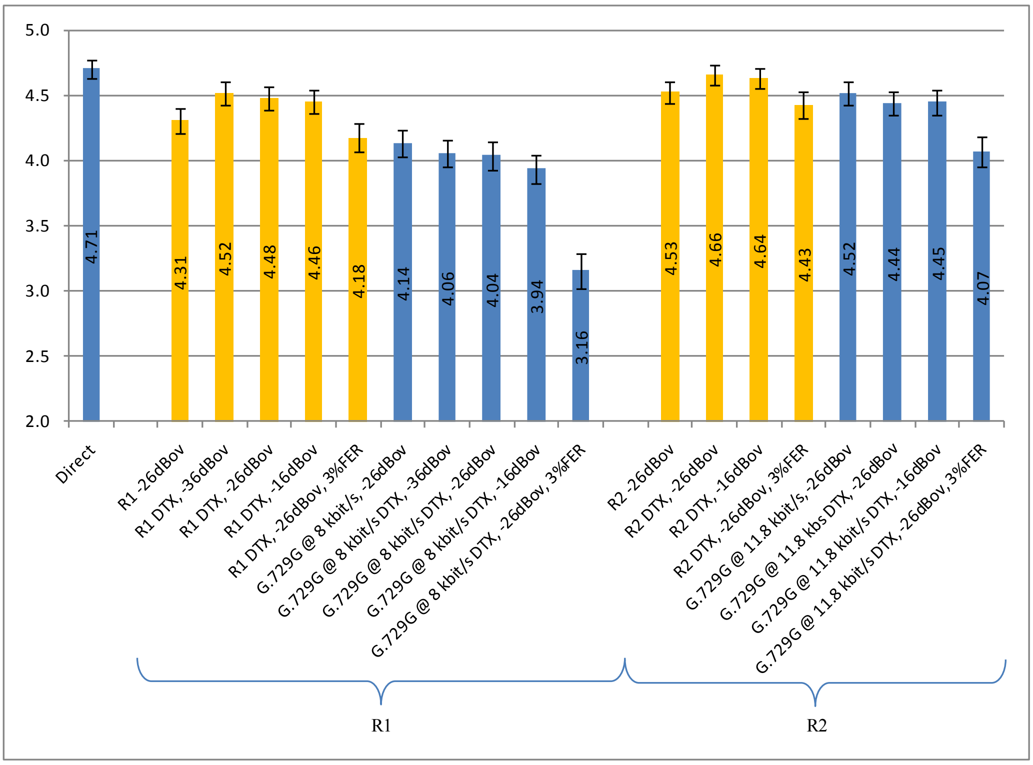

Figure 34: Experiment 3a results for Lab B, Finnish

Table 44: Experiment 3a results for both testing laboratories

Condition	Lab A			Lab B		
	MOS	SD	VAF	MOS	SD	VAF
Direct	4.73	0.55	-	4.71	0.50	-
R1, -26 dBov	4.27	0.87	-	4.31	0.68	-
R1 DTX, -36 dBov	4.39	0.80	61.8	4.52	0.63	55.9
R1 DTX, -26 dBov	4.34	0.82	62.3	4.48	0.65	56.3
R1 DTX, -16 dBov	4.43	0.75	62.7	4.46	0.64	56.5
R1 DTX, -26 dBov, 3%FER	4.11	0.94	-	4.18	0.80	-
R2, -26 dBov	4.46	0.75	-	4.53	0.61	-
R2 DTX, -26 dBov	4.65	0.63	62.3	4.66	0.55	56.3
R2 DTX, -16 dBov	4.57	0.68	62.7	4.64	0.53	56.6
R2 DTX, -26 dBov, 3%FER	4.30	0.88	-	4.43	0.70	-
G.729G @ 8 kbit/s, -26 dBov	4.02	0.95	-	4.14	0.74	-
G.729G @ 8 kbit/s DTX, -36 dBov	3.90	1.01	92.3	4.06	0.75	95.3
G.729G @ 8 kbit/s DTX, -26 dBov	3.94	0.96	93.1	4.04	0.74	96.9
G.729G @ 8 kbit/s DTX, -16 dBov	3.87	0.95	93.3	3.94	0.78	96.7
G.729G @ 8 kbit/s DTX, -26 dBov, 3%FER	2.83	0.96	-	3.16	0.95	-
G.729G @ 11.8 kbit/s, -26 dBov	4.51	0.69	-	4.52	0.62	-
G.729G @ 11.8 kbit/s DTX, -26 dBov	4.46	0.70	94.7	4.44	0.64	97.0
G.729G @ 11.8 kbit/s DTX,, -16 dBov	4.35	0.84	94.1	4.45	0.71	97.0
G.729G @ 11.8 kbit/s DTX, -26 dBov, 3%FER	3.77	0.95	-	4.07	0.83	-

Table 45: Verification against terms of reference - Experiment 3a

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R1, -26 dBov	G.729G @ 8 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1 DTX, -26 dBov	G.729G @ 8 kbit/s DTX, -26 dBov	NWT	**PASS**	**PASS**
	R2, -26 dBov	G.729G @ 11.8 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R2 DTX, -26 dBov	G.729G @ 11.8 kbit/s DTX, -26 dBov	NWT	**PASS**	**PASS**
Informative	R1 DTX, -36 dBov	G.729G @ 8 kbit/s DTX, -36 dBov	NWT	PASS	PASS
	R1 DTX, -16 dBov	G.729G @ 8 kbit/s DTX, -16 dBov	NWT	PASS	PASS
	R1 DTX, -26 dBov, 3%FER	G.729G @ 8 kbit/s DTX, -26 dBov, 3%FER	NWT	PASS	PASS
	R2 DTX, -16 dBov	G.729G @ 11.8 kbit/s DTX, -16 dBov	NWT	PASS	PASS
	R2 DTX, -26 dBov, 3%FER	G.729G @ 11.8 kbit/s DTX, -26 dBov, 3%FER	NWT	PASS	PASS

10.2.6 Experiment 3b: Narrowband noisy speech (office)

Experiment 3b has been run twice, once in Chinese (Lab A) and once in Finnish (Lab B). The purpose of this experiment was to evaluate the quality of the codec for narrowband noisy speech. One type of background noise was considered, namely office noise at 20 dB SNR. The codec was tested at two rates: R1 at 8 kbit/s and R2 at 12 kbit/s both, in clean channel and frame erasure conditions. Also, this experiment evaluated the performance of the codec in DTX operation. The test method used was DCR. The measured voice activity factors (VAF) are also reported.

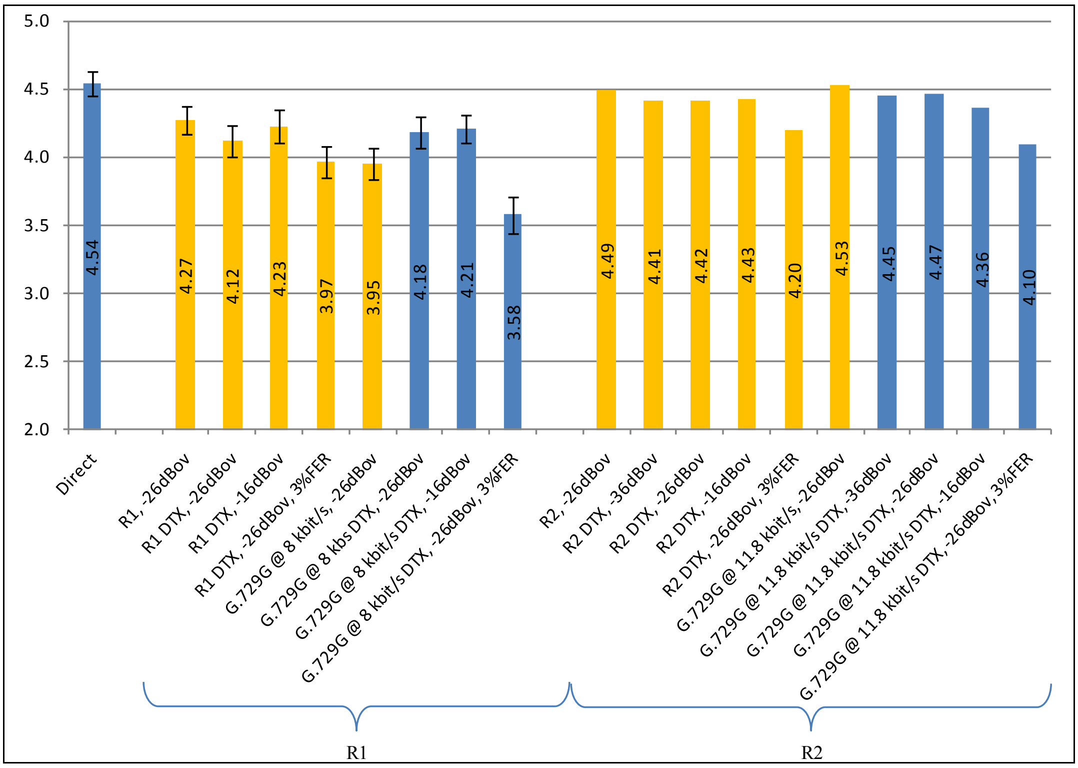

Figure 35: Experiment 3b results for Lab A, Chinese

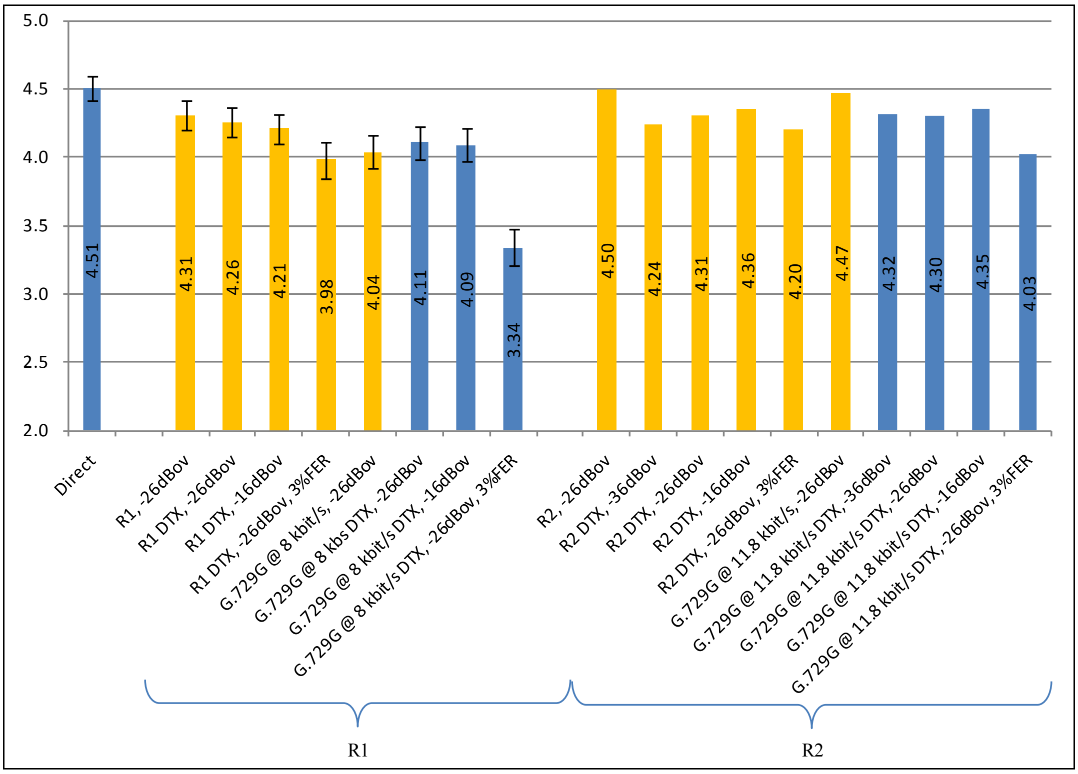

Figure 36: Experiment 3b results for Lab B, Finnish

Table 46: Experiment 3b results for both testing laboratories

Condition	Lab A			Lab B		
	MOS	SD	VAF	MOS	SD	VAF
Direct	4.54	0.63	-	4.51	0.66	-
R2, -26 dBov	4.49	0.63	-	4.50	0.69	-
R2 DTX, -36 dBov	4.41	0.75	60.3	4.24	0.76	58.7
R2 DTX, -26 dBov	4.42	0.72	63.2	4.31	0.73	60.8
R2 DTX, -16 dBov	4.43	0.77	64.6	4.36	0.74	62.3
R2 DTX, -26 dBov, 3%FER	4.20	0.75	-	4.20	0.82	-
R1, -26 dBov	4.27	0.73	-	4.31	0.78	-
R1 DTX, -26 dBov	4.12	0.85	63.2	4.26	0.75	60.8
R1 DTX, -16 dBov	4.23	0.83	64.6	4.21	0.80	62.3
R1 DTX, -26 dBov, 3%FER	3.97	0.81	-	3.98	0.94	-
G.729G @ 11.8 kbit/s, -26 dBov	4.53	0.61	-	4.47	0.71	-
G.729G @ 11.8 kbit/s DTX, -36 dBov	4.45	0.67	57.1	4.32	0.76	51.3
G.729G @ 11.8 kbit/s DTX, -26 dBov	4.47	0.71	57.9	4.30	0.84	51.2
G.729G @ 11.8 kbit/s DTX, -16 dBov	4.36	0.76	56.9	4.35	0.75	51.0
G.729G @ 11.8 kbit/s DTX, -26 dBov, 3%FER	4.10	0.80	-	4.03	0.88	-
G.729G @ 8 kbit/s, -26 dBov	3.95	0.82	-	4.04	0.85	-
G.729G @ 8 kbit/s DTX, -26 dBov	4.18	0.82	54.2	4.11	0.83	51.2
G.729G @ 8 kbit/s DTX, -16 dBov	4.21	0.75	54.0	4.09	0.86	51.0
G.729G @ 8 kbit/s DTX, -26 dBov, 3%FER	3.58	0.94	-	3.34	0.95	-

Table 47: Verification against terms of reference - Experiment 3b

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req	R1, -26 dBov	G.729G @ 8 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1 DTX, -26 dBov	G.729G @ 8 kbit/s DTX, -26 dBov	NWT	**PASS**	**PASS**
	R2, -26 dBov	G.729G @ 11.8 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R2 DTX, -26 dBov	G.729G @ 11.8 kbit/s DTX, -26 dBov	NWT	**PASS**	**PASS**
Informative	R1 DTX, -36 dBov	G.729G @ 8 kbit/s DTX, -36 dBov	NWT	PASS	PASS
	R1 DTX, -16 dBov	G.729G @ 8 kbit/s DTX, -16 dBov	NWT	PASS	PASS
	R1 DTX, -26 dBov, 3%FER	G.729G @ 8 kbit/s DTX, -26 dBov, 3%FER	NWT	PASS	PASS
	R2 DTX, -16 dBov	G.729G @ 11.8 kbit/s DTX, -16 dBov	NWT	PASS	PASS
	R2 DTX, -26 dBov, 3%FER	G.729G @ 11.8 kbit/s DTX, -26 dBov, 3%FER	NWT	PASS	PASS

10.2.7 Experiment 4a: Wideband noisy speech (car)

Experiment 4a has been run twice, once in Chinese (Lab A) and once in Swedish (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband noisy speech. One type of background noise was considered, namely car noise at 15 dB SNR. The codec was tested at various rates both, in clean channel and frame erasure conditions. Also, this experiment evaluated the performance of the codec in DTX operation and its interoperability with G.722.2/AMR-WB. The test method used was DCR. The measured voice activity factors (VAF) are also reported.

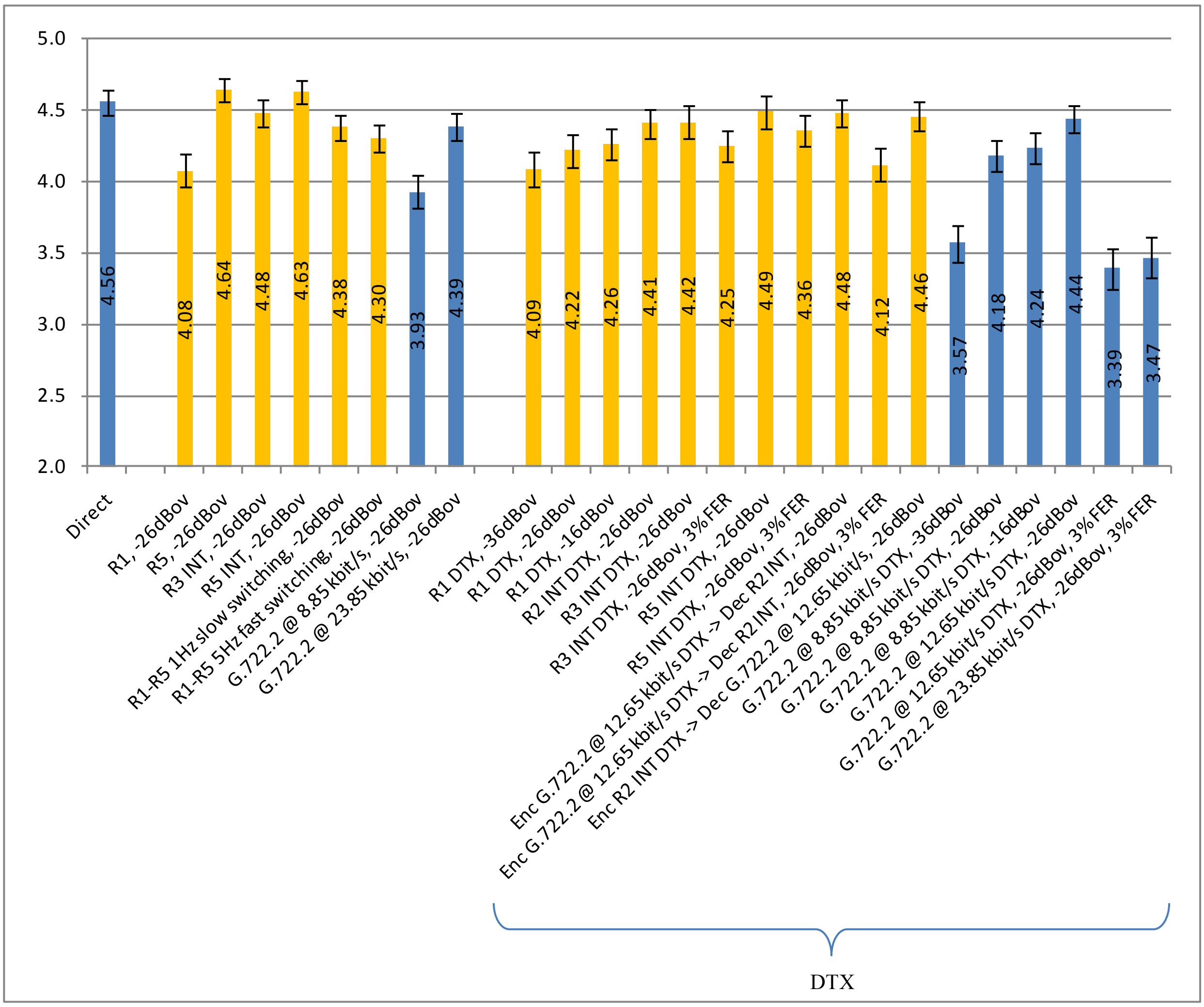

Figure 37: Experiment 4a results for Lab A, Chinese

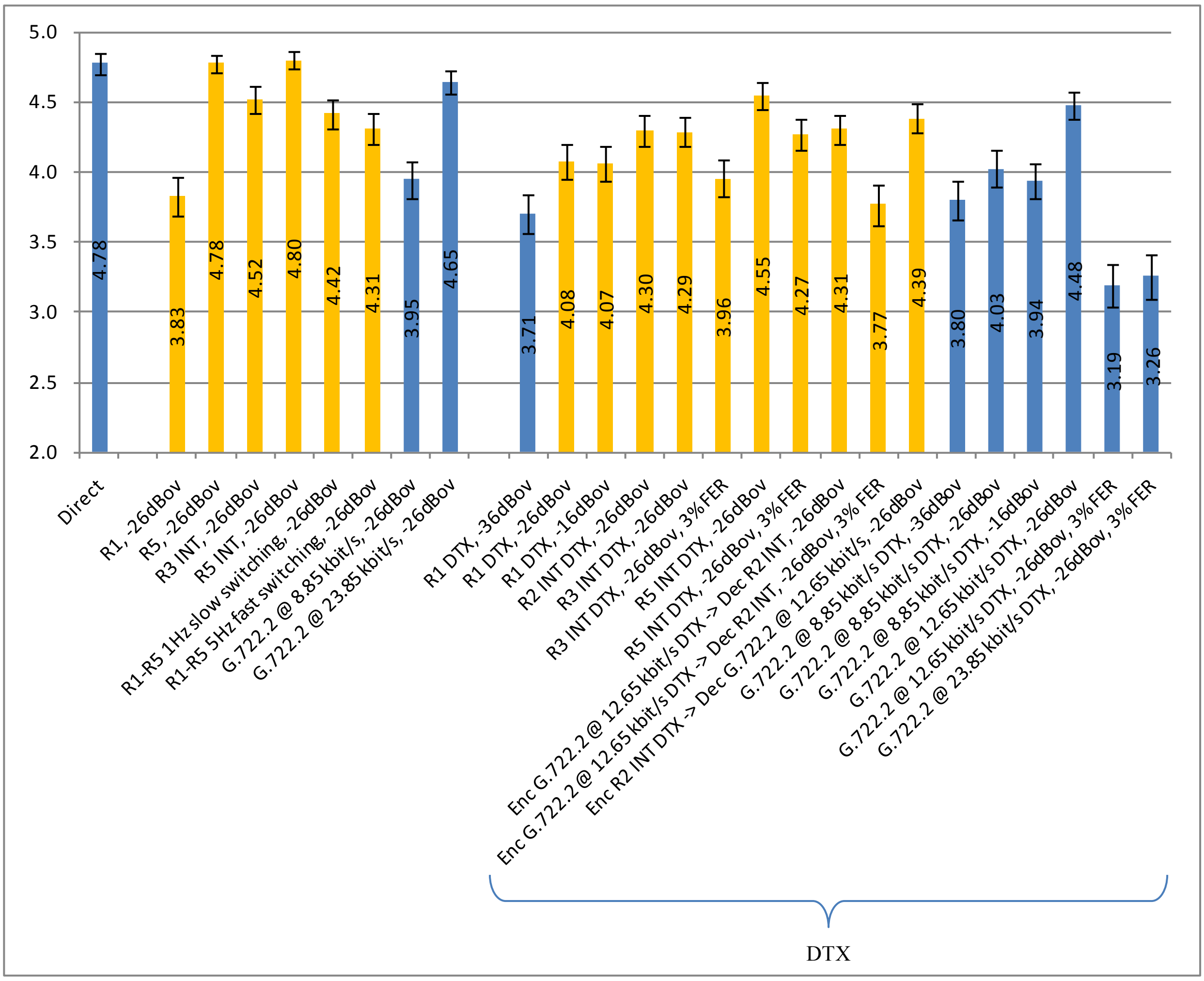

Figure 38: Experiment 4a results for Lab B, Swedish

Table 48: Experiment 4a results for both testing laboratories

Condition	Lab A			Lab B		
	MOS	SD	VAF	MOS	SD	VAF
Direct	4.56	0.63	-	4.78	0.52	-
R1, -26 dBov	4.08	0.81	-	3.83	0.98	-
R1 DTX, -36 dBov	4.09	0.88	52.6	3.71	0.98	49.5
R1 DTX, -26 dBov	4.22	0.83	53.5	4.08	0.87	50.4
R1 DTX, -16 dBov	4.26	0.76	57.8	4.07	0.89	54.6
R5, -26 dBov	4.64	0.57	-	4.78	0.47	-
R1-R5 5Hz fast switching, -26 dBov	4.30	0.67	-	4.31	0.79	-
R2 INT DTX, -26 dBov	4.41	0.73	53.5	4.30	0.77	50.4
Enc G.722.2 @ 12.65 kbit/s DTX → Dec R2 INT, -26 dBov	4.48	0.69	-	4.31	0.76	53.9
Enc G.722.2 @ 12.65 kbit/s DTX → Dec R2 INT, -26 dBov, 3% FER	4.12	0.82	-	3.77	1.01	-
Enc R2 INT DTX → Dec G.722.2 @ 12.65 kbit/s, -26 dBov	4.46	0.70	-	4.39	0.74	50.4
R3 INT DTX, -26 dBov	4.42	0.78	53.5	4.29	0.72	50.4
R3 INT, -26 dBov	4.48	0.65	-	4.52	0.65	-
R3 INT DTX, -26 dBov, 3% FER	4.25	0.80	-	3.96	0.89	-
R5 INT DTX, -26 dBov	4.49	0.81	53.5	4.55	0.65	50.4
R5 INT, -26 dBov	4.63	0.55	-	4.80	0.44	-
R5 INT DTX, -26 dBov, 3% FER	4.36	0.79	-	4.27	0.79	-
R1-R5 1Hz slow switching, -26 dBov	4.38	0.62	-	4.42	0.72	-
G.722.2 @ 8.85 kbit/s, -26 dBov	3.93	0.82	-	3.95	0.95	-
G.722.2 @ 8.85 kbit/s DTX, -36 dBov	3.57	0.91	56.4	3.80	0.98	52.0
G.722.2 @ 8.85 kbit/s DTX, -26 dBov	4.18	0.78	57.5	4.03	0.91	53.9
G.722.2 @ 8.85 kbit/s DTX, -16 dBov	4.24	0.78	58.9	3.94	0.86	54.7
G.722.2 @ 23.85 kbit/s, -26 dBov	4.39	0.65	57.5	4.65	0.59	-
G.722.2 @ 12.65 kbit/s DTX, -26 dBov	4.44	0.68	53.5	4.48	0.69	53.9
G.722.2 @ 12.65 kbit/s DTX, -26 dBov, 3% FER	3.39	1.02	-	3.19	1.10	-
G.722.2 @ 23.85 kbit/s DTX, -26 dBov, 3% FER	3.47	1.03	-	3.26	1.13	-

Table 49: Verification against terms of reference - Experiment 4a

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R1, -26 dBov	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1 DTX, -26 dBov	G.722.2 @ 8.85 kbit/s DTX, -26 dBov	NWT	**PASS**	**PASS**
	R5, -26 dBov	G.722.2 @ 23.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
Objectives	R2 INT DTX, -26 dBov	G.722.2 @ 12.65 kbit/s DTX, -26 dBov	NWT	PASS	FAIL
	Enc G.722.2 @ 12.65 kbit/s DTX → Dec R2 INT, -26 dBov	G.722.2 @ 12.65 kbit/s DTX, -26 dBov	NWT	PASS	FAIL
	Enc R2 INT DTX → Dec G.722.2 @ 12.65 kbit/s, -26 dBov	G.722.2 @ 12.65 kbit/s DTX, -26 dBov	NWT	PASS	PASS
	R3 INT, -26 dBov	G.722.2 @ 12.65 kbit/s DTX, -26 dBov	NWT	PASS	PASS
	R3 INT DTX, -26 dBov	G.722.2 @ 12.65 kbit/s DTX, -26 dBov	NWT	PASS	FAIL
	R5 INT, -26 dBov	G.722.2 @ 23.85 kbit/s, -26 dBov	NWT	PASS	PASS
	R5 INT DTX, -26 dBov	G.722.2 @ 23.85 kbit/s, -26 dBov	NWT	PASS	FAIL
Inf.	R1 DTX, -36 dBov	G.722.2 @ 8.85 kbit/s DTX, -36 dBov	NWT	PASS	PASS
	R1 DTX, -16 dBov	G.722.2 @ 8.85 kbit/s DTX, -16 dBov	NWT	PASS	PASS
	Enc G.722.2 @ 12.65 kbit/s DTX → Dec R2 INT, -26 dBov, 3% FER	G.722.2 @ 12.65 kbit/s DTX, -26 dBov, 3% FER	NWT	PASS	PASS
	R3 INT DTX, -26 dBov, 3% FER	G.722.2 @ 12.65 kbit/s DTX, -26 dBov, 3% FER	NWT	PASS	PASS
	R5 INT DTX, -26 dBov, 3% FER	G.722.2 @ 23.85 kbit/s DTX, -26 dBov, 3% FER	NWT	PASS	PASS
	R1-R5 5Hz fast switching, -26 dBov	G.722.2 @ 23.85 kbit/s, -26 dBov	NWT	PASS	FAIL
	R1-R5 1Hz slow switching, -26 dBov	G.722.2 @ 23.85 kbit/s, -26 dBov	NWT	PASS	FAIL

10.2.8 Experiment 4b: Wideband noisy speech (street)

Experiment 4b has been run twice, once in French (Lab A) and once in Canadian English (Lab B). The purpose of this experiment was to evaluate the quality of the codec for wideband noisy speech. One type of background noise was considered, namely street noise at 20 dB SNR. The codec was tested at various rates. Also, this experiment evaluated the performance of the codec in DTX operation and the noise reduction algorithm. The test method used was the one in ITU-T Recommendation P.835. The measured voice activity factors (VAF) are also reported.

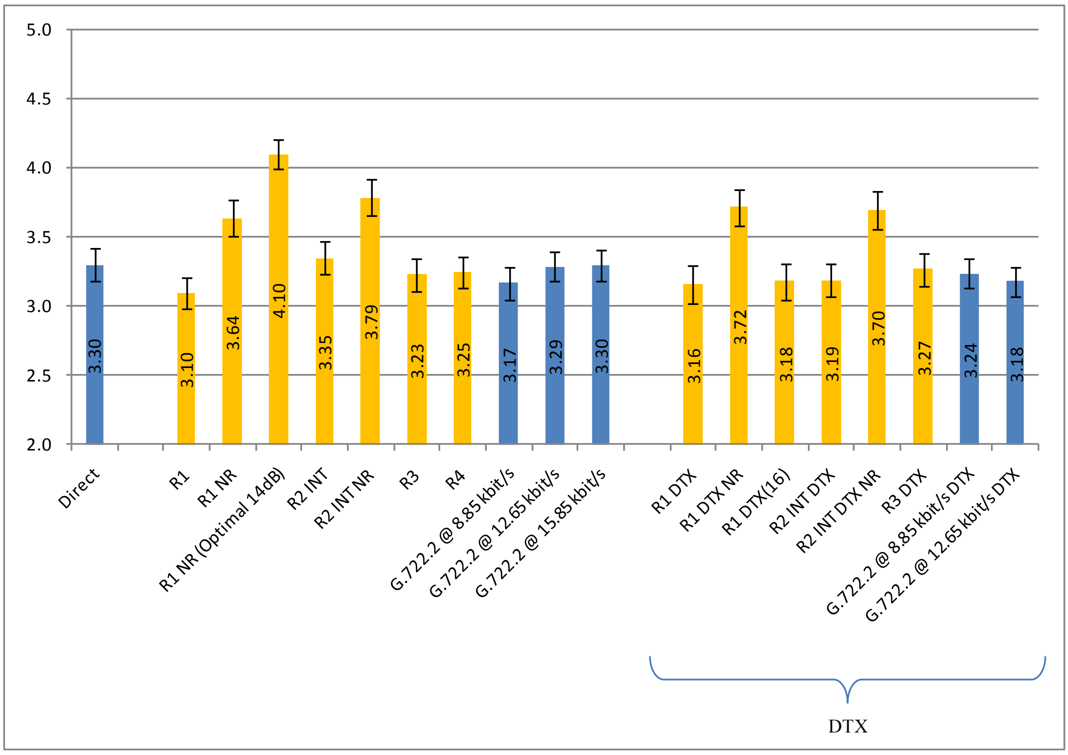

Figure 39: Experiment 4b results for Lab A, French

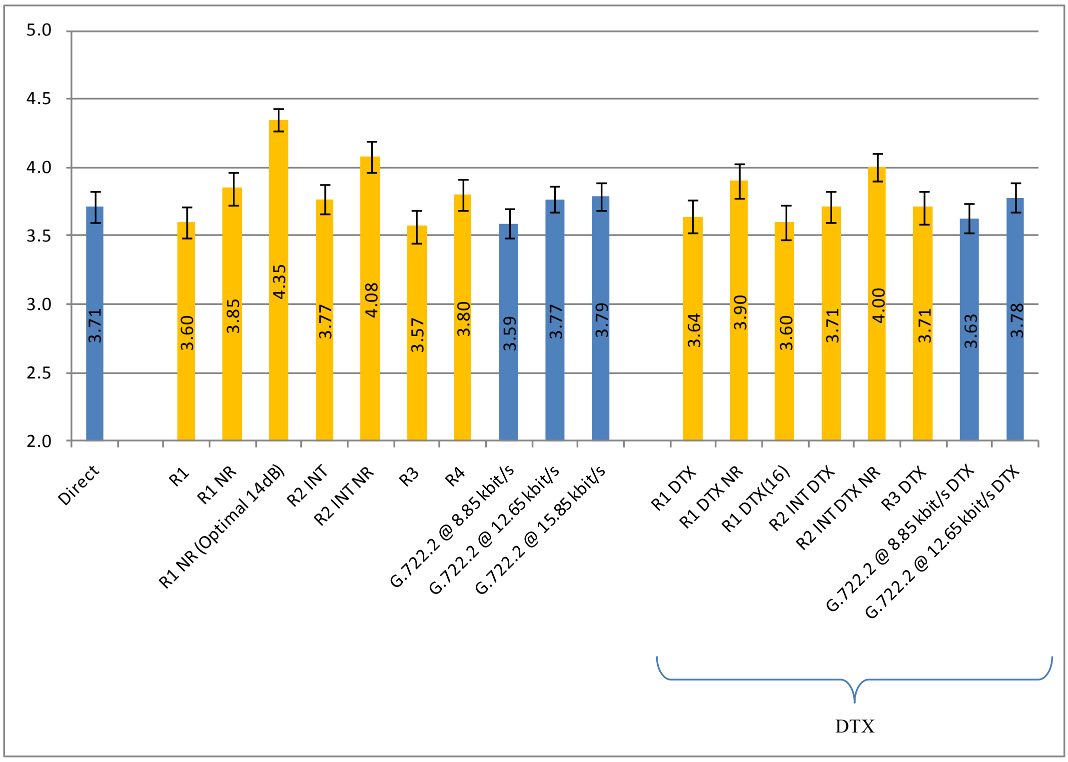

Figure 40: Experiment 4b results for Lab B, Canadian English

Table 50: Experiment 4b results for both testing laboratories

Condition	Lab A			Lab B		
	MOS	SD	VAF	MOS	SD	VAF
Direct	3.30	0.85	-	3.71	0.79	-
R1, -26 dBov	3.10	0.80	-	3.60	0.78	-
R1 DTX, -26 dBov	3.16	0.96	53.7	3.64	0.84	65.0
R1 NR, -26 dBov	3.64	0.94	-	3.85	0.84	-
R1 DTX NR, -26 dBov	3.72	0.93	53.8	3.90	0.87	64.8
R1 DTX(16), -26 dBov	3.18	0.95	53.7	3.60	0.92	65.0
R1 NR (Optimal 14dB), -26 dBov	4.10	0.76	-	4.35	0.60	-
R2 INT, -26 dBov	3.35	0.84	-	3.77	0.75	-
R2 INT DTX, -26 dBov	3.19	0.84	53.6	3.71	0.80	64.7
R2 INT NR, -26 dBov	3.79	0.93	-	4.08	0.77	-
R2 INT DTX NR, -26 dBov	3.70	0.97	53.8	4.00	0.73	64.8
R3, -26 dBov	3.23	0.84	-	3.57	0.82	-
R3 DTX, -26 dBov	3.27	0.85	53.7	3.71	0.84	65.0
R4, -26 dBov	3.25	0.77	-	3.80	0.78	-
G.722.2 @ 8.85 kbit/s, -26 dBov	3.17	0.84	-	3.59	0.77	-
G.722.2 @ 8.85 kbit/s DTX, -26 dBov	3.24	0.78	63.0	3.63	0.79	75.3
G.722.2 @ 12.65 kbit/s DTX, -26 dBov	3.18	0.77	63.0	3.78	0.76	75.3
G.722.2 @ 15.85 kbit/s, -26 dBov	3.30	0.78	-	3.79	0.73	-
G.722.2 @ 12.65 kbit/s, -26 dBov	3.29	0.75	-	3.77	0.68	-

Table 51: Verification against terms of reference - Experiment 4b

	Test Condition	Reference Condition	Criterion	Result Lab A	Result Lab B
Req.	R1, -26 dBov	G.722.2 @ 8.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
	R1 DTX, -26 dBov	G.722.2 @ 8.85 kbit/s DTX, -26 dBov	NWT	**PASS**	**PASS**
	R4, -26 dBov	G.722.2 @ 15.85 kbit/s, -26 dBov	NWT	**PASS**	**PASS**
Obj.	R2 INT, -26 dBov	G.722.2 @ 12.65 kbit/s, -26 dBov	NWT	PASS	PASS
	R2 INT DTX, -26 dBov	G.722.2 @ 12.65 kbit/s DTX, -26 dBov	NWT	PASS	PASS
	R3, -26 dBov	G.722.2 @ 15.85 kbit/s, -26 dBov	NWT	PASS	FAIL
Informative	R1 NR, -26 dBov	R1, -26 dBov	NWT	PASS	PASS
	R1 NR, -26 dBov	R1 NR (Optimal 14 dB), -26 dBov	NWT	FAIL	FAIL
	R1 DTX NR, -26 dBov	R1 NR, -26 dBov	NWT	PASS	PASS
	R1 DTX(16), -26 dBov	R1 DTX, -26 dBov	NWT	PASS	PASS
	R2 INT NR, -26 dBov	R2 INT, -26 dBov	NWT	PASS	PASS
	R2 INT DTX NR, -26 dBov	R2 INT NR, -26 dBov	NWT	PASS	PASS
	R3 DTX, -26 dBov	R3, -26 dBov	NWT	PASS	PASS

10.3 Summary of non bit-exact corrections to the source code after Step 2

10.3.1 Missing initializations in transition coding (TC mode)

Description: In the excitation coding of the Transition mode (TC mode), several variables were uninitialized. Although the uninitialized variables were always multiplied by zero in the subsequent processing, this problem could result in instability on some platforms. See [29] for more information.

Performance Impact: None.

10.3.2 Missing initialization of the correlation vector in the algebraic codebook search

Description: In general, the algebraic codebook search benefits from the joint optimization approach where a correlation vector between the target vector and the impulse response is computed. However, in some cases of the excitation search in the transition mode (TC mode), the joint optimization was not used, and the correlation vector was not calculated. However, the uninitialized correlation vector was then used for further processing. While it was multiplied exclusively by zero, this problem could result in codec instability on some platforms. For more information, see [29].

Performance Impact: None.

10.3.3 Out-of-memory access in narrowband post filter

Description: The narrowband post filter takes as input argument the fractional pitch lag value. In very rare occasions of long pitch lags, it may happen that this value is 231.5. Inside the narrowband post filter, this pitch lag value is rounded to 232. While searching for pitch delay around this value, the maximum pitch lag of 234 samples is evaluated. However, the signal array has only 233 elements, and out of memory access occurs. For more information, see [29].

Performance Impact: This problem did not affect wideband outputs. For narrowband outputs, no perceptual impact was observed.

10.3.4 Possible saturation in the autocorrelation function

Description: In the fixed point implementation, in case of narrowband input, and only for inactive input (voice activity detector indicates inactive signal), a normalization operation on the correlation vector of the LP filter may lead to a saturation in case of very low-level frames. For more information, see [29].

Perceptual impact: An occurrence of this problem is highly unlikely. However, when the problem occurs, it causes highly annoying saturation. The problem did not manifest itself during the Characterization Tests, Step 2.

11 Objective evaluation of the floating point implementation

The objective scores using the P.862 PESQ [31] / P862.2 WB-PESQ [32] algorithm have been computed for each sentence-pair/condition following the Characterization Tests, Step 1. The scores have been computed for the reference configuration of fixed-point encoder/fixed-point-decoder, and for each of the tested configurations: fixed-point encoder/floating-point decoder, floating-point encoder/ fixed-point decoder and floating-point encoder/floating-point decoder. Also, difference values between the tested configurations and the reference configuration have been computed for all sentence-pairs/conditions. The difference values and the correlations have been further computed per condition. Finally, the correlation has been computed also per experiment.

It was observed that overall the tested configurations perform similar as the reference configuration. In clean speech, clean channel conditions, i.e. in conditions where the P.862.2 scores are expected to be most reliable, the floating point implementation seemed to give slightly higher scores. It can be seen that while the difference value between the reference and the tested configuration is several times lower than -0.2 (scenario where the tested configuration performs better), it is never higher than 0.2 (scenario where the reference is better). The detailed report on the floating point implementation validation can be found in [30].