

International Telecommunication Union

ITU-T

Technical Paper

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(30 July 2010)

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

GSTP-ACP1

**Selection Test Results for G.718 Baseline and
Qualification Phase Test Results for G.729.1**

ITU-T

Summary

This technical paper presents the Selection Test Results of ITU-T G.EV-VBR (later G.718) Baseline and Qualification Test Results for G.729EV (later G.729.1) speech and audio codecs. Information on algorithmic complexity, memory requirements and algorithmic delay of the candidates is collected as well. The purpose of this Technical Paper is to allow interested parties finding the corresponding test results at a conveniently accessible place.

Change Log

This document contains Version 1 of the ITU-T Technical Paper on “*Selection Test Results for G.718 Baseline and Qualification Phase Test Results for G.729.1*” approved at the ITU-T Study Group 16 meeting held in Geneva, 19-30 July 2010.

Editor: Imre Varga
Qualcomm
USA

Tel: + 49 89 614 694 0015
Fax: + 49 89 614 694 0001
Email: ivarga@qualcomm.com

Contents

	Page
1 SCOPE	1
2 REFERENCES.....	1
3 ABBREVIATIONS	3
4 SCOPE OF G.718 (G.EV-VBR).....	3
5 G.EV-VBR CANDIDATES IN SELECTION PHASE: ALGORITHMIC OVERVIEW, DELAY AND COMPLEXITY.....	5
5.1 ERICSSON, MOTOROLA, TEXAS INSTRUMENTS CANDIDATE [16]	5
5.1.1 Encoder Overview.....	6
5.1.2 Decoder Overview	8
5.1.3 Complexity	9
5.2 HUAWEI CANDIDATE [17]	10
5.2.1 Description of the encoder.....	10
5.2.2 Frame erasure concealment	12
5.2.3 Description of the decoder.....	12
5.2.4 Frame Format.....	13
5.2.5 Algorithmic delay.....	14
5.2.6 Effective bandwidth.....	14
5.3 PANASONIC CANDIDATE [18].....	16
5.3.1 Encoder description	16
5.3.2 CELP module.....	18
5.3.3 BWE module	19
5.3.4 TRC module	20
5.3.5 Decoder description.....	20
5.3.6 The case of receiving no bits.....	21
5.3.7 The case of receiving R1 bits	21
5.3.8 The case of receiving R1 to R2 bits.....	21
5.3.9 The case of receiving R1 to R3 or the higher layer bits	21
5.3.10 Algorithmic delay.....	22
5.3.11 Complexity	22
5.4 NOKIA AND VOICEAGE CANDIDATE [19].....	22
5.4.1 Layer Structure	23
5.4.2 Encoder Overview.....	24
5.4.3 Classification Based Core Layer (Layer 1)	24
5.4.4 Second Stage ACELP encoding (Layer 2)	25
5.4.5 Frame Erasure Concealment Side Information (Layer 3)	25
5.4.6 Transform Coding of Higher Layers (Layers 3, 4, 5)	26
5.4.7 Decoder Overview	27
5.4.8 Complexity evaluation	27
6 TEST RESULTS IN G.EV-VBR SELECTION PHASE	28
6.1 GLOBAL ANALYSES FOR THE SELECTION PHASE FOR THE EMBEDDED-VBR SPEECH CODEC [21].....	28
6.1.1 Organization of the Selection Test.....	28
6.1.2 The Global Analysis Laboratory.....	29
6.1.3 Test Results	29
6.1.4 Requirements in Terms of Reference	36
6.1.5 Weighted Requirement ToR Passes.....	39
6.2 FIGURE OF MERIT.....	43
6.3 OBJECTIVE TERMS OF REFERENCE	45
6.4 SUMMARY ON G.EV-VBR SELECTION TEST RESULTS.....	46
6.5 EV-VBR COMPUTATION ON FREQUENCY ATTENUATION	47
7 SCOPE OF G.729.1 (G.729EV).....	54

8	G.729EV CANDIDATES: ALGORITHMIC OVERVIEW, DELAY AND COMPLEXITY FIGURES.....	54
8.1	FRANCE TELECOM CANDIDATE [10]	54
8.1.1	<i>Description of the encoder part</i>	55
8.1.2	<i>Description of the decoder part</i>	59
8.1.3	<i>Frame erasure concealment</i>	61
8.1.4	<i>Algorithmic delay</i>	62
8.1.5	<i>Complexity evaluation</i>	62
8.1.6	<i>Frequency response</i>	62
8.2	ETRI CANDIDATE [11], [12]	65
8.2.1	<i>Encoder description</i>	65
8.2.2	<i>Decoder description</i>	68
8.2.3	<i>Frame Format</i>	68
8.2.4	<i>Algorithmic delay</i>	69
8.2.5	<i>Effective bandwidth</i>	69
8.2.6	<i>Complexity and memory</i>	72
8.3	SAMSUNG CANDIDATE [13].....	72
8.3.1	<i>Main feature</i>	72
8.3.2	<i>Frame structure</i>	73
8.3.3	<i>Block diagram of encoder</i>	74
8.3.4	<i>Block diagram of decoder</i>	74
8.3.5	<i>Bit rate granularity</i>	75
8.3.6	<i>Algorithm delay</i>	75
8.3.7	<i>Effective bandwidth at all supported bit rates (Codec frequency response)</i>	75
8.3.8	<i>Complexity evaluation</i>	76
8.4	VOICEAGE CANDIDATE [14]	76
8.4.1	<i>Coding paradigm</i>	76
8.4.2	<i>Encoder description</i>	76
8.4.3	<i>Decoder description</i>	77
8.4.4	<i>Algorithmic delay</i>	79
8.4.5	<i>Effective bandwidth at supported bit rates</i>	79
8.4.6	<i>Frame structure</i>	79
8.4.7	<i>Complexity</i>	80
8.5	MATSUSHITA, MINDSPEED, SIEMENS CANDIDATE [15].....	80
8.5.1	<i>High Level Description</i>	80
8.5.2	<i>Frame size, lookahead, and delay</i>	81
8.5.3	<i>Core part</i>	82
8.5.4	<i>The 12 kbit/s enhancement layer</i>	82
8.5.5	<i>The 14 kbit/s layer</i>	82
8.5.6	<i>The 32 kbit/s layer</i>	82
8.5.7	<i>Pre echo Reduction and post processing schemes</i>	83
8.5.8	<i>Bit Rate Granularity</i>	83
8.5.9	<i>Complexity</i>	83
8.5.10	<i>Effective Bandwidth</i>	83
9	TEST RESULTS IN THE QUALIFICATION PHASE OF G729EV	85
9.1	EXPERIMENTS 1-4	85
9.1.1	<i>Test results – Experiment 1a</i>	86
9.1.2	<i>Test Results – Experiment 1b</i>	87
9.1.3	<i>Test Results – Experiment 2a</i>	88
9.1.4	<i>Test Results – Experiment 2b</i>	88
9.1.5	<i>Test Results – Experiment 2c</i>	89
9.1.6	<i>Test Results – Experiment 3a</i>	89
9.1.7	<i>Test Results – Experiment 3b</i>	90
9.1.8	<i>Test Results – Experiment 3c</i>	90
9.1.9	<i>Test Results – Experiment 4</i>	90
9.1.10	<i>Test Results Summary – Codec Comparison and MOS Analysis of the Candidates</i>	91
9.2	EXPERIMENT 5 (CLEAN SPEECH; WIDE BAND CASE, BIT RATE GRANULARITY).....	102
9.2.1	<i>Test Organization</i>	102
9.2.2	<i>Test Results – Experiment 5</i>	103
9.3	FREQUENCY RESPONSES OF CANDIDATES	108

List of Tables

	Page
TABLE 1: BIT ALLOCATION IN R1 AND L2	7
TABLE 2: COMPUTATIONAL COMPLEXITY OF THE CODEC	22
TABLE 3: MEMORY REQUIREMENTS FOR THE CODEC	22
TABLE 4: LAYER STRUCTURE FOR DEFAULT OPERATION	23
TABLE 5: ORGANIZATION OF THE SELECTION TEST.....	28
TABLE 6: ACR EXPERIMENT 1 – CLEAN NARROWBAND SPEECH ON R1 AND R2	30
TABLE 7: ACR EXPERIMENT 2A - CLEAN WIDEBAND SPEECH ON R1 AND R2	31
TABLE 8: ACR EXPERIMENT 2B - CLEAN WIDEBAND SPEECH ON R3 AND R4	32
TABLE 9: ACR EXPERIMENT 2C - CLEAN WIDEBAND SPEECH ON R5.....	33
TABLE 10: ACR EXPERIMENT 3 - MUSIC.....	34
TABLE 11: DCR EXPERIMENT 4 – NARROWBAND SPEECH WITH CAR NOISE	34
TABLE 12: DCR EXPERIMENT 5 – NARROWBAND SPEECH WITH STREET NOISE	35
TABLE 13: DCR EXPERIMENT 6 –WIDEBAND SPEECH WITH INTERFERING TALKER NOISE	35
TABLE 14: DCR EXPERIMENT 7 –WIDEBAND SPEECH WITH OFFICE NOISE	36
TABLE 15: RESULTS OF ToR TESTS FOR INPUT LEVEL CONDITIONS	37
TABLE 16: RESULTS OF ToR TESTS FOR ERROR CONDITIONS.....	38
TABLE 17: RESULTS OF ToR TESTS FOR MUSIC CONDITIONS	38
TABLE 18: RESULTS OF ToR TESTS FOR NOISE CONDITIONS.....	39
TABLE 19: NUMBER OF ToR FAILURES BY CONDITION CATEGORY	39
TABLE 20: ToR WEIGHTS	40
TABLE 21: WEIGHTED NUMBER OF PASSES.....	40
TABLE 22: WEIGHTED NUMBER OF PASSES FOR INPUT LEVEL ToRS	41
TABLE 23: WEIGHTED NUMBER OF PASSES FOR ERROR ToRS	42
TABLE 24: WEIGHTED NUMBER OF PASSES FOR MUSIC+NOISE ToRS	43
TABLE 25: FIGURE OF MERIT – WEIGHTED AVERAGE QUALITY SCORES	44
TABLE 26: FIGURE OF MERIT BY ToR CATEGORY.....	45
TABLE 27: NUMBER OF OBJECTIVE ToR FAILURES BY CODEC.....	45
TABLE 28: OBJECTIVE ToRS FOR INPUT LEVEL CONDITIONS.....	45
TABLE 29: OBJECTIVE ToRS FOR MUSIC CONDITIONS	46
TABLE 30: OBJECTIVE ToRS FOR NOISE CONDITIONS	46
TABLE 31: SUMMARY OF SELECTION CRITERIA	47
TABLE 32: RAM/ROM EVALUATION	62
TABLE 33: CONTRIBUTION OF DIFFERENT LAYERS TO OVERALL COMPLEXITY.....	62
TABLE 34: WORST CASE COMPUTATIONAL COMPLEXITY OF THE PROPOSED CODEC	72
TABLE 35: MEMORY REQUIREMENT OF THE PROPOSED CODEC.....	72
TABLE 36: COMPUTATIONAL COMPLEXITY.....	76
TABLE 37: ESTIMATED MEMORY SIZE.....	76

TABLE 38: COMPLEXITY AND MEMORY FIGURES.....	80
TABLE 39: COMPLEXITY ESTIMATES	83
TABLE 40: PAIRWISE COMPARISONS FOR EXPERIMENT 1A	86
TABLE 41: PAIRWISE COMPARISONS FOR EXPERIMENT 1B.....	87
TABLE 42: PAIRWISE COMPARISONS FOR EXPERIMENT 2A	88
TABLE 43: PAIRWISE COMPARISONS FOR EXPERIMENT 2B.....	88
TABLE 44: PAIRWISE COMPARISONS FOR EXPERIMENT 2C.....	89
TABLE 45: PAIRWISE COMPARISONS FOR EXPERIMENT 3A	89
TABLE 46: PAIRWISE COMPARISONS FOR EXPERIMENT 3B.....	90
TABLE 47: PAIRWISE COMPARISONS FOR EXPERIMENT 3C.....	90
TABLE 48: PAIRWISE COMPARISONS FOR EXPERIMENT 4.....	90
TABLE 49: CANDIDATE CODEC COMPARISON	91
TABLE 50: CANDIDATE A	92
TABLE 51: CANDIDATE A - CROSSCHECK.....	93
TABLE 52: CANDIDATE B	94
TABLE 53: CANDIDATE B - CROSSCHECK.....	95
TABLE 54: CANDIDATE C	96
TABLE 55: CANDIDATE C - CROSSCHECK.....	97
TABLE 56: CANDIDATE D.....	98
TABLE 57: CANDIDATE D - CROSSCHECK.....	99
TABLE 58: CANDIDATE E	100
TABLE 59: CANDIDATE E - CROSSCHECK	101
TABLE 60: FACTORS FOR EXPERIMENT 5.....	102
TABLE 61: CONDITIONS FOR EXPERIMENT 5.....	103
TABLE 62: EXPERIMENT 5 RESULTS FOR CODER A (WB-PESQ SCORES).....	107
TABLE 63: EXPERIMENT 5 RESULTS FOR CODER B (WB-PESQ SCORES).....	107
TABLE 64: EXPERIMENT 5 RESULTS FOR CODER C (WB-PESQ SCORES).....	107
TABLE 65: EXPERIMENT 5 RESULTS FOR CODER D (WB-PESQ SCORES).....	108
TABLE 66: EXPERIMENT 5 RESULTS FOR CODER E (WB-PESQ SCORES).....	108

List of Figures

Page

FIGURE 1 – THE EMT EV CODER BIT-STREAM	5
FIGURE 2 – ENCODING R1 AND L2 (ACELP)	6
FIGURE 3 – ENCODING L3, L4, AND L5 (MDCT).....	8
FIGURE 4 – DECODING R1 THROUGH R5	9
FIGURE 5 – BLOCK DIAGRAM OF THE PROPOSED ENCODER	11
FIGURE 6 – EXCITATION OF R1 R2AND R3	11
FIGURE 7 – BLOCK DIAGRAM OF THE PROPOSED DECODER.....	13
FIGURE 8 – FRAME FORMAT	14
FIGURE 9 – FEMALE SPEECH (EN01F01.PCM)	15
FIGURE 10 – MALE SPEECH (EN01M01.PCM).....	15
FIGURE 11 – BLOCK DIAGRAM OF THE ENCODER.....	17
FIGURE 12 – BIT STREAM STRUCTURE	18
FIGURE 13 – BLOCK DIAGRAM OF THE BWE MODULE.....	19
FIGURE 14 – BLOCK DIAGRAM OF THE MDCT-BASED ENCODER	20
FIGURE 15 – BLOCK DIAGRAM OF THE CANDIDATE DECODER.....	21
FIGURE 16 – CELP ENCODER SCHEME.....	24
FIGURE 17 – SECOND STAGE ACELP CODING	25
FIGURE 18 – MDCT ENCODING OF HIGHER LAYERS	26
FIGURE 19 – DECODER OVERVIEW.....	27
FIGURE 20 – WEIGHTED NUMBER OF PASSES BY ToR CATEGORY	40
FIGURE 21 – FIGURE OF MERIT BY ToR CATEGORY.....	45
FIGURE 22(A) – MALE SPEECH (P50M.16K), CuT @ 8 KBIT/S	49
FIGURE 22(B) – MALE SPEECH (P50M.16K), CuT @ 12 KBIT/S	49
FIGURE 22(C) – MALE SPEECH (P50M.16K), CuT @ 16 KBIT/S	50
FIGURE 22(D) – MALE SPEECH (P50M.16K), CuT @ 24 KBIT/S	50
FIGURE 22(E) – MALE SPEECH (P50M.16K), CuT @ 32 KBIT/S.....	51
FIGURE 23(A) – FEMALE SPEECH, CuT @ 8 KBIT/S	51
FIGURE 23(B) – FEMALE SPEECH, CuT @ 12 KBIT/S.....	52
FIGURE 23(C) – FEMALE SPEECH, CuT @ 16 KBIT/S.....	52
FIGURE 23(D) – FEMALE SPEECH, CuT @ 24 KBIT/S.....	53
FIGURE 23(E) – FEMALE SPEECH, CuT @ 32 KBIT/S	53
FIGURE 24 – ENCODER BLOCK DIAGRAM FOR 8, 12, 14 KBIT/S	56
FIGURE 25 – ENCODER DIAGRAM FOR 16 KBIT/S AND ABOVE.....	57
FIGURE 26 – CELP 8-12 KBIT/S.....	58
FIGURE 27 – DECODER FOR 8, 12, 14 KBIT/S.....	60
FIGURE 28 – DECODER FOR 16 KBIT/S AND ABOVE.....	60
FIGURE 29 – BIT RATES FROM 14 TO 22 KBIT/S.....	63
FIGURE 30 – BIT RATES FROM 24 TO 32 KBIT/S.....	63

FIGURE 31 – BIT RATES FROM 14 TO 22 KBIT/S.....	64
FIGURE 32 – BIT RATES FROM 24 TO 32 KBIT/S.....	64
FIGURE 33 – ENCODER BLOCK DIAGRAM OF THE PROPOSED CODEC	65
FIGURE 34 – CORE LAYER WITH CELP ENHANCEMENT LAYER.....	66
FIGURE 35 – WIDEBAND EXTENSION LAYER OF THE PROPOSED CODER	67
FIGURE 36 – DECODER BLOCK DIAGRAM OF THE PROPOSED CODEC	69
FIGURE 37 – FRAME FORMAT	69
FIGURE 38 – FREQUENCY RESPONSE FOR FEMALE SPEECH SAMPLE (12 KBIT/S ~ 20 KBIT/S).....	70
FIGURE 39 – FREQUENCY RESPONSE FOR FEMALE SPEECH SAMPLE (22 KBIT/S ~ 32 KBIT/S).....	70
FIGURE 40 – FREQUENCY RESPONSE FOR MALE SPEECH SAMPLE (12 KBIT/S ~ 20 KBIT/S)	71
FIGURE 41 – FREQUENCY RESPONSE FOR MALE SPEECH SAMPLE (22 KBIT/S ~ 32 KBIT/S)	71
FIGURE 42 – BANDWIDTH DEFINITION	73
FIGURE 43 – FRAME STRUCTURE	73
FIGURE 44 – BLOCK DIAGRAM OF THE ENCODER.....	74
FIGURE 45 – BLOCK DIAGRAM OF THE DECODER.....	74
FIGURE 46 – BIT STREAM STRUCTURE FOR THE SPEECH MODE	75
FIGURE 47 – BITSTREAM STRUCTURE FOR THE MUSIC MODE	75
FIGURE 48 – BLOCK DIAGRAM OF THE ENCODER.....	76
FIGURE 49 – BLOCK DIAGRAM OF THE DECODER.....	78
FIGURE 50 – EFFECTIVE BANDWIDTH AT 8 KBIT/S	79
FIGURE 51 – EFFECTIVE BANDWIDTH AT 12 KBIT/S	79
FIGURE 52 – EFFECTIVE BANDWIDTH AT 14 KBIT/S	79
FIGURE 53 – EFFECTIVE BANDWIDTH AT 32 KBIT/S	79
FIGURE 54 – FRAME STRUCTURE.....	80
FIGURE 55 – BLOCK DIAGRAM OF THE G.729EV ENCODER.....	81
FIGURE 56 – BLOCK DIAGRAM OF THE G.729EV DECODER.....	81
FIGURE 57 – FRAME STRUCTURE	83
FIGURE 58 – EFFECTIVE BANDWIDTH	84
FIGURE 59 – WB-PESQ SCORES ON ALL DATABASES.....	104
FIGURE 60 – WB-PESQ SCORES ON FRENCH DATABASE.....	104
FIGURE 61 – WB-PESQ SCORES ON KOREAN DATABASE.....	105
FIGURE 62 – WB-PESQ SCORES ON GERMAN DATABASE	105
FIGURE 63 – WB-PESQ SCORES ON JAPANESE DATABASE.....	106
FIGURE 64 – WB-PESQ SCORES ON NORTH AMERICAN ENGLISH DATABASE.....	106
FIGURE 65(A) – MALE SPEECH (P50M.16k), 14 KBIT/S	109
FIGURE 65(B) – MALE SPEECH (P50M.16k), 16 KBIT/S	109
FIGURE 65(C) – MALE SPEECH (P50M.16k), 18 KBIT/S	110
FIGURE 65(D) – MALE SPEECH (P50M.16k), 20 KBIT/S	110
FIGURE 65(E) – MALE SPEECH (P50M.16k), 22 KBIT/S	111

FIGURE 65(F) – MALE SPEECH (P50M.16K), 24 KBIT/S	111
FIGURE 65(G) – MALE SPEECH (P50M.16K), 26 KBIT/S	112
FIGURE 65(H) – MALE SPEECH (P50M.16K), 28 KBIT/S	112
FIGURE 65(I) – MALE SPEECH (P50M.16K), 30 KBIT/S	113
FIGURE 65(J) – MALE SPEECH (P50M.16K), 32 KBIT/S	113
FIGURE 66(A) – FEMALE SPEECH, 14 KBIT/S	114
FIGURE 66(B) – FEMALE SPEECH, 16 KBIT/S	114
FIGURE 66(C) – FEMALE SPEECH, 18 KBIT/S	115
FIGURE 66(D) – FEMALE SPEECH, 20 KBIT/S	115
FIGURE 66(E) – FEMALE SPEECH, 22 KBIT/S	116
FIGURE 66(F) – FEMALE SPEECH, 24 KBIT/S	116
FIGURE 66(G) – FEMALE SPEECH, 26 KBIT/S	117
FIGURE 66(H) – FEMALE SPEECH, 28 KBIT/S	117
FIGURE 66(I) – FEMALE SPEECH, 30 KBIT/S	118
FIGURE 66(J) – FEMALE SPEECH, 32 KBIT/S	118

ITU-T Technical Paper GSTP-ACP1

Selection Test Results for G.718 Baseline and Qualification Phase Test Results for G.729.1

Summary

This technical paper presents the Selection Test Results of ITU-T G.EV-VBR (later G.718) Baseline and Qualification Test Results for G.729EV (later G.729.1) speech and audio codecs. Information on algorithmic complexity, memory requirements and algorithmic delay of the candidates is collected as well. The purpose of this Technical Paper is to allow interested parties finding the corresponding test results at a conveniently accessible place.

1 Scope

This document collects the selection test results of ITU-T Rec. G.718 and the qualification test results for ITU-T Rec. G.729.1. Information on algorithmic complexity, memory requirements and algorithmic delay of the candidates is collected as well.

2 References

- [1] ITU-T Recommendation G.729.1 (2006), G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729
- [2] ITU-T Recommendation G.718 (2008), Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s
- [3] G.EV-VBR selection related documents in Q7/12:
- [4] TD 68 (WP3/16), “G.729EV qualification phase test results: subjective and objective (WB-PESQ) scores for Experiment 1b wideband conditions”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [5] TD 69 (WP3/16), “G.729EV qualification phase test results: Experiment 5 (Clean Speech; wide band case, Bit rate granularity)”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [6] TD 71 (WP3/16), “G.729EV Qualification Test Results (Exp 1-4)”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [7] TD 73 (WP3/16), “G.729EV qualification phase test results: Experiment 6 (Clean Speech; bit rate switching case)”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [8] TD 76 (WP3/16), “G.729EV qualification phase test results: Complexity evaluation”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [9] TD 81 (WP3/16), “G.729EV qualification phase test results: Frequency responses of G.729EV candidates”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [10] COM16-D135-E, “France Telecom G729EV Candidate: High level description and complexity evaluation”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [11] COM16-D151-E, “High-level description of ETRI candidate codec for G.729EV”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [12] COM16-D152-E, “Complexity evaluation of ETRI candidate codec for G.729EV”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005

- [13] COM16-D176-E, “High-level description of Samsung candidate algorithm for G.729 EV codec”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [14] COM16-D189-E, “High-level description of VoiceAge candidate for G.729EV”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [15] COM16-D214-E, “High level description of the scalable 8-32 kbit/s algorithm submitted to the Qualification Test by Matsushita, Mindspeed and Siemens”, ITU-T SG16 Meeting, Geneva, 26 July – 5 August, 2005
- [16] AC-0703-Q9-04, “High Level Description of the Ericsson/Motorola/TI EV coder candidate”, ITU-T SG16 Meeting, Geneva, 21—30 March, 2007
- [17] AC-0703-Q9-07R1, “High-level description of Huawei’s candidate Codec for G.VBR”, ITU-T SG16 Meeting, Geneva, 21—30 March, 2007
- [18] AC-0703-Q9-09, “High level description of G.EV candidate codec algorithm proposed by Panasonic”, ITU-T SG16 Meeting, Geneva, 21—30 March, 2007
- [19] AC-0703-Q9-10, “High-level description of the Nokia/VoiceAge candidate for EV-VBR Codec”, ITU-T SG16 Meeting, Geneva, 21—30 March, 2007
- [20] AC-0703-Q9-14, “Comments of Q9/16 EV-VBR codec standardisation”, ITU-T SG16 Meeting, Geneva, 21—30 March, 2007
- [21] AC-0703-Q9-22R2, “Report of the Global Analysis Laboratory for the EV-VBR Selection Phase”, ITU-T SG16 Meeting, Geneva, 21—30 March, 2007
- [22] AC-0703-Q9-24, “Supplemental Analyses from Global Analysis Laboratory for the EV-VBR Selection Phase”, ITU-T SG16 Meeting, Geneva, 21—30 March, 2007
- [23] AC-0703-Q9-25, “Summary of the Results of the 4 Candidate Baseline Codecs”, ITU-T SG16 Meeting, Geneva, 21—30 March, 2007
- [24] AH-07-10 - France Telecom host laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A1!MSW-E
- [25] AH-07-12 - ARCON Corporation host laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A2!MSW-E
- [26] AH-07-02 - Dynastat listening laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A3!MSW-E
- [27] AH-07-04 - Nokia listening laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A4!MSW-E
- [28] AH-07-05 - VoiceAge listening laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A5!MSW-E
- [29] AH-07-06 - BIT listening laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A6!MSW-E
- [30] AH-07-07 - France Telecom listening laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A7!MSW-E
- [31] AH-07-11 - NTT-AT listening laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A8!MSW-E
- [32] AH-07-13 - ARCON listening laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A9!MSW-E

- [33] AH-07-03R1 - Global analysis laboratory report
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A10!MSW-E
- [34] AH-07-16 - Supplemental Analyses from Global Analysis Laboratory for the EV-VBR Selection Phase (Part I)
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A11!MSW-E
- [35] AH-07-18 - Supplemental Analyses from Global Analysis Laboratory for the EV-VBR Selection Phase (Part II)
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A12!MSW-E
- [36] AH-07-08 - Report on computation of gains
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A13!MSW-E
- [37] AH-07-17 - EV-VBR computation on frequency attenuation
http://itu.int/md/dologin_md.asp?id=T05-SG16-070626-TD-GEN-0344!A14!MSW-E

3 Abbreviations

ACR	Absolute Category Rating
CCR	Comparison Category Rating
CNG	Comfort Noise Generator
DCR	Degradation Category Rating
DTX	Discontinuous Transmission
NB	Narrowband
NT	Non Transmitted
PESQ	Perceptual evaluation of speech quality
SID	Silence Insertion Descriptor
SWB	Superwideband
VAD	Voice Activity Detection
WB	Wideband
WB-PESQ	Wideband extension to PESQ
WMOPS	Weighted Million Operations Per Second

4 Scope of G.718 (G.EV-VBR)

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 are proposed 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 are proposed 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

Based on previous discussions, some general guidelines have been derived that have driven the drafting of the preliminary ToR:

- primary signals of interest are speech but in high quality audio conferencing, background signals shall be considered, not as the noise anymore, but as a part of the signals that convey information
- to cope with heterogeneous accesses and terminals, it is important to consider bit-rate scalability but also bandwidth scalability and complexity scalability
- narrowband/wideband signal capability with HiFi bandwidth as a requirement and stereo/multi-channel capability as 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 (\cong 32 kbit/s); for mobile users, it is highly desirable to introduce bitrates compatible with mobile links
- it will be attractive to provide fine-grain bit-rate scalability to allow trade-off between speech and audio quality and the quality of other services (e.g., video)
- to maintain a good quality of services requiring interactivity, it is necessary to maintain the overall delay as low as possible (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 must be 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 20ms frames and comprises 5 fixed-rate layers 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.

5 G.EV-VBR Candidates in Selection Phase: Algorithmic Overview, Delay and Complexity

5.1 Ericsson, Motorola, Texas Instruments Candidate [16]

The EMT EV coder is designed to process narrowband signals at 8 kHz and wideband signals at 16 kHz, and provide an embedded bit-stream with bit-rates ranging from 8 to 32 kbit/s (Figure 1). With an 8 kHz narrowband input, the EV encoder produces a 12 kbit/s bit-stream, R2, comprising of an 8 kbit/s core-layer, R1, and a 4 kbit/s enhancement layer, L2. With a 16 kHz wideband input, the EV encoder produces a 32 kbit/s bit-stream, R5, which includes an 8 kbit/s core-layer, R1, two 4 kbit/s enhancement layers, L2 and L3, and two 8 kbit/s enhancement layers, L4 and L5. The EMT EV decoder can produce an 8 kHz narrowband-output from R1 or R2 bit-streams obtained from an 8 kHz narrowband-input, or an 8 kHz narrowband or 16 kHz wideband-output from R1 through R5 bit-streams obtained from a 16 kHz wideband-input. This design satisfies the Q9/16 ToR with respect to the input and output sampling rates and signal bandwidth, coder bit-rates, and embedded bit-stream.

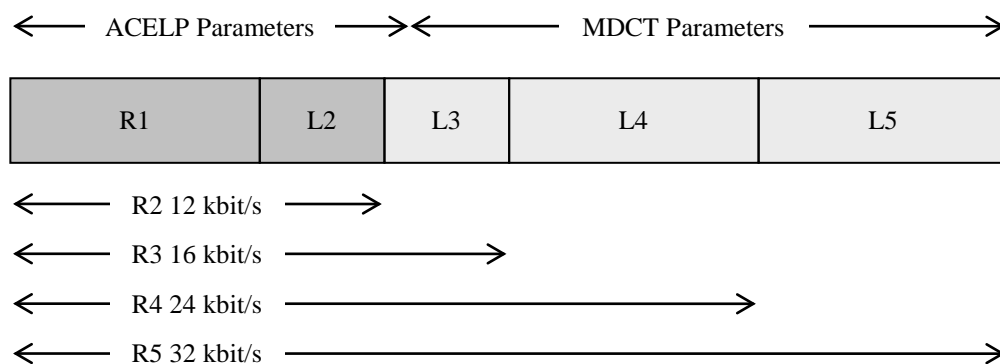


Figure 1 – The EMT EV coder bit-stream

The lower two layers, R1 and L2, are encoded with an Algebraic CELP (ACELP) coder based on a modified 3GPP2 VMR-WB standard. For 8 and 16 kHz sampling-rate inputs, the signal is up and down-sampled to 12.8 kHz, respectively, and a VMR-based ACELP coder is applied to the 12.8 kHz re-sampled signal. At the decoder, for wideband output, the 6.4 to 7 kHz bandwidth is reconstructed as in the G722.2 (AMR-WB) standard. The encoding of the upper three layers, L3 through L5, is performed only for a 16 kHz wideband input; it is applied to the difference between the original signal and the R2-encoded signal. The L3 through L5-layer encoding is performed in a perceptually weighted Modified Discrete Cosine Transform (MDCT) domain and is based on scalable algebraic Vector Quantization (VQ) used in the 3GPP AMR-WB+ standard. For improved performance in channel errors, frame-erasure concealment (FEC) is also implemented. The bits carrying the additional ACELP FEC information are placed in L3.

The maximum delay of the EMT EV coder is 55.75 ms (a sum of the coder processing frame, look-ahead, and the delay of re-sampling filters) which satisfies the Q9/16 ToR maximum algorithmic delay requirement of no more than 60 ms. For R1 and L2 processing only, the maximum coder delay could be 35.75 ms as the extra 20 ms used in higher layers for the FEC and MDCT-coding overlap-add is not needed. A reduced-delay processing of the R1 and L2 layers is included in the EMT EV coder.

The processing frame of the EMT EV coder is 20 ms which meets the corresponding requirement.

The EMT EV codec includes also some of the additional features present in the VMR-WB coder, e.g., discontinuous transmission mode (DTX) and interoperability with the AMR-WB coder. The availability of an integrated DTX and AMB-WB interoperability may be beneficial in the EV coder standardization moving forward.

5.1.1 Encoder Overview

The EMT EV codec encodes the core layer and the first enhancement layer, R1 and L2, with an ACELP coder (Figure 2) based on the VMR-WB standard. For improved performance in channel errors, memory-less quantization is applied in some of the CELP frames. Additional FEC is also implemented with the corresponding bits included in the L3 layer. The upper three layers, L3 through L5, are encoded with an MDCT transform coder (Figure 2) based on scalable algebraic VQ used in the AMR-WB+ standard.

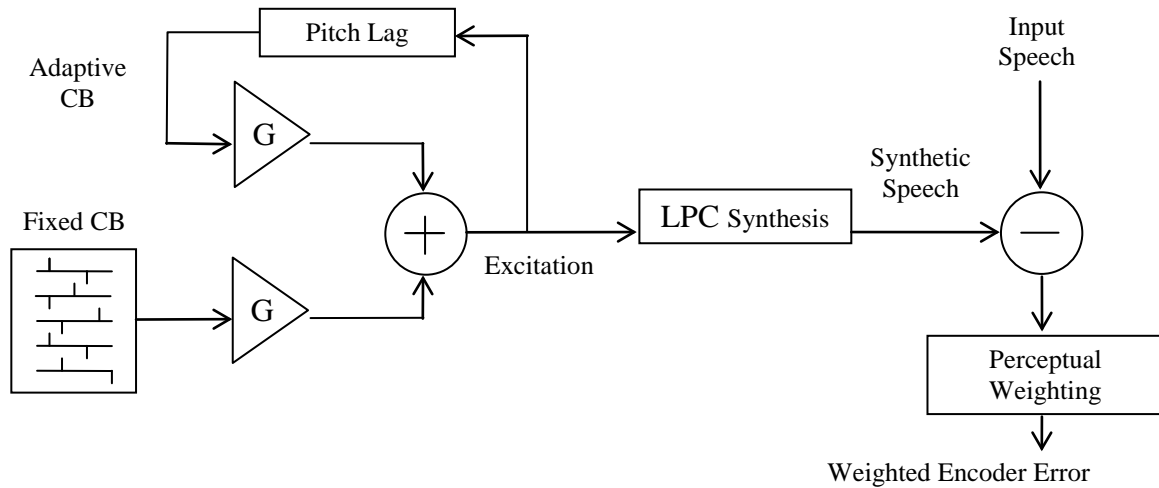


Figure 2 – Encoding R1 and L2 (ACELP)

The EMT EV coder accepts input signals at 8 and 16 kHz sampling rates. In both cases, the input signal is re-sampled to 12.8 kHz sampling rate before ACELP processing. However, different encoding optimization techniques are applied when the 12.8 kHz signal is considered to represent narrowband or wideband signal, and one bit is used to indicate this to the decoder. In particular, signal pre- and de-emphasis, Linear Prediction Coefficients (LPCs) quantization codebooks, and post-filtering differs for narrowband and wideband inputs.

In R1 and L1, similarly to the VMR-WB coder, different encoding modes are used to represent efficiently distinctive input-signal features. In the EMT EV coder, we classify the input signal into unvoiced frames, generic frames, and voiced frames. To encode unvoiced frames, we use two random codebooks whose entries are scaled and summed; an adaptive codebook is not used in these frames. A tilt is applied to the random codebook entries to improve spectral matching with the original signal. Generic frames use an adaptive codebook and a fixed codebook as in a standard ACELP coder. In voiced frames, the pitch variation is limited and a reduced number of bits are used to encode the adaptive codebook so that more bits may be allocated to the fixed-codebook parameters. For improved performance in frame erasures, the EMT EV coder identifies certain error-sensitive voiced frames (which typically follow signal transitions) and encodes them with a memory-less quantization scheme using a pulse-shape codebook. This memory-less quantization is designed to limit propagation of frame-erasure errors which produce audible artifacts when onset frames of a steady-voiced speech are erased and, subsequently, the decoder's adaptive-codebook memory is not initialized properly.

The LPCs are estimated and encoded twice per frame in all modes using a 20 ms analysis window. The two sets of LPC parameters, one for frame-end and one for mid-frame, are transformed into Immitance Spectral Frequencies (ISFs). The frame-end ISFs are quantized with a switched-predictive Multi-Stage Vector Quantizer (MSVQ). The ISFs are predicted from previous frame's quantized ISFs using a switched autoregressive (AR) predictor; two codebooks matched to two predictors (corresponding to weakly and strongly predictive frames, respectively) are searched to

find the predictor/codebook and the codebook entry that minimize the distortion with respect to the estimated ISF vector. As frame-erasures related errors propagate in decoded ISFs due to the AR prediction, the weak-predictor coefficients are selected such that the decoded error decays rapidly; for some frames, the weak-predictor coefficients are set to zero. To provide additional error robustness, the weakly-predictive codebook is always chosen when its quantization distortion is sufficiently close to that of the strongly-predictive codebook, or the ISFs change significantly from one frame to another. Mid-frame ISFs are encoded with an interpolative split VQ; for each ISF subgroup, a linear interpolation coefficient is found so that the difference between the estimated and the interpolated quantized ISFs is minimized.

The open-loop pitch search and pitch tracking are performed on a de-noised signal available from the integrated VMR-WB noise suppressor. Compliant with the Q9/16 ToR, noise suppression is not used, however, to modify characteristics of the encoded input signal so that the noise level of the decoded output matches that of the input.

The search of the fixed and adaptive codebooks is jointly optimized using a codebook orthogonalization technique as in the VMR-WB coder.

In the L2 layer, the EMT EV encoder adds another fixed CB and modifies the adaptive CB to include not only the past R1 contribution, but also the past L2 contribution. The adaptive-CB pitch-lag is the same in R1 and L2 to maintain time synchronization between the layers. The adaptive and fixed CB gains in the L2 layer are then re-optimized to minimize the perceptually weighted coding error. With this approach, the two CB gains used in R1 (for adaptive and fixed CBs, respectively) are replaced in L2 by four gains corresponding to: R1 adaptive-CB contribution to R2, R1 fixed-CB contribution to R2, L2 adaptive-CB contribution, and L2 fixed-CB. In a similar manner to the codebook search for the R1 layer, the L2 fixed-CB search is performed with an optimal-gain assumption. The four L2 CB gains are predictively vector-quantized with respect to the two R1 CB gains encoded in the R1 layer.

The bit allocation for the unvoiced, generic, and voiced frames in R1 and L2 is included in Table 1. The bit-allocation for the memory-less voiced frames is a derivative of the voiced-frames bit-allocation.

Table 1: Bit allocation in R1 and L2

Layer	Parameter	Unvoiced	Generic	Voiced
R1	Frame Type	3	2	2
	NB / WB Input	1	1	1
	LPCs	46	36	34
	Gains	24	23	23
	Adaptive CB	-	26	20
	Fixed CB	52	72	80
	Spectral Tilt	17	-	-
	Unused	17	-	-
	Total	160	160	160
L2	Gains	8	16	16
	Fixed CB	72	64	64
	Total	80	80	80

In layers L3, L4, and L5, the error between the original speech and the R2 ACELP-synthesized signal is encoded in the perceptually-weighted MDCT domain (Figure 3). The MDCT components are estimated once per frame using 40 ms analysis windows. They are quantized with a scalable algebraic VQ similar to the one used in the AMR-WB+ coder. To enhance performance of the MDCT quantizer, perceptual weighting is applied prior to the quantization; the perceptual-weighting filter is derived from the ISF coefficients extrapolated from the 12.8 kHz ACELP to the 16 kHz MDCT sampling-rate domain. To better represent characteristics of the original input signal, additional quantization-noise shaping and high frequency compensation are also performed.

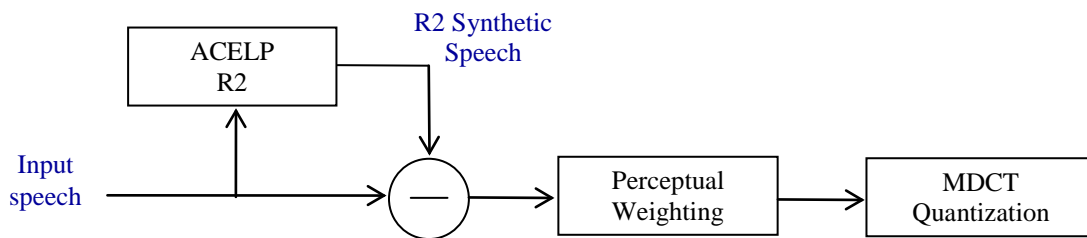


Figure 3 – Encoding L3, L4, and L5 (MDCT)

The MDCT parameters are encoded with 62 bits per frame (3.1 kbit/s) in the L3 layer, and with 160 bits per frame (8 kbit/s) in the L4 and L5 layers. Note that the L3 layer, in addition to the MDCT parameters, also includes the ACELP FEC parameters (described earlier in this section).

5.1.2 Decoder Overview

Depending on the number of bit-stream layers received, the EMT EV decoder generates a synthetic signal according to the procedure outlined in Figure 4. The decoder produces an 8 kHz output from a bit-stream corresponding to a narrowband input, or an 8 or 16 kHz output from a bit-stream corresponding to a wideband input, as requested.

The ACELP decoder converts the R1 and, if available, L2 bits into synthetic speech at a 12.8 kHz sampling rate. If no higher bit-stream layers are available, formant and pitch-sharpening post-filters are applied (different for narrowband and wideband signals) to enhance the synthetic speech. Resampling of the 12.8 kHz synthetic speech provides 8 kHz or 16 kHz signal, depending on the requested output sampling rate. For 16 kHz wideband output, 6.4 to 7 kHz bandwidth regeneration is performed as in the G722.2 standard.

In case of frame erasures, FEC is applied to the ACELP parameters. Enhanced FEC is performed when L3-layer bits are available. The additional FEC bits carry information about frame type, signal energy, and glottal-pulse position enabling the decoder to faster match the original signal characteristics after missing data.

When enhancement layers above L2 are available, the EMT EV decoder proceeds with decoding the MDCT parameters in a perceptually-weighted domain. Temporal noise shaping (similar to the one used in the AMR-WB+ coder) is performed to enhance temporal signal characteristics that may have been modified by the MDCT transformation. An overlap-add synthesis is performed to avoid abrupt transitions, and the decoded signal is added to the ACELP-decoded R2-layer speech. A post-filter is then applied to enhance low frequencies of the generated output.

When decoding received bit streams corresponding to rates R1 or R2, the delay of the decoder may be 20 ms shorter (35.75 ms maximum coder algorithmic delay) than when decoding bit streams corresponding to rates R3, R4, or R5. The additional 20 ms delay in the upper-layer decoding is used for the enhanced ACELP FEC and MDCT overlap-add. A reduced-delay processing mode of the R1 and L2 layer decoder is built-in into the EMT EV code functionality.

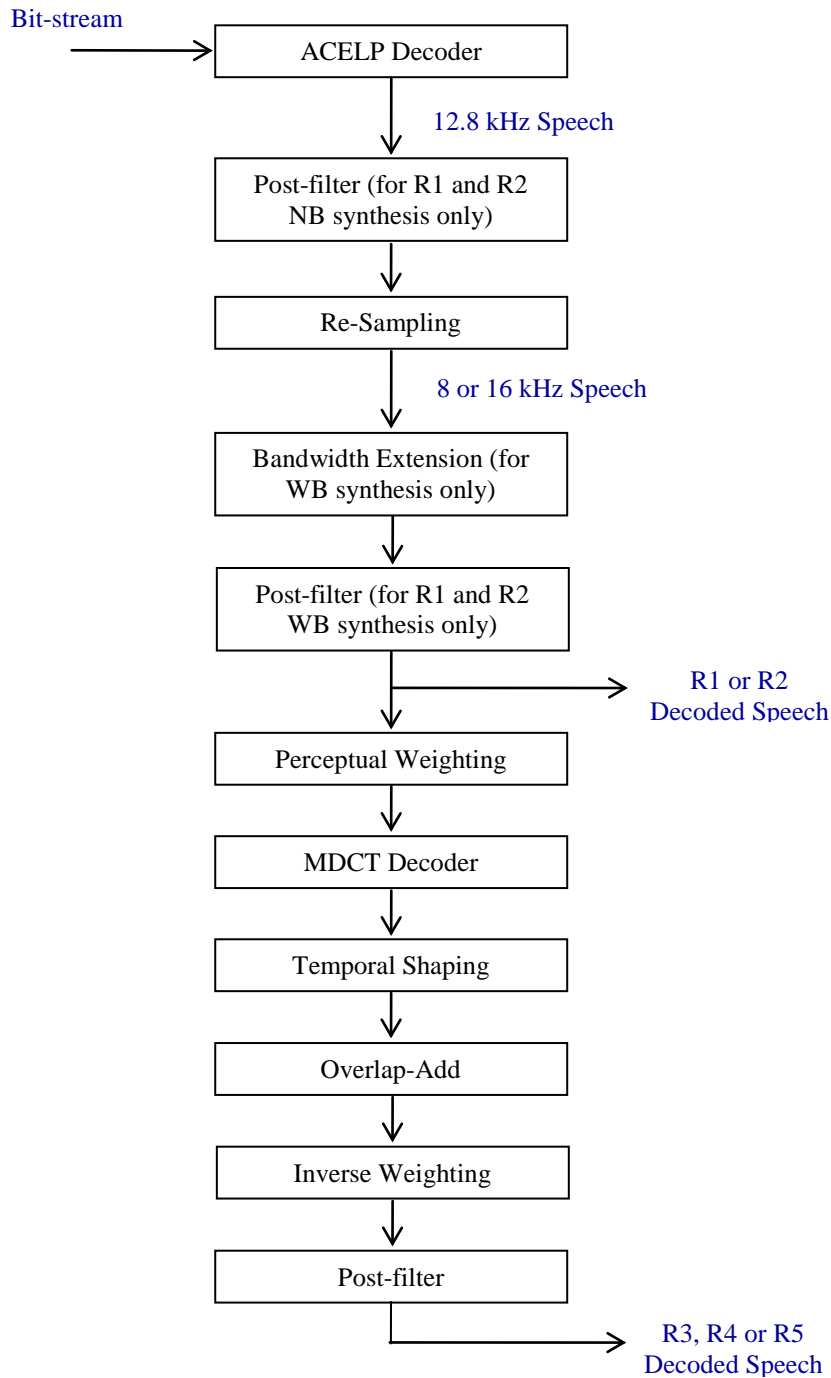


Figure 4 – Decoding R1 through R5

5.1.3 Complexity

The computational complexity estimates of the EMT EV encoder and decoder are around 54 and 16 WMOPS, respectively. This estimate conforms to the Q9/16 ToR complexity requirement of implementability on a single commercially available fixed-point DSP.

5.2 Huawei Candidate [17]

Huawei's candidate codec is an embedded variable bit-rate speech codec, which is designed to support bandwidth scalability (narrow band and wide band signal) and bit-rate scalability (8~32 kbit/s). In the default mode, the codec accepts input signals sampled at 16 kHz on the encoder side and generates output signals sampled at 16 kHz on the decoder side. The codec is also compatible with input signals sampled at 8 kHz and outputs signals sampled at 8 kHz.

The output bitstream produced by the encoder is scalable, consisting of a core layer and four enhancement layers at 8 kbit/s, 12 kbit/s, 16 kbit/s, 24 kbit/s and 32 kbit/s individually.

The core layer uses the ACELP (algebraic code excited linear prediction) technique, and the enhancement layers are designed to improve the speech quality by adopting CELP enhancement layer and TCX (transform coding excitation) techniques. From 12 kbit/s up to 16 kbit/s, layer 2 and layer 3 are CELP enhancement coding which give additional information about the algebraic codebook and gain. TCX is applied to the higher two enhancement layers above 16 kbit/s.

5.2.1 Description of the encoder

The block diagram of the proposed encoder is given in Figure 5. The sampling rate of the input signals should be selected before the signals are processed by the encoder. If the sampling rate is set to 16 kHz, the input signal is down-sampled to 12.8 kHz. If the input signal is a narrowband signal sampled at 8 kHz, the input signal is up-sampled to 12.8 kHz. The encoder operates on 20ms input frame, i.e. 256 samples of input signal, which is pre-processed by a high-pass filter with 50Hz cut-off frequency. The processing starts with ACELP codec:

Core layer: ACELP structure is applied to the 8 kbit/s layer. LP (linear prediction) analysis and quantization, open-loop pitch analysis, and perceptual weighting filter operate on every 20ms frame. Adaptive-codebook, algebraic codebook search and gain parameters estimation rely on 5ms sub-frame. LP analysis is performed once per speech frame using the autocorrelation approach with 30 ms asymmetric windows. The LP coefficients are obtained by Levinson-Durbin algorithm, and then quantized in ISFs domain by a new Unequal Coefficients Interframe Predictive Split Vector Quantizer. Adaptive codebook, algebraic codebook and gain parameters, are encoded on every 5 ms sub-frame so that perceptually weighted encoding distortion becomes minimum. Meanwhile, the gain of adaptive codebook is restricted in order to serve for frame erasure concealment.

12-16 kbit/s CELP enhancement layers: Each layer uses an additional algebraic codebook to refine the excitation. The target signal to search the enhancement layer 1 is equal to the difference of the original target signal and the contribution of the current layer's adaptive codebook and the core layer's algebraic codebook, and the target signal to the enhancement layer 2 is equal to the difference of the original target signal and the contribution of the current layer's adaptive codebook and the algebraic codebook of the core layer and the enhancement layer 1. So the algebraic codebooks of the lower three layers are not independent, but embedded each other. In each layer, the algebraic codebook signal of the lower three layers is represented by three algebraic pulses from four tracks every 5 ms. The signs and positions of the pulses are quantized with 18 bits, and one bit is used to represent the track of the third pulse.

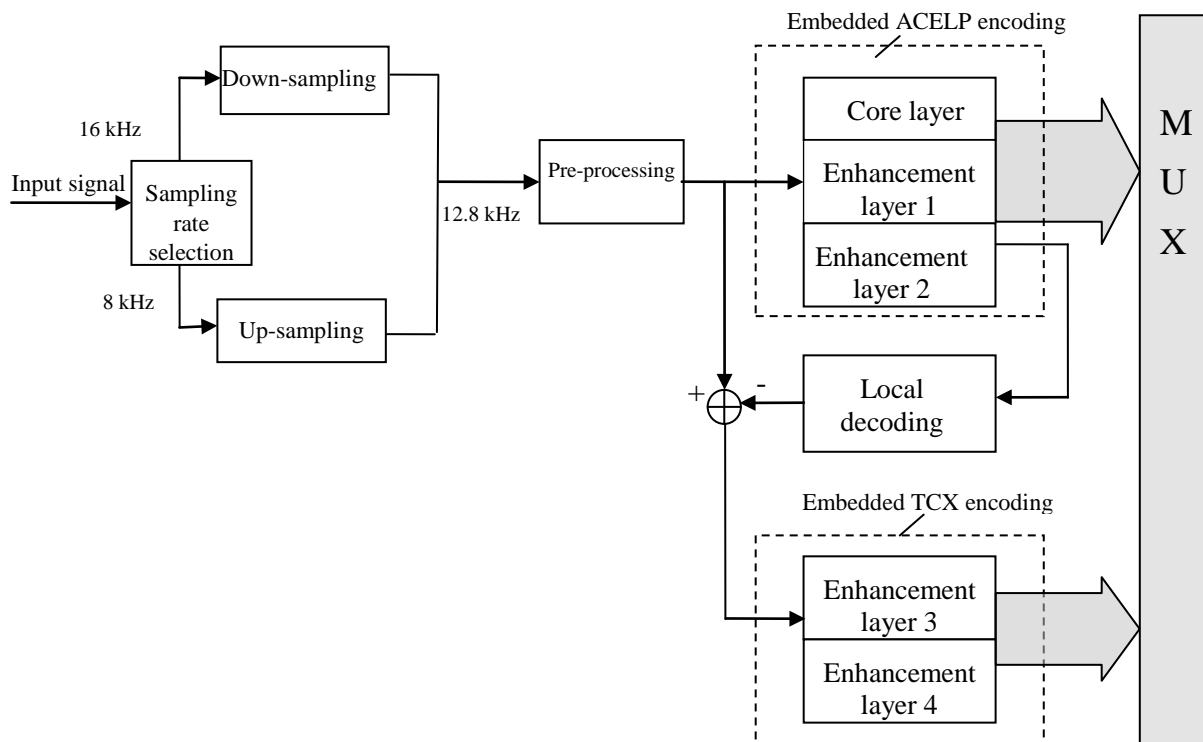


Figure 5 – Block diagram of the proposed encoder

In the lower three layers, adaptive codebook of each layer is obtained by interpolating the past excitation of each layer at the selected fractional pitch lag, but different layer has different excitation (Figure 6), so the candidate codec proposes a method that sets one adaptive codebook buffer for each layer and using each layer's excitation to update its adaptive codebook. That is to say, different layer has different adaptive codebook, and the filter's memory of the different layer is updated respectively. This kind of adaptive codebook structure is superior to finding the best coding parameters for each layer and make different layer independent relatively. The embedded speech coder using this structure can achieve good performance both in core layer and enhancement layers.

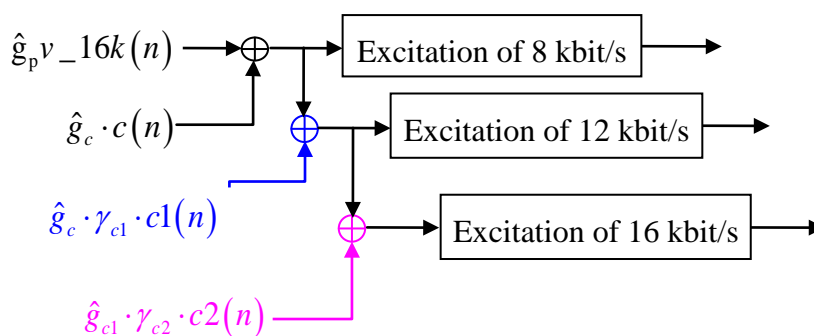


Figure 6 – Excitation of R1 R2and R3

24-32 kbit/s TCX enhancement layers: The difference between the pre-processed signal and the local synthesis speech signal of 16 kbit/s is the target signal of enhancement layer 3 as $e3(n)$, and then coded by TCX coding. When the difference is coded by TCX coding, the coefficient β of the perceptually weighted filter $w(z) = \hat{A}(z/\gamma)/(1 - \beta z^{-1})$ is 0.8. After the TCX coding, a bit-stream of 160 bits is obtained, 10 bits represented noise factor and global gain, the rest of bits (150bits) are used to quantize the pre-shaped spectrum using split multi-rate lattice VQ. Meanwhile, the quantized signal

of the difference is also obtained as $\hat{e}_3(n)$. TCX operates on 288 samples. Since the overlap-and-add method is adopted. Thus, when ACELP coding is used in the former three layers, the loop of sub-frames run five times, however, only first four sub-frames information is written into bit-stream.

The difference between the target signal of enhancement layer 3 ($e_3(n)$) and the quantized signal of the target signal of enhancement layer 3 ($\hat{e}_3(n)$) is the target signal of enhancement layer 4 as $e_4(n)$, that is to say, $e_4(n) = e_3(n) - \hat{e}_3(n)$, and then is coded by TCX coding. At this time, the coefficient of the perceptually weighted filter $W(z) = \hat{A}(z/\gamma)/(1 - \beta z^{-1})$ is 0.84. After the TCX coding, a bit-stream of 160 bits is obtained, the method of bit allocation is the same to the enhancement layer 3.

5.2.2 Frame erasure concealment

When a frame erasure occurs, a frame erasure concealment algorithm is invoked to improve the synthesized quality in frame erasure condition.

The last speech frame of the erased frame is classified as VOICED, UNVOICED, SILENCE, UNVOICED TRANSIT TO VOICED, and VOICED TRANSIT TO UNVOICED at decoder. The parameters of the erased frame are properly recovered based on the parameters from past frames. The energy of excitations is carefully controlled depending on the classification of the speech. To match with the configuration of the embedded speech codec, the recovery of erased frame's adaptive codebook will depend on the length of bit-stream of the last frame. In addition, to increasing the robustness of the codec, the contribution of adaptive codebook is properly constrained at encoder.

5.2.3 Description of the decoder

Figure 7 shows the block diagram of the decoder. The decoder also consists of five layers: core layer, two CELP enhancement layers and two TCX enhancement layers. In each 20-ms frame, the decoder can receive any of the supported bit rates, from 8 kbit/s up to 32 kbit/s. Thus, the operation of each layer depends on the size of the received bit-stream. After bit-rate layer extraction, the decoding of each layer depends on the number of bits that have been received.

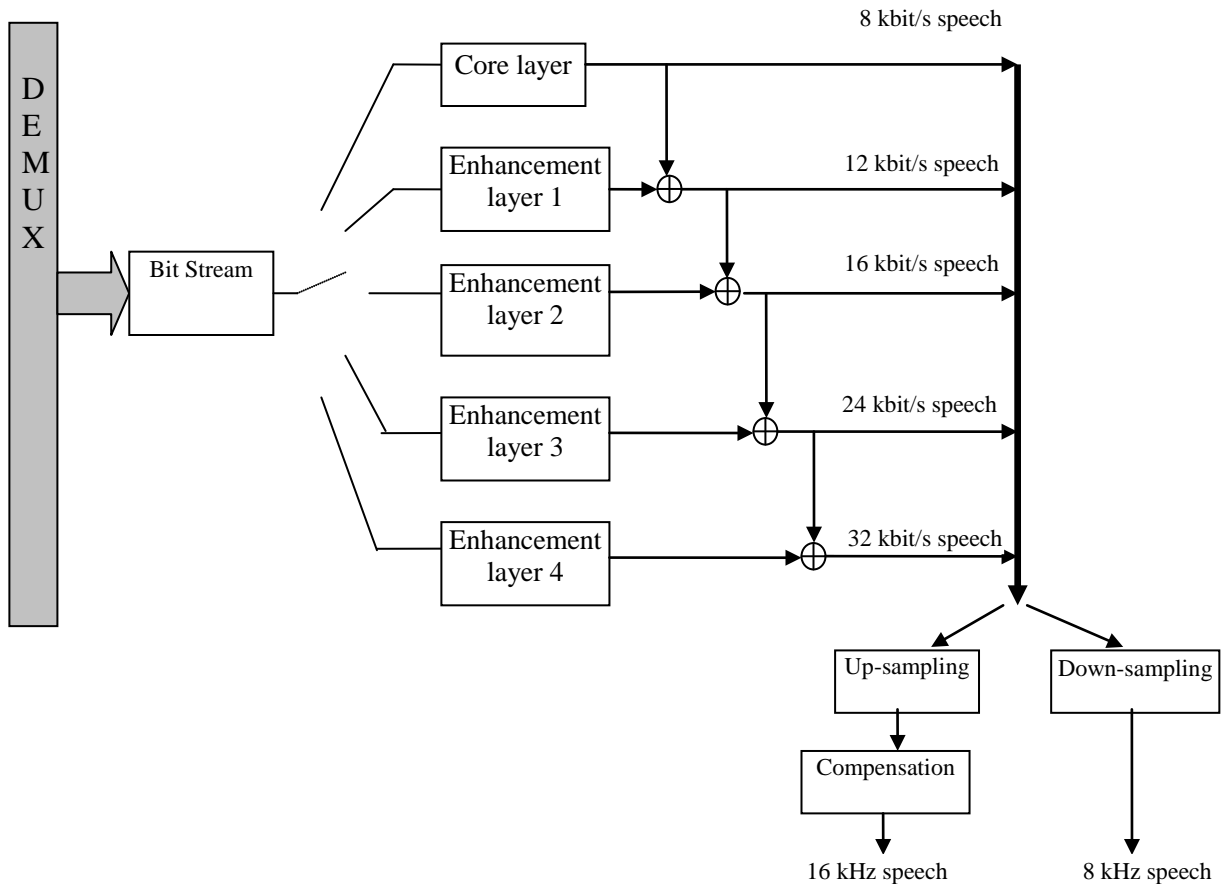


Figure 7 – Block diagram of the proposed decoder

1. If the received bit rate at decoder side is equal 8 kbit/s, the decoding of core layer is performed. The quantized ISF (Immittance Spectral Frequency) is decoded and converted to LPC coefficients. The adaptive codebook and algebraic excitations are generated by decoding the index of pitch delay and algebraic codebook. The gains of adaptive and algebraic codebook are also obtained by decoding the index of the gains.
2. If the received bit rate at decoder side is 12 or 16 kbit/s, both core layer and CELP enhancement layer decoding are performed. The bit-stream of the enhancement layer contains extra algebraic excitation from additional algebraic codebook.
3. If the received bit rate at decoder side is 24 or 32 kbit/s, the CELP 8-12 kbit/s decoder (as in case 2) is first activated, followed by TCX enhancement layers decoding.
4. Then, the post-processed 12.8 kHz synthesis signal is re-sampled and filtered to produce 8 kHz or 16 kHz signal.
5. If the option of the sampling rate is set to 8 kHz, the 8 kHz re-sampled synthesis is output signal. If the option is set to 16 kHz, high frequency compensation is operated on the synthesis signal. Firstly, generate a modified Gaussian white noise as high-band excitation signal, and then the excitation signal is filtered by a modified synthesis filter and band-pass filter to obtain synthesis signal. Finally, the synthesis signal is summed the 16 kHz re-sampled synthesis signal (as case 4) to generate the 16 kHz output signal.

When a frame erasure occurs, a frame error concealment algorithm is invoked to improve the synthesized quality in frame erasure condition.

5.2.4 Frame Format

The frame structure is shown in Figure 8.

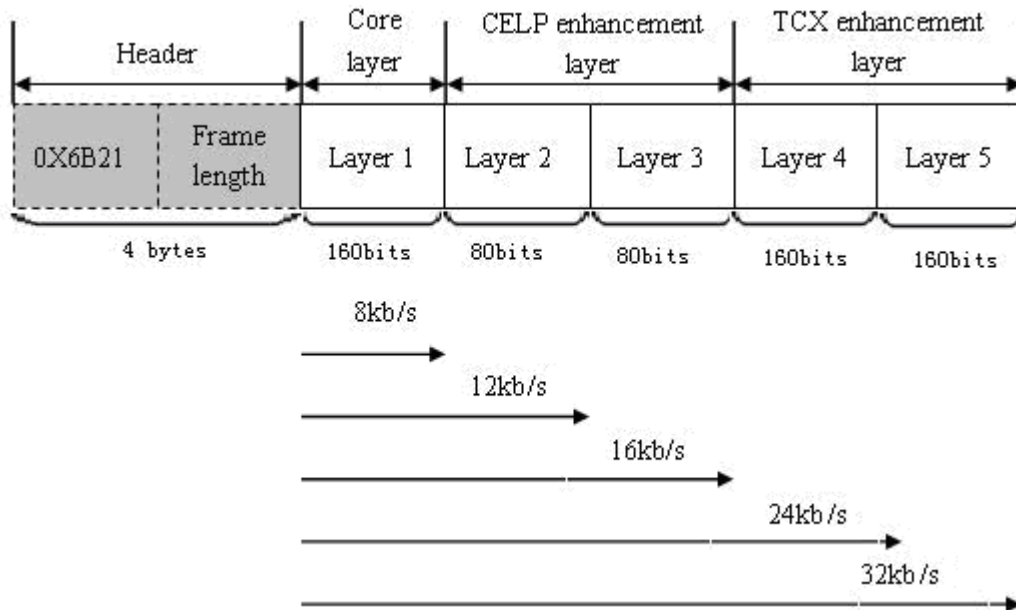


Figure 8 – Frame format

5.2.5 Algorithmic delay

(1) *16 kHz input and output signal*

- The encoder algorithmic delay: $20+5=25$ ms
where 20ms is frame size, 5ms is look-head.
- The decoder algorithmic delay: 0.9375 ms
where 0.9375 ms is re-sampling delay (12.8 kHz up to 16 kHz)
- Total algorithmic delay: 25.9375

(2) *8 kHz input and output signal*

- The encoder algorithmic delay: $20+5+2=27$ ms
where 2ms is re-sampling delay (8kHz up to 12.8kHz)
- The decoder algorithmic delay: 1.875 ms
- Total algorithmic delay: 28.875 ms

5.2.6 Effective bandwidth

The bitrates from 8 kbit/s to 32 kbit/s provide narrowband and wideband outputs. Frequency response was computed with the freqresp STL 2005 tool for two files sampled at 16 kHz (EN01F01.pcm and EN01M01.pcm from NTT speech database), see Figures 9 and 10.

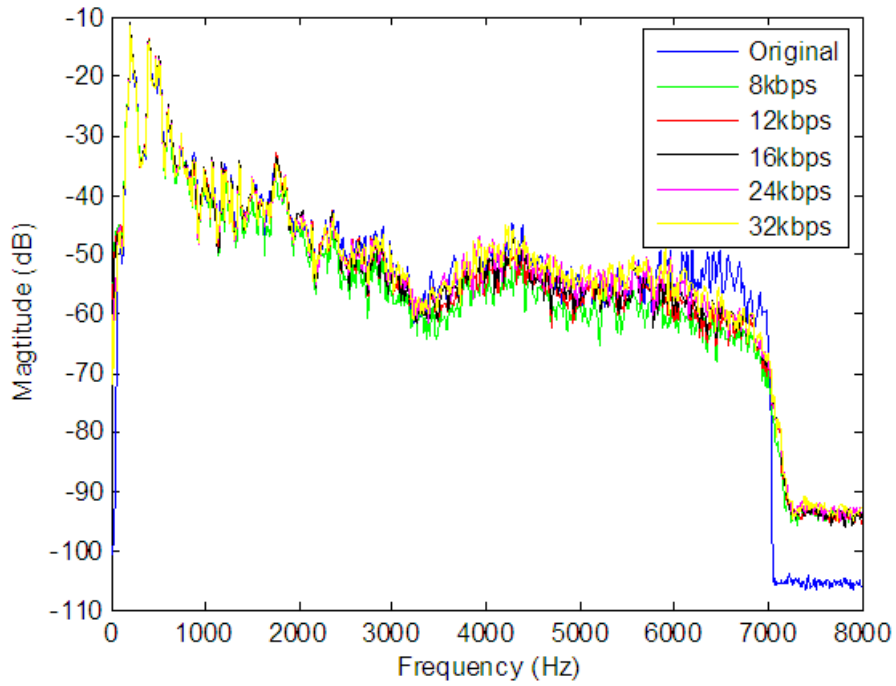


Figure 9 – Female speech (EN01F01.pcm)

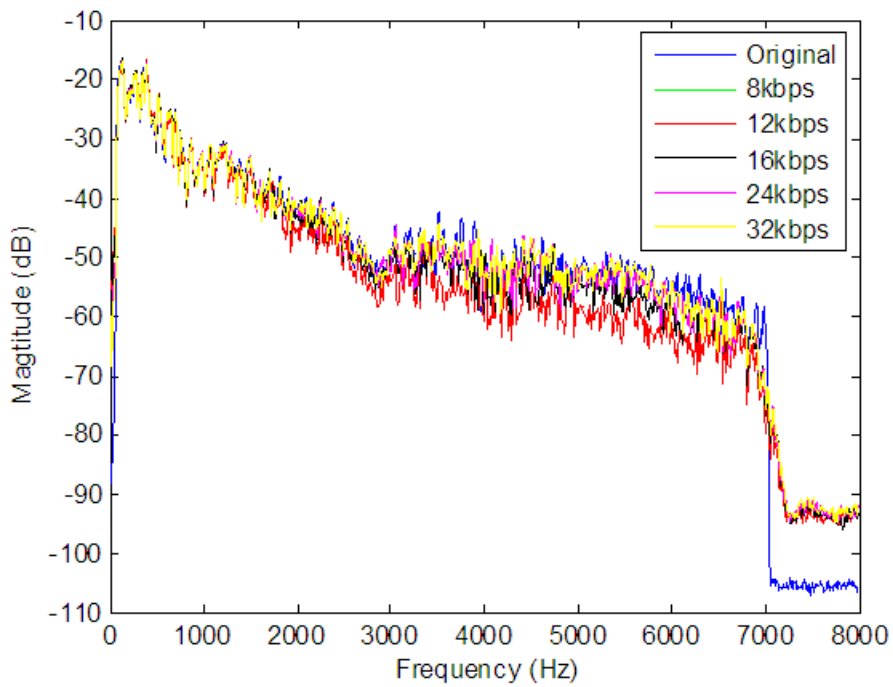


Figure 10 – Male speech (EN01M01.pcm)

5.3 Panasonic Candidate [18]

The codec operates on the 20 ms frame basis. The encoder supports 8 and 16 kHz sampling inputs and always produces a 32 kbit/s bit stream. The codec comprises a 6.8 kbit/s core layer and eight enhancement layers with the bit rates of 0.8, 1.2, 3.2 or 4 kbit/s, and therefore the bit stream can be truncated at the points of 6.8 kbit/s, 8 kbit/s, 12 kbit/s, 15.2 kbit/s, 16 kbit/s, 20 kbit/s, 24 kbit/s, 28 kbit/s and 32 kbit/s. The truncated bit stream is fed to the decoder, which outputs a decoded signal at the sampling rate of 16 or 8 kHz.

The codec has three main modules; a code-excited-linear-prediction (CELP) module, a bandwidth-extension (BWE) module and a transform-coding (TRC) module. The CELP module is designed to encode narrowband signal with high efficiency. Accompanied with the BWE module, the CELP module becomes capable to work as a wideband codec, and this combination forms the R1 codec at 8 kbit/s. The BWE module extends the bandwidth from narrowband to wideband in frequency domain using the modified-discrete-cosine-transform (MDCT). The TRC module encodes the fine spectral structure of wideband signal using an MDCT-based codec on top of the CELP and BWE modules. In addition to the three main modules, the codec has a module for encoding parameters used in frame erasure concealment (FEC) in order to obtain robustness against frame erasure conditions.

The codec algorithm meets the requirements on algorithmic delay, computational complexity and memory requirements by 50 ms, 80.4 WMOPS and 22.5 kword data-ROM + 7.3 kword RAM, respectively.

This codec is designed to have some special features in its algorithmic flexibilities as follows.

- Capability of finer increment of bit-rate is provided with its 9-layer structure. This feature somewhat meets the Objective of the parameter 1 in the ToR. This feature would be desirable for some of foreseen applications.
- The core layer works as a 6.8 kbit/s CELP coder for narrowband input signal. This feature nearly meets the Objective of parameter 2 for the rate R0. This is also beneficial to some narrowband applications.
- The actual bit rates for R3, R4 and R5 are 15.95 kbit/s, 23.85 kbit/s and 31.75 kbit/s, respectively, and one or two bits are reserved for these layers for future algorithmic extension. This would be useful for the collaboration phase and may help to realize an easy integration of new technologies.
- Surplus of the 10 ms is gained with its 50 ms algorithmic delay. The super-wideband and stereo extensions can exploits this benefit in their algorithm developments.

5.3.1 Encoder description

Block diagram of the encoder is shown in Figure 11. A 16 kHz sampled wideband signal is inputted to down-sampler and the down-sampler converts the input signal to an 8 kHz sampled narrowband signal. Direct input of the 8 kHz sampled narrowband signal is also supported. The narrowband signal is fed to CELP module that outputs R1a, R2 and FEC bits. The R1a and R2 bits are generated by a cascaded CELP encoder in the CELP module. The cascaded CELP has two layers, a core layer and an enhancement layer, and they generate the R1a bits and R2 bits respectively. The FEC bits are generated by an FEC encoder in the CELP module.

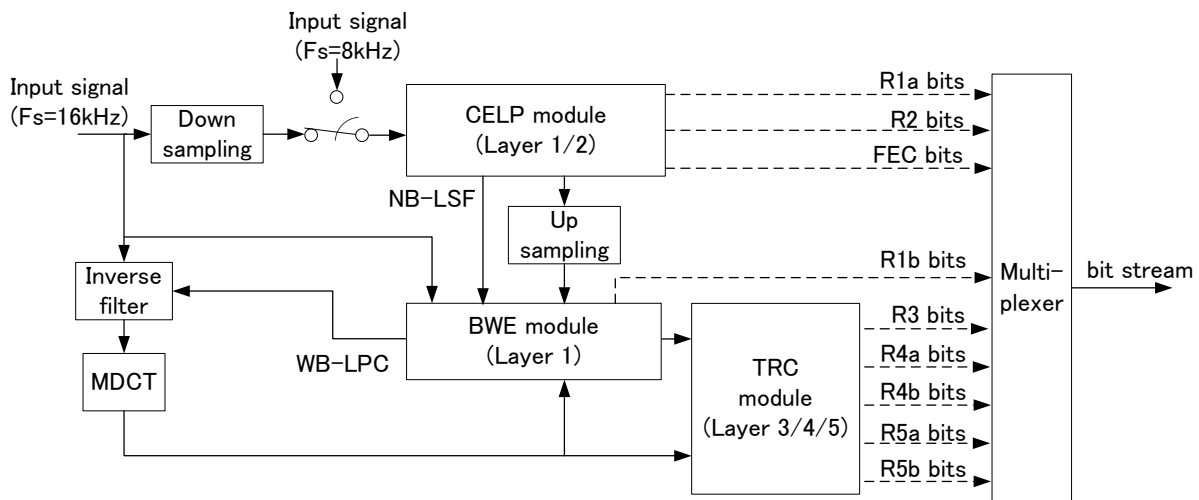


Figure 11 – Block diagram of the encoder.

A local-decoded signal of the CELP module is given to up-sampler to convert its sampling rate to 16 kHz. An upsampled signal is provided to the BWE module. Line spectral frequencies (LSF) quantized by the CELP module and the 16 kHz sampling input signal is also inputted to the BWE module.

In the BWE module, a spectral envelope component and a spectrum fine structure component of the local-decoded signal are separately extended to wideband from narrowband. The spectral envelope component is represented with LSF, and the spectrum fine structure component is represented with the spectrum of a linear prediction (LP) residual signal. The extension of the LSF is realized by a wideband LSF (WB-LSF) quantization based on a predictive quantization using quantized narrowband LSF (NB-LSF). The extension of the LP residual spectrum is performed by a kind of pitch filtering in the frequency domain. The BWE module produces R1b bits comprising extension information on both of the spectral envelope component and the spectrum fine structure component. The BWE module outputs quantized LP coefficients and an extended spectrum of the LP residual signal to an inverse filter and the TRC module respectively.

The inverse filter generates an LP residual signal of the 16 kHz sampling input signal and provides it to an MDCT section. The MDCT is performed every 40 ms windowed signal with 50 % overlap.

The TRC module encodes the errors between the MDCT coefficients obtained from the input signal and the signal decoded by the BWE module with 64 (R3) or 80 (R4a, R4b, R5a and R5b) bits steps by using a gain-shape vector quantization.

Bit stream structure of the codec is shown in Figure 12. Each bit stream is packetized frame by frame with a header comprising two 16-bit words. The header is used for synchronization and indicating the number of bits in the bit stream.

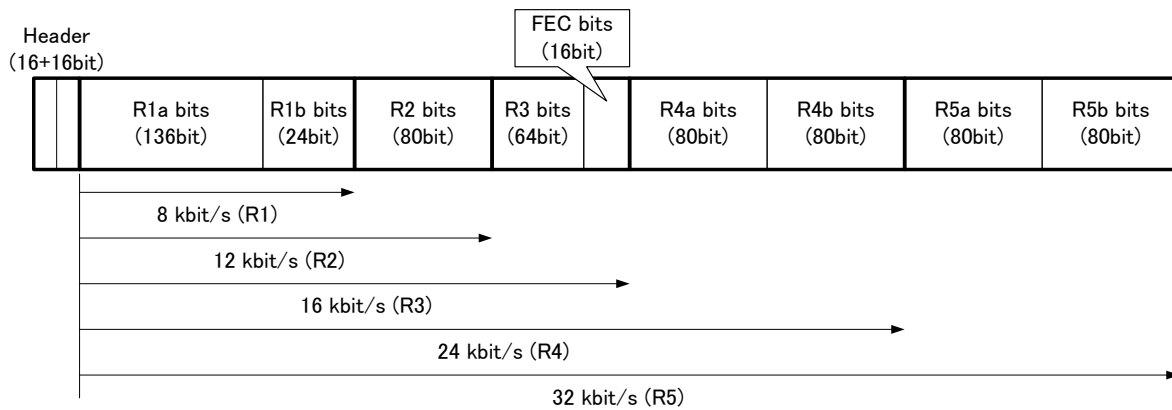


Figure 12 – Bit stream structure

5.3.2 CELP module

The CELP module processes an 8 kHz sampled narrowband signal. This module includes a cascaded CELP codec having two layers, Layer-1 and Layer-2. An FEC encoder is also included in this module for encoding the information served for frame erasure concealment.

Layer-1 encodes the input signal with the bit rate of 6.8 kbit/s. An error signal between the input signal and a signal decoded locally by Layer-1 becomes the subject to be encoded by Layer-2 with the bit rate of 4 kbit/s.

In Layer-1, a 12-order LP analysis is performed on every 20 ms frame. After converting to LSF, 12 LP coefficients are quantized in the LSF domain. The NB-LSF is quantized by an MA-predictive two-stage vector quantizer. Excitation parameters, adaptive codebook (ACB), fixed codebook (FCB) and gain parameters, are encoded on every 5 ms subframe so that perceptually weighted encoding distortion becomes minimum. Perceptually weighting is performed by a typical pole-zero filter but having an adaptive tilt control function. Differential quantization is exploited to quantize the ACB lags for even subframes with using 4 bits. The lags for odd subframes are combined and quantized with using 15 bits. This results in the 7.5 bits assignment for the 1st and 3rd subframes. The FCB is a multi-dispersed-pulse based codebook, which generates an FCB vector composed of one or more dispersed pulses. 16 bits are used for encoding combinations between pulses and their positions. An orthogonalized FCB search procedure is used to realize simultaneous optimization of ACB and FCB vectors. A pair of ACB and FCB gains is quantized by a 7-bit MA-predictive vector quantizer. The FCB vector and the gain pair are quasi-simultaneously optimized using a joint codebook search approach.

The Layer-2 has a similar structure of the Layer-1. That is, the encoding error of Layer-1 is further quantized by another CELP layer. However, the Layer-2 utilizes the LSF and ACB lag quantized by the Layer-1 as Layer-2 parameters, and therefore an FCB vector and a gain pair are the subject to be encoded in the Layer-2. The FCB vector is encoded with 14 bits, and the gain pair is encoded with 6 bits.

The FEC encoder quantizes the following parameters; the position, sign and amplitude of the last pitch pulse in the previous frame, the energy of the excitation signal in the previous frame, and the prediction gain of the LP filter in the previous frame. The pitch pulse amplitude, excitation energy, and the filter prediction gain are jointly vector-quantized in a logarithmic scale with using 8 bits. 7 bits and 1 bit are assigned respectively to encode the position and sign of the pitch pulse.

5.3.3 BWE module

Block diagram of the BWE module is shown in Figure 13. The BWE module extends the bandwidth of the local-decoded narrowband signal from the CELP module to wideband. In BWE module, extension processing is performed separately on a spectral envelope component and a spectrum fine structure component. The spectral envelope component is quantized as LSF, and the spectrum fine structure component is encoded as an LP residual spectrum.

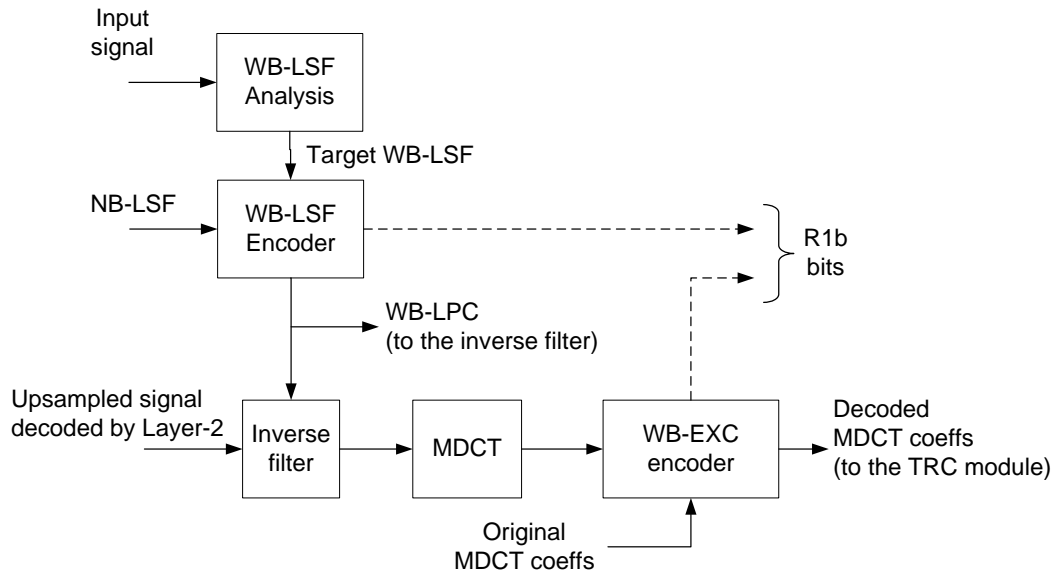


Figure 13 – Block diagram of the BWE module

A WB-LSF encoder is used to extend the NB-LSF quantized by the CELP module into WB-LSF. A total of 11 bits are used by the WB-LSF encoder for quantizing the inputted WB-LSF with predictive vector quantization in which the quantized NB-LSF is exploited efficiently for predicting the WB-LSF. A quantized WB-LSF is converted to LP coefficients (WB-LPC), which are provided to two inverse filters.

The inverse filter in the BWE module performs inverse filtering on the upsampled signal decoded by the Layer-2 to flatten its spectrum. The resultant signal is transformed into MDCT coefficients. At this stage, the valid bandwidth of the MDCT coefficients is still narrowband although the sampling rate of the upsampled signal decoded by the Layer-2 is 16 kHz.

A WB-EXC encoder performs to extend the narrowband MDCT coefficients into wideband MDCT coefficients. High frequency band MDCT coefficients of the original MDCT coefficients, which are obtained by applying the inverse filter to the 16 kHz sampling input signal, are approximated using the narrowband MDCT coefficients. The approximation algorithm has two main steps; a step of modification to the narrowband MDCT spectrum and a step of pitch filtering on the narrowband MDCT spectrum. The modification step is aimed at fitting the dynamic range of the narrowband MDCT spectrum to that of the high frequency band spectrum of the original. 4 bits are assigned to encode the information for this step. In the next step, single- or multi-tap pitch filtering is performed on the modified narrowband MDCT spectrum to generate the high frequency band MDCT coefficients. The basic concept of this pitch filtering is to preserve a harmonics structure of the high frequency band spectrum by utilizing the narrowband spectrum. The parameters, a pitch lag and gains for the approximated MDCT coefficients are sequentially encoded so that an encoding error is minimized; 4 and 5 bits are assigned to encode the lag and gains, respectively.

5.3.4 TRC module

The TRC module is a cascaded MDCT-based codec that has five layers; Layer 3, Layer 4a, Layer 4b, Layer 5a and Layer 5b. Each layer has similar structure as shown in Figure 14.

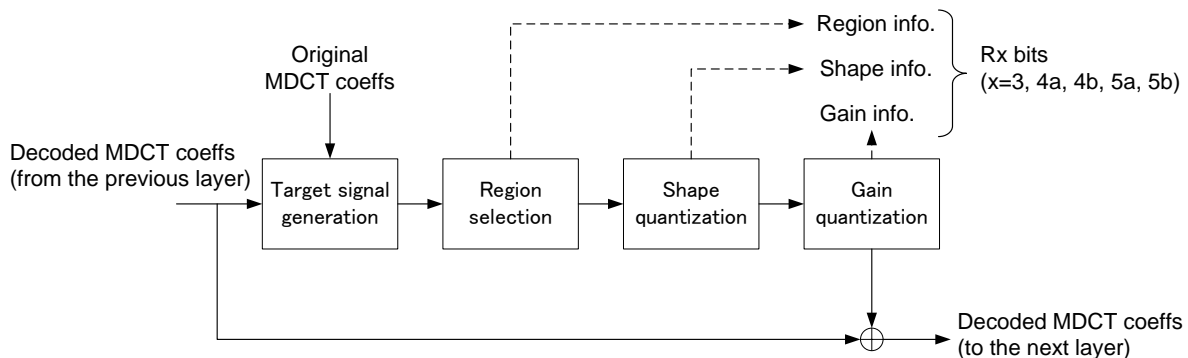


Figure 14 – Block diagram of the MDCT-based encoder

The target vector to be encoded by each layer consists of errors between the original MDCT coefficients and the MDCT coefficients decoded by the layer prior to each layer. The target vector is divided into 17 uniform sub-bands. We defined 8 regions which consist of five consecutive sub-bands. The eight regions are arranged to be overlapped with each neighbouring regions and to cover full-band of 7 kHz. Each layer of the TRC module encodes one of the eight regions. Using 3 bits, the region selection is done by referring the energy of the target vector in the region. Once the region is selected, the target vector in the region is quantized using a shape-gain vector quantization in which a shape and a gain of the target vector are sequentially quantized. Shape-vectors are formed from 5 main and 4 sub pulses or 5 main and 6 sub pulses, and the combination of those pulse positions and signs are encoded with 50 or 60 bits. Following to the shape quantization, a gain quantization is performed on calculated sub-band gains. Since the region contains 5 sub-band, 5 gains are obtained for the region and vector quantized using 10 or 16 bits. The vector quantization exploits a switched prediction scheme.

5.3.5 Decoder description

Block diagram of the decoder is shown in Figure 15. The input bit stream is de-multiplexed and the resultant bits are provided to corresponding modules i.e. CELP, BWE and TRC modules. There are two kinds of post-processing units in the decoder. One is an NB post-processing unit operated on a 8 kHz sampling signal, and the other is a WB post-processing unit operated on a 16 kHz sampling signal. Both of them have similar structure and comprise a pitch postfilter, a formant postfilter, a spectral tilt controller and a noise post-processor. These two post-processors do not activate simultaneously, and one of two is selectively enabled according to the number of receiving bits.

Since the input bit stream may be truncated, there are four possible configurations of decoding process depending on how many bits are received by the decoder. Each configuration is described below.

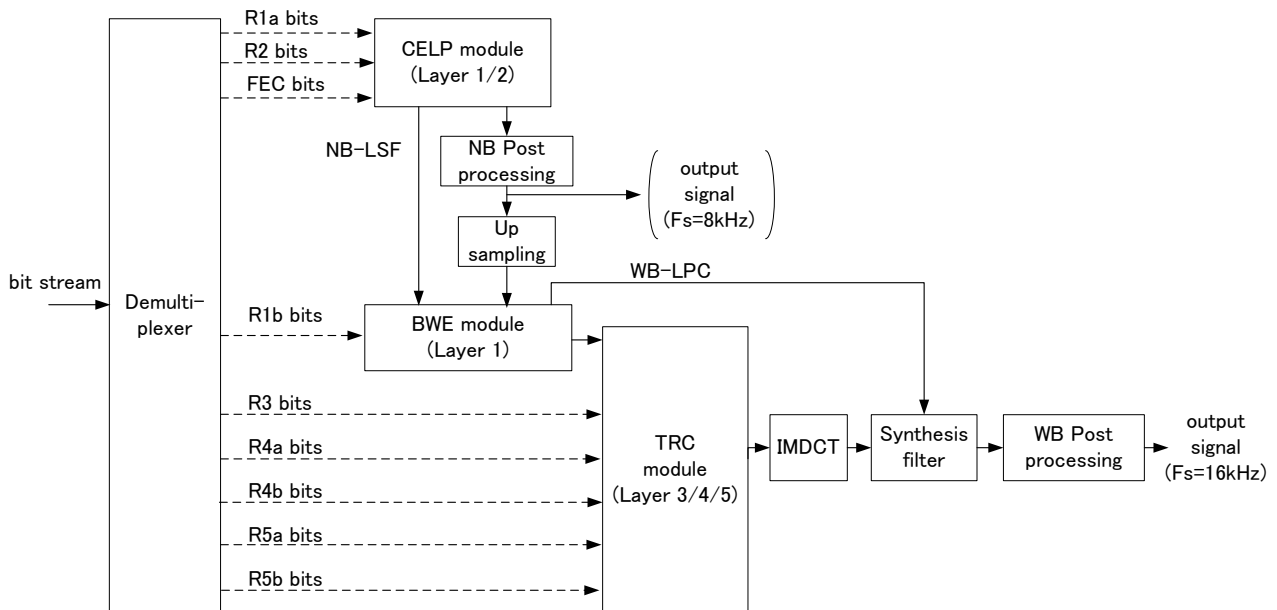


Figure 15 – Block diagram of the candidate decoder

5.3.6 The case of receiving no bits

In this case, a frame erasure concealment process is performed using the FEC bits received in the next frame or using the parameters decoded in the past. Since the next frame is supposed to be received and decoded for overlapping in the inverse MDCT operation, this procedure does not introduce any additional algorithmic delay.

5.3.7 The case of receiving R1 bits

In this case, the CELP and BWE modules and NB post-processing unit are activated. The CELP module generates a signal decoded by the Layer-1 from R1a bits and the decoded signal is inputted to the BWE module after passing through the NB post-processing unit and a up-sampler.

The BWE module extends the bandwidth of the signal imputed from the up-sampler to wideband by using parameters decoded from R1b bits.

5.3.8 The case of receiving R1 to R2 bits

In this case, the CELP and BWE modules and NB post-processing unit are activated as in the case of Section 3.2. But the CELP module generates a signal decoded by the Layer-2 from R1a bits and R2 bits. Therefore, the signal to be extended by the BWE module is the signal decoded by the Layer 2 after passing through the NB post-processing unit and a up-sampler.

5.3.9 The case of receiving R1 to R3 or the higher layer bits

In this case, all the modules except the NB post-processing unit are enabled. The CELP module generates the signal decoded by the Layer-2. After passing through the up-sampler, the decoded signal is inputted to the BWE module. The BWE module extends the bandwidth of the Layer-2 decoded signal to wideband.

The TRC module piles decoded residual MDCT coefficients on top of the extended Layer-2 MDCT coefficients. The fidelity of the decoded residual MDCT coefficients depends on the number of bits received by the decoder.

5.3.10 Algorithmic delay

The overall algorithmic delay is 50 ms. The details are as follows.

- Frame length is 20 ms.
- Look-ahead for LPC analysis is 5 ms.
- Delay for overlapping in Inverse MDCT is 20 ms.
- Delay for sampling rate conversions is 5 ms (including upsampling and downsampling).

5.3.11 Complexity

The complexity of the candidate codec is shown in Table 2.

The memory size of the candidate codec is shown in Table 3.

Table 2: Computational complexity of the codec

Layer	Encoder (WMOPS)	Decoder (WMOPS)
R1	39.5	2.9
R2	8.5	1.0
R3	1.5	0.1
R4	2.5	0.3
R5	2.5	0.3
Others	7.2 (*1)	14.0 (*2)
Total	61.7	18.7

(*1) It contains two MDCTs, sampling rate conversion and two inverse filters.

(*2) It contains MDCT/IMDCT, sampling rate conversion, inverse/synthesis filter and post-processing.

Table 3: Memory requirements for the codec.

Data-ROM	22.5 kwords
RAM	7.3 kwords

5.4 Nokia and VoiceAge Candidate [19]

The Q9 codec is an embedded codec comprising 5 layers where higher layer bitstreams can be discarded without affecting the decoding of the lower layers. The layers will be referred to as L1 (core layer) through L5 (the highest extension layer).

The codec can accept both wideband (WB) signals sampled at 16 kHz, and narrowband (NB) signals sampled at 8 kHz. Similarly, the codec output can be wideband or narrowband. While the WB rendering is provided for all the layers, the narrowband rendering is implemented only for L1 and L2. Independently of the input signal sampling rate, L1 and L2 internal sampling is 12.8 kHz.

The codec has been designed with the primary objective of a high-performance wideband speech coding for error prone telecommunications channels, without compromising the quality for narrowband speech signals or wideband music signals.

The codec delay is variable depending on the sampling rate of the input and the output, not exceeding 56 ms. For example for WB input and WB output, the overall algorithmic delay is 54.75 ms. The delay consists of a 20 ms frame, 1.875 ms delay of input and output resampling filters, 11.875 ms for the encoder lookahead, 1 ms of post-filtering delay, and one 20 ms frame delay at decoder needed for overlap-add operation of higher layer transform coding. The one-frame transform coding delay is optional for L1 and L2 and can be deactivated at the decoder using `-ld` command-line option.

The codec is equipped with a basic discontinuous transmission (DTX) scheme. The DTX operation can be activated through `-dtx` encoder command-line option.

In the Q9 Terms of Reference, it is stated that: “Interoperability with other ITU-T speech encoding standards and Interoperability with 2G and 3G mobile radio systems is desirable. Interoperability with G.722.2 at 12.65 kbit/s is of particular interest”. In order to meet this objective, the Nokia/VoiceAge candidate is equipped with an option to allow for interoperability with G.722.2 at 12.65 kbit/s. . When invoked at the encoder, this option allows using G.722.2 mode 2 (12.65 kbit/s) to replace L1 and L2. The enhancement layers L3, L4 and L5 are similar to the default operation. The addition of this option is straight forward since the core ACELP layer is similar to G.722.2 (e.g. operation at 12.8 kHz internal sampling, use of the same pre-emphasis and perceptual weighting)

5.4.1 Layer Structure

The 5-layer structure is outlined in Table 4. The internal sampling of the first two layers is 12.8 kHz, independently of the input signal sampling rate. Higher layer coding for WB rendering is then done at 16 kHz sampling rate.

Table 4: Layer structure for default operation

Layer	Bitrate (kbit/s)	Technique		Sampling rate (kHz)	
L1	8	Classification-based ACELP core layer		12.8	
L2	+4	Algebraic codebook layer		12.8	
L3*	+4	FEC	MDCT	12.8	16
L4*	+8	MDCT		16	
L5*	+8	MDCT		16	

* Not implemented for narrowband I/O

The core layer encoding takes advantage of the performance of signal classification based encoding. Four distinct signal classes are considered for different encoding of each frame: Unvoiced coding optimized for encoding unvoiced speech frames, Voiced coding optimized for encoding quasi-periodic segments with smooth pitch evolution, Transition mode for encoding frames following voiced onsets designed to minimize error propagation in case of frame erasures, and Generic coding for remaining frames. Further, bit allocation and quantization is optimized separately for NB and WB inputs. All core layer coding modes are waveform encoded and except for the Unvoiced coding they use ACELP technology. Unvoiced frames are coded by means of a Gaussian codebook.

Layer 2 uses algebraic codebooks to further minimize the perceptually weighted coding error from L1. As the sampling rate is 12.8 kHz, for WB rendering, the AMR-WB bandwidth extension is used to generate the missing 6.4-7 kHz spectral band.

To enhance the codec frame erasure concealment (FEC), side information is computed and transmitted in L3. Independently of the core layer coding mode, the side information includes signal

classification. In Transition coding mode, the two first ISF indices of the previous frame are also transmitted for better estimation of the erased frame LP synthesis filter. In other coding modes, the side information further consists of pitch-synchronous synthesized signal energy and basic phase information.

For WB output, the weighted error signal after L2 encoding is coded using overlap-add transform coding based on the modified discrete cosine transform (MDCT). The transform coding is done at 16 kHz sampling rate, i.e. covering the whole WB band.

5.4.2 Encoder Overview

Prior to encoding, the input signal is high-pass filtered at 25 Hz for WB input and 100 Hz for NB input. The sampling rate is then converted to 12.8 kHz for L1 and L2 encoding, and the input signal is pre-emphasized to attenuate low frequencies. An FFT-based spectral analysis is then performed twice per frame for use in voice activity detection (VAD) and signal classification algorithms. The signal energy is computed for each perceptual critical band [1].

The core layer is based on the Code-excited Linear Prediction (CELP) technology where speech signal is modeled by an excitation signal passed through a linear prediction (LP) synthesis filter. The LP filter is quantized in Immitance spectral frequencies (ISFs) [2] domain using Safety-Net [3] approach and a multi-stage vector quantization (MSVQ) for generic and voiced modes. The unvoiced and transition modes are quantized without prediction in order to reduce the error propagation. The effect of the Safety-Net approach is to reduce the error propagation due to ISF prediction in case of frame erasures hitting segments where the speech spectral envelope evolves rapidly.

The open-loop (OL) pitch analysis is done by a pitch-tracking algorithm to insure a smoothness of the pitch contour by exploiting adjacent values. Further, two concurrent pitch evolution contours are compared and the track that yields smoother contour is selected.

5.4.3 Classification Based Core Layer (Layer 1)

To get maximum speech coding performance at 8 kbit/s, the core layer uses signal classification and four distinct coding modes tailored for each class of speech signal, namely Unvoiced coding, Voiced coding, Transition coding and Generic coding. All the coding modes are based on the standard CELP paradigm (Figure 16).

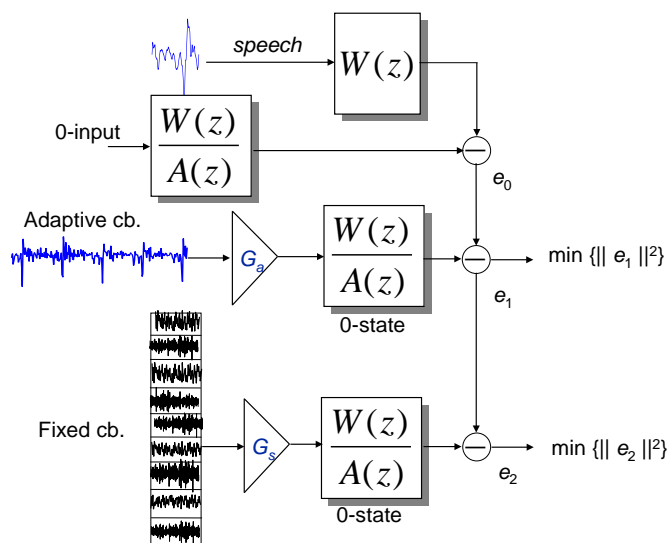


Figure 16 – CELP encoder scheme

The frames to be encoded with Unvoiced coding mode are selected first. Unvoiced coding is designed to encode unvoiced speech frames and, in the absence of DTX, most of the inactive frames. In Unvoiced coding, the adaptive codebook is not used and the excitation is selected from a Gaussian codebook. The excitation gain is coded with a memoryless scalar quantizer. Quasi-periodic segments are encoded with Voiced coding mode. Voiced coding selection is conditioned by a smooth pitch evolution. The Transition coding mode has been designed to enhance the codec performance in presence of frame erasures by limiting past frame information usage. To minimize at the same time its impact on clean channel performance, it is used only on most critical frames from frame erasure point of view. In Transition coding frame, the adaptive codebook in the subframe containing the glottal impulse of the first pitch period is replaced with a fixed codebook. In the preceding subframes, the adaptive codebook is omitted. In the following subframes, a legacy Algebraic CELP (ACELP) is used. All other frames (in absence of DTX) are processed through a Generic ACELP.

To further reduce frame error propagation in case of frame erasures, gain coding does not use prediction from previous frames. Instead, the algebraic codebook frame gain for Voiced, Transition or Generic coding modes is first estimated and coded for the whole frame using three bits and the gain quantization is further refined for each subframe. This gain quantization coding is similar to the scheme used in the AMR-WB+ codec.

5.4.4 Second Stage ACELP encoding (Layer 2)

In L2, the quantization error from the core layer is encoded using again the algebraic codebooks. The principle is outlined in Figure 17.

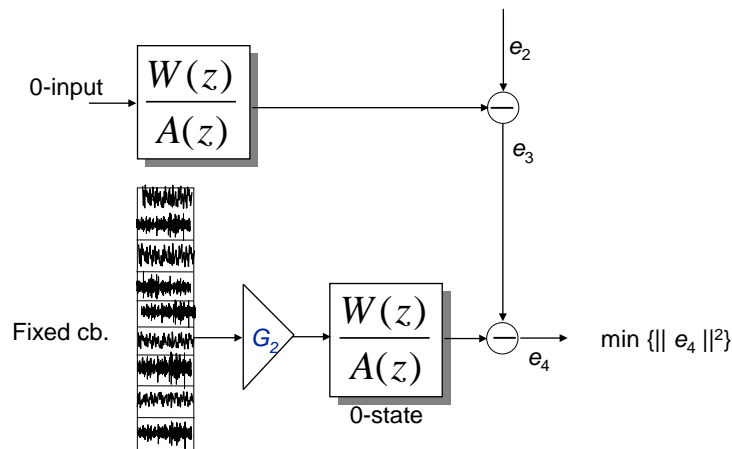


Figure 17 – Second stage ACELP coding

The ACELP layers (L1 and L2) operate at 12.8 kHz sampling rate. The output from L2 thus consists of a synthesized signal encoded in 0-6.4 kHz frequency band. The AMR-WB bandwidth extension is used to generate the missing 6.4-7 kHz bandwidth.

5.4.5 Frame Erasure Concealment Side Information (Layer 3)

The codec has been designed with the focus on performance in frame erasure conditions. As mentioned previously, several techniques limiting the frame error propagation have been implemented, namely the Transition coding mode, the Safety-Net approach for ISF coding and memoryless gain quantization.

To further enhance the performance in frame erasure conditions, side information is sent in layer 3. The side information consists of class information for all coding modes. Previous frame spectral

envelope information is further transmitted for core layer Transition coding. For other core layer coding modes, a phase information and the pitch-synchronous energy of the synthesized signal are also sent. These parameters help the concealment of the erased frame and, more importantly, the recovery of the decoder after the erasure.

The class information helps mainly to enhance the concealed frame in the sense that if the class indicates a quasi stationary segment, the waveform is basically repeated from the previous frame. On the contrary, the energy is attenuated rapidly during transient events. The phase information helps to maintain the phase synchrony between the encoder and the decoder during a voiced frame erasure. Further this phase information is also used for artificial onset reconstruction in case a voiced onset is lost and the first frame after the erasure is not encoded with the Transition mode. The energy parameter is used to scale the synthesis in the first good frame after an erasure [5, 6]. The previous frame spectral envelope information, represented by the two first ISF indices, is used for better LP synthesis filter estimation at the decoder.

5.4.6 Transform Coding of Higher Layers (Layers 3, 4, 5)

The error resulting from the 2nd stage CELP coding in L2 is further quantized in L3, L4 and L5 using MDCT with 50% over-lap add. The transform coding is performed at 16 kHz sampling frequency and it is implemented only for WB rendering.

The higher layer encoding is outlined in Figure 18. The de-emphasized synthesis from L2 is resampled to 16 kHz sampling rate and high-pass filtered, then the bandwidth extension is added to generate the 6.4-7 kHz bandwidth. The resulting signal is then subtracted from the high-pass filtered input signal to obtain the error signal which is weighted and encoded using MDCT. The MDCT coefficients are quantized using scalable algebraic vector quantization. The MDCT is computed every 20 ms, and its spectral coefficients are quantized in 8-dimensional blocks. An audio cleaner is applied, derived from the spectrum of the original signal.

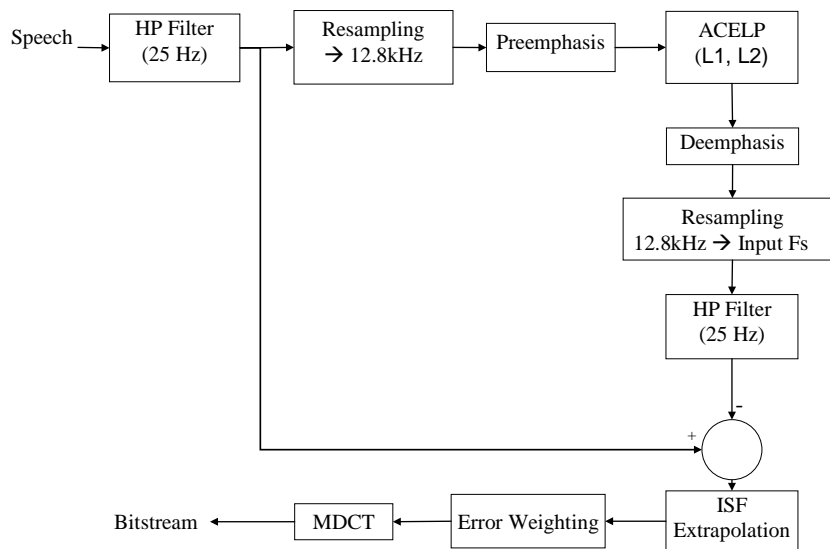


Figure 18 – MDCT encoding of higher layers

The transform coefficients are quantized in the following way. Global gains are transmitted in L3. Further, few bits are used for high-frequency compensation. Remaining L3 bits are used for quantization of MDCT coefficients. The L4 and L5 bits are used such that the performance is maximized independently at L4 and L5 levels.

5.4.7 Decoder Overview

Figure 19 shows the block diagram of the decoder. In each 20-ms frame, the decoder can receive any of the supported bit rates, from 8 kbit/s up to 32 kbit/s. This means that the decoder operation is conditional to the number of bits, or layers, received in each frame. In Figure 19, we assume that the output is WB and that at least Layers 1, 2, 3 have been received at the decoder.

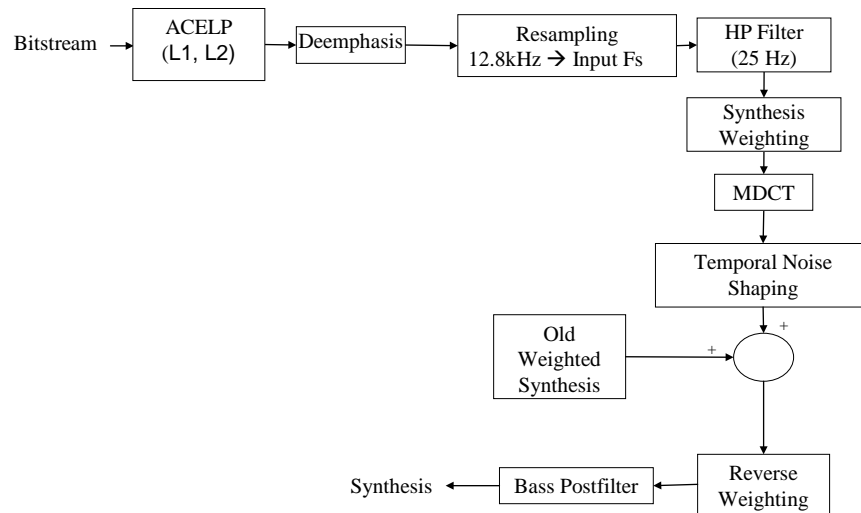


Figure 19 – Decoder overview

First, the core layer and the ACELP enhancement layer (L1 and L2) are decoded. The synthesized signal is then de-emphasized, resampled to 16 kHz and high-pass filtered. Transform coding enhancement layers are added to the perceptually weighted synthesis and temporal noise shaping is applied. The weighted synthesis is then added to the synthesis of the previous frame with 50% overlap. Reverse perceptual weighting is applied to restore the synthesized WB signal, followed by a pitch post-filter [4]. If higher layers (L3-L5) are not available or the quality increment of these layers is not needed, lower delay output can be commanded by the decoder.

In NB case only L1 and L2 are supported. The one-frame decoder delay serves here to improve the concealment of frame erasures and is optional at the decoder. Further, legacy pitch and formant post filter is used for NB output.

In case of frame erasures and if the one-frame delay is available at the decoder, the construction of the missing frames is delayed by 20 ms. The pitch evolution in those frames can thus be estimated more accurately.

5.4.8 Complexity evaluation

The complexity is estimated to be around 70 WMOPS.

6 Test Results in G.EV-VBR Selection Phase

Q7/12 reviewed the work done by the two contracted host laboratories (ARCON Corp. and France Telecom). The host laboratory reports are [AH-07-10 and AH-07-12].

France Telecom provided the report on computation of gains [AH-07-08], the largest absolute value was 1.3dB. There was no evidence of impact of gains on the results.

France Telecom provided the frequency response of the four codecs under test [AH-07-17]. Two candidates i.e. CuT2 and CuT4 showed noticeable spectral components above 7kHz, another candidate CuT3 also showed some spectral components above 7kHz. Q7/12 was unable to determine whether this impacted on speech and/or music quality without further investigation. Candidate CuT1 showed attenuation up to 20dB around 3.5kHz at low bit rate.

Q7/12 reviewed the reports of the listening laboratories [AH-07-02 from Dynastat, AH-07-04 from Nokia, AH-07-05 from VoiceAge, AH-07-06 from BIT, AH-07-07 from France Telecom, AH-07-11 from NTT-AT and AH-07-13 from ARCON Corp.].

6.1 Global Analyses for the Selection Phase for the Embedded-VBR Speech Codec [21]

The Selection Test Plan included nine subjective tests designed to evaluate the performance of four candidate codecs relative to that of standard reference codecs. Dynastat performed the activities of the Global Analysis Laboratory as described in the Selection Test Plan. This document provides a description and the results of those activities.

6.1.1 Organization of the Selection Test

The selection test involved nine subjective experiments in seven categories:

- Experiment 1: Input Level performance for narrowband speech (ACR)
- Experiment 2: Clean wideband speech performance
 - Exp. 2a: Clean wideband speech performance on R1, R2 (ACR)
 - Exp. 2b: Clean wideband speech performance on R3, R4 (ACR)
 - Exp. 2c: Clean wideband speech performance on R5 (ACR)
- Experiment 3: Music performance (ACR)
- Experiment 4: Car noise performance for narrowband speech (DCR)
- Experiment 5: Street noise performance for narrowband speech (DCR)
- Experiment 6: Interfering Talker performance for wideband speech (DCR)
- Experiment 7: Office noise performance for wideband speech (DCR)

Each of the nine tests was conducted in two Listening Laboratories where each lab used a different language. Table 5 summarizes the distribution of the subjective tests and the Listening Labs.

Table 5: Organization of the Selection Test

ID	Listening Lab	Experiments	Language
A	Arcon	2c, 6	North American English
B	BIT	1, 2a, 7	Chinese
D	Dynastat	2a, 4	North American English
F	France Telecom	2c, 3	French
J	NTT-AT	2b, 4, 5	Japanese
N	Nokia	2b, 5, 7	Finnish
V	VoiceAge	1, 3, 6	Canadian French

The test methodologies included the Absolute Category Rating (ACR) method and the Degradation Category Rating (DCR) method, both described in ITU-T Recommendation P.800. All of the subjective tests involved the use of four listening panels of eight subjects each. Eight of the nine tests involved the evaluation of speech and one test, Exp.3, involved music signals. The speech tests involved six talkers, three males and three females, and the music test involved four music genres.

6.1.2 The Global Analysis Laboratory

The test plan was originally written and organized for five candidate codecs. However, after the withdrawal of one of the candidates, there was a need to replace the test conditions in each experiment originally reserved for the fifth candidate. It was agreed that a reference codec recently standardized in ITU-T/SG16, G.729.1, would be used to fill the test conditions originally intended for the fifth candidate. The question was then raised as to whether the G.729.1 codec should be included in the Dunnett's ToR tests as an additional treatment condition. At first, the GAL recommended that the replacement codec, G.729.1, should not be used in the Dunnett's tests since it was not one of the candidate codecs and its inclusion might contaminate the statistical tests for the four candidates. However, an extensive comparison of the results of the Dunnett's ToR tests with five codecs (reference plus four candidates) and with six codecs (reference plus four candidates plus G.729.1) has revealed only minor differences between the two sets of analyses. Appendix C contains a comparison of the Dunnett's ToR tests using five codecs and six codecs in the statistical analyses. The Dunnett's ToR tests reported in the remainder of this document are based on the scores for six codecs — the control condition (the reference codec) plus four candidate codecs (CuT1, CuT2, CuT3, CuT4) plus G.729.1.

6.1.3 Test Results

Mean Scores by Experiment and Listening Lab

Table 6, Table 7, Table 8, Table 9, Table 10, and Table 11 present the scores for the nine subjective experiments for each of the two listening labs. Each table shows the test conditions and overall Means and Standard Deviations for each of the two Listening Labs that conducted the experiment. For the experiments involving speech signals (i.e., all experiments except Exp.3) the results are based on 192 votes — 32 subjects x 6 talkers. For Exp.3, the results are based on 128 votes — 32 subjects x 4 music genres. Complete results by experiment are contained in an Excel file imbedded in Appendix A of the original contribution.

Table 6: ACR Experiment 1¹ – Clean Narrowband Speech on R1 and R2

Exp.1	NB-Clean	Input level	FER	Lab B		Lab V	
				Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 8 dB	Nom	0%	1.599	0.709	1.104	0.306
c02	MNRU, Q = 14 dB	Nom	0%	2.083	0.808	1.609	0.595
c03	MNRU, Q = 20 dB	Nom	0%	2.833	0.795	2.224	0.676
c04	MNRU, Q = 26 dB	Nom	0%	3.287	0.756	3.052	0.811
c05	MNRU, Q = 32 dB	Nom	0%	3.609	0.830	3.787	0.838
c06	Direct	Nom	0%	3.583	0.795	4.495	0.655
c07	G.729	Nom	0%	3.703	0.793	4.083	0.740
c08	G.729E	Nom	0%	3.552	0.784	4.287	0.684
c09	G.729E	High	0%	3.557	0.836	4.333	0.697
c10	G.729E	Low	0%	3.380	0.848	3.958	0.751
c11	G.729E	Nom	3%	3.255	0.820	4.016	0.755
c12	CuT1 @R1	Nom	0%	3.922	0.868	3.984	0.918
c13	CuT1 @R2	Nom	0%	3.849	0.852	4.302	0.740
c14	CuT1 @R1	High	0%	3.854	0.818	3.958	0.920
c15	CuT1 @R1	Low	0%	3.953	0.852	3.943	0.851
c16	CuT1 @R1	Nom	3%	3.495	0.954	3.682	0.903
c17	CuT2 @R1	Nom	0%	3.781	0.828	4.307	0.748
c18	CuT2 @R2	Nom	0%	3.662	0.853	4.297	0.656
c19	CuT2 @R1	High	0%	3.688	0.823	4.365	0.696
c20	CuT2 @R1	Low	0%	3.646	0.831	4.208	0.722
c21	CuT2 @R1	Nom	3%	3.531	0.862	4.047	0.821
c22	CuT3 @R1	Nom	0%	3.594	0.826	4.037	0.762
c23	CuT3 @R2	Nom	0%	3.583	0.814	4.297	0.648
c24	CuT3 @R1	High	0%	3.516	0.806	4.068	0.738
c25	CuT3 @R1	Low	0%	3.073	0.841	3.573	0.762
c26	CuT3 @R1	Nom	3%	3.323	0.844	3.823	0.831
c27	CuT4 @R1	Nom	0%	3.703	0.806	4.318	0.765
c28	CuT4 @R2	Nom	0%	3.672	0.832	4.271	0.709
c29	CuT4 @R1	High	0%	3.740	0.762	4.391	0.678
c30	CuT4 @R1	Low	0%	3.662	0.796	4.208	0.771
c31	CuT4 @R1	Nom	3%	3.484	0.831	4.125	0.690
c32	G.729.1 @8kb/s	Nom	0%	4.177	0.904	4.240	0.727
c33	G.729.1 @12kb/s	Nom	0%	3.901	1.011	4.349	0.715
c34	G.729.1 @8kb/s	High	0%	4.094	0.887	4.214	0.725
c35	G.729.1 @8kb/s	Low	0%	3.875	0.968	3.995	0.822
c36	G.729.1 @8kb/s	Nom	3%	3.823	1.121	3.917	0.846

¹ For Lab B most (20 out of 29) of the test conditions scored higher than the unprocessed source — Direct = 3.583. Furthermore, the correlation between the scores for the test-conditions (excluding direct and MNRU conditions) across the two listening labs that conducted Exp.1a was 0.43. For the other eight experiments the minimum correlation was 0.79. These observations raise concerns about the validity of the test scores for Exp.1 conducted in Lab B.

Table 7: ACR Experiment 2a - Clean Wideband Speech on R1 and R2

Exp.2a	WB-Clean on R1, R2	Input level	FER	Switching	Lab B		Lab D	
					Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 5 dB	Nom	0%	no	1.224	0.454	1.135	0.514
c02	MNRU, Q = 13 dB	Nom	0%	no	2.037	0.789	1.443	0.558
c03	MNRU, Q = 21 dB	Nom	0%	no	3.141	0.995	2.453	0.849
c04	MNRU, Q = 29 dB	Nom	0%	no	3.797	0.913	3.224	1.022
c05	MNRU, Q = 37 dB	Nom	0%	no	4.021	0.844	3.974	0.929
c06	MNRU, Q = 45 dB	Nom	0%	no	4.037	0.821	4.297	0.717
c07	Direct	Nom	0%	no	4.089	0.778	4.271	0.716
c08	G.722.2 – 8.85k	Nom	0%	no	3.677	0.943	3.802	0.788
c09	G.722.2 – 8.85k	High	0%	no	3.693	0.847	3.635	0.858
c10	G.722.2 – 8.85k	Low	0%	no	3.391	0.798	3.323	0.938
c11	G.722.2 – 8.85k	Nom	3%	no	3.182	1.015	3.214	0.845
c12	CuT1 @R1	Nom	0%	no	4.099	0.872	4.026	0.821
c13	CuT1 @R1	High	0%	no	3.807	1.018	3.870	0.779
c14	CuT1 @R1	Low	0%	no	4.234	0.911	3.839	0.862
c15	CuT1 @R1-R5	Nom	0%	fast	3.885	0.942	4.162	0.766
c16	CuT1 @R1	Nom	3%	no	3.401	1.163	3.495	0.868
c17	CuT2 @R1	Nom	0%	no	4.073	0.841	4.104	0.799
c18	CuT2 @R1	High	0%	no	3.984	0.822	4.016	0.847
c19	CuT2 @R1	Low	0%	no	3.938	0.854	4.010	0.812
c20	CuT2 @R1-R5	Nom	0%	fast	3.938	0.896	4.156	0.757
c21	CuT2 @R1	Nom	3%	no	3.625	0.924	3.807	0.850
c22	CuT3 @R1	Nom	0%	no	3.859	0.952	3.839	0.825
c23	CuT3 @R1	High	0%	no	3.781	0.906	3.807	0.818
c24	CuT3 @R1	Low	0%	no	3.380	0.952	3.557	0.890
c25	CuT3 @R1-R5	Nom	0%	fast	3.766	0.905	3.906	0.826
c26	CuT3 @R1	Nom	3%	no	3.396	1.018	3.688	0.790
c27	CuT4 @R1	Nom	0%	no	4.010	0.773	4.151	0.801
c28	CuT4 @R1	High	0%	no	3.880	0.807	3.938	0.803
c29	CuT4 @R1	Low	0%	no	3.880	0.893	4.005	0.762
c30	CuT4 @R1-R5	Nom	0%	fast	3.865	0.864	4.083	0.747
c31	CuT4 @R1	Nom	3%	no	3.729	0.943	3.880	0.826
c32	G.729.1 @14kb/s	Nom	0%	no	3.552	0.948	3.885	0.804
c33	G.729.1 @14kb/s	High	0%	no	3.583	0.906	3.865	0.801
c34	G.729.1 @14kb/s	Low	0%	no	3.208	1.043	3.714	0.842
c35	G.729.1 @14, (14), 16, 24 & 32kb/s	Nom	0%	fast	3.698	0.894	4.031	0.779
c36	G.729.1 @14kb/s	Nom	3%	no	3.052	0.936	3.599	0.863

Table 8: ACR Experiment 2b - Clean Wideband Speech on R3 and R4

Exp.2b	WB-Clean on R3, R4	Input level	FER	Lab J		Lab N	
				Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 5 dB	Nom	0%	1.005	0.072	1.005	0.072
c02	MNRU, Q = 15 dB	Nom	0%	1.266	0.455	1.323	0.551
c03	MNRU, Q = 25 dB	Nom	0%	2.010	0.522	1.922	0.825
c04	MNRU, Q = 35 dB	Nom	0%	2.984	0.935	2.984	0.895
c05	MNRU, Q = 45 dB	Nom	0%	3.964	0.858	4.031	0.824
c06	Direct	Nom	0%	4.130	0.792	4.276	0.787
c07	G.722.2 -8.85k	Nom	0%	3.063	0.860	3.276	0.851
c08	G.722.2 -12.65k	Nom	0%	3.385	0.891	3.729	0.806
c09	G.722.2 -12.65k	High	0%	3.583	0.858	3.865	0.833
c10	G.722.2 -12.65k	Low	0%	2.469	0.862	2.589	0.972
c11	G.722.2 -15.85k	Nom	0%	3.438	0.907	3.776	0.810
c12	CuT1 @R3	Nom	0%	4.005	0.871	3.974	0.815
c13	CuT1 @R3	High	0%	4.042	0.867	3.938	0.878
c14	CuT1 @R3	Low	0%	4.063	0.848	3.839	0.831
c15	CuT1 @R3	Nom	6%	2.974	0.984	3.203	0.990
c16	CuT1 @R4	Nom	0%	4.021	0.856	4.099	0.809
c17	CuT2 @R3	Nom	0%	3.917	0.827	4.010	0.812
c18	CuT2 @R3	High	0%	3.901	0.822	3.958	0.778
c19	CuT2 @R3	Low	0%	3.630	0.923	3.802	0.827
c20	CuT2 @R3	Nom	6%	3.365	0.905	3.682	0.811
c21	CuT2 @R4	Nom	0%	3.880	0.832	4.073	0.762
c22	CuT3 @R3	Nom	0%	3.510	0.874	3.635	0.839
c23	CuT3 @R3	High	0%	3.552	0.896	3.734	0.823
c24	CuT3 @R3	Low	0%	2.542	0.920	2.760	0.834
c25	CuT3 @R3	Nom	6%	3.073	0.918	3.203	0.853
c26	CuT3 @R4	Nom	0%	3.188	0.803	3.167	0.900
c27	CuT4 @R3	Nom	0%	3.938	0.777	4.068	0.766
c28	CuT4 @R3	High	0%	3.813	0.890	3.901	0.809
c29	CuT4 @R3	Low	0%	3.641	0.881	3.792	0.824
c30	CuT4 @R3	Nom	6%	3.396	0.926	3.755	0.836
c31	CuT4 @R4	Nom	0%	4.005	0.877	4.037	0.858
c32	G.729.1 @16kb/s	Nom	0%	3.563	0.848	3.797	0.816
c33	G.729.1 @16kb/s	High	0%	3.635	0.876	3.771	0.792
c34	G.729.1 @16kb/s	Low	0%	2.927	0.952	2.865	0.928
c35	G.729.1 @16kb/s	Nom	6%	2.807	0.874	3.318	0.830
c36	G.729.1 @24kb/s	Nom	0%	3.547	0.867	3.813	0.784

Table 9: ACR Experiment 2c - Clean Wideband Speech on R5

Exp.2c	WB-Clean on R5	Input level	FER	Lab A		Lab F	
				Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 5 dB	Nom	0%	1.219	0.451	1.037	0.188
c02	MNRU, Q = 13 dB	Nom	0%	1.964	0.726	1.307	0.495
c03	MNRU, Q = 21 dB	Nom	0%	2.719	0.741	1.880	0.664
c04	MNRU, Q = 29 dB	Nom	0%	3.505	0.812	2.662	0.705
c05	MNRU, Q = 37 dB	Nom	0%	4.078	0.818	3.620	0.816
c06	MNRU, Q = 45 dB	Nom	0%	4.313	0.699	4.250	0.671
c07	Direct	Nom	0%	4.417	0.642	4.370	0.634
c08	G.722.2 – 12.65k	Nom	0%	4.031	0.716	3.995	0.667
c09	G.722.2 – 23.85k	Nom	0%	4.063	0.728	4.078	0.701
c10	G.722.2 – 23.85k	High	0%	4.260	0.659	4.083	0.711
c11	G.722.2 – 23.85k	Low	0%	3.354	0.779	3.458	0.811
c12	CuT1 @R5	Nom	0%	4.542	0.629	4.365	0.633
c13	CuT1 @R5	High	0%	4.505	0.614	4.333	0.608
c14	CuT1 @R5	Low	0%	4.323	0.647	4.359	0.623
c15	CuT1 @R5	Nom	3%	4.401	0.664	3.953	0.852
c16	CuT1 @R5	Nom	0,2,4,6,10%	4.531	0.622	4.266	0.692
c17	CuT2 @R5	Nom	0%	4.297	0.671	4.307	0.601
c18	CuT2 @R5	High	0%	4.287	0.706	4.188	0.644
c19	CuT2 @R5	Low	0%	4.193	0.655	4.224	0.669
c20	CuT2 @R5	Nom	3%	4.193	0.752	3.958	0.730
c21	CuT2 @R5	Nom	0,2,4,6,10%	4.313	0.652	4.214	0.664
c22	CuT3 @R5	Nom	0%	3.724	0.739	3.677	0.786
c23	CuT3 @R5	High	0%	3.839	0.766	3.547	0.751
c24	CuT3 @R5	Low	0%	3.005	0.835	3.115	0.791
c25	CuT3 @R5	Nom	3%	3.583	0.747	3.203	0.803
c26	CuT3 @R5	Nom	0,2,4,6,10%	3.771	0.752	3.526	0.779
c27	CuT4 @R5	Nom	0%	4.370	0.689	4.292	0.638
c28	CuT4 @R5	High	0%	4.219	0.719	4.255	0.657
c29	CuT4 @R5	Low	0%	4.125	0.734	4.177	0.694
c30	CuT4 @R5	Nom	3%	4.229	0.671	4.052	0.722
c31	CuT4 @R5	Nom	0,2,4,6,10%	4.307	0.667	4.297	0.648
c32	G.729.1 @32kb/s	Nom	0%	4.177	0.679	4.135	0.657
c33	G.729.1 @32kb/s	High	0%	4.224	0.707	4.271	0.647
c34	G.729.1 @32kb/s	Low	0%	3.495	0.844	3.682	0.798
c35	G.729.1 @32kb/s	Nom	3%	3.969	0.758	3.885	0.771
c36	G.729.1 @14,16,24 & 32kb/s	Nom	0,2,4,6,10%	4.141	0.756	4.245	0.729

Table 10: ACR Experiment 3 - Music

Exp.3	WB-Music	Lab F		Lab V	
		Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 5 dB	1.086	0.281	1.070	0.336
c02	MNRU, Q = 15 dB	1.617	0.834	1.688	0.707
c03	MNRU, Q = 25 dB	2.625	1.035	2.539	0.850
c04	MNRU, Q = 35 dB	3.844	1.038	3.766	0.926
c05	MNRU, Q = 45 dB	4.133	0.899	4.359	0.729
c06	Direct	4.227	0.844	4.367	0.720
c07	G.722.2 – 12.65k	3.289	0.871	3.422	0.961
c08	G.722 – 48k	3.805	0.989	3.414	0.856
c09	G.722 – 56k	3.891	0.872	3.484	0.972
c10	CuT1 R3	3.836	0.876	3.836	0.867
c11	CuT1 R4	4.148	0.785	4.328	0.785
c12	CuT1 R5	4.313	0.718	4.391	0.713
c13	CuT2 R3	3.281	0.947	3.336	0.899
c14	CuT2 R4	4.188	0.771	4.234	0.789
c15	CuT2 R5	4.297	0.691	4.141	0.801
c16	CuT3 R3	3.148	0.989	3.203	0.891
c17	CuT3 R4	3.516	0.964	3.383	0.897
c18	CuT3 R5	3.539	0.921	3.438	0.885
c19	CuT4 R3	3.367	1.003	3.547	0.930
c20	CuT4 R4	4.234	0.726	4.289	0.733
c21	CuT4 R5	4.242	0.718	4.211	0.780
c22	G.729.1–16kb/s	3.016	1.027	3.297	0.934
c23	G.729.1–24kb/s	3.570	0.945	3.531	0.878
c24	G.729.1–32kb/s	4.320	0.720	4.031	0.832

Table 11: DCR Experiment 4 – Narrowband Speech with Car Noise

Exp.4	NB-Car15	Lab D		Lab J	
		Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 8 dB	1.323	0.773	1.141	0.391
c02	MNRU, Q = 14 dB	1.938	0.848	1.563	0.706
c03	MNRU, Q = 20 dB	2.740	0.853	2.510	0.910
c04	MNRU, Q = 26 dB	3.839	0.786	3.901	0.890
c05	MNRU, Q = 32 dB	4.287	0.803	4.677	0.588
c06	Direct	4.396	0.738	4.880	0.326
c07	G.729	3.844	0.810	3.693	0.906
c08	G.729E	4.099	0.720	4.302	0.794
c09	CuT1 @R1	3.979	0.709	3.854	0.909
c10	CuT1 @R2	4.266	0.764	4.417	0.704
c11	CuT2 @R1	4.214	0.702	4.042	0.903
c12	CuT2 @R2	4.302	0.703	4.458	0.662
c13	CuT3 @R1	4.000	0.799	3.766	0.928
c14	CuT3 @R2	4.177	0.779	4.422	0.719
c15	CuT4 @R1	4.208	0.708	4.109	0.858
c16	CuT4 @R2	4.271	0.792	4.505	0.647
c17	G.729.1 @8kb/s	3.797	0.776	3.484	0.898
c18	G.729.1 @12kb/s	4.104	0.779	4.287	0.756

Table 12: DCR Experiment 5 – Narrowband Speech with Street Noise

Exp.5	NB-Street15	Lab J		Lab N	
		Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 8 dB	1.073	0.280	1.193	0.409
c02	MNRU, Q = 14 dB	1.531	0.646	1.813	0.714
c03	MNRU, Q = 20 dB	2.412	0.864	2.630	0.789
c04	MNRU, Q = 26 dB	3.620	1.047	3.797	0.859
c05	MNRU, Q = 32 dB	4.542	0.646	4.438	0.749
c06	Direct	4.844	0.378	4.646	0.588
c07	G.729	3.672	0.928	3.875	0.877
c08	G.729E	4.276	0.863	4.406	0.739
c09	CuT1 @R1	3.781	0.973	4.089	0.798
c10	CuT1 @R2	4.219	0.828	4.432	0.675
c11	CuT2 @R1	4.120	0.838	4.391	0.737
c12	CuT2 @R2	4.479	0.731	4.537	0.646
c13	CuT3 @R1	3.833	0.911	4.037	0.871
c14	CuT3 @R2	4.302	0.788	4.500	0.686
c15	CuT4 @R1	4.203	0.828	4.333	0.658
c16	CuT4 @R2	4.542	0.685	4.537	0.646
c17	G.729.1 @8kb/s	3.635	0.934	3.813	0.823
c18	G.729.1 @12kb/s	4.193	0.779	4.302	0.807

Table 13: DCR Experiment 6 – Wideband Speech with Interfering Talker Noise

Exp.6	WB-IntTIkr15	Lab A		Lab V	
		Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 5 dB	1.063	0.391	1.000	0.000
c02	MNRU, Q = 15 dB	1.688	0.699	1.380	0.487
c03	MNRU, Q = 25 dB	2.563	0.784	2.104	0.686
c04	MNRU, Q = 35 dB	3.787	0.773	3.255	0.858
c05	MNRU, Q = 45 dB	4.380	0.691	4.385	0.714
c06	Direct	4.563	0.668	4.573	0.574
c07	G.722.2 –8.85k	4.135	0.718	3.734	0.817
c08	G.722.2 –12.65k	4.349	0.700	4.260	0.734
c09	G.722.2 –15.85k	4.474	0.655	4.427	0.659
c10	G.722.2 –23.85k	4.401	0.664	4.385	0.707
c11	CuT1 @R1	4.094	0.787	3.333	1.015
c12	CuT1 @R3	4.432	0.720	4.063	0.878
c13	CuT1 @R4	4.594	0.648	4.443	0.750
c14	CuT1 @R5	4.620	0.611	4.599	0.542
c15	CuT2 @R1	4.443	0.721	4.052	0.836
c16	CuT2 @R3	4.458	0.678	4.307	0.748
c17	CuT2 @R4	4.635	0.649	4.708	0.510
c18	CuT2 @R5	4.583	0.650	4.625	0.592
c19	CuT3 @R1	3.880	0.819	3.255	0.894
c20	CuT3 @R3	4.156	0.757	3.969	0.792
c21	CuT3 @R4	3.927	0.734	3.698	0.858
c22	CuT3 @R5	3.818	0.747	3.401	0.832
c23	CuT4 @R1	4.464	0.693	4.047	0.858
c24	CuT4 @R3	4.526	0.655	4.412	0.718
c25	CuT4 @R4	4.656	0.620	4.714	0.487
c26	CuT4 @R5	4.641	0.623	4.599	0.588
c27	G.729.1 @14kb/s	4.229	0.709	3.906	0.800
c28	G.729.1 @16kb/s	4.172	0.810	4.156	0.706
c29	G.729.1 @24kb/s	4.396	0.671	4.479	0.655
c30	G.729.1 @32kb/s	4.516	0.716	4.516	0.597

Table 14: DCR Experiment 7 –Wideband Speech with Office Noise

Exp.7	WB-Office20	Lab B		Lab N	
		Mean	Stdev	Mean	Stdev
c01	MNRU, Q = 5 dB	1.484	0.793	1.047	0.212
c02	MNRU, Q = 15 dB	2.318	1.082	1.448	0.539
c03	MNRU, Q = 25 dB	3.651	0.996	2.151	0.719
c04	MNRU, Q = 35 dB	4.458	0.670	3.146	0.880
c05	MNRU, Q = 45 dB	4.542	0.654	4.380	0.706
c06	Direct	4.557	0.676	4.688	0.557
c07	G.722.2 –8.85k	3.662	0.929	3.682	0.903
c08	G.722.2 –12.65k	4.115	0.830	4.313	0.777
c09	G.722.2 –15.85k	4.307	0.796	4.443	0.669
c10	G.722.2 –23.85k	4.427	0.762	4.438	0.684
c11	CuT1 @R1	3.516	0.874	3.412	0.999
c12	CuT1 @R3	4.214	0.807	4.208	0.778
c13	CuT1 @R4	4.234	0.826	4.458	0.744
c14	CuT1 @R5	4.427	0.776	4.620	0.566
c15	CuT2 @R1	3.901	0.958	3.948	0.969
c16	CuT2 @R3	4.297	0.786	4.323	0.759
c17	CuT2 @R4	4.531	0.638	4.682	0.530
c18	CuT2 @R5	4.589	0.649	4.656	0.518
c19	CuT3 @R1	3.318	0.931	3.135	0.999
c20	CuT3 @R3	3.969	0.843	3.922	0.874
c21	CuT3 @R4	4.318	0.778	3.766	0.928
c22	CuT3 @R5	4.234	0.788	3.620	0.854
c23	CuT4 @R1	4.037	0.900	3.984	0.918
c24	CuT4 @R3	4.370	0.795	4.313	0.777
c25	CuT4 @R4	4.547	0.654	4.651	0.540
c26	CuT4 @R5	4.516	0.752	4.677	0.522
c27	G.729.1 @14kb/s	3.844	0.908	3.969	0.862
c28	G.729.1 @16kb/s	3.943	0.881	4.115	0.729
c29	G.729.1 @24kb/s	4.271	0.868	4.406	0.672
c30	G.729.1 @32kb/s	4.464	0.772	4.573	0.610

6.1.4 Requirements in Terms of Reference

In total, 35 requirement ToRs were specified to be tested in the following four categories:

- 14 input level conditions – Experiments 1, 2a, 2b, 2c
- 6 error (FER) conditions – Experiments 1, 2a, 2b, 2c
- 3 Music conditions – Experiment 3
- 12 noise conditions – Experiments 4, 5, 6, 7

Moreover, each ToR was tested in two Listening Labs for a total of 70 ToR tests for each codec.

Table 12 shows the results of the 14 ToR Tests for the Input level category.

The description that follows the table also applies to the ToR tests for Error conditions (Table 13), Music conditions (Table 14), and Noise conditions (Table 15).

Table 15: Results of ToR tests for Input Level Conditions

Clean - Input Level		Reference		CuT1	CuT2	CuT3	CuT4	G.729.1
R1 <i>nwt</i> G.729E, Nom level	Exp. 1	c08		c12	c17	c22	c27	c32
	B	3.552	0.207	3.922	3.781	3.594	3.703	4.177
	V	4.287	0.142	3.984	4.307	4.037	4.318	4.240
R2 <i>bt</i> G.729, Nom level	Exp. 1	c07		c13	c18	c23	c28	c33
	B	3.703	0.203	3.849	3.662	3.583	3.672	3.901
	V	4.083	0.141	4.302	4.297	4.297	4.271	4.349
R1 <i>nwt</i> G.729E, High level	Exp. 1	c09		c14	c19	c24	c29	c34
	B	3.557	0.201	3.854	3.688	3.516	3.740	4.094
	V	4.333	0.138	3.958	4.365	4.068	4.391	4.214
R1 <i>nwt</i> G.729E, Low level	Exp. 1	c10		c15	c20	c25	c30	c35
	B	3.380	0.210	3.953	3.646	3.073	3.662	3.875
	V	3.958	0.165	3.943	4.208	3.573	4.208	3.995
R1 <i>nwt</i> G.722.2 at 8.85k, Nom level	Exp. 2a	c08		c12	c17	c22	c27	c32
	B	3.677	0.161	4.099	4.073	3.859	4.010	3.552
	D	3.802	0.137	4.026	4.104	3.839	4.151	3.885
R1 <i>nwt</i> G.722.2 at 8.85k, High level	Exp. 2a	c09		c13	c18	c23	c28	c33
	B	3.693	0.177	3.807	3.984	3.781	3.880	3.583
	D	3.635	0.149	3.870	4.016	3.807	3.938	3.865
R1 <i>nwt</i> G.722.2 at 8.85k, Low level	Exp. 2a	c10		c14	c19	c24	c29	c34
	B	3.391	0.196	4.234	3.938	3.380	3.880	3.208
	D	3.323	0.202	3.839	4.010	3.557	4.005	3.714
R3 <i>nwt</i> G.722.2 at 12.65k, Nom level	Exp. 2b	c08		c12	c17	c22	c27	c32
	J	3.385	0.205	4.005	3.917	3.510	3.938	3.563
	N	3.729	0.147	3.974	4.010	3.635	4.068	3.797
R3 <i>nwt</i> G.722.2 at 12.65k, High level	Exp. 2b	c09		c13	c18	c23	c28	c33
	J	3.583	0.162	4.042	3.901	3.552	3.813	3.635
	N	3.865	0.150	3.938	3.958	3.734	3.901	3.771
R3 <i>nwt</i> G.722.2 at 12.65k, Low level	Exp. 2b	c10		c14	c19	c24	c29	c34
	J	2.469	0.213	4.063	3.630	2.542	3.641	2.927
	N	2.589	0.187	3.839	3.802	2.760	3.792	2.865
R4 <i>nwt</i> G.722.2 at 15.85k, Nom level	Exp. 2b	c11		c16	c21	c26	c31	c36
	J	3.438	0.173	4.021	3.880	3.188	4.005	3.547
	N	3.776	0.189	4.099	4.073	3.167	4.037	3.813
R5 <i>nwt</i> G.722.2 at 23.85k, Nom level	Exp. 2c	c09		c12	c17	c22	c27	c32
	A	4.063	0.139	4.542	4.297	3.724	4.370	4.177
	F	4.078	0.149	4.365	4.307	3.677	4.292	4.135
R5 <i>nwt</i> G.722.2 at 23.85k, High level	Exp. 2c	c10		c13	c18	c23	c28	c33
	A	4.260	0.131	4.505	4.287	3.839	4.219	4.224
	F	4.083	0.155	4.333	4.188	3.547	4.255	4.271
R5 <i>nwt</i> G.722.2 at 23.85k, Low level	Exp. 2c	c11		c14	c19	c24	c29	c34
	A	3.354	0.165	4.323	4.193	3.005	4.125	3.495
	F	3.458	0.200	4.359	4.224	3.115	4.177	3.682

Each ToR test involves three rows in the table where the first row shows the experiment and condition labels, the second row shows results for one lab that conducted the test, the third row shows results for the other lab. For example, the first three rows of Table 12 show the ToR test results for: “R1 nwt G.729E, Nominal level” tested in Exp.1 in Lab B and Lab V

- Row 1 – the reference codec was condition c08, the test codecs were c12, c17, c22, c27, c32 for CuT1, CuT2, CuT3, CuT4, G.729.1, respectively.
- Row 2 – in Lab B the score for the reference codec was 3.552 and the Dunnett’s Test yielded a Critical Significant Difference of 0.207, scores for the test codecs were 3.992, 3.782, 3.594, 3.703, 4.177.
- Row 3 – in Lab V the score for the reference codec was 4.287 and the Dunnett’s Test yielded a Critical Significant Difference of 0.142, scores for the test codecs were 3.984, 4.307, 4.037, 4.318, 4.240.

There were two failures of the nwt (not worse than) ToR indicated by the highlighted scores for CuT1 (3.984) and CuT3 (4.037) in Lab V. These two scores failed the ToR because they were lower than the Reference minus the Critical Significant Difference [$4.287 - 0.142 = 4.145$].

Table 16 shows the results for the ToR tests in Error conditions. Table 17 is organized in the same manner as described for Table 16.

Table 16: Results of ToR tests for Error Conditions

Clean - Error		Reference		CuT1	CuT2	CuT3	CuT4	G.729.1
R1 <i>nwt</i> G.729E, 3% FER	Exp. 1	c11		c16	c21	c26	c31	c36
	B	3.255	0.207	3.495	3.531	3.323	3.484	3.823
	V	4.016	0.148	3.682	4.047	3.823	4.125	3.917
R1-R5 <i>nwt</i> G.722.2,8.85k, fast switching	Exp. 2a	c08		c15	c20	c25	c30	c35
	B	3.677	0.169	3.885	3.938	3.766	3.865	3.698
	D	3.802	0.142	4.162	4.156	3.906	4.083	4.031
R1 <i>nwt</i> G.722.2 8.85k, 3%FER	Exp. 2a	c11		c16	c21	c26	c31	c36
	B	3.182	0.194	3.401	3.625	3.396	3.729	3.052
	D	3.214	0.157	3.495	3.807	3.688	3.880	3.599
R3 <i>nwt</i> G.722.2 8.85k, 6%FER	Exp. 2b	c07		c15	c20	c25	c30	c35
	J	3.063	0.212	2.974	3.365	3.073	3.396	2.807
	H	3.276	0.176	3.203	3.682	3.203	3.755	3.318
R5 <i>nwt</i> G.722.2 12.55k, 3%FER	Exp. 2c	c08		c15	c20	c25	c30	c35
	A	4.031	0.135	4.401	4.193	3.583	4.229	3.969
	F	3.995	0.157	3.953	3.958	3.203	4.052	3.885
R5 <i>nwt</i> G.722.2 at 12.65k, 0,2,4,6,10% FER	Exp. 2c	c08		c16	c21	c26	c31	c36
	A	4.031	0.148	4.531	4.313	3.771	4.307	4.141
	F	3.995	0.146	4.266	4.214	3.526	4.297	4.245

Table 17 shows the results for the ToR tests for Music conditions. Table 17 is organized in the same manner as described for Table 15.

Table 17: Results of ToR tests for Music Conditions

Music		Reference		CuT1	CuT2	CuT3	CuT4	G.729.1
R3 <i>nwt</i> G.722.2 at 12.65k	Exp. 3	c07		c10	c13	c16	c19	c22
	F	3.289	0.186	3.836	3.281	3.148	3.367	3.016
	V	3.422	0.161	3.836	3.336	3.203	3.547	3.297
R4 <i>nwt</i> G.722 at 48k	Exp. 3	c08		c11	c14	c17	c20	c23
	F	3.805	0.219	4.148	4.188	3.516	4.234	3.570
	V	3.414	0.203	4.328	4.234	3.383	4.289	3.531
R3 <i>nwt</i> G.722 at 56k	Exp. 3	c09		c12	c15	c18	c21	c24
	F	3.891	0.198	4.313	4.297	3.539	4.242	4.320
	V	3.484	0.207	4.391	4.141	3.438	4.211	4.031

Table 18 shows the results for the ToR tests for Noise conditions. Table 15 is organized in the same manner as described for Table 15.

Table 18: Results of ToR tests for Noise Conditions

Noise		Reference		CuT1	CuT2	CuT3	CuT4	G.729.1
R1 <i>nvt</i> G.729, Carl15	Exp. 4	c07		c09	c11	c13	c15	c17
	D	3.844	0.146	3.979	4.214	4.000	4.208	3.797
	J	3.693	0.153	3.854	4.042	3.766	4.109	3.484
R2 <i>nvt</i> G.729E, Carl15	Exp. 4	c08		c10	c12	c14	c16	c18
	D	4.099	0.147	4.266	4.302	4.177	4.271	4.104
	J	4.302	0.130	4.417	4.458	4.422	4.505	4.286
R1 <i>nvt</i> G.729, Street20	Exp. 5	c07		c09	c11	c13	c15	c17
	J	3.672	0.166	3.781	4.120	3.833	4.203	3.635
	N	3.875	0.158	4.089	4.391	4.037	4.333	3.813
R2 <i>nvt</i> G.729E, Street20	Exp. 5	c08		c10	c12	c14	c16	c18
	J	4.276	0.152	4.219	4.479	4.302	4.542	4.193
	N	4.406	0.140	4.432	4.537	4.500	4.537	4.302
R1 <i>nvt</i> G.722.2, 8.85k, IT15	Exp. 6	c07		c11	c15	c19	c23	c27
	A	4.135	0.141	4.094	4.443	3.880	4.464	4.229
	V	3.734	0.203	3.333	4.052	3.255	4.047	3.906
R3 <i>nvt</i> G.722.2, 12.65k, IT15	Exp. 6	c08		c12	c16	c20	c24	c28
	A	4.349	0.146	4.432	4.458	4.156	4.526	4.172
	V	4.260	0.172	4.063	4.307	3.969	4.412	4.156
R4 <i>nvt</i> G.722.2, 15.85k, IT15	Exp. 6	c09		c13	c17	c21	c25	c29
	A	4.474	0.151	4.594	4.635	3.927	4.656	4.396
	V	4.427	0.167	4.443	4.708	3.698	4.714	4.479
R5 <i>nvt</i> G.722.2, 23.85k, IT15	Exp. 6	c10		c14	c18	c22	c26	c30
	A	4.401	0.154	4.620	4.583	3.818	4.641	4.516
	V	4.385	0.160	4.599	4.625	3.401	4.599	4.516
R1 <i>nvt</i> G.722.2, 8.85k, Off20	Exp. 7	c07		c11	c15	c19	c23	c27
	B	3.662	0.180	3.516	3.901	3.318	4.037	3.844
	N	3.682	0.190	3.412	3.948	3.135	3.984	3.969
R3 <i>nvt</i> G.722.2, 12.65k, Off20	Exp. 7	c08		c12	c16	c20	c24	c28
	B	4.115	0.170	4.214	4.297	3.969	4.370	3.943
	N	4.313	0.162	4.208	4.323	3.922	4.313	4.115
R4 <i>nvt</i> G.722.2, 15.85k, Off20	Exp. 7	c09		c13	c17	c21	c25	c29
	B	4.307	0.142	4.234	4.531	4.318	4.547	4.271
	N	4.443	0.165	4.458	4.682	3.766	4.651	4.406
R5 <i>nvt</i> G.722.2, 23.85k, Off20	Exp. 7	c10		c14	c18	c22	c26	c30
	B	4.427	0.134	4.427	4.589	4.234	4.516	4.464
	N	4.438	0.161	4.620	4.656	3.620	4.677	4.573

Table 19 summarizes the ToR results presented in this section. None of the codecs passed all 70 of the “Requirement” ToR tests involved in the EV-VBR Selection Phase. However, two of the candidates, CuT2 and CuT4, failed only one ToR and failed it in only one lab — Lab B. Results for the other three codecs were, in order: 7 failures for CuT1, 8 failures for G.729.1, and 35 failures for CuT3.

Table 19: Number of ToR Failures By Condition Category

Category	# ToR's	CuT1	CuT2	CuT3	CuT4	G.729.1
Input Level	28	3	1	13	1	1
Error	12	1	0	5	0	1
Music	6	3	0	3	0	2
Noise	24	0	0	14	0	4
Total	70	7	1	35	1	8

6.1.5 Weighted Requirement ToR Passes

The group responsible for the standardization of the EV-VBR codec, ITU-T Q9/SG16, determined that the four categories of ToRs were not of equal importance for the targeted applications for the

EV-VBR codec. Therefore, neither the categories nor the ToRs within each category should be given equal weight in the selection process. It was determined that the Input Level and Error categories should have equal weight and a third category made up of Music+Noise should have half the weight of the other two. Therefore, using a 100-point scale, the Input Level and Error Categories have weight of 40 points each and the Music+Noise category has a weight of 20 points. Furthermore, each ToR within each category doesn't have equal importance for the intended applications of the codec. Accordingly, each ToR within each category was assigned a rating associated with its importance where High=3, Medium=2, Low=1. A table of ToR weights was developed under these two constraints, category and importance, so that a candidate that passes all 70 ToRs receives a score of 100 points. Table 20 shows the table of ToR weights.

Table 20: ToR Weights²

Category (points)	Importance					
	High		Medium		Low	
	#	Wt.	#	Wt.	#	Wt.
Input Level (40)	13	1.463	1	0.976	0	0.488
Error (40)	6	3.333	0	2.222	0	1.111
Music+Noise (20)	1	1.000	13	0.667	1	0.333

Using the weights from Table 20 and applying them to the results of the ToR tests (where Pass = 1 and Fail = 0) results in a weighted number of passes score for each test codec. These scores are based on a 100-point scale, i.e., 0 failures results in a score of 100. The weighted number of passes scores, by ToR Category and Total, are presented in Table 21 and in Figure 20. Table 22, Table 23, and Table 24 provide the details of the computations for the weighted number of passes for the ToR categories Input Level, Error, and Music+Noise, respectively.

Table 21: Weighted Number of Passes

Category	(##)	CuT1	CuT2	CuT3	CuT4	G.729.1
Input Level	(40)	35.61	38.54	21.95	38.54	38.54
Error	(40)	36.67	40.00	23.33	40.00	36.67
Music+Noise	(20)	18.00	20.00	8.67	20.00	16.33
Total	(100)	90.28	98.54	53.95	98.54	91.54

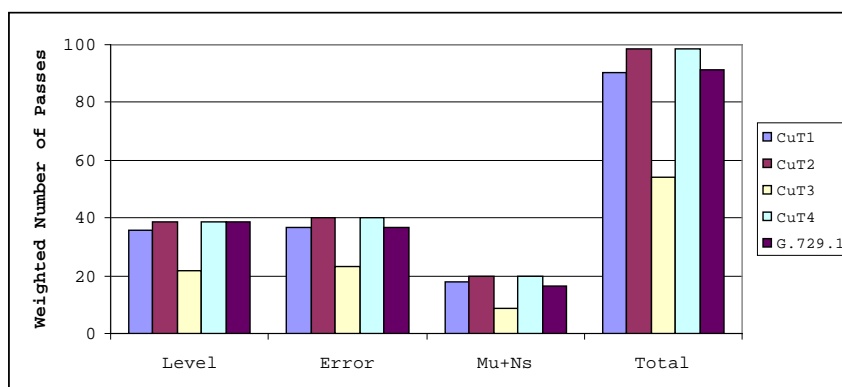


Figure 20 – Weighted Number of Passes by ToR Category

² The entries in the *Wt.* columns have been adjusted to take into account that each ToR was tested in two labs.

Table 22: Weighted Number of Passes for Input Level ToRs

Input Level			CuT1	CuT2	CuT3	CuT4	G. 729.1
R1 <i>nvt</i> G.729E, Nom level	Exp.1	High					
	B	1.463	1	1	1	1	1
	V	1.463	0	1	0	1	1
R2 <i>bt</i> G.729, Nom level	Exp.1	High					
	B	1.463	0	0	0	0	0
	V	1.463	1	1	1	1	1
R1 <i>nvt</i> G.729E, High level	Exp.1	High					
	B	1.463	1	1	1	1	1
	V	1.463	0	1	0	1	1
R1 <i>nvt</i> G.729E, Low level	Exp.1	High					
	B	1.463	1	1	0	1	1
	V	1.463	1	1	0	1	1
R1 <i>nvt</i> G.722.2 at 8.85k, Nom level	Exp.2a	High					
	B	1.463	1	1	1	1	1
	D	1.463	1	1	1	1	1
R1 <i>nvt</i> G.722.2 at 8.85k, High level	Exp.2a	High					
	B	1.463	1	1	1	1	1
	D	1.463	1	1	1	1	1
R1 <i>nvt</i> G.722.2 at 8.85k, Low level	Exp.2a	High					
	B	1.463	1	1	1	1	1
	D	1.463	1	1	1	1	1
R3 <i>nvt</i> G.722.2 at 12.65k, Nom level	Exp.2b	High					
	J	1.463	1	1	1	1	1
	N	1.463	1	1	1	1	1
R3 <i>nvt</i> G.722.2 at 12.65k, High level	Exp.2b	High					
	J	1.463	1	1	1	1	1
	N	1.463	1	1	1	1	1
R3 <i>nvt</i> G.722.2 at 12.65k, Low level	Exp.2b	High					
	J	1.463	1	1	1	1	1
	N	1.463	1	1	1	1	1
R4 <i>nvt</i> G.722.2 at 15.85k, Nom level	Exp.2b	Med					
	J	0.976	1	1	0	1	1
	N	0.976	1	1	0	1	1
R5 <i>nvt</i> G.722.2 at 23.85k, Nom level	Exp.2c	High					
	A	1.463	1	1	0	1	1
	F	1.463	1	1	0	1	1
R5 <i>nvt</i> G.722.2 at 23.85k, High level	Exp.2c	High					
	A	1.463	1	1	0	1	1
	F	1.463	1	1	0	1	1
R5 <i>nvt</i> G.722.2 at 23.85k, Low level	Exp.2c	High					
	A	1.463	1	1	0	1	1
	F	1.463	1	1	0	1	1
Input Level			CuT1	CuT2	CuT3	CuT4	G. 729.1
Points (out of 40)			35.61	38.54	21.95	38.54	38.54

Table 23: Weighted Number of Passes for Error ToRs

Error			CuT1	CuT2	CuT3	CuT4	G. 729.1
R1 <i>nvt</i> G.729E, 3% FER	Exp. 1	High					
	B	3.333	1	1	1	1	1
	V	3.333	0	1	0	1	1
R1-R5 <i>nvt</i> G.722.2,8.85k, fast switching	Exp. 2a	High					
	B	3.333	1	1	1	1	1
	D	3.333	1	1	1	1	1
R1 <i>nvt</i> G.722.2 8.85k, 3%FER	Exp. 2a	High					
	B	3.333	1	1	1	1	1
	D	3.333	1	1	1	1	1
R3 <i>nvt</i> G.722.2 8.85k, 6%FER	Exp. 2b	High					
	J	3.333	1	1	1	1	0
	N	3.333	1	1	1	1	1
R5 <i>nvt</i> G.722.2 12.55k, 3%FER	Exp. 2c	High					
	A	3.333	1	1	0	1	1
	F	3.333	1	1	0	1	1
R5 <i>nvt</i> G.722.2 at 12.65k, 0,2,4,6,10% FER	Exp. 2c	High					
	A	3.333	1	1	0	1	1
	F	3.333	1	1	0	1	1
Error			CuT1	CuT2	CuT3	CuT4	G. 729.1
Points (out of 40)			36.67	40.00	23.33	40.00	36.67

Table 24: Weighted Number of Passes for Music+Noise ToRs

Music			CuT1	CuT2	CuT3	CuT4	G.729.1
R3 <i>nvt</i> at 12.65k	Exp. 3	Low					
	V	0.333	1	1	1	1	0
	F	0.333	1	1	0	1	1
R4 <i>nvt</i> at 48k	Exp. 3	Med					
	V	0.667	1	1	0	1	0
	F	0.667	1	1	1	1	1
R3 <i>nvt</i> at 56k	Exp. 3	High					
	V	1.000	1	1	0	1	1
	F	1.000	1	1	1	1	1
Noise			CuT1	CuT2	CuT3	CuT4	G.729.1
R1 <i>nvt</i> G.729, Carl5	Exp. 4	Med					
	D	0.667	1	1	1	1	1
	J	0.667	1	1	1	1	0
R2 <i>nvt</i> G.729E, Carl5	Exp. 4	Med					
	D	0.667	1	1	1	1	1
	J	0.667	1	1	1	1	1
R1 <i>nvt</i> G.729, Street20	Exp. 5	Med					
	J	0.667	1	1	1	1	1
	N	0.667	1	1	1	1	1
R2 <i>nvt</i> G.729E, Street20	Exp. 5	Med					
	J	0.667	1	1	1	1	1
	N	0.667	1	1	1	1	1
R1 <i>nvt</i> G.722.2, 8.85k, IT15	Exp. 6	Med					
	A	0.667	1	1	0	1	1
	V	0.667	0	1	0	1	1
R3 <i>nvt</i> G.722.2, 12.65k, IT15	Exp. 6	Med					
	A	0.667	1	1	0	1	0
	V	0.667	0	1	0	1	1
R4 <i>nvt</i> G.722.2, 15.85k, IT15	Exp. 6	Med					
	A	0.667	1	1	0	1	1
	V	0.667	1	1	0	1	1
R5 <i>nvt</i> G.722.2, 23.85k, IT15	Exp. 6	Med					
	A	0.667	1	1	0	1	1
	V	0.667	1	1	0	1	1
R1 <i>nvt</i> G.722.2, 8.85k, Off20	Exp. 7	Med					
	B	0.667	1	1	0	1	1
	N	0.667	0	1	0	1	1
R3 <i>nvt</i> G.722.2, 12.65k, Off20	Exp. 7	Med					
	B	0.667	1	1	1	1	0
	N	0.667	1	1	0	1	0
R4 <i>nvt</i> G.722.2, 15.85k, Off20	Exp. 7	Med					
	B	0.667	1	1	1	1	1
	N	0.667	1	1	0	1	1
R5 <i>nvt</i> G.722.2, 23.85k, Off20	Exp. 7	Med					
	B	0.667	1	1	0	1	1
	N	0.667	1	1	0	1	1
Music+Noise			CuT1	CuT2	CuT3	CuT4	G.729.1
Points (out of 20)			18.00	20.00	8.67	20.00	16.33

6.2 Figure of Merit

The Test Plan specified that the weights from Table 20 should be applied to the subjective scores for each requirement ToR to produce a Figure of Merit (FoM) score for each codec. Table 25 presents the appropriate scores for all 70 Requirement ToR conditions with the associated weights. The bottom row of the table also shows the FoM score for each codec. Table 26 and Figure 21 show the FoM scores by ToR Category. While the codecs can be ranked on the basis of the FoM scores, there is no measure of variability and therefore no statistical tests among codecs based on the FoM.

Table 25: Figure of Merit – Weighted Average Quality Scores

Terms of Reference		Wt.	Lab	Ref	CuT 1	CuT 2	CuT 3	CuT 4	G.729.1		
Input Level	Exp. 1	R1 <i>nwt</i> G.729E, Nom level	1.463	B	3.552	3.922	3.781	3.594	3.703	4.177	
			1.463	V	4.287	3.984	4.307	4.037	4.318	4.240	
		R2 <i>bt</i> G.729, Nom level	1.463	B	3.703	3.849	3.662	3.583	3.672	3.901	
			1.463	V	4.083	4.302	4.297	4.297	4.271	4.349	
		R1 <i>nwt</i> G.729E, High level	1.463	B	3.557	3.854	3.688	3.516	3.740	4.094	
			1.463	V	4.333	3.958	4.365	4.068	4.391	4.214	
	Exp. 2a	R1 <i>nwt</i> G.729E, Low level	1.463	B	3.380	3.953	3.646	3.073	3.662	3.875	
			1.463	V	3.958	3.943	4.208	3.573	4.208	3.995	
		R1 <i>nwt</i> G.722.2 at 8.85k, Nom level	1.463	B	3.677	4.099	4.073	3.859	4.010	3.552	
			1.463	D	3.802	4.026	4.104	3.839	4.151	3.885	
		R1 <i>nwt</i> G.722.2 at 8.85k, High level	1.463	B	3.693	3.807	3.984	3.781	3.880	3.583	
			1.463	D	3.635	3.870	4.016	3.807	3.938	3.865	
	Exp. 2b	R1 <i>nwt</i> G.722.2 at 8.85k, Low level	1.463	B	3.391	4.234	3.938	3.380	3.880	3.208	
			1.463	D	3.323	3.839	4.010	3.557	4.005	3.714	
		R3 <i>nwt</i> G.722.2 at 12.65k, Nom level	1.463	J	3.385	4.005	3.917	3.510	3.938	3.563	
			1.463	N	3.729	3.974	4.010	3.635	4.068	3.797	
		R3 <i>nwt</i> G.722.2 at 12.65k, High level	1.463	J	3.583	4.042	3.901	3.552	3.813	3.635	
			1.463	N	3.865	3.938	3.958	3.734	3.901	3.771	
	Exp. 2c	R3 <i>nwt</i> G.722.2 at 12.65k, Low level	1.463	J	2.469	4.063	3.630	2.542	3.641	2.927	
			1.463	N	2.589	3.839	3.802	2.760	3.792	2.865	
		R4 <i>nwt</i> G.722.2 at 15.85k, Nom level	0.976	J	3.438	4.021	3.880	3.188	4.005	3.547	
			0.976	N	3.776	4.099	4.073	3.167	4.037	3.813	
		R5 <i>nwt</i> G.722.2 at 23.85k, Nom level	1.463	A	4.063	4.542	4.297	3.724	4.370	4.177	
			1.463	F	4.078	4.365	4.307	3.677	4.292	4.135	
	Error	Exp. 1	R5 <i>nwt</i> G.722.2 at 23.85k, High level	1.463	A	4.260	4.505	4.287	3.839	4.219	4.224
				1.463	F	4.083	4.333	4.188	3.547	4.255	4.271
		Exp. 2a	R5 <i>nwt</i> G.722.2 at 23.85k, Low level	1.463	A	3.354	4.323	4.193	3.005	4.125	3.495
				1.463	F	3.458	4.359	4.224	3.115	4.177	3.682
		Exp. 2b	R1 <i>nwt</i> G.729E, 3% FER	3.333	B	3.255	3.495	3.531	3.323	3.484	3.823
				3.333	V	4.016	3.682	4.047	3.823	4.125	3.917
		Exp. 2c	R1-R5 <i>nwt</i> G.722.2, 8.85k, fast switching	3.333	B	3.677	3.885	3.938	3.766	3.865	3.698
				3.333	D	3.802	4.162	4.156	3.906	4.083	4.031
		Exp. 2d	R1 <i>nwt</i> G.722.2 8.85k, 3%FER	3.333	B	3.182	3.401	3.625	3.396	3.729	3.052
				3.333	D	3.214	3.495	3.807	3.688	3.880	3.599
		Exp. 2e	R3 <i>nwt</i> G.722.2 8.85k, 6%FER	3.333	J	3.063	2.974	3.365	3.073	3.396	2.807
				3.333	N	3.276	3.203	3.682	3.203	3.755	3.318
		Exp. 2f	R5 <i>nwt</i> G.722.2 12.65k, 3%FER	3.333	A	4.031	4.401	4.193	3.583	4.229	3.969
				3.333	F	3.995	3.953	3.958	3.203	4.052	3.885
Exp. 2g		R5 <i>nwt</i> G.722.2 at 12.65k, 0,2,4,6,10% FER	3.333	A	4.031	4.531	4.313	3.771	4.307	4.141	
			3.333	F	3.995	4.266	4.214	3.526	4.297	4.245	
Music		Exp. 3	R3 <i>nwt</i> G.722.2 at 12.65k	0.333	F	3.289	3.836	3.281	3.148	3.367	3.016
				0.333	V	3.422	3.836	3.336	3.203	3.547	3.297
			R4 <i>nwt</i> G.722 at 48k	0.667	F	3.805	4.148	4.188	3.516	4.234	3.570
		Exp. 4		0.667	V	3.414	4.328	4.234	3.383	4.289	3.531
			R3 <i>nwt</i> G.722 at 56k	1.000	F	3.891	4.313	4.297	3.539	4.242	4.320
				1.000	V	3.484	4.391	4.141	3.438	4.211	4.031
Noise		Exp. 4	R1 <i>nwt</i> G.729, Car15	0.667	D	3.844	3.979	4.214	4.000	4.208	3.797
				0.667	J	3.693	3.854	4.042	3.766	4.109	3.484
		Exp. 5	R2 <i>nwt</i> G.729E, Car15	0.667	D	4.099	4.266	4.302	4.177	4.271	4.104
				0.667	J	4.302	4.417	4.458	4.422	4.505	4.286
		Exp. 6	R1 <i>nwt</i> G.729, Street20	0.667	J	3.672	3.781	4.120	3.833	4.203	3.635
				0.667	N	3.875	4.089	4.391	4.037	4.333	3.813
		Exp. 7	R2 <i>nwt</i> G.729E, Street20	0.667	J	4.276	4.219	4.479	4.302	4.542	4.193
				0.667	N	4.406	4.432	4.537	4.500	4.537	4.302
	Exp. 8	R1 <i>nwt</i> G.722.2, 8.85k, IT15	0.667	A	4.135	4.094	4.443	3.880	4.464	4.229	
			0.667	V	3.734	3.333	4.052	3.255	4.047	3.906	
		R3 <i>nwt</i> G.722.2, 12.65k, IT15	0.667	A	4.349	4.432	4.458	4.156	4.526	4.172	
			0.667	V	4.260	4.063	4.307	3.969	4.412	4.156	
		R4 <i>nwt</i> G.722.2, 15.85k, IT15	0.667	A	4.474	4.594	4.635	3.927	4.656	4.396	
			0.667	V	4.427	4.443	4.708	3.698	4.714	4.479	
R5 <i>nwt</i> G.722.2, 23.85k, IT15		0.667	A	4.401	4.620	4.583	3.818	4.641	4.516		
		0.667	V	4.385	4.599	4.625	3.401	4.599	4.516		
Exp. 9	R1 <i>nwt</i> G.722.2, 8.85k, Off20	0.667	B	3.662	3.516	3.901	3.318	4.037	3.844		
		0.667	N	3.682	3.412	3.948	3.135	3.984	3.969		
	R3 <i>nwt</i> G.722.2, 12.65k, Off20	0.667	B	4.115	4.214	4.297	3.969	4.370	3.943		
		0.667	N	4.313	4.208	4.323	3.922	4.313	4.115		
	R4 <i>nwt</i> G.722.2, 15.85k, Off20	0.667	B	4.307	4.234	4.531	4.318	4.547	4.271		
		0.667	N	4.443	4.458	4.682	3.766	4.651	4.406		
	R5 <i>nwt</i> G.722.2, 23.85k, Off20	0.667	B	4.427	4.427	4.589	4.234	4.516	4.464		
		0.667	N	4.438	4.620	4.656	3.620	4.677	4.573		
Weighted Average (FoM)				3.725	3.982	4.037	3.586	4.050	3.822		

Table 26: Figure of Merit by ToR Category

Category	Ref.	CuT 1	CuT 2	CuT 3	CuT 4	G.729.1
Input level	3.662	4.073	4.028	3.543	4.016	3.808
Error	3.628	3.787	3.902	3.522	3.934	3.707
Music	3.606	4.228	4.065	3.423	4.110	3.798
Noise	4.155	4.179	4.387	3.893	4.411	4.149
Total	3.725	3.982	4.037	3.586	4.050	3.822

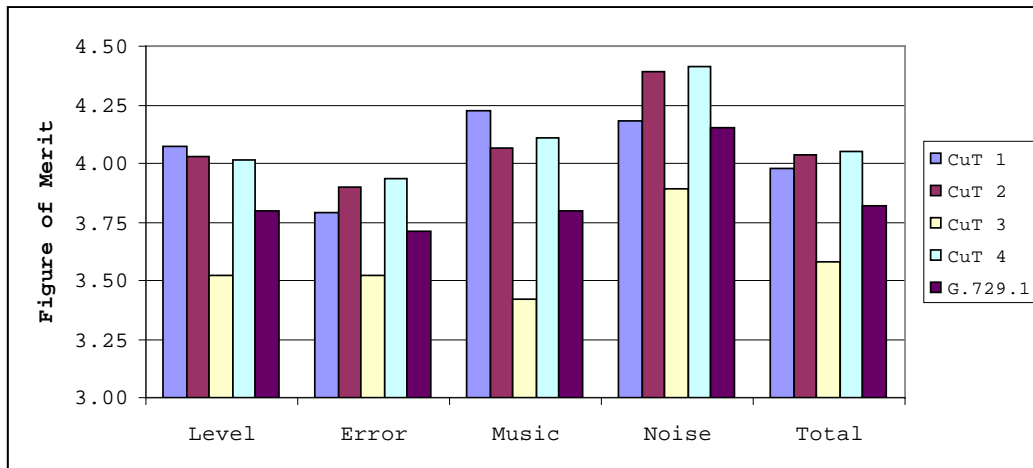


Figure 21 – Figure of Merit by ToR Category

6.3 Objective Terms of Reference

For the Objectives for the EV-VBR codec, there were 14 ToRs: four Input Level conditions, two Music conditions, and eight Noise conditions. Table 27 shows the number of Objective ToR failures by codec. The details of the ToR tests are shown in Table 28, Table 29, and Table 30 for the three ToR categories.

Table 27: Number of Objective ToR Failures by Codec.

Category	# ToR's	CuT1	CuT2	CuT3	CuT4	G.729.1
Input Level	8	0	1	6	1	4
Music	4	0	1	2	1	1
Noise	16	7	1	14	2	12
Total	28	7	3	22	4	17

Table 28: Objective ToRs for Input Level Conditions

Input Level	Reference	CuT1	CuT2	CuT3	CuT4	G.729.1		
R3 <i>nr</i> G.722.2 at 15.85k, Nom level	Exp. 2b	c11	c12	c17	c22	c27	c32	
	J	3.438	0.212	4.005	3.917	3.510	3.938	3.563
	N	3.776	0.167	3.974	4.010	3.635	4.068	3.797
R5 <i>bt</i> G.722.2 at 23.85k, Nom level	Exp. 2c	c09	c12	c17	c22	c27	c32	
	A	4.063	0.143	4.542	4.297	3.724	4.370	4.177
	F	4.078	0.153	4.365	4.307	3.677	4.292	4.135
R5 <i>bt</i> G.722.2 at 23.85k, High level	Exp. 2c	c10	c13	c18	c23	c28	c33	
	A	4.260	0.135	4.505	4.287	3.839	4.219	4.224
	F	4.083	0.160	4.333	4.188	3.547	4.255	4.271
R5 <i>bt</i> G.722.2 at 23.85k, Low level	Exp. 2c	c11	c14	c19	c24	c29	c34	
	A	3.354	0.169	4.323	4.193	3.005	4.125	3.495
	F	3.458	0.205	4.359	4.224	3.115	4.177	3.682

Table 29: Objective ToRs for Music Conditions

Music		Reference		CuT1	CuT2	CuT3	CuT4	G.729.1
R3 <i>nvt</i> G.722 at 48k	Exp. 3	c08		c10	c13	c16	c19	c22
	F	3.805	0.191	3.836	3.281	3.148	3.367	3.016
	V	3.414	0.165	3.836	3.336	3.203	3.547	3.297
R4 <i>nvt</i> G.722 at 56k	Exp. 3	c09		c11	c14	c17	c20	c23
	F	3.891	0.224	4.148	4.188	3.516	4.234	3.570
	V	3.484	0.208	4.328	4.234	3.383	4.289	3.531

Table 30: Objective ToRs for Noise Conditions

Noise		Reference		CuT1	CuT2	CuT3	CuT4	G.729.1
R1 <i>nvt</i> G.729E, Car15	Exp. 4	c08		c09	c11	c13	c15	c17
	D	4.099	0.147	3.979	4.214	4.000	4.208	3.797
	J	4.302	0.162	3.854	4.042	3.766	4.109	3.484
R1 <i>nvt</i> G.729E, Street15	Exp. 5	c08		c09	c11	c13	c15	c17
	J	4.276	0.174	3.781	4.120	3.833	4.203	3.635
	H	4.406	0.165	4.089	4.391	4.037	4.333	3.813
R3 <i>nvt</i> G.722.2, 15.65k, IT15	Exp. 6	c09		c12	c16	c20	c24	c28
	A	4.474	0.155	4.432	4.458	4.156	4.526	4.172
	V	4.427	0.183	4.063	4.307	3.969	4.412	4.156
R4 <i>nvt</i> G.722.2, 23.85k, IT15	Exp. 6	c10		c13	c17	c21	c25	c29
	A	4.401	0.152	4.594	4.635	3.927	4.656	4.396
	V	4.385	0.173	4.443	4.708	3.698	4.714	4.479
R5 <i>bt</i> G.722.2, 23.85k, IT15	Exp. 6	c10		c14	c18	c22	c26	c30
	A	4.401	0.158	4.620	4.583	3.818	4.641	4.516
	V	4.385	0.164	4.599	4.625	3.401	4.599	4.516
R3 <i>nvt</i> G.722.2, 15.65k, Off20	Exp. 7	c09		c12	c16	c20	c24	c28
	A	4.427	0.175	4.214	4.297	3.969	4.370	3.943
	V	4.438	0.171	4.208	4.323	3.922	4.313	4.115
R4 <i>nvt</i> G.722.2, 23.85k, Off20	Exp. 7	c10		c13	c17	c21	c25	c29
	A	4.307	0.144	4.234	4.531	4.318	4.547	4.271
	V	4.443	0.172	4.458	4.682	3.766	4.651	4.406
R5 <i>bt</i> G.722.2, 23.85k, Off20	Exp. 7	c10		c14	c18	c22	c26	c30
	A	4.427	0.137	4.427	4.589	4.234	4.516	4.464
	V	4.438	0.166	4.620	4.656	3.620	4.677	4.573

6.4 Summary on G.EV-VBR Selection Test Results

Q7/12 reviewed the reports of the listening laboratories [AH-07-02 from Dynastat, AH-07-04 from Nokia, AH-07-05 from VoiceAge, AH-07-06 from BIT, AH-07-07 from France Telecom, AH-07-11 from NTT-AT and AH-07-13 from ARCON Corp.].

The following relevant comments were made on the report of the global analysis laboratory:

- On the report from Lab B for experiment 1, some concerns on the results were raised from a number of companies, e.g.
 - Direct MOS was felt too low, as it was lower than some references, and lower than most tested conditions including CuTs (20 out of 29, not all significantly different).
 - The MNRU scores were felt too high for the lower Q values with unexpected high standard deviation.
 - The MOS of G.729 and G.729E were unexpected, since G.729E scored significantly lower than G.729. Also G.729.1 low bit rate scored significantly better than its higher bit rate. G.729.1 high MOS were felt due to the silence cleaning effect.

- The inter-correlation between results for the two test labs contracted for each experiment were computed, the correlation for experiment 1 was 0.43 while for the other experiments was 0.79 and higher.

Nevertheless MNRU curve was S-shaped as expected, relative performance of fixed point codecs (G.729E and G.729.1 at 8 kbit/s) low input level compared to nominal level was as expected; the work was endorsed since compliant with the test plan and Lab B results were agreed to be included in the global analysis.

- Some requirements were evaluated in the saturation region where the "better than" criterion on its own would be inappropriate, while "better than" or "equivalent to direct" would compensate the fact that the scores are in the saturation region.

Table 31 summarizes the results of the EV-VBR Selection Phase. Two of the candidate codecs showed superior performance on the subjective tests compared to the remaining codecs (including G.729.1). These two codecs, CuT2 and CuT4, were virtually identical on all selection criteria. Where they differed, the difference was small – CuT4 was 0.014 higher on FoM, CuT2 passed one more Objective ToR than CuT4. The prescribed selection criteria cannot distinguish between the two best performing candidate codecs – CuT2 and CuT4.

Table 31: Summary of Selection Criteria

Criterion	Note	CuT1	CuT2	CuT3	CuT4	G.729.1
Requirement ToR Passes	out of 70 ToR's	63	69	35	69	62
Weighted ToR Passes	100-point scale	90.28	98.54	53.95	98.54	92.20
Figure of Merit	5-point scale	3.982	4.037	3.578	4.050	3.817
Objective ToR Passes	out of 28 ToR's	21	25	6	24	11

Note: G.729.1 was included in the test plan to cope with a withdrawal of one candidate. G.729.1 was used with higher bit rates than the candidate codecs in

- Experiment 2a,
- the 0, 2, 4, 6 & 10% FER condition in Experiment 2c,
- R1 background noise conditions in Experiment 6 and Experiment 7.

The effect of these bit rate revisions has not been reflected in the above table.

The global analysis laboratory [AH-07-16] supplementary analysis comparing the two best performing codecs on an experiment by experiment basis. The global analysis laboratory provided [AH-07-18] supplementary analysis comparing the three best performing codecs on an experiment by experiment basis. The results of those analyses showed that:

- CuT2 and CuT4 showed equivalent performance in eight (out of nine) experiments.
- CuT4 was significantly better than CuT2 in experiment 3.
- CuT2 was significantly better than CuT1 in experiments 2a, 4, 5, 6 and 7.
- CuT1 was significantly better than CuT2 in experiment 2c and 3.
- CuT4 was significantly better than CuT1 in experiments 4, 5, 6 and 7.
- CuT1 was significantly better than CuT4 in experiments 2c and 3.

6.5 EV-VBR computation on Frequency Attenuation

France Telecom provided the frequency response of the four codecs under test [AH-07-17]. Two candidates i.e. CuT2 and CuT4 showed noticeable spectral components above 7kHz, another

candidate CuT3 also showed some spectral components above 7kHz. Q7/12 was unable to determine whether this impacted on speech and/or music quality without further investigation. Candidate CuT1 showed attenuation up to 20dB around 3.5kHz at low bit rate.

The attenuations were obtained by subtracting the frequency response of P50 speech file to the frequency responses of the output of each CuT. The frequency responses were computed using software tool freqresp. This tool computes and outputs the average amplitude spectra in ASCII and also produces a bitmap file. P50 test signals which are representative of speech signals were used to compute the frequency response: P50m.16k for male speech and P50f.16k for female speech.

Figure 22 and Figure 23 show the average attenuation of the four candidate coders at the 5 bit rates (8, 12, 16, 24, and 32 kbit/s) with P50m.16k and P50f.16k input files, respectively.

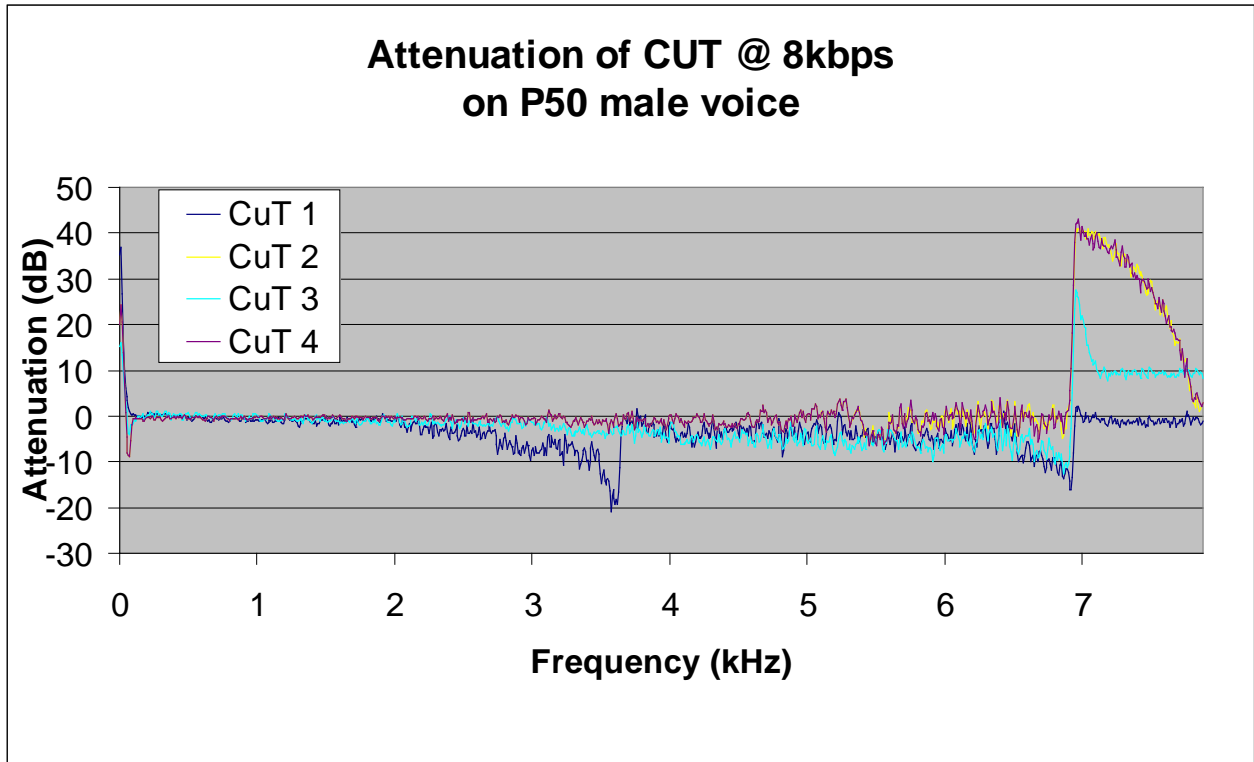


Figure 22(a) – Male speech (P50m.16k), CuT @ 8 kbit/s

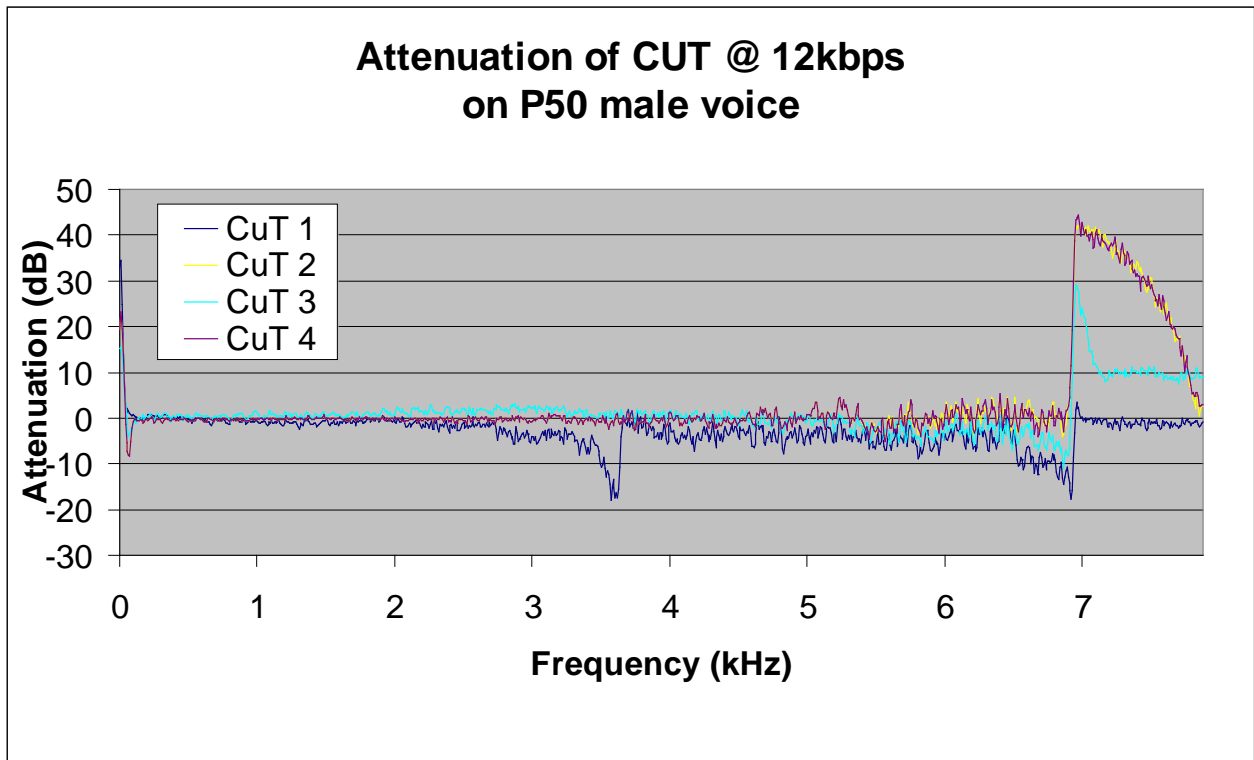


Figure 22(b) – Male speech (P50m.16k), CuT @ 12 kbit/s

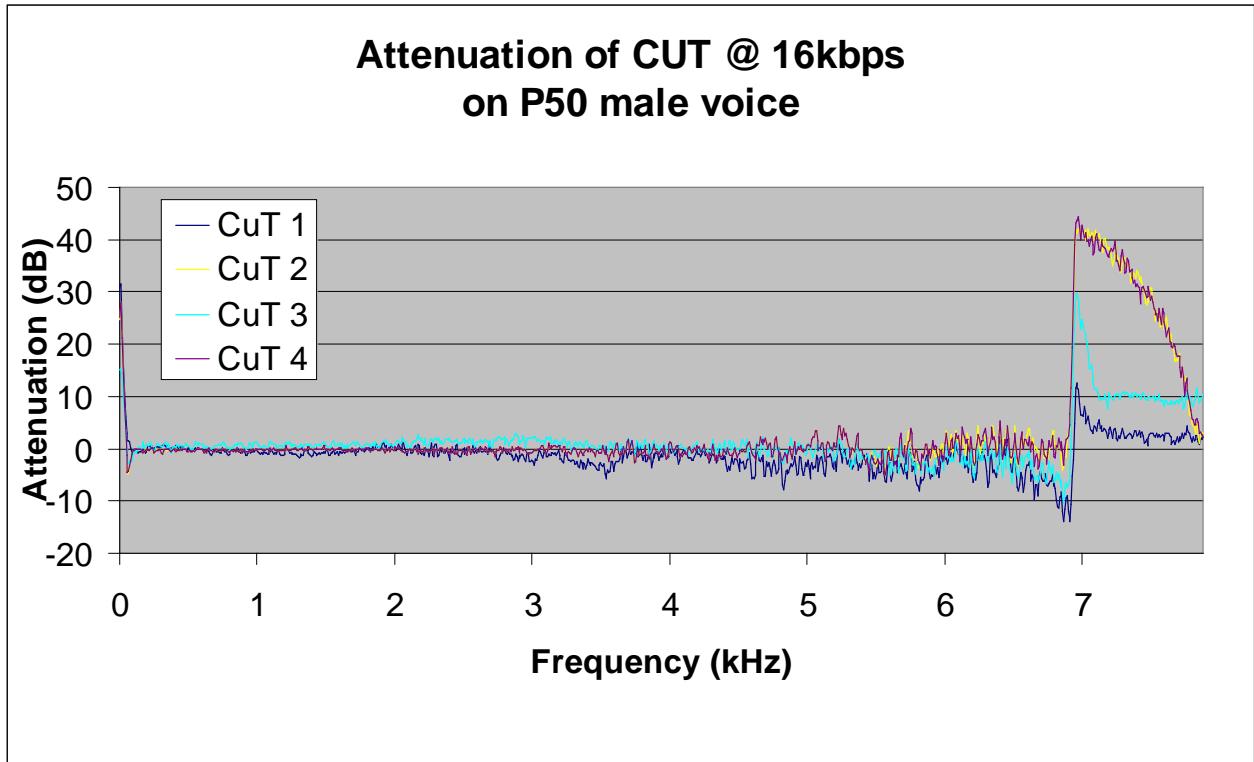


Figure 22(c) – Male speech (P50m.16k), CuT @ 16 kbit/s

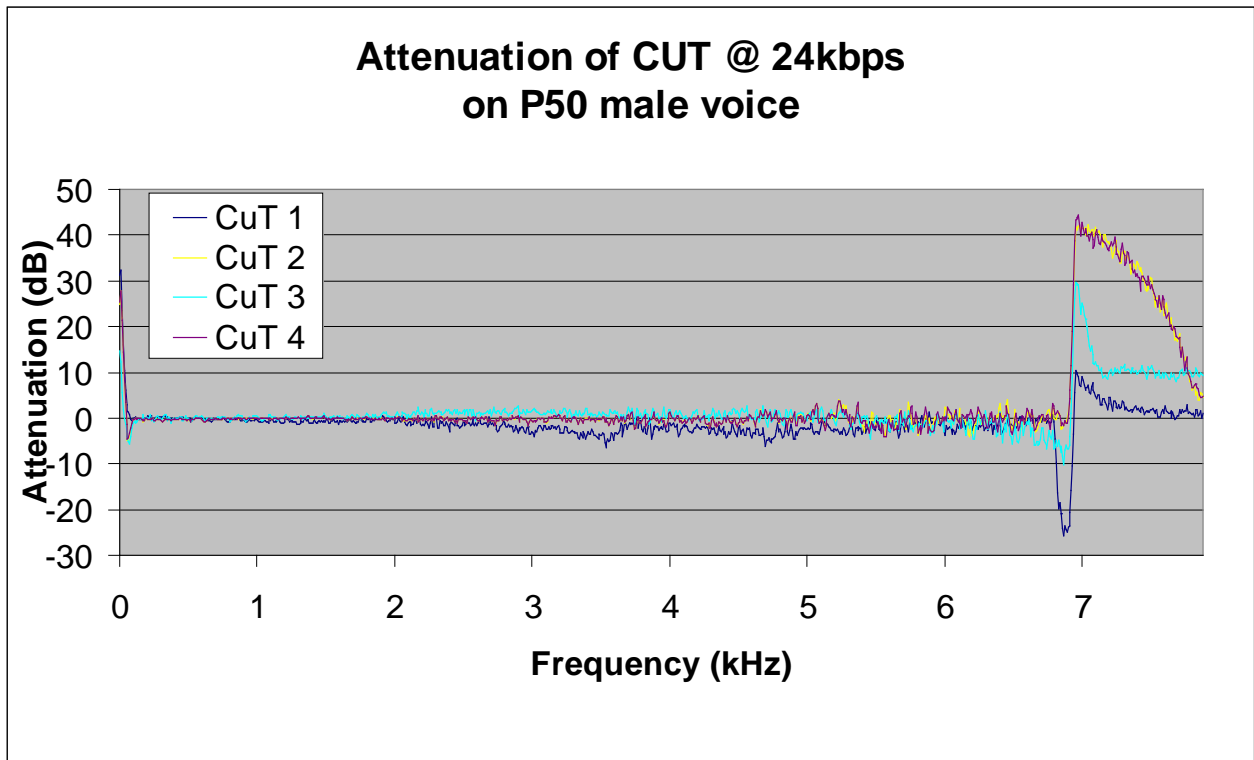


Figure 22(d) – Male speech (P50m.16k), CuT @ 24 kbit/s

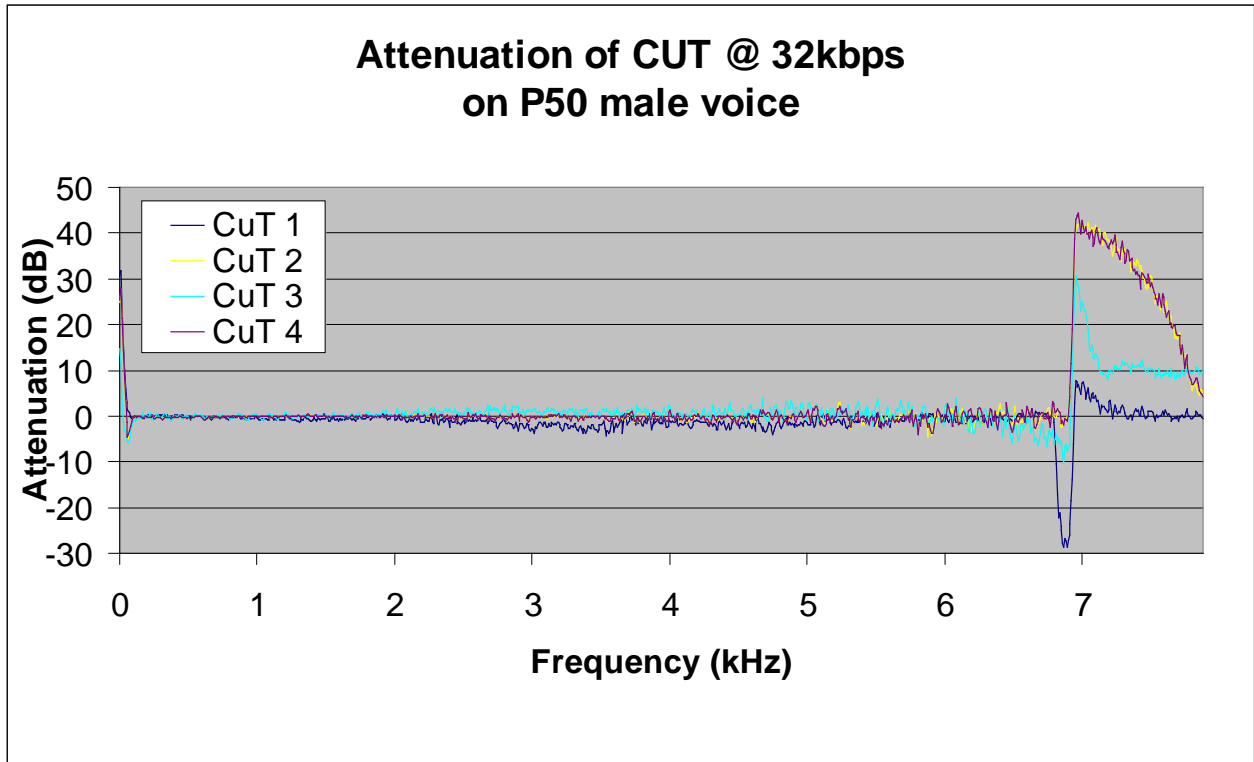


Figure 22(e) – Male speech (P50m.16k), CuT @ 32 kbit/s

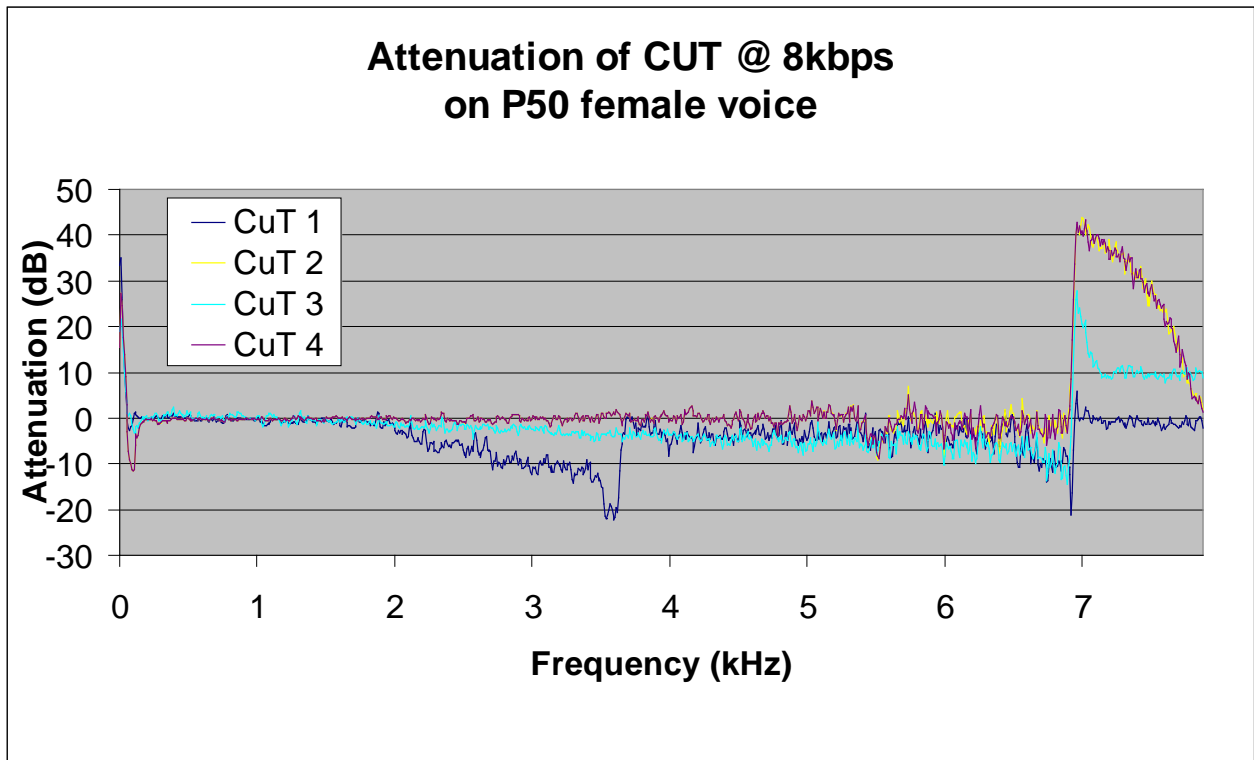


Figure 23(a) – Female speech, CuT @ 8 kbit/s

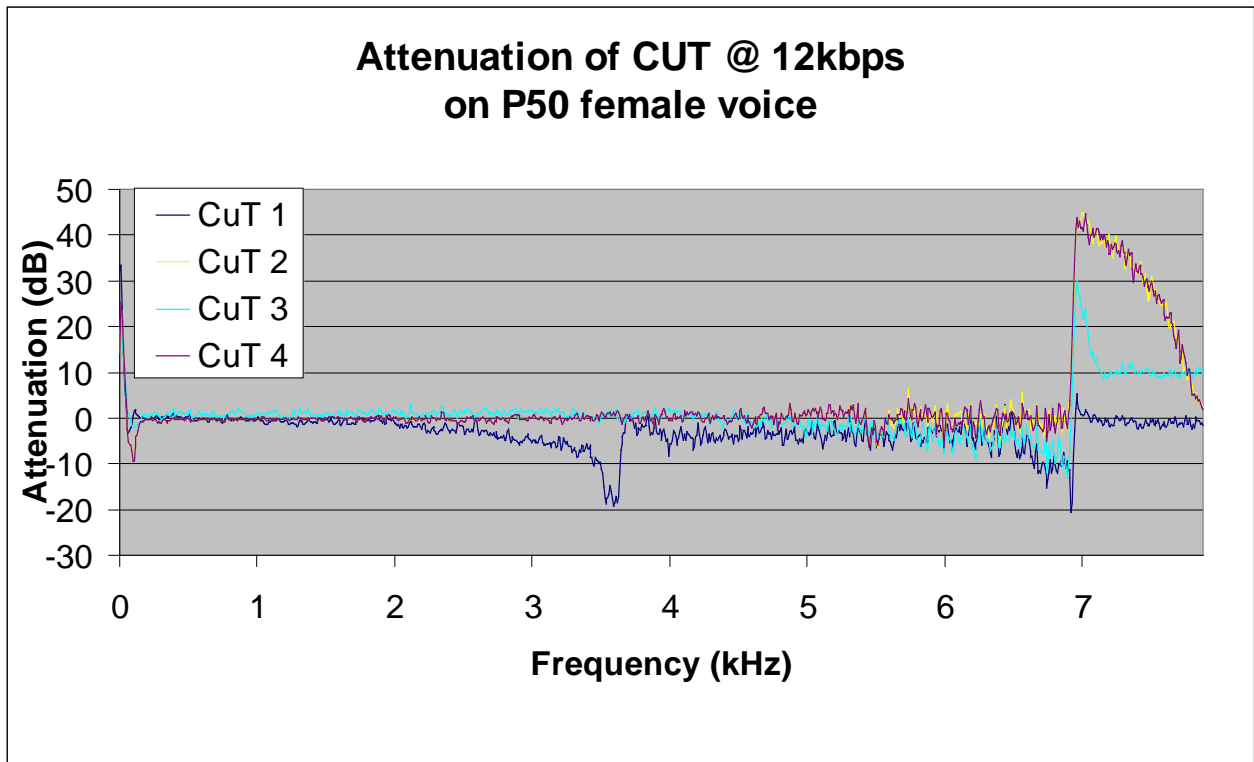


Figure 23(b) – Female speech, CuT @ 12 kbit/s

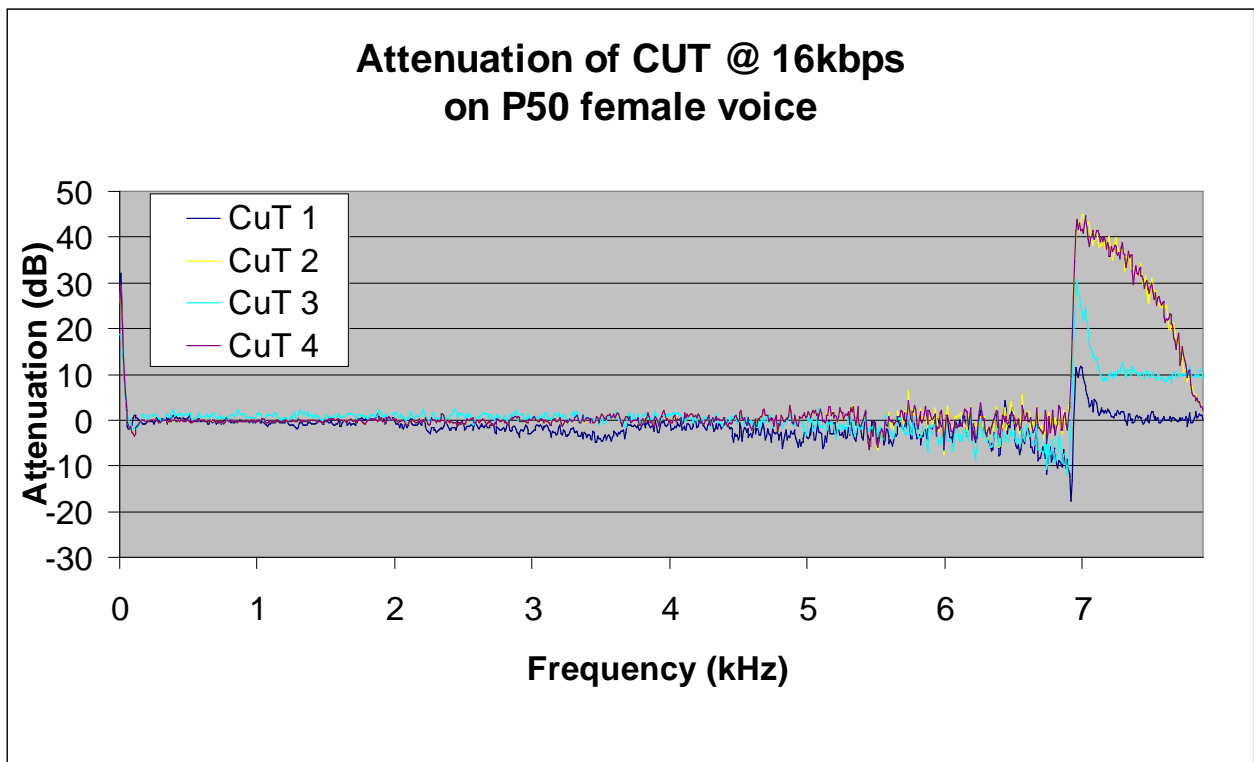


Figure 23(c) – Female speech, CuT @ 16 kbit/s

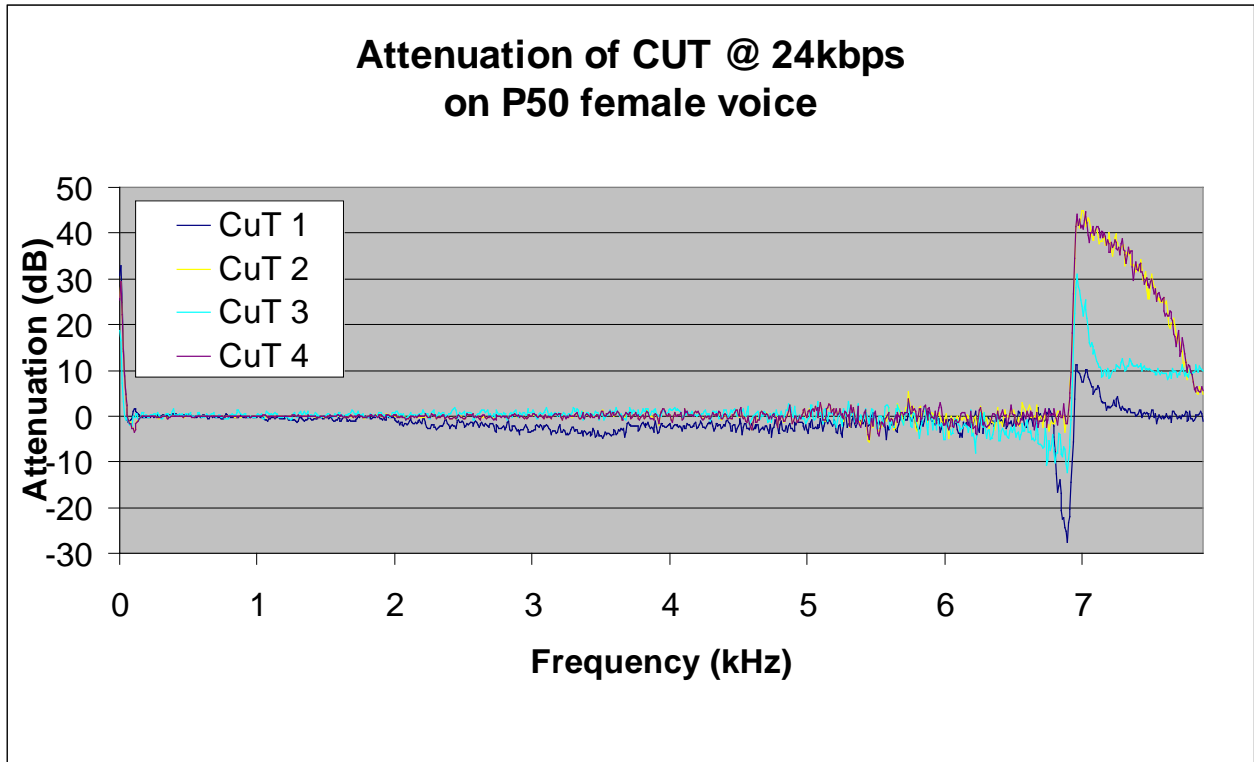


Figure 23(d) – Female speech, CuT @ 24 kbit/s

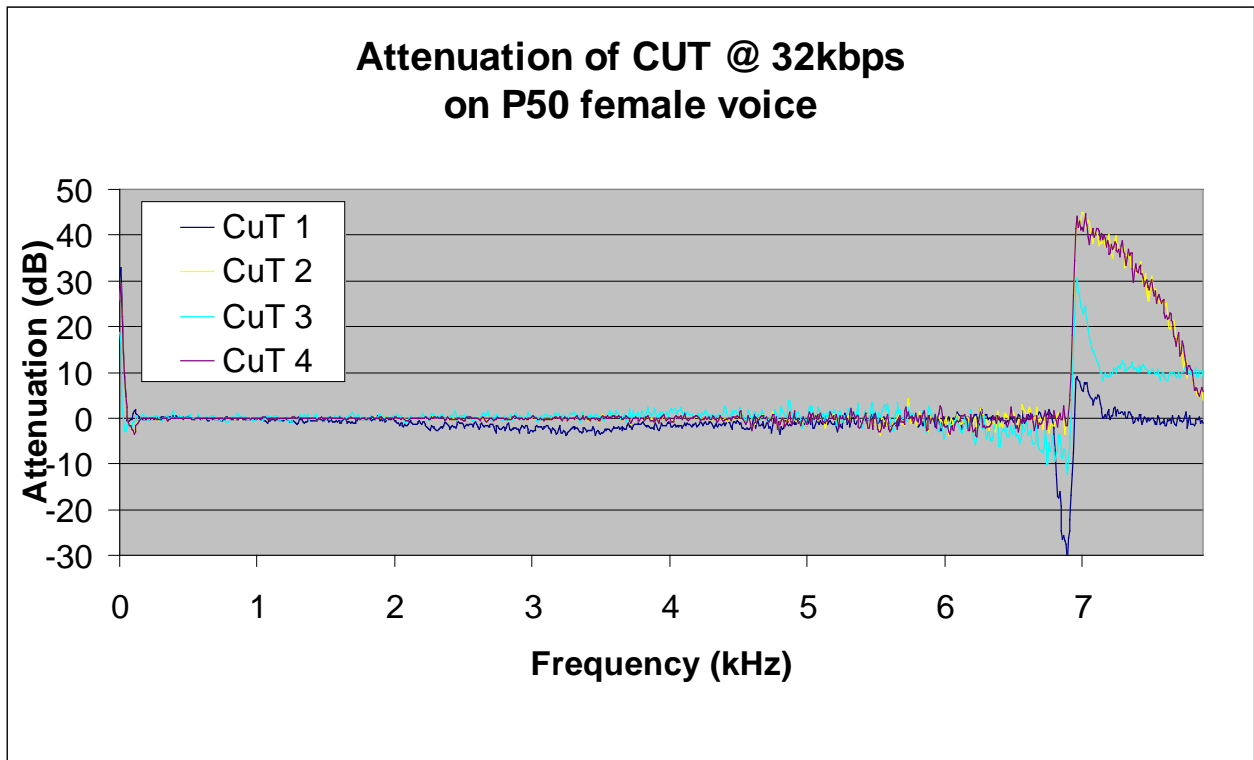


Figure 23(e) – Female speech, CuT @ 32 kbit/s

7 Scope of G.729.1 (G.729EV)

As mentioned in the ToRs of G.729.1, G.729.1 was prepared in a timely fashion, while maintaining speech quality requirements. So the work was focused on main application constraints (e.g. NB to WB only, bit-rate range limited to 8-32 kbit/s).

The targeted applications can be classified into two types:

- Packetized wideband voice (VoIP, VoATM, ToIP, IP phone, private networks) – this does not prevent from having access to the wireless world through a gateway
 - designed for applications requiring scalable wideband on top of G.729, in particular for residential and corporate services such as providing mono or multi-lines
 - designed for easy integration with existing VOIP infrastructure and services allowing for a fast deployment
 - designed to cope with other services such as videoconferencing and VOD
 - scalability used for:
 - gateways or other devices that multiplex or combine data streams (including audio)
 - handling heterogeneous accesses/terminals
 - Examples are residential gateways, IPBX, CME/Trunking equipment (optimization of bitrate allocation and network congestion handling) and voice messaging, which requires capacity vs quality tradeoff optimization and access adaptation (in terms of bitrate and format, for heterogeneous accesses)
- High quality audio/video conferencing
 - graceful quality degradation from wideband (50-7000 Hz) to narrowband (100-3400 Hz)
 - stereo capability was a desirable feature

8 G.729EV Candidates: Algorithmic Overview, Delay and Complexity Figures

8.1 France Telecom Candidate [10]

The FT codec accepts input signals sampled at 16 kHz on the encoder side and generates output signals sampled at 16 kHz on the decoder side. The output bitstream produced by the encoder is scalable, consisting of 3 layers at 8 kbit/s, 12 kbit/s and 14 kbit/s and 45 fine granularity layers from 14 kbit/s to 32 kbit/s by steps of 0.4 kbit (one byte per 20ms frame) including the 10 layers (by steps of 2 kbit/s) required.

Figure 24 and Figure 25 illustrate the encoder part, whereas Figure 27 and Figure 28 illustrate the decoder part.

The FT codec structure relies on three components: first CELP coding, then bandwidth extension and finally predictive transform coding similar to the TCX (Transform Coded eXcitation) or TPC (Transform Predictive Coding) techniques.

The CELP coding stage is a two-layer CELP coder that produces an embedded 8 and 12 kbit/s bitstream, and will therefore be called "CELP 8-12k codec".

The bandwidth extension is an additional layer that raises the bitrate to 14 kbit/s and provides a wideband signal ([50-7000] Hz) as an output. In what follows the CELP 8-12k together with the extension layer will be denoted as "8-12-14k codec".

8.1.1 Description of the encoder part

8.1.1.1 Overall description: Encoder at 8, 12 and 14 kbit/s (Figure 24)

The encoder operates on 20 ms input frames, i.e. 320 samples of input signal at 16 kHz. The processing starts with the 8-12-14k codec, presented on Figure 24. A high pass filter with 50 Hz cutoff frequency is first applied to the input signal, producing signal S_{WB} .

S_{WB} is low-pass filtered and sub-sampled by a factor of 2, yielding the 8 kHz narrow band signal S_{LB} , which is processed by the CELP 8-12k coder.

The first layer of the CELP 8-12k coder is bitstream interoperable with ITU-T G.729 at 8 kbit/s. Close to G.729 main body, this coder part is described in Section 2.1

The second layer of the CELP 8-12k coder is a CELP enhancement layer as described in Section 2.2. The bitrate of this second layer is 4 kbit/s. At 12 kbit/s, the decoder produces a narrow band output signal of a higher quality than the 8 kbit/s output.

Signal S_{LB} is encoded and decoded by the CELP 8-12k to obtain synthesized signal \tilde{S}_{12}^{LB} which is up-sampled and low-pass filtered, yielding the 16 kHz sampling frequency signal \tilde{S}_{12}^{WB} .

The bandwidth extension layer is the last part of the 8-12-14k codec. An 18-order LPC analysis of the wideband filtered input signal S_{WB} is performed. Prior to this analysis, a pre-emphasis filter is applied (using a pre-emphasis factor $\mu=0.68$). LPC parameters are vector quantized, exploiting the fact that the narrow band part of the signal has already been analysed and LPC coded. The description of the multi-stage VQ used for the quantization of WB LPC parameters is given in Section 2.3.

A wideband excitation signal \tilde{e}_{14}^{WB} is then derived from the narrow band excitation signal calculated by the CELP 8-12k codec. Section 2.4 describes the process used to generate this signal.

The excitation signal \tilde{e}_{14}^{WB} is filtered by the (decoded) synthesis filter $\frac{1}{\hat{A}_{WB}(z)}$, then a de-emphasis filter dual to the pre-emphasis filter mentioned above is applied, yielding the signal \bar{S}_{14}^{WB} . A gain adjustment is performed to align the energy level of the upper band (3400-7000 Hz) of \bar{S}_{14}^{WB} with the one of the input signal. A high-pass filtered version \bar{S}_{14}^{UB} of \bar{S}_{14}^{WB} is compared to the corresponding high pass filtered version r_{12}^{UB} of the residual error in wideband (r_{12}^{WB}) of the CELP 8-12k layer, to compute gain g_{WB} , which is 4 bits quantized. This takes into account aliasing components in the neighbourhood of the band split border. The level adjustment is performed on 5 ms blocks of samples (i.e. 80 samples long), the adjusted signal \tilde{S}_{14}^{UB} is then added to the CELP 8-12k upsampled output \tilde{S}_{12}^{WB} to obtain the final 14 kbit/s output signal \tilde{S}_{14}^{WB} .

The last layer of the FT codec operates in the transform domain using a transform predictive coding scheme. Figure 25 describes this part of the encoding process. The delayed filtered input signal S_{WB} and the 8-12-14k codec synthesis signal \tilde{S}_{14}^{WB} are both filtered by a perceptual weighting filter $\hat{W}_{WB}(z) = \hat{A}_{WB}(z/\gamma) \times (1 - \mu z)$ with $\gamma = 0.92$ and $\mu = 0.68$. Those two signals are then transformed using a MDCT (Modified Discrete Cosine Transform) performed on 40 ms windows with 50% overlap.

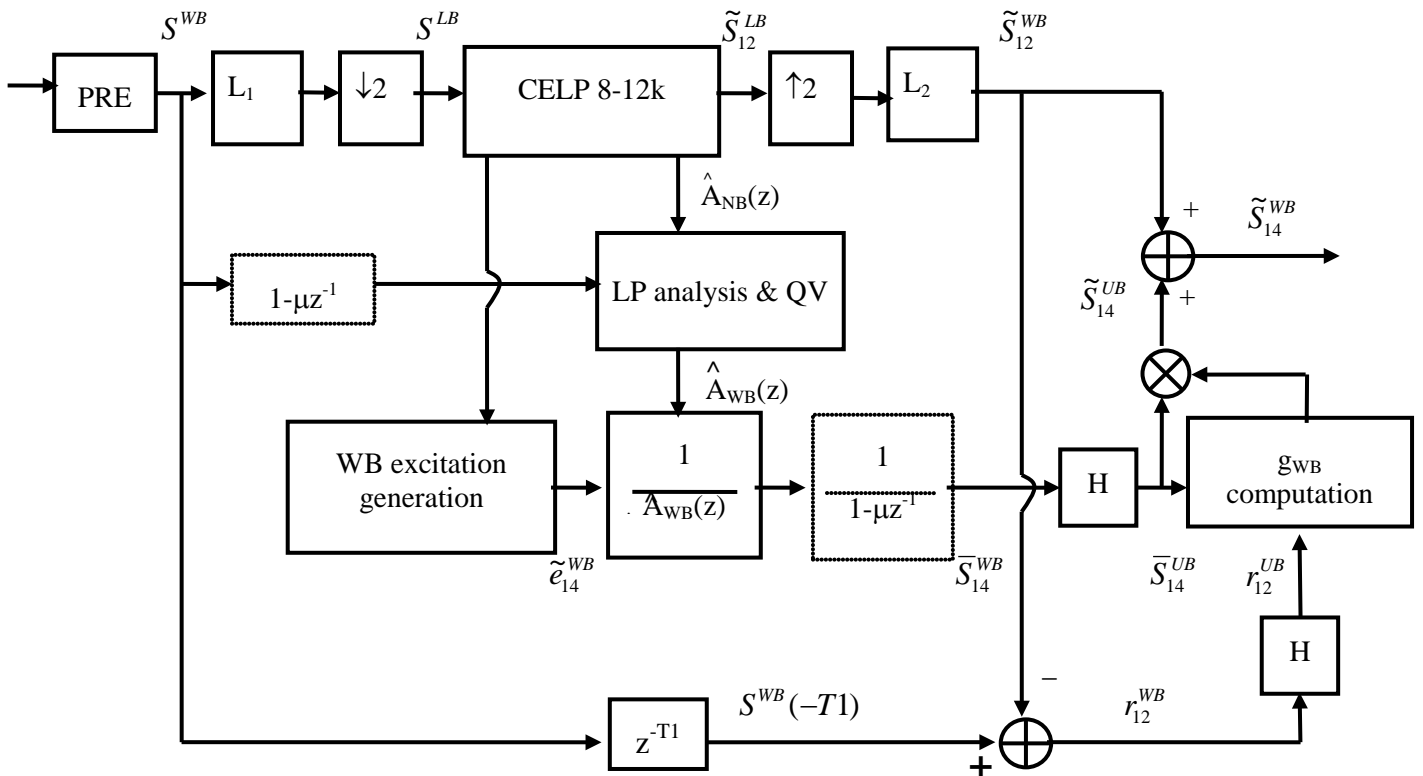


Figure 24 – Encoder block diagram for 8, 12, 14 kbit/s

8.1.1.2 Overall description: Encoder at 16 kbit/s and above (Figure 25)

MDCT coefficients corresponding to the difference between the weighted input signal and the 8-12-14 kbit/s codec for the [0, 3400 Hz] band, and to the weighted input signal for the [3400, 7000 Hz] band are encoded. The last 40 coefficients (corresponding to the [7000, 8000 Hz] band) are set to zero.

The spectrum is encoded by the coder part referred as "TDAC codec" in Figure 25, where TDAC stands for Time Domain Aliasing Cancellation. The spectrum is divided into 18 bands, the first band is composed of 8 coefficients and all the other bands are 16 coefficients long.

For each band, the R.M.S of the signal to be encoded is calculated, quantized and Huffman coded.

The number of bits allocated to each band is then computed, relying on the quantized energy of the band. Therefore the bit allocation can be also calculated at the decoder level and no other side information is requested. The bit allocation algorithm is presented in Section 2.5.

Then the normalized coefficients of the bands are vector quantized, using codebooks that are embedded in size and composed of an union of permutation codes.

Finally, all the information from the CELP 8k, 12k, extension layer, spectrum envelope and normalized coefficients is multiplexed and transmitted to the decoder.

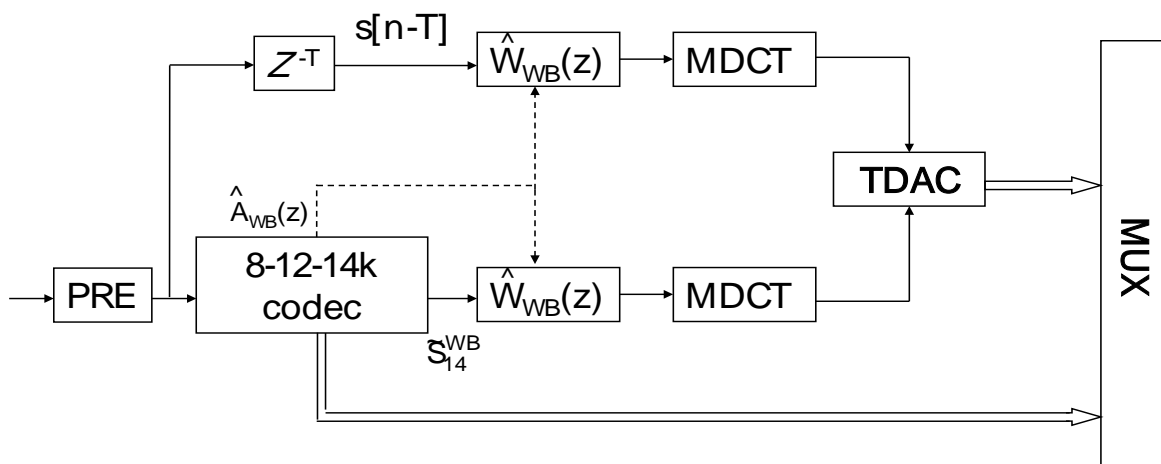


Figure 25 – Encoder diagram for 16 kbit/s and above

CELP 8k part

The 8 kbit/s part of the FT codec is similar to G.729 main body except that:

- The pre-filtering of G729 has been suppressed.
- Post-filtering and post-processing have been suppressed, unless explicitly mentioned in the description (case of 8 kbit/s and 12 kbit/s outputs).
- The fixed codebook search is replaced by a less complex one based on global pulse replacement method. This algorithm consists of two stages: initial codevector determination and pulse replacement. The initial codevector of each track is determined using pulse-position likelihood-estimate vector. Then a new pulse is searched sequentially by a pulse basis replacement from all tracks.

CELP 8-12k part

The second layer of the CELP 8-12k codec and the connection to the 8k CELP layer is described on Figure 26.

It relies on an innovation codebook that, when combined with the excitation from the core layer, will produce a richer excitation signal.

The second fixed codebook is a derivative of the algebraic codebook used in the 8 kbit/s stage. Every 5 ms subframe is split into 5 tracks in a similar way as in G.729.

Four positions are selected in tracks 1, 2, 3 and 4/5. The innovative codebook search is based on the same algorithm as described in the previous sections. The resulting innovative codeword is scaled with a gain factor $\mathbf{g}'\mathbf{c}$.

The coefficient $\mathbf{gamma} = \mathbf{g}'\mathbf{c} / \mathbf{g}\mathbf{c}$ is quantized using a 3 bit scalar quantizer.

The bitrate is 4 kbit/s and the associated parameters are pulse positions and signs and subframe gains.

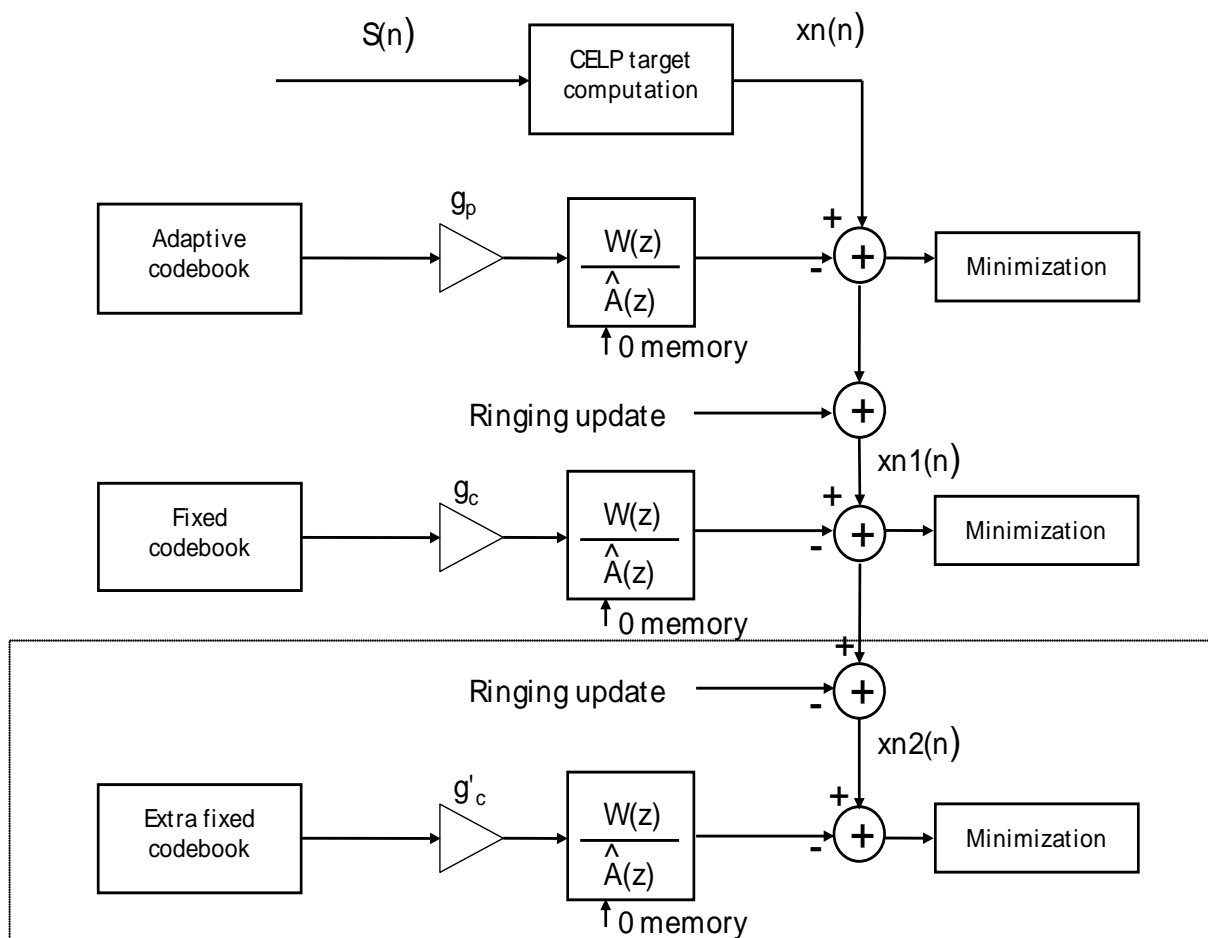


Figure 26 – CELP 8-12 kbit/s

WB LPC Quantization

An 18-order LPC analysis is performed on the pre-emphasized wideband input signal. LPC parameters are converted into LSF parameters and quantized with 2-stage switched predictive VQ. The 2-stage VQ codebook is split into 3 sub-codebooks in its second stage. A 6-bit 18-dimension VQ codebook is exploited in the first stage, and 17 bits are assigned to the 3 sub-codebooks of the second stage. One bit is used for switching two predictors. As a result, the 2-stage VQ uses 24 bits for quantizing the 18-order LPC parameters. One of the two predictors utilizes both of intra-frame and inter-frame predictions, while the other predictor exploits intra-frame prediction only.

14 kbit/s WB excitation generation

The wideband excitation signal \tilde{z}_{14}^{WB} is obtained from excitation parameters of the CELP 8-12k codec: pitch lag, pitch gain and fixed codebook excitations from the 8k and 12k layers with associated gains. This excitation is obtained by addition of two wideband excitation signals:

- the first one is an adaptive excitation obtained by transposing to wideband the pitch parameters (pitch lag and interpolation filter)
- the second one is an innovation part produced by upsampling the excitation of the CELP 8-12k.

TDAC bit allocation

The bit allocation in the TDAC codec part is divided in two steps. At first, the number of bits allocated to each band is computed according to:

$$b_{opt}(i) = \frac{1}{2} \log_2 [e_q^2(i)] + C, 0 \leq i \leq M - 1$$

where

$$C = \frac{B}{M} - \frac{1}{2M} \sum_{i=0}^{M-1} \log_2 [e_q^2(i)]$$

is a constant factor, B being the total number of bits available (or bit budget), M the number of bands, $e_q^2(i)$ the decoded value of the spectrum envelope for the band number i .

Each obtained value is rounded to the rate of the nearest possible codebook. If the total allocated rate is not exactly equal to the sum of the available codebooks rates, an iterative procedure is employed to get closer to the total bit budget.

8.1.2 Description of the decoder part

Overall Description

After de-multiplexing the bitstream, the decoder operates as follows depending on the number of bits that have been received. Four cases can be identified. The first 3 cases are described on Figure 27 and the last one is illustrated by Figure 28.

1. If the received bit rate is 8 kbit/s, only the G.729 bitstream is available and G.729 main body decoding (including post filtering and post processing) is applied. The narrow band output signal is up-sampled and filtered to produce a 16 kHz output signal.
2. If the received bit rate is 12 kbit/s, both 8k (G729) decoding and 12k enhancement layer decoding are performed. The obtained synthesis is also post filtered and post processed using the G.729 main body algorithms. Then the signal is up-sampled and filtered to produce a 16 kHz output signal.
3. If the received bit rate is 14 kbit/s, the CELP 8-12k decoder (as in case 2) is first activated, followed by bandwidth extension using transmitted WB LPC parameters and adjustment gains. The wide band excitation is generated using the same procedure as in the encoder part, then the 14k synthesis is performed and finally the level is adjusted like is done in the encoder part. To obtain a [50,7000 Hz] output signal, the filter bank of the transform predictive layer is used with the 40 last coefficients set to zero (i.e. the 14k output signal is weighted in the perceptual domain, MDCT transformed then MDCT inverse transform is done followed by inverse perceptual weighting filtering).
4. For cases when the received bitrate is above 14 kbit/s (i.e. from 16 to 32 kbit/s) step 3 (8-12-14k decoding) is first performed. Then, depending on the number of extra bits received, the decoding scheme is adapted:
 - If the total number of bits received corresponds to the whole or part only of the spectrum envelope (no bits from the fine structure having been received), then the received partial or full spectral envelope is used to re-adjust the energy of the MDCT bands corresponding to the 3400 Hz to 7000 Hz range (i.e. where the bandwidth extension has been used), to gradually improve the upper band quality.

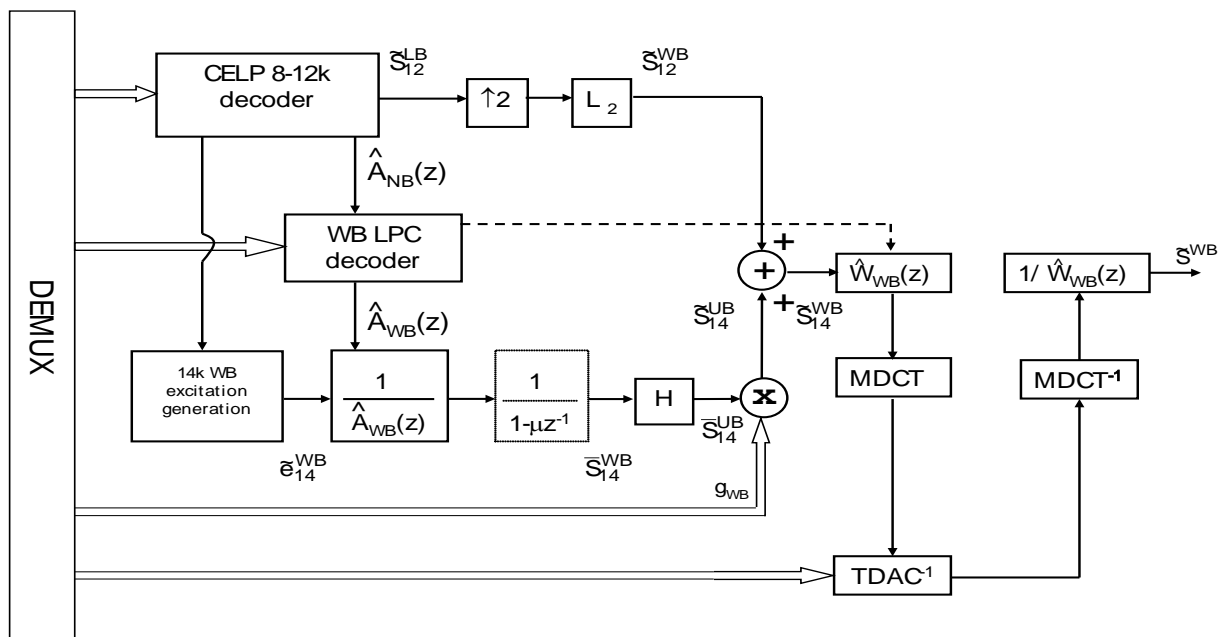


Figure 27 – Decoder for 8, 12, 14 kbit/s

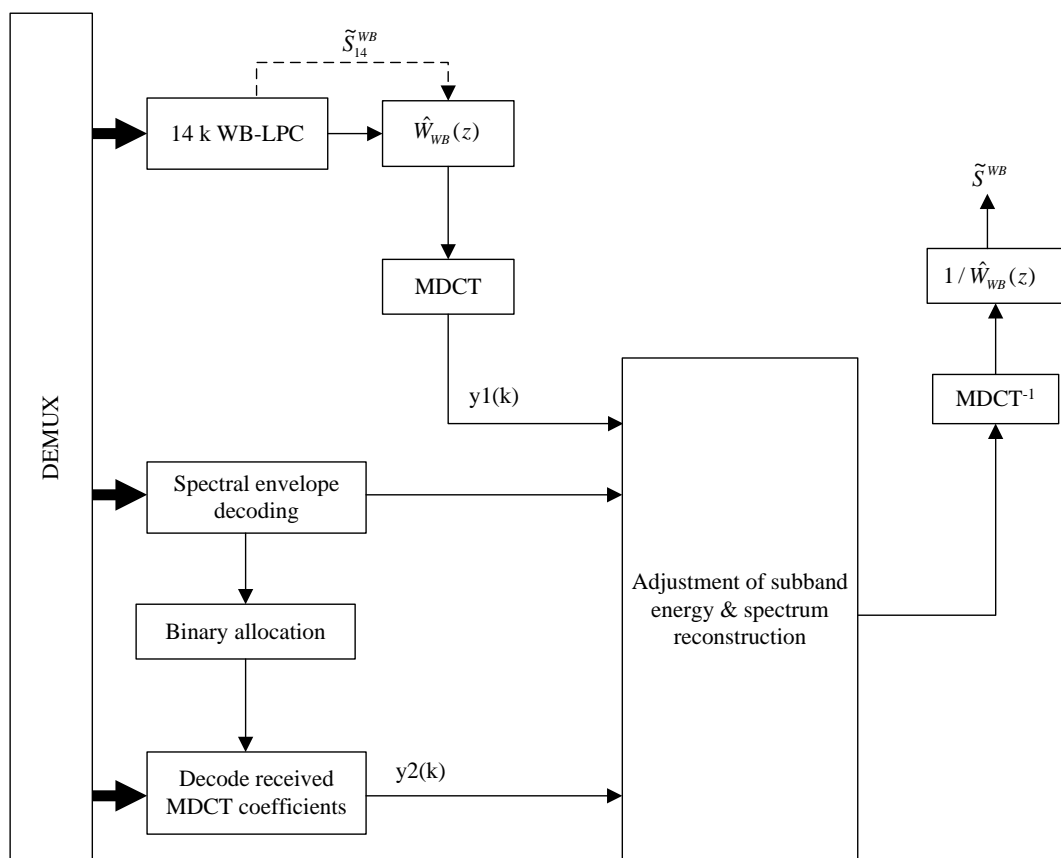


Figure 28 – Decoder for 16 kbit/s and above

- If the total number of bits received corresponds to the full spectrum envelope plus part or the whole fine structure, then the bit allocation is performed in the same manner as at the encoder side. For the bands where the fine structure has been received, MDCT coefficients are calculated using the decoded fine structure and spectrum envelope. When the fine structure has not been received, the processing depends on which part of the spectrum the missing band is part of. If it is an upper band (i.e. between 3400 Hz and 7000 Hz), as in the previous paragraph, level adjustment of the 14k synthesis signal is performed using the knowledge of the spectrum envelope for the band. If it is a lower band, coefficients are just set to zero.

Therefore the MDCT spectrum is composed of:

- for the lower band the decoded CELP 8-12k MDCT coefficients plus (when received) the decoded coefficients of the CELP 8-12k error signal,
- and for the upper band, either (if received) the decoded coefficients of the input signal or the 8-12-14k codec MDCT coefficients adjusted in energy, all those signals being in the perceptually weighted domain.

The inverse MDCT is performed and the inverse weighting filter is applied to obtain the final output signal. The reconstruction of the signal from the MDCT coefficients uses a technique to detect possible pre-echoes and reduce them, as presented in Section 3.2.

When a frame erasure occurs, a frame error concealment algorithm is invoked as described in Section 3.3.

Pre-echo reduction

The inverse MDCT involves: first a conversion to the time domain, then overlap-add of the windowed signal of the previous and current windows. The current frame is in the overlap section, corresponding to the last portion of the previous window and the first portion of the current one. The energies of the last portion of the previous window and the last portion of the current one are compared. If they are significantly different the shape of the first portion of the current window is modified to reduce pre-echo energy.

8.1.3 Frame erasure concealment

The frame erasure concealment method is independent of the bitrate used for the frame preceding the erasure. It uses a source-filter model as in the CELP layer. The filter is a wideband LPC filter, the WB LPC filter of the last non erased frame being used when available. If the last non erased frame was an 8 kbit/s or a 12 kbit/s frame, this LPC filter is recalculated from the last received samples. A voiced/unvoiced detection is applied to this past signal. The excitation generation is different depending on the voiced or unvoiced nature of this signal. In voiced cases, the excitation is obtained through adaptive prediction. For this, a LTP analysis of the last received samples is performed using the knowledge of the CELP 8k transmitted pitch lag. The past excitation is then calculated and LT filtered to obtain the current excitation. In unvoiced cases, a non harmonic component is generated based on the preceding excitation. Synthesis is then performed and the energy of the output signal is controlled using a multi-criteria analysis of the past signal components (mainly voiced/unvoiced, pitch and energy slope). This analysis is done at the first erased frame, and in case of several erased frames, the same parameters are kept, except the energy term which decreases according the multi-criteria analysis performed previously.

Finally, memories of different decoder parts are updated with the produced output signal.

8.1.4 Algorithmic delay

The algorithmic delay is **48.75 ms**. The following contributions to this delay are identified below:

- MDCT window: 640 samples @ 16kHz → 40 ms
- Upsampling filter (L2): 20 samples @ 16kHz → 1.25ms
- Lookahead G729A: 80 samples @ 16kHz → 5 ms
- Downsampling filter (L1): 20 samples @ 16kHz → 1.25 ms
- High pass FIR filter for g_{WB} computation (H): 20 samples @ 16 kHz → 1.25 ms

8.1.5 Complexity evaluation

Table 32 show the RAM/ROM requirements and Table 33 has the contribution of different layers to overall complexity. The total WMOPS estimation is **38.95 WMOPS**.

Table 32: RAM/ROM evaluation

	Total 16 bits Words	Encoder 16 bits Words	Decoder 16 bits Words
Static RAM	9305	3843	5462
Dynamic RAM	5000	-	-
Tables (ROM)	17384	-	-
Program ROM	23000	-	-

Table 33: Contribution of different layers to overall complexity

	Encoder (WMOPS)		Decoder (WMOPS)		Encoder + Decoder (WMOPS)	
8kbit/s	-	10.65	-	3.45	-	14.10
12kbit/s	+ 2.3	12.95	+ 0.05	3.5	+ 2.35	16.45
14 kbit/s	+ 8.9	21.95	+ 2.6	6.1	+ 11.5	27.95
32 kbit/s	+ 5.6	27.45	+ 5.4	11.5	+ 11	38.95

8.1.6 Frequency response

The following frequency response curves were computed using STL2005 tool *freqresp*, with P50 test signals: P50m.16k for male speech and P50f.16k for female speech.

Male speech (P50m.16k)

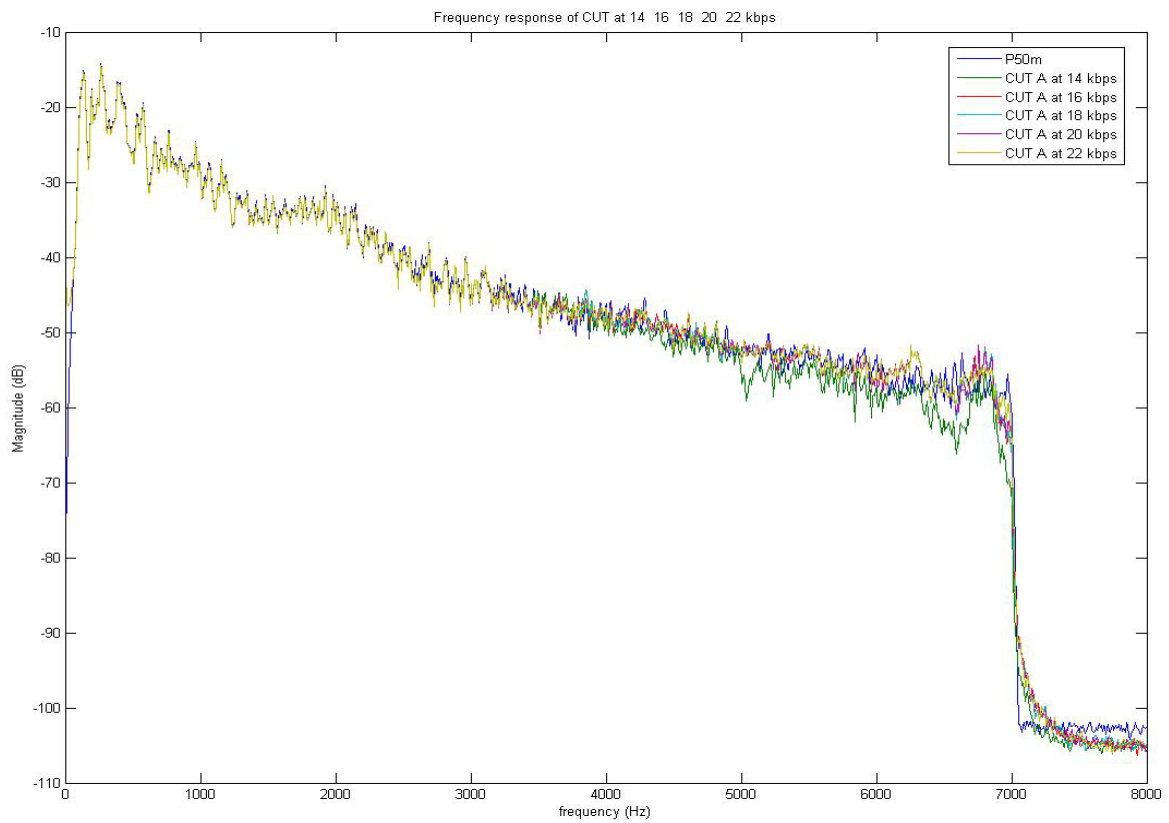


Figure 29 – Bit rates from 14 to 22 kbit/s

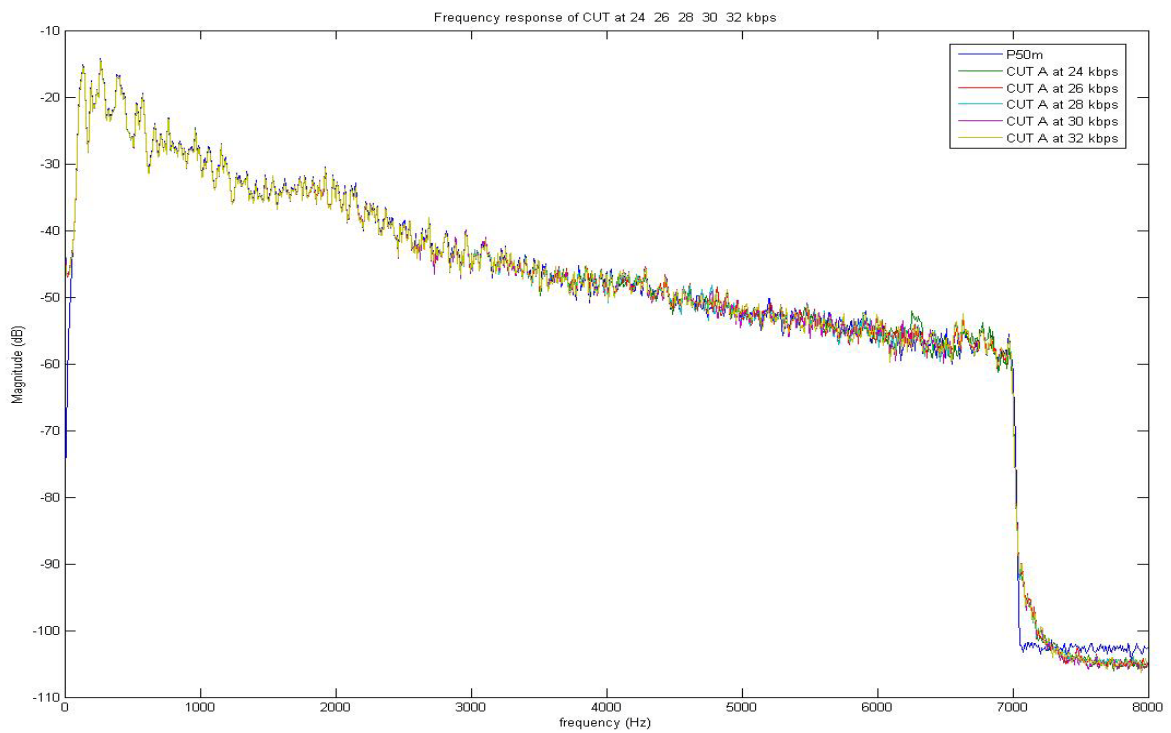


Figure 30 – Bit rates from 24 to 32 kbit/s

Female speech (P50m.16k)

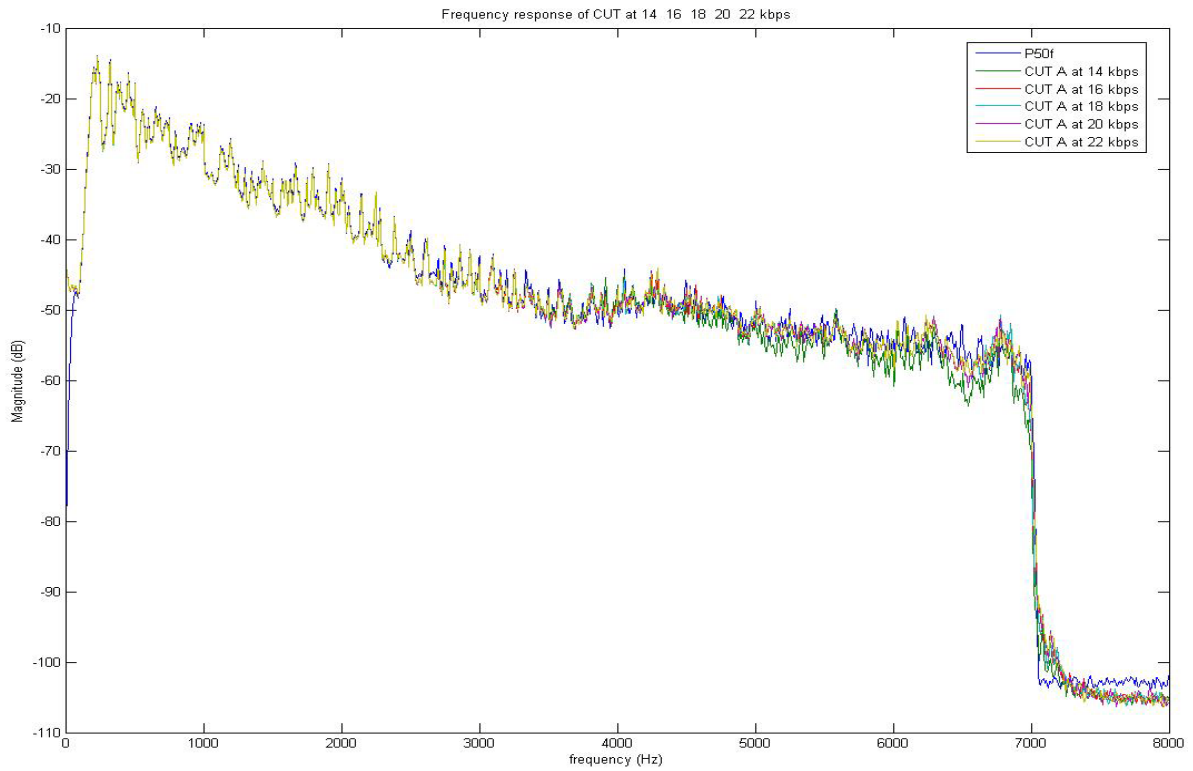


Figure 31 – Bit rates from 14 to 22 kbit/s

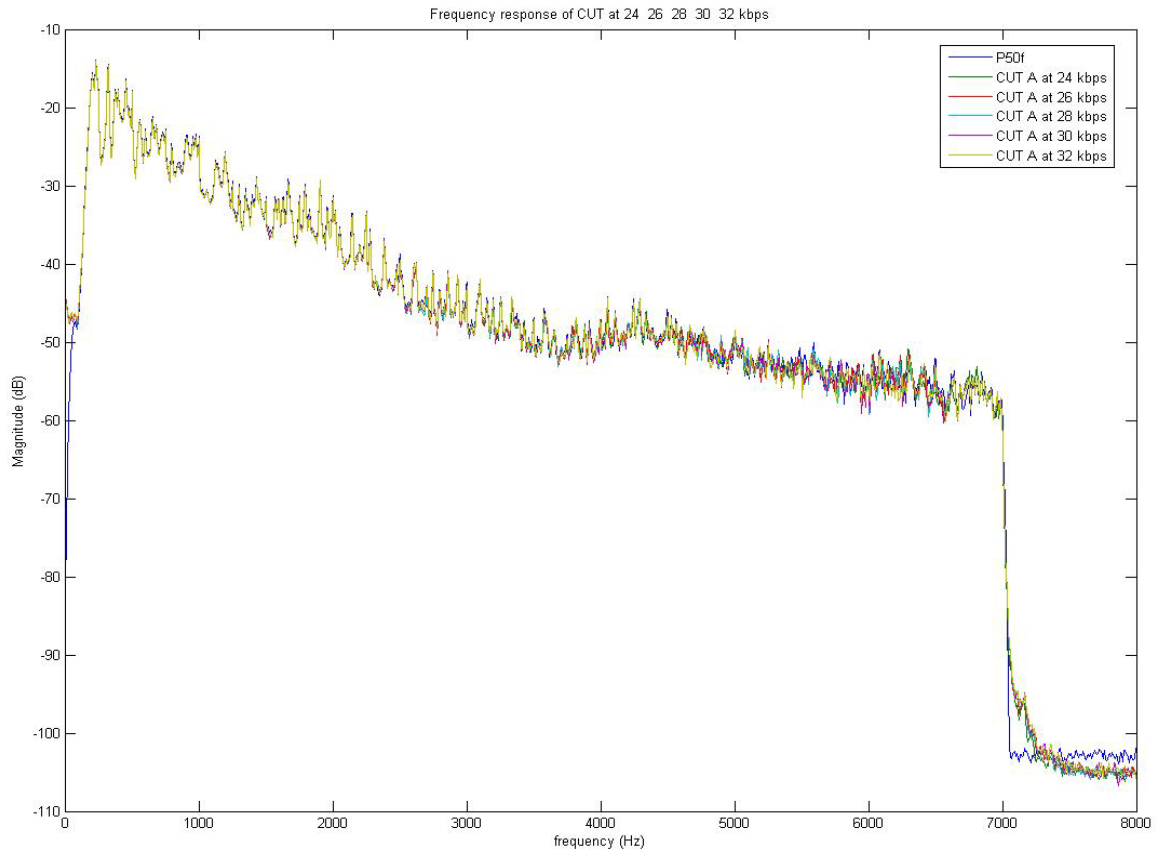


Figure 32 – Bit rates from 24 to 32 kbit/s

8.2 ETRI Candidate [11], [12]

The proposed candidate codec is an embedded variable bitrate speech codec based on G.729 which designed to support bandwidth scalability and bitrate scalability. The proposed codec consist of three layers: core layer, CELP enhancement layer, and wideband extension layer.

The core layer is designed to be interoperable with ITU-T G.729B/G.729AB standard codec. The speech quality of the core layer is further improved by using CELP enhancement layer. Thus the output of core layer and CELP enhancement layer is narrowband signal and both layers operate on every 10 ms.

The wideband extension layer provides bandwidth scalability and bitrate scalability. This layer operates on every 20 ms in transform domain. The output of the wideband extension layer is wideband or intermediate bandwidth signal.

This contribution is organized as follows. Section 2 and 3 describes the proposed encoder and decoder respectively. Section 4 presents bit allocation and frame format of the proposed codec. The algorithmic delay is given in section 5. Section 6 shows the effective bandwidth of output signal at each layer.

8.2.1 Encoder description

The block diagram of the proposed encoder is given in Figure 33. The 16 kHz wideband input is low-pass filtered and then down sampled to 8 kHz. This narrowband signal is encoded by core layer and CELP enhancement layer. The core layer is based on ITU-T G.729 standard codec. The G.729' shown in Figure 34 means modified version of G.729. This layer is bitstream interoperable with ITU-T G.729 standard codec.

The fixed codebook error signal of the core layer is processed in CELP enhancement layer to improve the quality of core layer. Thus the output of this layer is narrowband signal and the bitrate is 1.5 kbit/s.

The difference signal between the delayed wideband input and up-sampled output of local decoder is processed in wideband extension layer. The difference signal is transformed using a MDCT. The coefficients are divided into several bands. The scale factor and normalized shape vector of each band are quantized respectively.

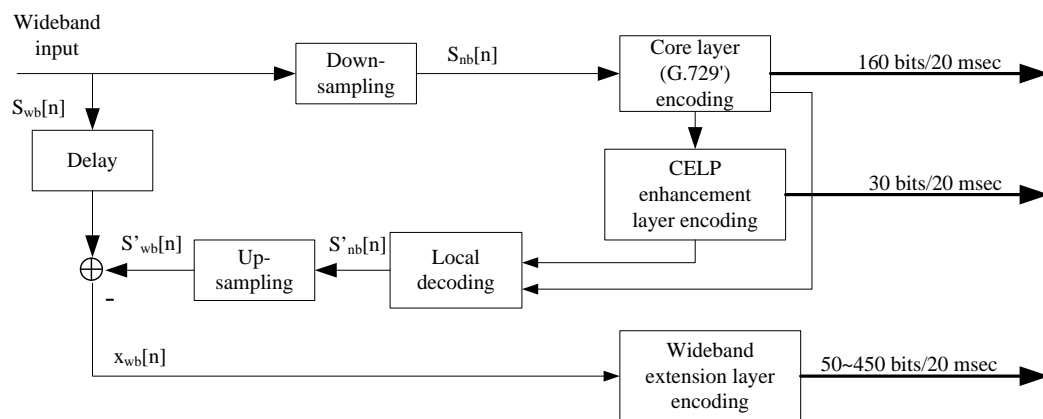


Figure 33 – Encoder block diagram of the proposed codec

8.2.1.1 Core layer

The core layer is similar to ITU-T G.729 standard codec except the LPC analysis window, pre-filtering, and post-filtering. The pre-filtering and post-filtering are suppressed. The length of the

cosine part and the center location of the LPC analysis window are changed as follows. Also, the look-ahead size is increased from 5 ms to 10 ms.

$$\begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{279}\right), & \text{for } n = 0, 1, \dots, 139 \\ \cos\left(\frac{2\pi(n-140)}{399}\right), & \text{for } n = 140, 141, \dots, 239 \end{cases}$$

8.2.1.2 CELP enhancement layer

The CELP enhancement layer is used to improve the quality of core layer. Figure 34 shows the connection of core layer with CELP enhancement layer. In this layer, the fixed codebook error signal of core layer is represented by two algebraic pulses every 10 ms. The signs and positions of the pulses are quantized with 15 bits. The pulses are scaled with the fixed codebook gain of core layer.

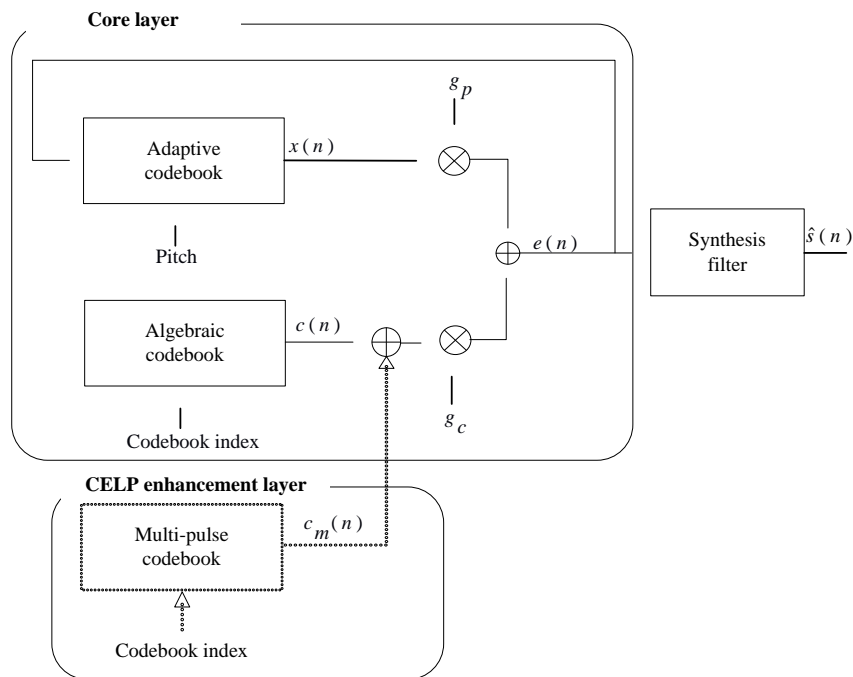


Figure 34 – Core layer with CELP enhancement layer

8.2.1.3 Wideband extension layer

The overall encoding processing of wideband extension layer is given in Figure 35. At this layer, an input signal, $x_{WB}(n)$, is the difference between the delayed wideband input signal and the up-sampled version of locally decoded narrowband signal, and the signal is processed on every 20 ms frames. At first, the signal $x_{WB}(n)$ is transformed using a MDCT (Modified Discrete Cosine Transform). The MDCT is performed on 40 ms windowed signal with 20 ms overlap.

The MDCT coefficients, $X(k)$, is split into two typical bands, one for [0, 2.7 kHz] and the other for [2.7, 7.0 kHz]. The coefficients of the first band are quantized on MDCT domain and the coefficients of the second band are quantized on LPC (Linear Predictive Coding) residual domain.

The MDCT coefficients between 0 kHz and 2.7 kHz, $X(k)|_{0,0-2.7\text{kHz}}$, are split into four bands again, and then the frequency information of each band is quantized by a gain-shape quantizer (Type 3). The

Type 3 gain-shape quantizer encodes the MDCT coefficients corresponding to each band respectively as following steps:

- Partition each band to several sub-bands
- Compute a scale factor and quantize it at each sub-band
- Compute a normalized shape vector and quantize it at each sub-band
- All of the quantized scale factors and some of the normalized shape vectors which have higher scale factor are coded

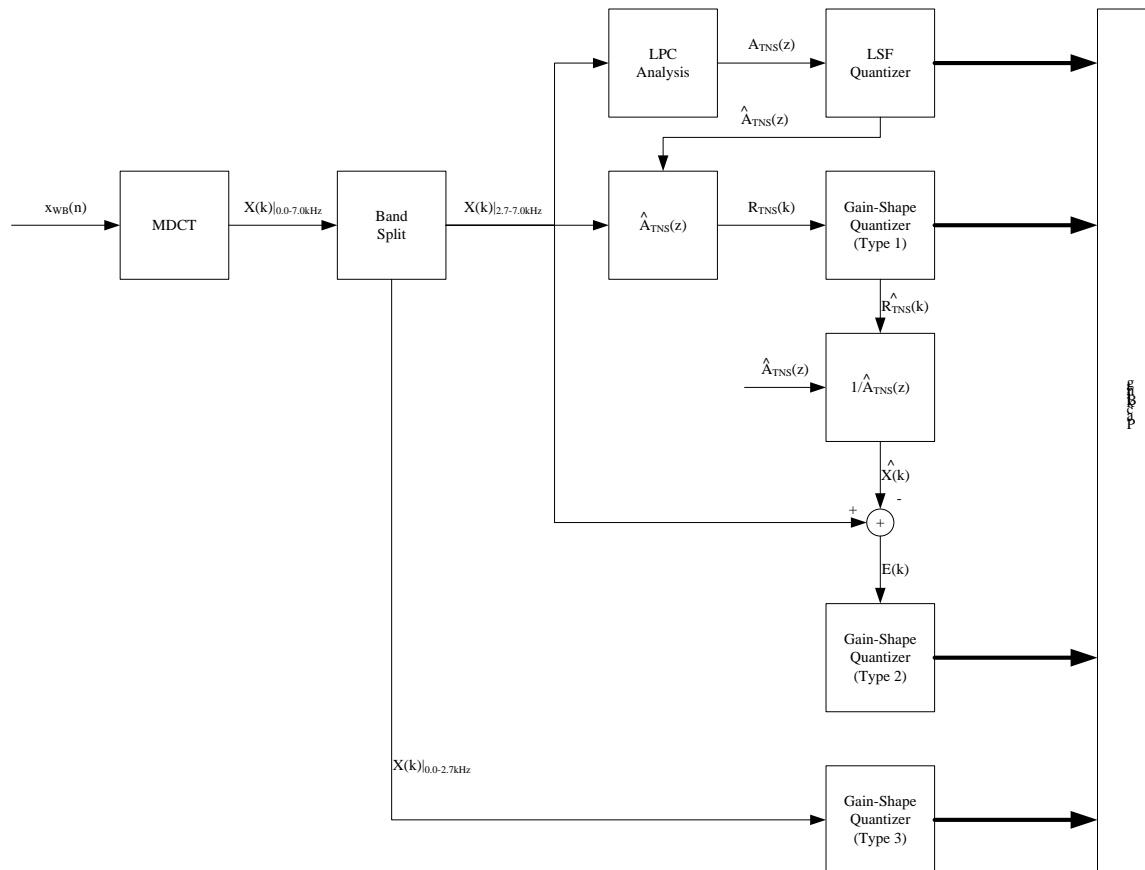


Figure 35 – Wideband extension layer of the proposed coder

The MDCT coefficients corresponding to [2.7, 7.0 kHz], $X(k)|_{2.7-7.0\text{kHz}}$, are encoded using the gain-shape quantizer in two stages. The Temporal Noise Shaping technique is used in this band.

At the first stage, a 10th order LPC coefficients are computed from $X(k)|_{2.7-7.0\text{kHz}}$. The LPC coefficients, $A_{TNS}(z)$, are converted into LSFs and quantized with 10 bits. The LPC residual coefficients, $R_{TNS}(k)$, are split into four bands, which are quantized by a gain-shape quantizer (Type 1). The LPC residual coefficients of each band are divided into several sub-bands. A scale factor is calculated and quantized on each sub-band. A shape vectors are computed by normalizing the LPC residual coefficients by the quantized scale factors. The normalized shape vectors are quantized using a weighted interleaving VQ or a conventional VQ dependent on bands.

At the second stage, the difference between the original MDCT coefficients and the quantized MDCT coefficients is quantized by a gain-shape quantizer (Type 2). The Type 2 quantizer encodes the error coefficients, $E(k)$, split into three bands. The error coefficients of each band are divided into several sub-bands and each sub-band vectors are quantized based on cross-correlation criterion, e.g. finding a codeword and its gain simultaneously by maximizing the following term:

$$C(m) = \frac{\sum_k E(k) \cdot S^m(k)}{\sqrt{\sum_k S^m(k) \cdot S^m(k)}}, \quad \text{for all codeword vectors}$$

where, $S^m(k)$ is m^{th} codeword vector. The corresponding gain is computed as:

$$g = \frac{\sum_k E(k) \cdot S'(k)}{\sum_k S'(k) \cdot S'(k)},$$

where $S'(k)$ is the best codeword.

Finally, all of the quantized parameters are coded and packed into a bitstream at the bit-packing block according to the predefined order.

8.2.2 Decoder description

Figure 36 shows the block diagram of the proposed decoder. The decoder also consists of three layers: core layer, CELP enhancement layer and wideband extension layer. The operation of each layer depends on the size of the received bitstream.

If only 160 bits have been received, the core layer is operated to reconstruct the narrowband signal. If the number of bits received is equal or above 240 bits, then all three layers are operated. If 240 bits have been received, the wideband extension layer and CELP enhancement layer is operated with the core layer to synthesize the output signal band-limited to 5.3 kHz. If the number of received bitstream is over 240 bits, the output is wideband signal (bandwidth between 0.5 kHz and 7.0 kHz).

A frame erasure concealment algorithm is applied to improve the synthesized quality in frame erasure condition. The frame erasure concealment algorithm of core layer is partly modified based on a state machine. The pitch gain and fixed codebook gain is reconstructed by an attenuated version of the previous pitch gain and fixed codebook gain respectively. The attenuation coefficient is depends on the state. In case of voiced frame, the fixed codebook gain is set to zero. In wideband extension layer, an erased frame is recovered by multiplying a randomly generated shape vector by the attenuated scale factor of the previous frame.

8.2.3 Frame Format

The analysis frame length of the proposed candidate codec is 20 ms. As shown in Figure 37, data frames are divided into four parts, : header, core layer, CELP enhancement layer and wideband extension layer. Header includes sync word and data frame length.

Core layer information part consists of two G.729 coded frames. The CELP enhancement layer contains the multi-pulse **excitation** parameters used to improve the quality of G.729. Finally, the wideband extension layer consists of LPC coefficients, scale factor, and shape vector parameters of each frequency band.

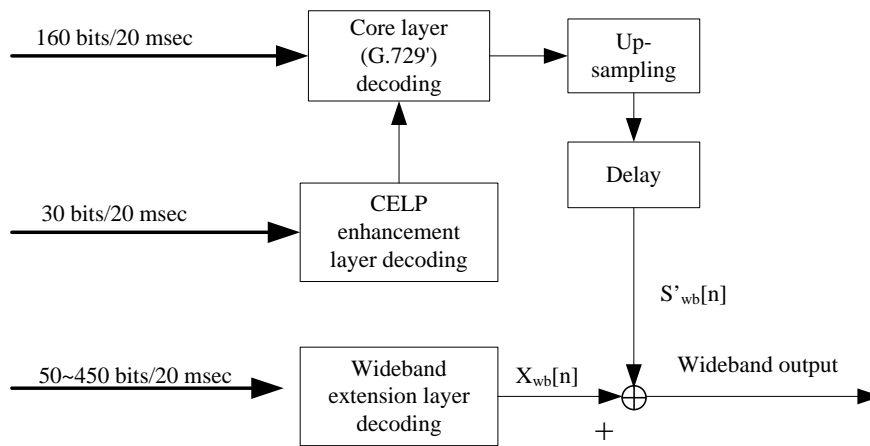


Figure 36 – Decoder block diagram of the proposed codec

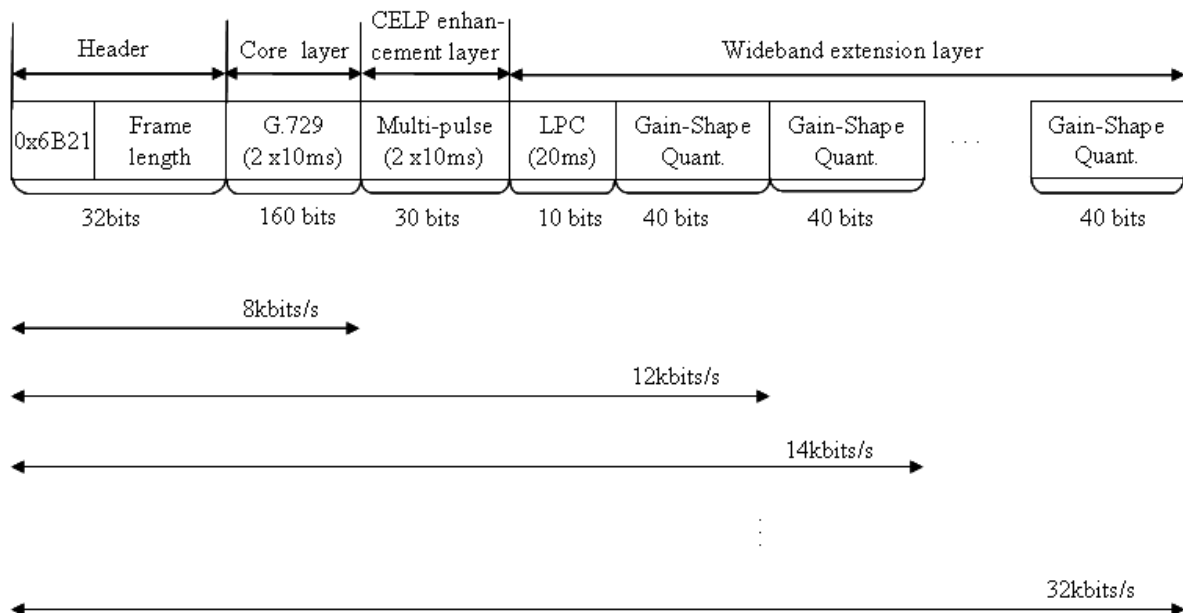


Figure 37 – Frame format

8.2.4 Algorithmic delay

The encoder algorithmic delay is 30.375 ms accounting for the frame length of 20 ms, 10 ms of look-ahead and 0.375 ms of re-sampling delay. The decoder algorithmic delay is 10.375 ms, which comprises MDCT overlap-add window and re-sampling filter. Thus the total algorithmic delay of the proposed codec is 40.75 ms.

8.2.5 Effective bandwidth

The output of 8 kbit/s layer is narrowband signal. The output signal at 12 kbit/s has a bandwidth between 0.05 kHz and 5.3 kHz. The output at 14 kbit/s and its above bitrates is wideband signals (bandwidth between 0.05 kHz and 7 kHz). Figure 38, Figure 39, Figure 40, and Figure 41 show the frequency response of the proposed codec. The frequency responses were computed using STL2005 tool with female and male speech samples which used in qualification test

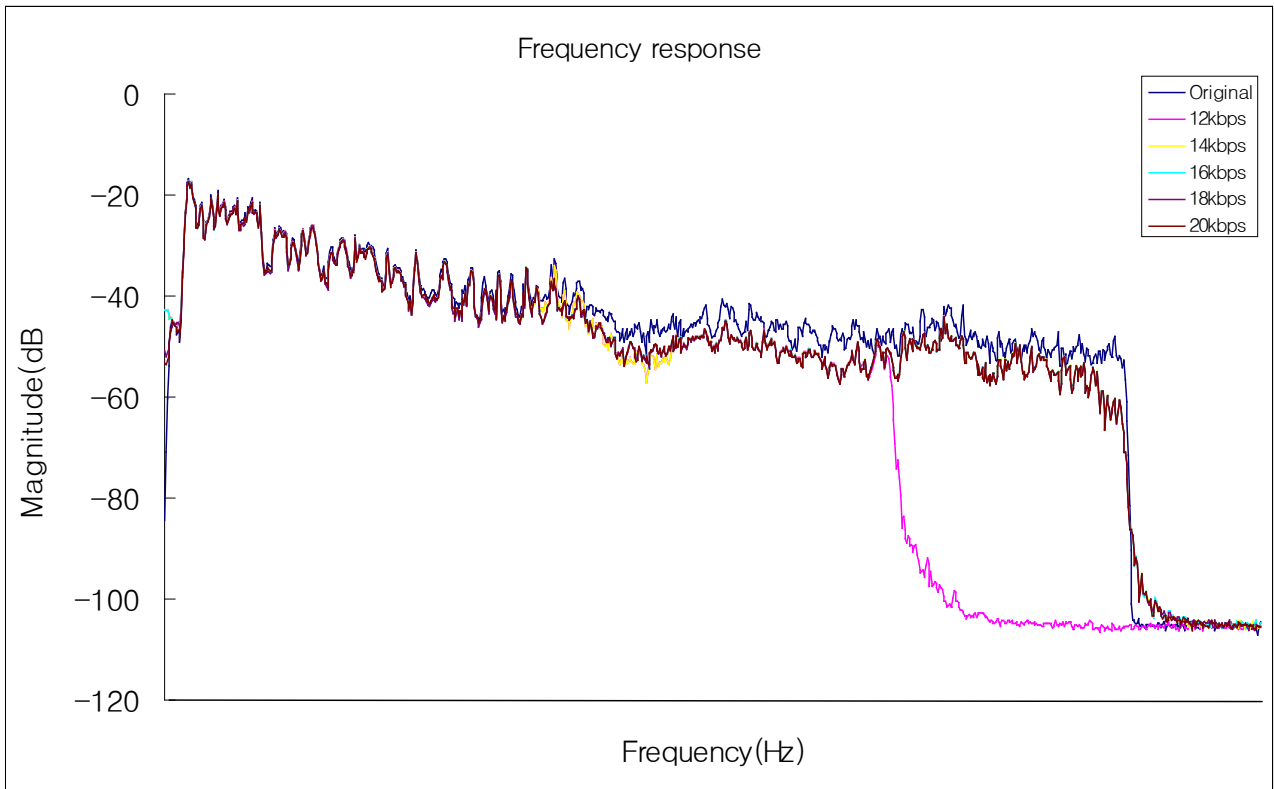


Figure 38 – Frequency response for female speech sample (12 kbit/s ~ 20 kbit/s)

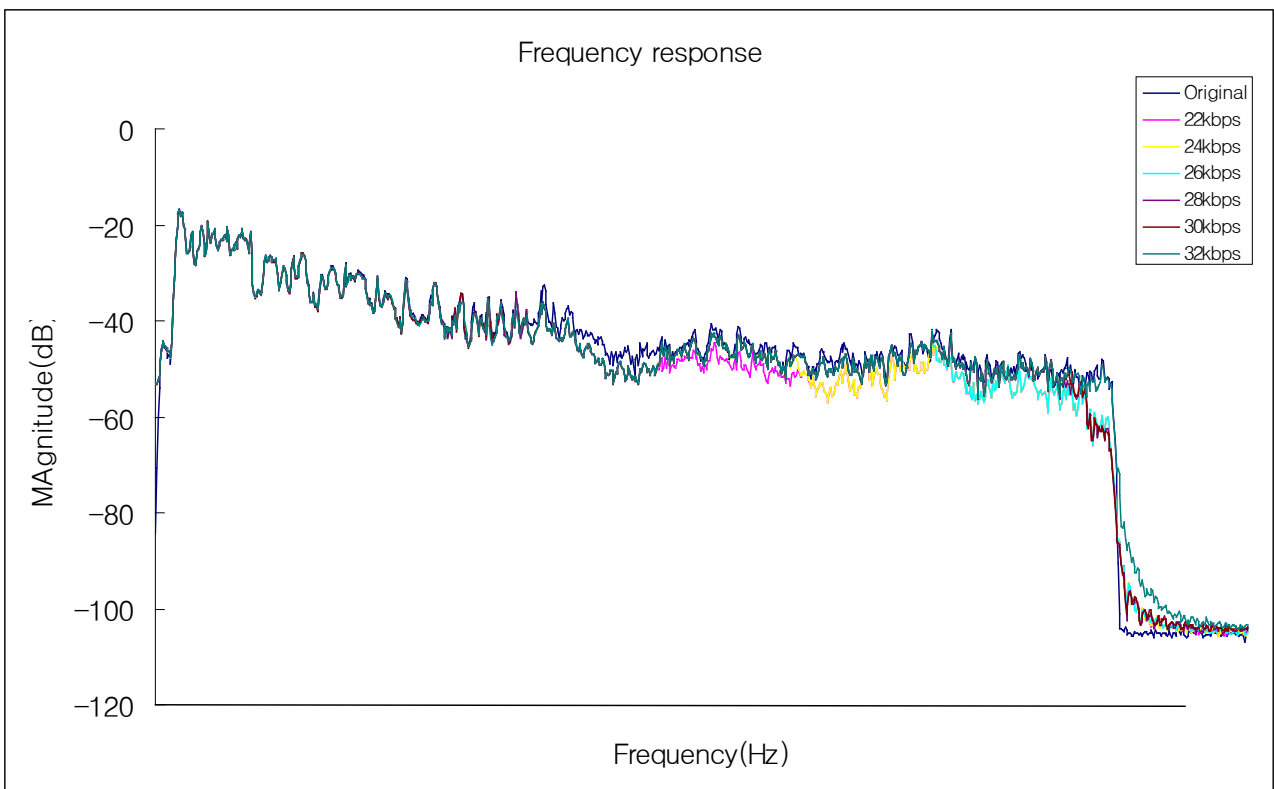


Figure 39 – Frequency response for female speech sample (22 kbit/s ~ 32 kbit/s)

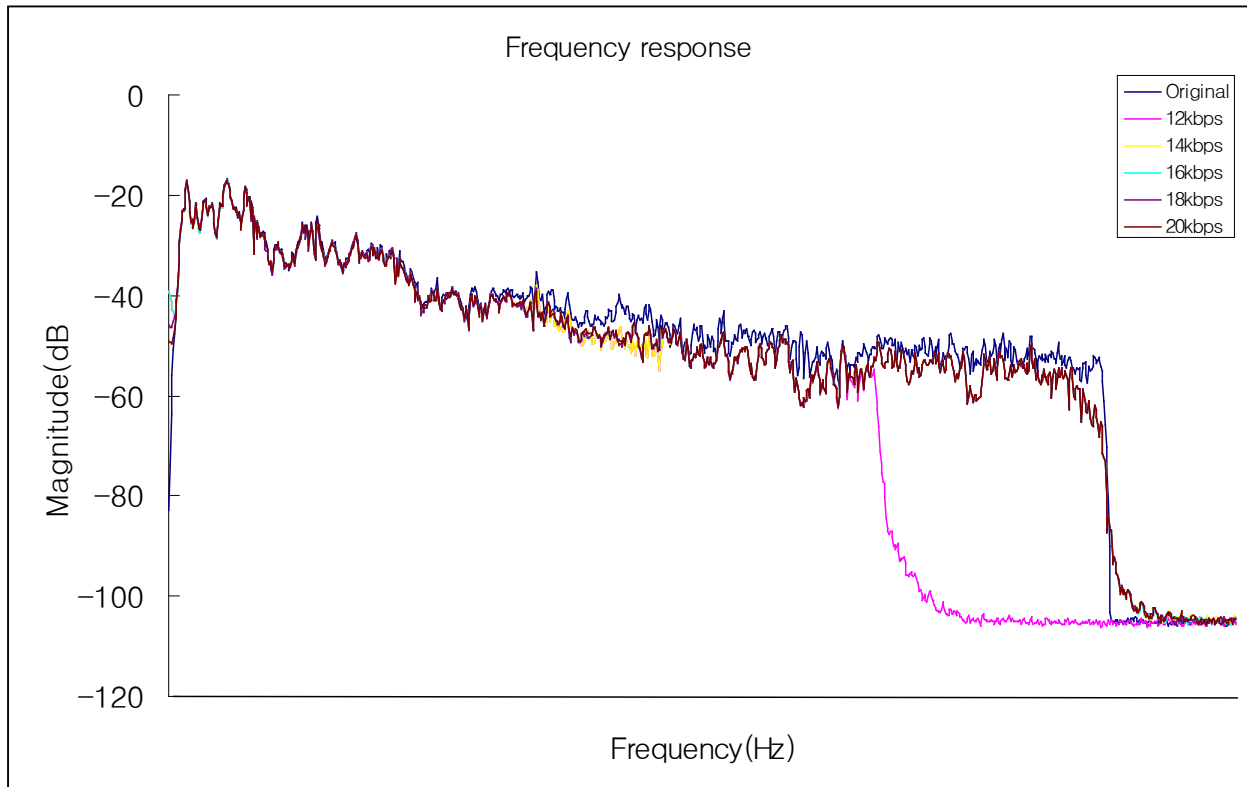


Figure 40 – Frequency response for male speech sample (12 kbit/s ~ 20 kbit/s)

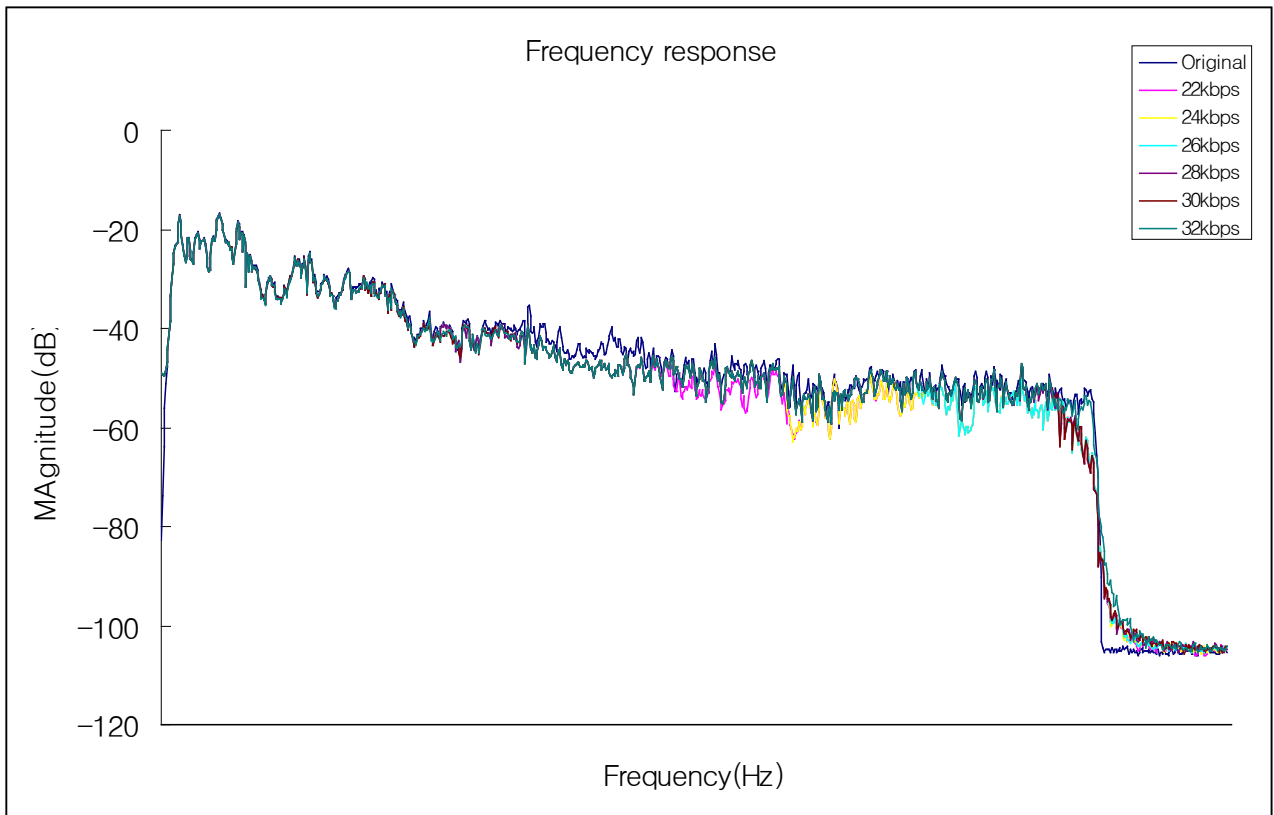


Figure 41 – Frequency response for male speech sample (22 kbit/s ~ 32 kbit/s)

8.2.6 Complexity and memory

The complexity and memory figures are summarized in the Table 34 and Table 35 respectively. In Table 34, the complexity is evaluated in the worst case. The total values are given by the sum of the three layers and other functions such as re-sampling. The overall complexity of the proposed codec is about 37.85 WMOPS.

The DROM takes into account all the constant tables. Same tables are used in both encoder and decoder. Thus the DROM in Table 35 is the sum of encoder and decoder. The DRAM corresponds to the memory of all the static variables plus the worst case of the dynamic RAM usage.

Table 34: Worst case computational complexity of the proposed codec

	Encoder (WMOPS)	Decoder (WMOPS)
Core layer	11.683	2.612
CELP enhancement layer	5.707	0.042
Wideband enhancement layer	9.692	4.265
Other functions	2.519	1.372
Total	29.601	8.249

Table 35: Memory requirement of the proposed codec

	Encoder (16-bit words)	Decoder (16-bit words)	Total (16-bit words)
PROM	3704	2943	6647
DROM	22865		22865
DRAM	4295	3897	8192

8.3 Samsung Candidate [13]

The core layer and the 1st enhancement layer employ a CELP based approach. The second enhancement layer is based on a parametric approach. Sinusoidal coding, in particular, is used to efficiently encode the higher band signal. Layers beyond 14 kbit/s are encoded using vector quantization of the residual signal in the transformed domain. Magnitude and phase value of the Fourier transformed residual signal is vector quantized.

8.3.1 Main feature

In order to reduce the complexity of conventional 8 kbit/s G.729, the core layer uses a fast fixed codebook search algorithm that enhances the search method of the ACELP fixed codebook. Layer 2, i.e. the 1st enhancement layer, uses an ATC (algebraic Trellis code). This functions as a 2nd fixed codebook and uses an overall bit rate of 3.4 kbit/s. In total, therefore, the core layer and the 1st enhancement layer use 11.4 kbit/s. To encode all the layers beyond the 1st enhancement layer, a WB error signal is computed by subtracting the contribution of the first two layers from the 16 kHz input signal. Subsequently, the WB error signal is decimated from 16 kHz down to 12.8 kHz. Linear prediction analysis is performed on the WB error signal once every speech frame. A WB residual signal is generated by filtering the WB error signal using a LPC analysis filter. The amplitude and

phase spectrum of the residual signal is calculated using a FFT and these spectral parameters are subsequently encoded.

From 14 kbit/s up to 32 kbit/s, FGS is supported. In layer 3, i.e. the layer contributing to 14 kbit/s, the pitch and LPC information of NB is used. This information helps to encode the amplitudes and phases efficiently. First, the positions of the harmonics, belonging to the HB (high band), are extracted using the pitch frequency obtained from NB parameters. Subsequently, the phase data of the harmonics is selectively encoded in the 14 kbit/s layer.

To support FGS up to 32 kbit/s, more and more amplitude and phase information is added to bitstream such that they can be encoded more and more precisely. After completing the encoding of HB, the amplitudes and phases for LB (low band) are encoded. We have divided the entire low band in to a lower LB and an upper LB.

Different bit allocation is used in case of speech and music signals. In order to differentiate between speech and music signals, a speech/music discriminator has been introduced.

8.3.2 Frame structure

The bandwidth definition is depicted in Figure 42 and the frame structure is depicted in Figure 43.

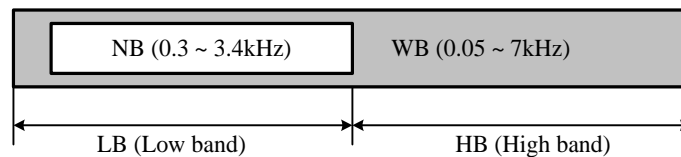


Figure 42 – Bandwidth definition

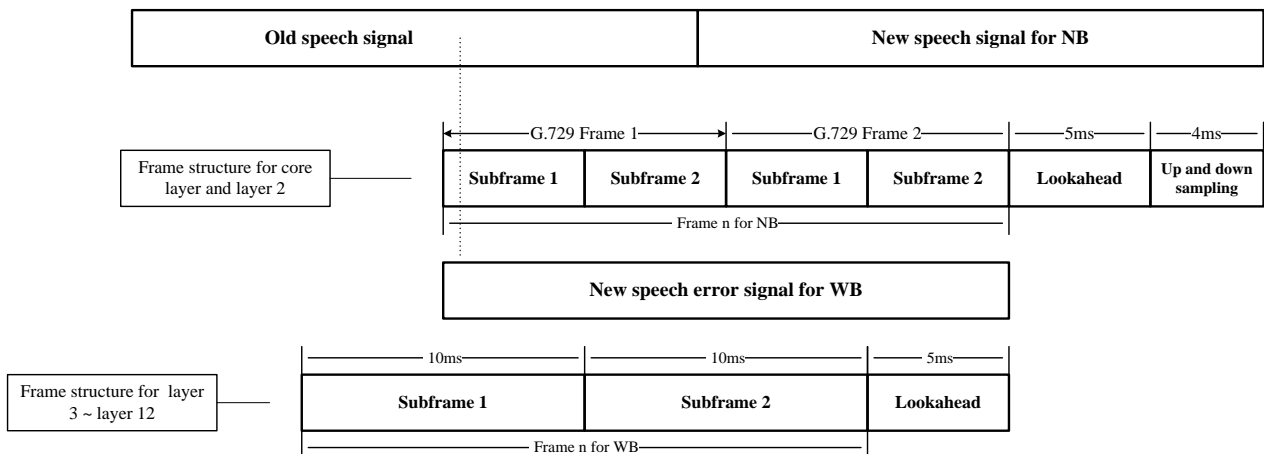


Figure 43 – Frame structure

8.3.3 Block diagram of encoder

The block diagram of the encoder is given in Figure 44.

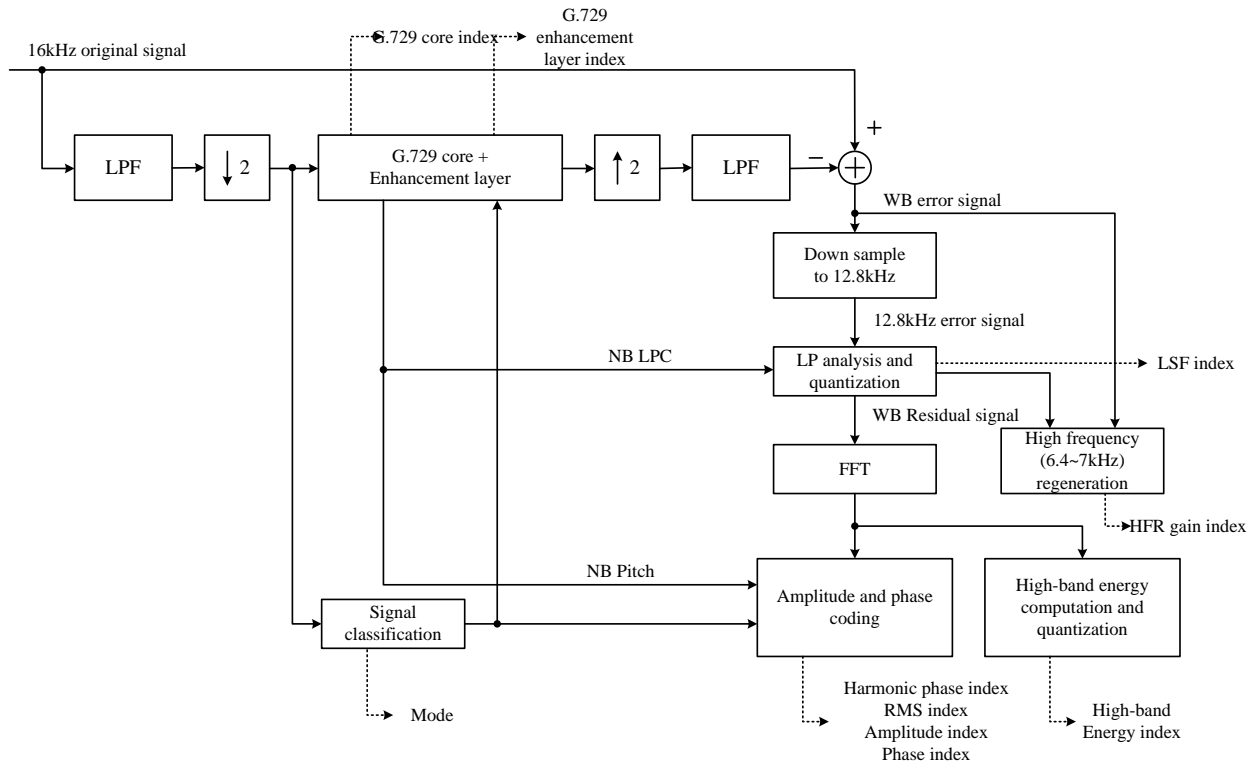


Figure 44 – Block diagram of the encoder.

8.3.4 Block diagram of decoder

The block diagram of the decoder is given in Figure 45.

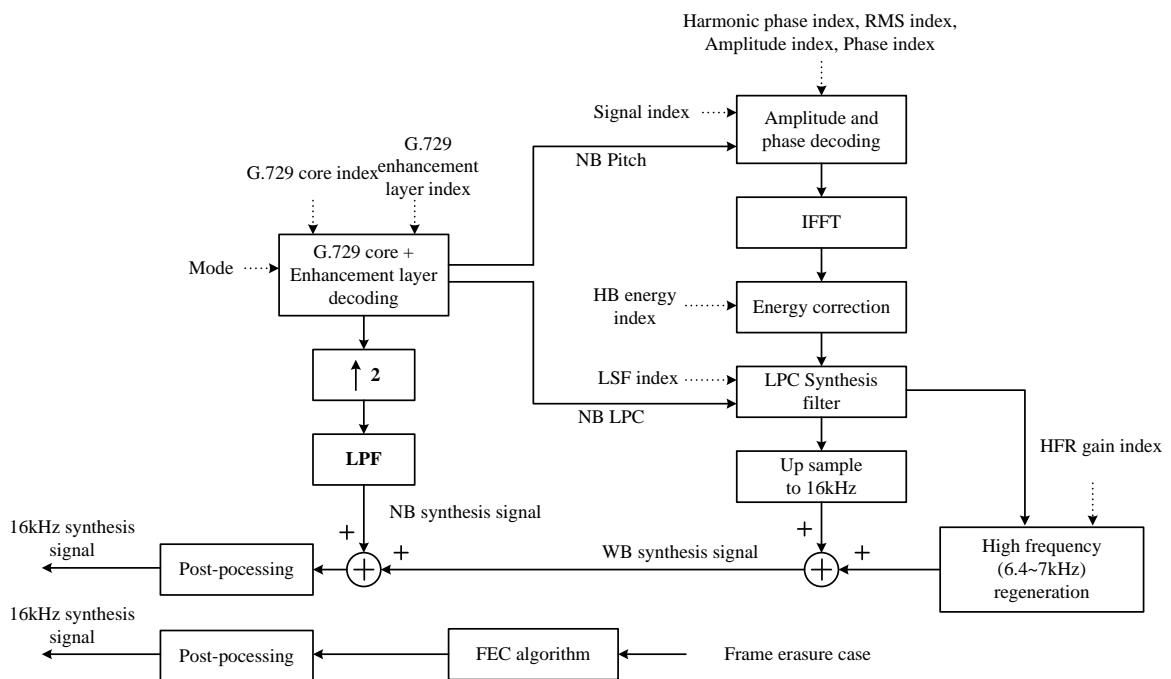


Figure 45 – Block diagram of the decoder.

8.3.5 Bit rate granularity

Figure 46 shows the bitstream structure for the speech mode illustrating the bit rate granularity, while Figure 47 shows the bitstream structure for the music mode.

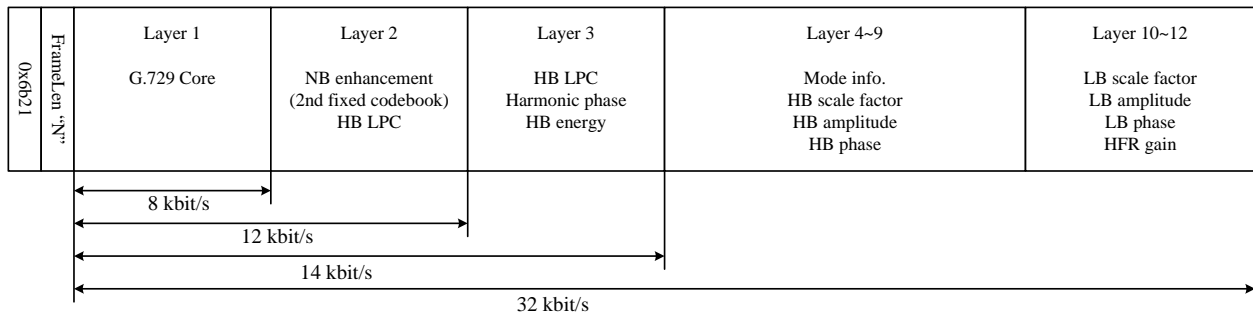


Figure 46 – Bit stream structure for the speech mode

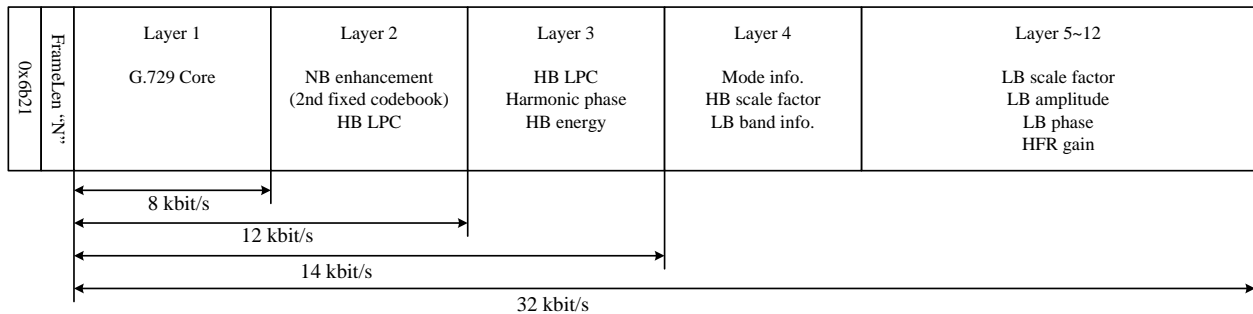


Figure 47 – Bitstream structure for the music mode

8.3.6 Algorithm delay

Overall delay: 46.75 ms

Coder:

- Down sampling 1(16 → 8 kHz): 2 ms
- Up sampling 1(8 → 16 kHz): 2 ms
- G.729 look-ahead: 5 ms
- Wide band coder look-ahead: 5 ms

Decoder:

- Up sampling 1(8 → 16 kHz): 2 ms
- Wide band coder over lap: 10 ms
- Up sampling 2(12.8 → 16 kHz): 0.75 ms

8.3.7 Effective bandwidth at all supported bit rates (Codec frequency response)

- 8 and 12 kbit/s: 300 ~ 3400 Hz
- 14 kbit/s ~ 32 kbit/s: 50 ~ 7000 Hz

8.3.8 Complexity evaluation

The estimated computational complexity and the estimated memory size are shown in Table 36 and Table 37, respectively.

Table 36: Computational complexity

	Complexity (WMOPS)
Maximum Encoder	29.933
Maximum Decoder	8.317
Maximum Total	38.25

Table 37: Estimated memory size

	Size (words)
Data RAM	13.268k
Data ROM	26.694k

8.4 VoiceAge Candidate [14]

8.4.1 Coding paradigm

The encoder uses predictive coding (CELP) in the first layers, and then quantizes in the frequency domain the coding error of the first layers. An MDCT is used to map the signal to the frequency domain. The MDCT coefficients are quantized using scalable algebraic vector quantization. To increase the audio bandwidth, parametric coding is applied to the high-frequencies.

The encoder uses 20 ms frames with 20 ms look-ahead due to the 50% overlap of the MDCT. The encoder produces an output at 32 kbit/s, which translates in 20-ms frames containing 640 bits each. The bits in each frame are arranged in embedded layers. Layer 1 has 160 bits representing 20 ms of standard G.729 at 8 kbit/s. Layer 2 has 80 bits, representing an additional 4 kbit/s. Then each additional layer (Layers 3 to 12) adds 2 kbit/s, up to 32 kbit/s.

8.4.2 Encoder description

Figure 48 shows a general block diagram of the embedded encoder.

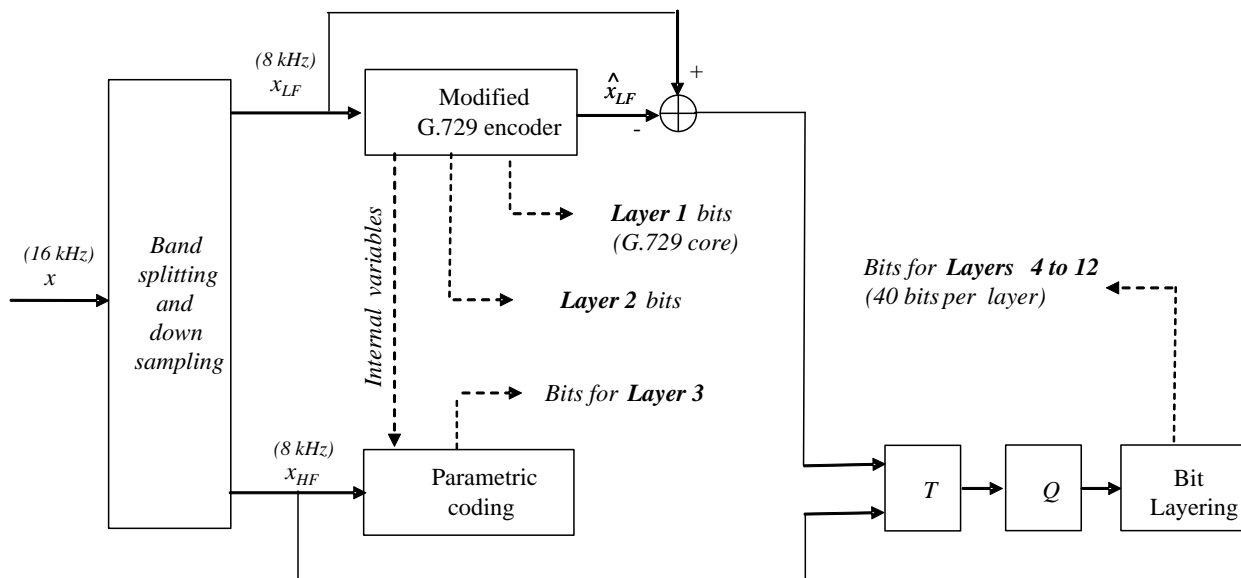


Figure 48 – Block diagram of the encoder

The original signal x , sampled at 16 kHz, is first split into two bands: 0-4000 Hz and 4000-8000 Hz. Band splitting is realized using a QMF filter bank with 64 coefficients. After band splitting, two signals are obtained, one covering the 0-4000 Hz zband (low band) and the other covering the 4000-8000 band (high band). The signals in each of these two bands are downsampled by a factor of 2. This yields 2 signals at 8 kHz sampling frequency: x_{LF} for the low band, and x_{HF} for the high band.

The low band signal x_{LF} is fed into a modified version of the G.729 encoder. This modified version first produces the standard G.729 bitstream at 8 kbit/s, which constitutes the bits for **Layer 1**. Note that the candidate operates on 20 ms frames, therefore the bits of the **Layer 1** correspond to two G.729 frames.

Then, the G.729 encoder is modified to include a second innovative (ACELP) codebook to enhance the low band signal. This second codebook is identical to the innovative codebook in G.729, and it requires 20 bits per 5-ms subframe to encode the codebook pulses and gain. This produces $4 \times 20 = 80$ bits for **Layer 2**. The target signal used for this second-stage innovative codebook is obtained by subtracting the contribution of the G.729 innovative codebook in the weighted speech domain.

The synthesis signal \hat{x}_{LF} of the modified G.729 encoder is obtained by adding the innovative excitation of the standard G.729 and the innovative excitation of the additional innovative codebook, and passing this enhanced excitation through the usual G.729 synthesis filter. This is the synthesis signal that the decoder will produce if it receives only Layer 1 and Layer 2 from the bitstream.

Layer 3 extends the bandwidth from narrowband to wideband quality. This is done by applying parametric coding to the high-frequency component x_{HF} . Only the spectral envelope and gain of x_{HF} are computed and transmitted for this layer. The spectral envelope is computed and transmitted once per 20-ms frame using a 10-th order linear prediction (LP) filter quantized in the LSF domain. The LSFs are quantized using 20 bits. The remaining 20 bits for this layer are used to encode the energy information of the frame.

Then, from Figure 48, the coding error from the modified G.729 encoder, along with the high-frequency signal x_{HF} , are both mapped into the frequency domain. The modified Discrete Cosine Transform (MDCT), with 50% overlap, is used for this time-frequency mapping. The MDCT coefficients are then quantized using scalable algebraic vector quantization in a manner similar to the quantization of the FFT coefficients in the 3GPP AMR-WB+ audio coder (3GPP TS 26.290). The total bit rate for this spectral quantization is 18 kbit/s, which amounts to a bit budget of 360 bits per 20-ms frame. After quantization, the corresponding bits are layered in steps of 2 kbit/s to form **Layers 4 to 12**. Each 2 kbit/s layer thus contains 40 bits per 20-ms frame.

The algorithmic extensions, compared to the core G.729 encoder, can be summarized as follows: 1) the innovative codebook of G.729 is repeated a second time (Layer 2); 2) parametric coding is applied to extend the bandwidth, where only the spectral envelope and gain information are computed and quantized (Layer 3); 3) an MDCT is computed every 20-ms, and its spectral coefficients are quantized in 8-dimensional blocks using scalable algebraic VQ; and 4) a bit layering routine is applied to format the 18 kbit/s stream from the algebraic VQ into layers of 2 kbit/s each (Layers 4 to 12);.

8.4.3 Decoder description

Figure 49 shows the block diagram of the decoder. In each 20-ms frame, the decoder can receive any of the supported bit rates, from 8 kbit/s up to 32 kbit/s. This means that the decoder operation is conditional to the number of bits, or layers, received in each frame. In Figure 49, we assume that at least Layers 1, 2, 3 and 4 have been received at the decoder. The cases of the lower bit rates will be described below.

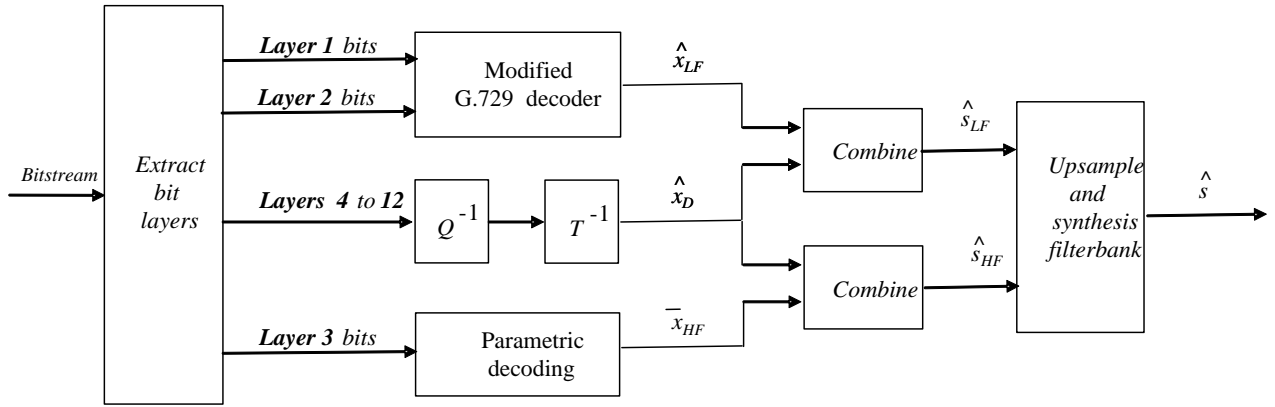


Figure 49 – Block diagram of the decoder.

From Figure 49, the received bitstream is first separated into bit Layers as produced by the encoder. Layers 1 and 2 form the input to the modified G.729 decoder, which produces a synthesis signal \hat{x}_{LF} for the lower band (0-4000 Hz, sampled at 8 kHz). Recall that Layer 2 essentially contains the bits for a second innovative codebook with the same structure as the G.729 innovative codebook.

Then, the bits from Layer 3 form the input to the parametric decoder. The Layer 3 bits give a parametric description of the high-band (4000-8000 Hz, sampled at 8 kHz). Specifically, the Layer 3 bits describe the high-band spectral envelope of the 20-ms frame, along with gain information. An excitation signal for the high-band is obtained from the low-band excitation. This synthetic excitation is adjusted in gain using the decoded gain information, and then passed through the LP filter describing the high-band envelope. The result is a parametric approximation of the high-band signal, called \bar{x}_{HF} in Figure 49.

Then, the bits from Layer 4 and up form the input of the inverse quantizer Q^{-1} . This inverse quantizer is a set of functions related to the algebraic structure of the 8-dimensional vector quantizer used. The output of Q^{-1} is a set of quantized spectral coefficients. These quantized coefficients form the input of the inverse transform T^{-1} , specifically an inverse MDCT with 50% overlap. The output of the inverse MDCT is the signal \hat{x}_D . This signal \hat{x}_D can be seen as the quantized coding error of the modified G.729 encoder in the low band, along with the quantized high band if any bits were allocated to the high band in the given frame.

The component of \hat{x}_D forming the quantized coding error of the modified G.729 encoder is then combined with (added to) \hat{x}_{LF} to form the low-band synthesis \hat{s}_{LF} . In the same manner, the component of \hat{x}_D (if any) forming the quantized high band is combined with the parametric approximation of the high band, \bar{x}_{HF} , to form the high band synthesis \hat{s}_{HF} . Signals \hat{s}_{LF} and \hat{s}_{HF} are passed through the synthesis QMF filterbank to form the total synthesis signal \hat{s} at 16 kHz sampling rate.

In the case where Layers 4 and up are not received, then \hat{x}_D is zero, and the output of the “Combine” boxes in Figure 49 are equal to their input, namely \hat{x}_{LF} and \bar{x}_{HF} . If only Layers 1 and 2 are received, then the decoder only has to apply the modified G.729 decoder to produce signal \hat{x}_{LF} . The high band component will be zero, and the upsampled signal at 16 kHz (if required) will have content only in the low band.

8.4.4 Algorithmic delay

The total algorithmic delay is 51 ms. The frames are 20-ms in duration, with a 20-ms lookahead required for the MDCT. The remaining delay is required for the QMF filter bank and lookahead of LP analysis.

8.4.5 Effective bandwidth at supported bit rates

The effective bandwidth of the embedded codec at different bit rates was measured by passing white Gaussian noise through the codec and observing the average power spectrum of the decoded signal. Figure 50, Figure 51, Figure 52, and Figure 53 show the average power spectrum of the decoded signal at 8, 12, 14, and 32 kbit/s respectively. The effective bandwidth at the other supported bit rates (16, 18, 20 kbit/s etc.) is similar to that pictured in Figure 53 for the 32 kbit/s case.

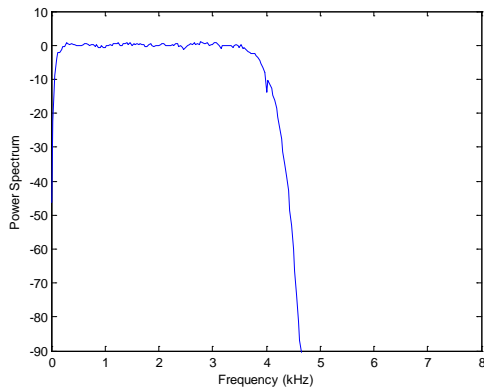


Figure 50 – Effective bandwidth at 8 kbit/s

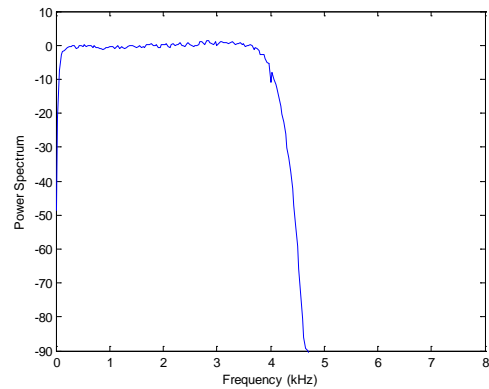


Figure 51 – Effective bandwidth at 12 kbit/s

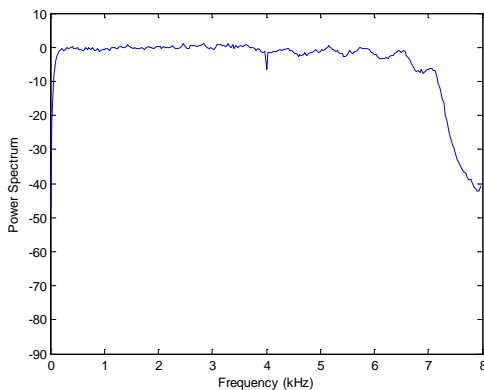


Figure 52 – Effective bandwidth at 14 kbit/s

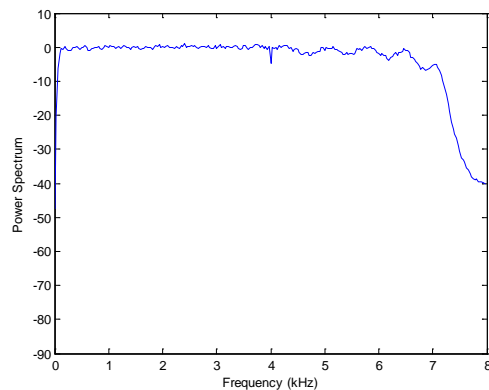


Figure 53 – Effective bandwidth at 32 kbit/s

8.4.6 Frame structure

The encoder outputs 20-ms frames comprising 640 bits which are packetized in 12 successive layers as shown in Figure 54. The coding algorithm used in each layer was explained in the previous sections. The frame header contains two 16-bit words. The first is a synchronisation word set at the value 0x6B21. The second word in the header is an integer indicating the number of bits in the frame. This integer is changed to the proper value when the frame is truncated to reduce the bit rate at the decoder. Note that the bits in Layer 3, corresponding to the parametric coding of the high frequencies, are labelled BWE for “bandwidth extension”. Finally, Layers 4 to 12 contain the bits encoding the MDCT coefficients quantized using algebraic vector quantization (AVQ).

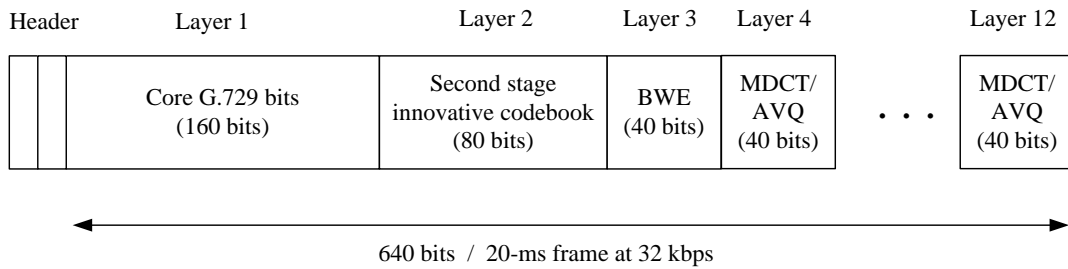


Figure 54 – Frame structure.

8.4.7 Complexity

The complexity and memory requirements of the encoder and decoder are given in Table 38. Note that the complexity of the decoder is estimated assuming decoding of the total bitstream at 32 kbit/s. Note also that the encoder and decoder share some of the tables in ROM. When implementing both the encoder and decoder, the total ROM is only about 8000 words.

Table 38: Complexity and memory figures.

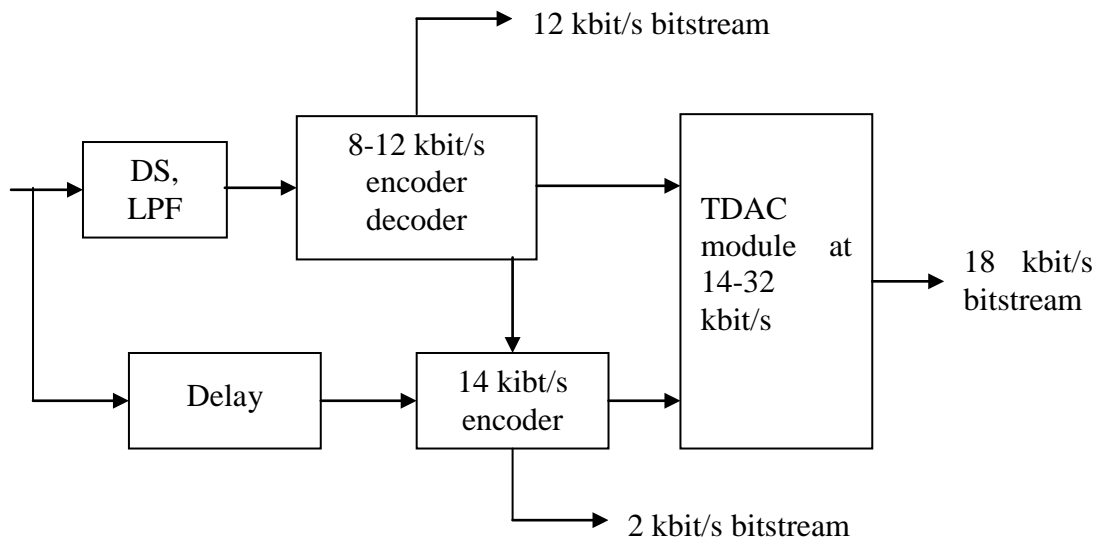
	Complexity (WMOPS)	ROM (words)	RAM (words)
Encoder	18	6000	6500
Decoder	8	6000	5000

8.5 Matsushita, Mindspeed, Siemens Candidate [15]

The proposed G729EV candidate was developed jointly by the members of the consortium. The core of this algorithm is based on the G.729 principle and is bitstream interoperable with it. The algorithm is further improved with a 4 kbit/s module in narrow band. ACELP coding techniques are used for these both modules. An additional module, using 2 kbit/s to reach 14 kbit/s, relies on bandwidth extension techniques and allows the 14 kbit/s output to be wideband. The layer between 14 and 32 kbit/s is using TDAC technique.

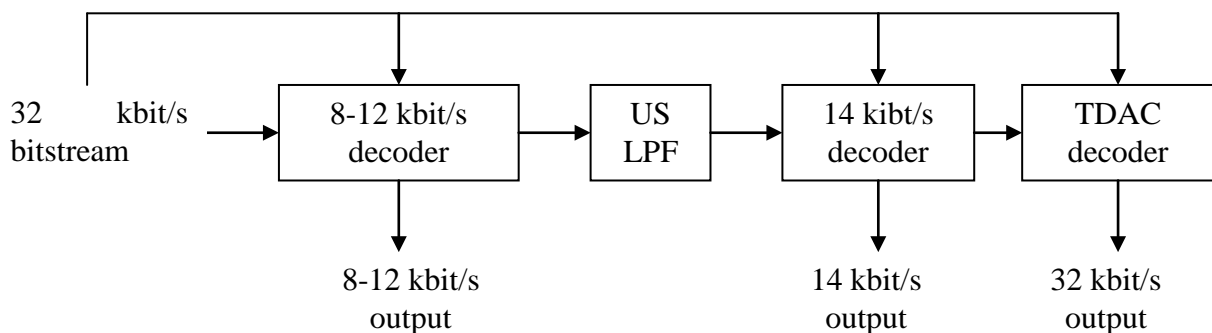
8.5.1 High Level Description

The first step (Figure 55) of the coding process is to down-sample the input signal to 8 kHz and to low-pass it. This low frequency signal is used as input signal of the 8-12 kbit/s codec. The 8 kbit/s core codec is not a pure G.729A codec but a modified version with some improvements. The bitstream stays compatible with G.729 codec. The additional 4 kbit/s bitrate is used for a second stage fixed codebook contribution. At the output of these modules the signal is up-sampled and is used as the input of the 14 kbit/s layer together with the delayed original signal. The 14 kbit/s module uses wideband extension technology by means of time and frequency envelope shaping. A difference signal between the delayed original signal and the output of the 12 kbit/s codec is computed. This signal is a real difference signal in low frequency and the original signal in high frequency. To increase the quality, a pre-echo reduction algorithm and a post-processing module are introduced and shortly described as well. The decoder is illustrated in Figure 56.



DS: Down sampling
 LPF: Low pass filtering

Figure 55 – Block diagram of the G.729EV encoder



US: upsampling
 LPF: low pass filtering

Figure 56 – Block diagram of the G.729EV decoder

8.5.2 Frame size, lookahead, and delay

The frame size of our G.729EV candidate is 20 ms. The different delays are the following:

- MDCT Transform coding with overlap-add of 50%, an additional delay of 20 ms is also added
- Each low pass and high pass filtering with 41 coefficients adds a delay of 3 times 20 samples
- A look ahead of 5 ms is added by G.729 codec

The overall delay is then of 48.75 ms (framing, overlap-add, three filterings, 5 ms look-ahead).

8.5.3 Core part

The core layer is based on the G.729 algorithm with replacement of its fixed-codebook (FCB) search with the one of the G.729A algorithm. Therefore, the coder operates on frames of 10 ms, using 5 ms look-ahead for linear prediction (LP) analysis, and the 10 ms frame is divided into two 5 ms subframes. It should be noted that the G.729 10 ms frame corresponds to the 10 ms subframe in the table of bit-allocation in Section 2.8. In order to improve the encoding performance, the FCB search is orthogonalized to the adaptive-codebook (ACB) vector, i.e. the FCB vector is searched in a way where the ACB and FCB vectors are jointly optimized.

8.5.4 The 12 kbit/s enhancement layer

The 12 kbit/s layer is using an additional FCB based on the 17-bit FCB of G.729 for quantizing the encoding error of the core layer. It is using the LP synthesis filter quantized in the core layer. The encoding process of the 12 kbit/s layer follows the one of the core layer process, thus it operates on 10 ms frames and 5 ms subframes of the G.729 algorithm. To improve codec performance, excitation pulses generated from the additional FCB are convoluted with some dispersion patterns that are obtained through off-line training. The gain of the FCB is quantized with a 3-bit predictive scalar quantizer.

8.5.5 The 14 kbit/s layer

For data rates of 14 kbit/s or higher, the transmitted signal bandwidth is from 50 Hz to 7 kHz. To accomplish the step from narrowband to wideband coding, the 14 kbit/s layer performs a kind of bandwidth extension in the decoder by means of time and frequency envelope shaping of synthetically generated extension band (EB) components.

In the encoder, first, the EB signal components are isolated by band-pass filtering (frequency range 3.4 kHz to 7 kHz) of the original wideband speech or audio signal. Then, both time envelope and frequency envelope of the EB signal components are calculated. The determined time and frequency envelopes are jointly quantized and encoded for transmission in the digital bitstream. Quantization is performed using split-VQ in a transformed domain. The gross bit rate of the time and frequency envelope information is 40 bit/frame or 2 kbit/s.

In the receiver, an “excitation signal” is produced synthetically. To do so, some of the already decoded parameters of the 8-12 kbit/s baseband decoder are taken into account. The generated excitation signal contributes to the spectral fine structure of the EB signal components. By decoding the bitstream the information on the time and frequency envelopes of the extension band signal is reconstructed. Then, time and frequency envelopes of the “excitation signal” are consecutively shaped by gain manipulations and filtering operations to match the reconstructed side information targets. Hence, while the fine details in time and frequency are given by the generated “excitation signal”, the time and frequency envelopes of the EB components in the wideband output signal match those of the original wideband input signal at the transmitter. The final shaping of the spectral envelope is performed in the MDCT domain, thus providing for a fine frequency resolution and for a smooth transition to the subsequent TDAC transform coding layer (14-32 kbit/s).

8.5.6 The 32 kbit/s layer

This layer is using a “classical” MDCT encoding technique. The difference signal is transformed with 20 ms frame length. 320 coefficients come out of this codec, there are gathered in 20 bands, each band having 14 coefficients. The coefficients between 280 and 320 are not taken into account (band 7-8 kHz is not transmitted). According to a psychoacoustically shaped energy of these bands a binary allocation is done and then the frequency coefficients are encoded using spherical vector quantization.

8.5.7 Pre echo Reduction and post processing schemes

Pre-echoes may be created in the TDAC layer due to the quantization of the transmitted spectral components. The pre-echo reduction scheme uses the fact that the energy envelope of the 12 kbit/s and 14 kbit/s layers are closer to the energy envelope of the original signal than the energy envelope of the TDAC layer. The pre-echo artifacts are reduced by applying the energy envelope of the CELP and 14 kbit/s layers to the output signal of the TDAC layer. The pre-echo reduction scheme runs entirely in the decoder, no additional information needs to be transmitted.

The narrow-band layers at 8 kbit/s and 12 kbit/s use time-domain short-term and long-term post filtering, similar to traditional CELP-type codecs for improving the perceptual quality of the decoded signals. The wideband layers from 14 kbit/s up to 32 kbit/s employ the spectral coefficients to achieve similar perceptual improvements.

8.5.8 Bit Rate Granularity

The frame structure is shown in Figure 57.

G.729 160 bits	4 kbit/s NB 80 bits	2 kbit/s WB 40 bits	TDAC Layer 360 bits
-------------------	------------------------	------------------------	------------------------

Figure 57 – Frame structure

8.5.9 Complexity

The narrow band codecs at 8-12 kbit/s have an estimated complexity of 17.8 WMOPS. The complexity of the 14 kbit/s module is around 3 WMOPS, where the encoder has a complexity of about 1 WMOPS and the decoder about 2 WMOPS. The TDAC module has a complexity of 12 WMOPS. The pre-echo scheme has an estimated complexity below 0.5 WMOPS. The overall complexity is estimated at around 34 WMOPS.

The ROM/RAM figures are listed in Table 39.

Table 39: Complexity estimates

Computational complexity	34 WMOPS
ROM	< 35.2 kwords
RAM	< 13.2 kwords

8.5.10 Effective Bandwidth

Figure 58 shows the frequency response of our candidate codec computed with the STL 2005 tool: freqresp and using two concatenated files (P50_AV_F.pcm and P50_AV_M.pcm that have been pre-filtered by P341 filter).

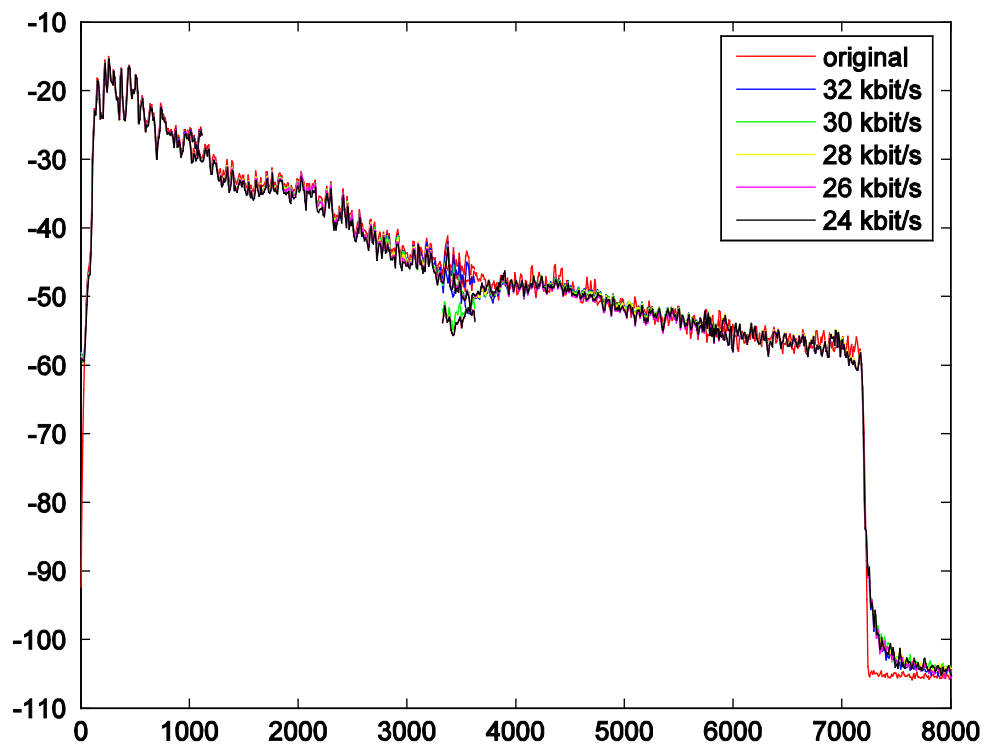
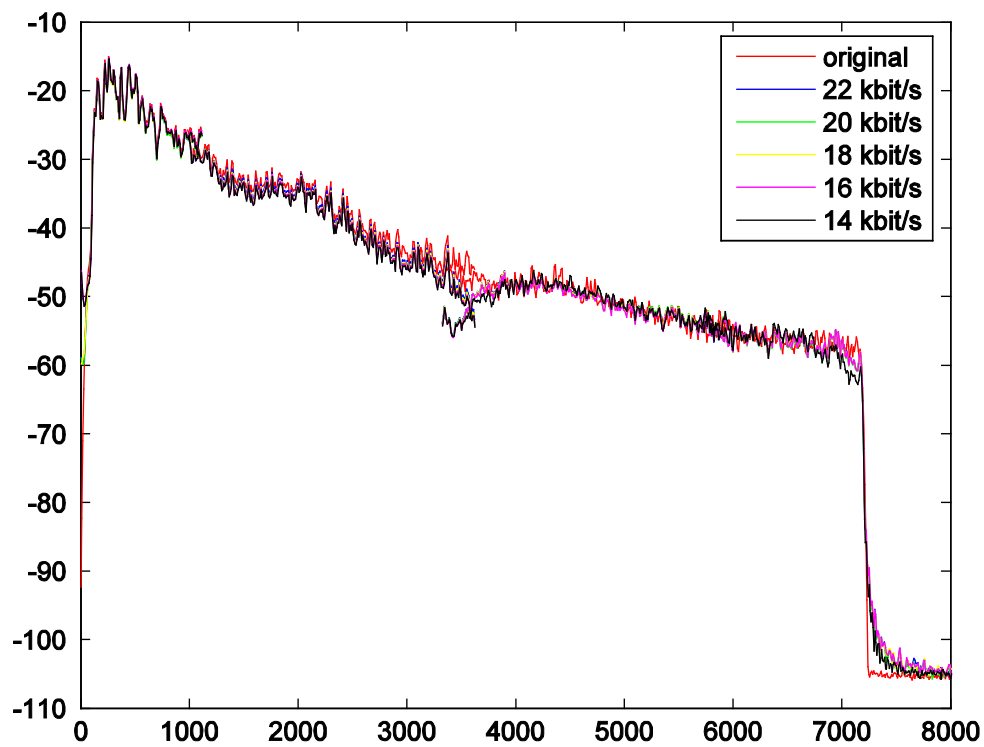


Figure 58 – Effective bandwidth

9 Test Results in the Qualification Phase of G729EV

9.1 Experiments 1-4

This section presents the results of subjective Experiments (Experiments 1 to 4) of the five G.729EV qualification phase candidates. Two candidates passed all requirements in all laboratories. These candidates are A (France Telecom) and D (Siemens-Matsushita-Mindspeed). The other candidates fail some requirements.

Analysis was performed at 95% Confidence interval and also on request of Q10/16 at 99% confidence interval for information.

Blinding of executable was the following:

- A: France Telecom
- B: ETRI
- C: VoiceAge
- D: Siemens Matsushita Mindspeed
- E: Samsung

The testing was performed in different languages:

- French for France Telecom laboratory (lab A),
- American English for ETRI (lab B),
- French for Voice Age (lab C),
- Korean for Samsung (lab E).
- For the consortium Siemens-Masushita-Mindspeed (lab D) the testing was split in 3 parts, each tested in a different language; exp 1a and exp 2 were tested in Japanese, exp 1b and 4 in American English and exp 3 in German.

The listening material depended on the test laboratories:

- Sennheiser HD25 for Lab A, B and E
- Beyerdynamic DT770 professional for Lab C
- Sennheiser HD25 for exp 1a, 2 and 3 in lab D

A pass/fail summary of subjective experiments can be found in Table 40, Table 41, Table 42, Table 43, Table 44, Table 45, Table 46, and Table 47. The detailed analysis is included in the excel sheets as well; these sheets are also available under the Q10/16 informal FTP area:

<http://ifa.itu.int/t/2005/sg16/xchange/wp3/q10/g729ev/exp1-4/>

Only two candidates pass all requirements in all laboratories and on all conditions.

In narrow band speech experiment, most failures concern CuT at 12 kbit/s. In wide band speech experiment, the failures concern CuT at 14 kbit/s. For wide band music, all candidates but one pass the requirement. All candidates pass all requirements in experiment 3: Wide band speech in background noise.

9.1.1 Test results – Experiment 1a

Table 40: Pairwise comparisons for Experiment 1a

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A		A (X-chk)		B		B (X-chk)		C		C (X-chk)		D		D (X-chk)		E		E (X-chk)	
CuT cond.	Ref cond.	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 8k, -26dBov	G.729A at 8k, -26dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 12k, -26dBov	G.729E, -26dBov	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Fail	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass
CuT at 8k, -16dBov	G.729A at 8k, -16dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 12k, -16dBov	G.729E, -16dBov	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 8k, -36dBov	G.729A at 8k, -36dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 12k, -36dBov	G.729E, -36dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 8k, 3% FER	G.729A at 8k, 3% FER	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 12k, 3% FER	G.729A at 8k, 3% FER	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

In this experiment, Candidates A and D pass all requirements. Candidate B fails one requirement in lab B and one requirement in crosschecked lab A. Candidate C fails one requirement in crosschecked lab B. Candidate E fails one requirement in crosschecked lab D.

Most failures are on condition CuT at 12k, -26dB with the reference G.729E, -26dB.

9.1.2 Test Results – Experiment 1b

Table 41: Pairwise comparisons for Experiment 1b

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A	99%	A (X-chk)	99%	B	99%	B (X-chk)	99%	C	99%	C (X-chk)	99%	D	99%	D (X-chk)	99%	E	99%	E (X-chk)	99%
CuT cond.	Reference cond.	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 14k, -26dBov	G.729A at 8k, -26dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Fail	Pass	Pass
CuT at 14k, -26dBov	G.722.2 at 8.85k, -26dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 24k, -26dBov	G.722 at 48k, -26dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 32k, -26dBov	G.722 at 56k, -26dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 14k, -16dBov	G.729A at 8k, -16dBov	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Fail	Pass	Pass
CuT at 14k, -16dBov	G.722.2 at 8.85k, -16dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 24k, -16dBov	G.722 at 48k, -16dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 32k, -16dBov	G.722 at 56k, -16dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 14k, -36dBov	G.729A at 8k, -36dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 14k, -36dBov	G.722.2 at 8.85k, -36dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 24k, -36dBov	G.722 at 48k, -36dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 32k, -36dBov	G.722 at 56k, -36dBov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 24k, 1% FER	G.722 at 48k, 0% FER	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 32k, 1% FER	G.722 at 56k, 0% FER	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

In this experiment, Candidates A, C and D pass all requirements. Candidate B fails one requirement in crosschecked lab E. Candidate E fails two requirements in lab E. Most failures are on condition CuT at 14k, -16dB with the reference G.729A at 8k, -16dB.

9.1.3 Test Results – Experiment 2a

Table 42: Pairwise comparisons for Experiment 2a

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A		B		C		D		E	
CuT conditions	Reference conditions	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 8k, Music Bkgr	G.729A at 8k, Music Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 12k, Music Bkgr	G.729A at 8k, Music Bkgr	Pass	Pass	Fail	<i>Fail</i>	Pass	Pass	Pass	Pass	Pass	Pass

In this experiment, Candidates A, C, D and E pass all requirements. Candidate B fails one requirement in lab B. The failure is on condition CuT at 12k with the reference G.729A.

9.1.4 Test Results – Experiment 2b

Table 43: Pairwise comparisons for Experiment 2b

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A		B		C		D		E	
CuT conditions	Reference conditions	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 8k, Office Bkgr	G.729A at 8k, Office Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 12k, Office Bkgr	G.729A at 8k, Office Bkgr	Pass	Pass	Pass	<i>Fail</i>	Pass	Pass	Pass	Pass	Pass	<i>Fail</i>

In this experiment, all Candidates pass all requirements at 95% confidence interval.

9.1.5 Test Results – Experiment 2c

Table 44: Pairwise comparisons for Experiment 2c

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A	99%	A (X-chk)	99%	B	99%	B (X-chk)	99%	C	99%	C (X-chk)	99%	D	99%	D (X-chk)	99%	E	99%	E (X-chk)	99%
CuT cond.	Reference cond.	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 8k, Babble Bkgr	G.729A at 8k, Babble Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 12k, Babble Bkgr	G.729A at 8k, Babble Bkgr	Pass	<i>Fail</i>	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	<i>Fail</i>	Pass	<i>Fail</i>	<i>Fail</i>	<i>Fail</i>

In this experiment, Candidates A, B, C and D pass all requirements at 95% confidence interval. Candidate E fails one requirement in crosschecked lab B. The failure is on condition CuT at 12k with the reference G.729A.

9.1.6 Test Results – Experiment 3a

Table 45: Pairwise comparisons for Experiment 3a

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A	99%	A (X-chk)	99%	B	99%	B (X-chk)	99%	C	99%	C (X-chk)	99%	D	99%	D (X-chk)	99%	E	99%	E (X-chk)	99%
CuT cond.	Reference cond.	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 24k, Music Bkgr	G.722 at 48k, Music Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 32k, Music Bkgr	G.722 at 56k, Music Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

In this experiment, all Candidates pass all requirements.

9.1.7 Test Results – Experiment 3b

Table 46: Pairwise comparisons for Experiment 3b

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A	99%	A (X-chk)	99%	B	99%	B (X-chk)	99%	C	99%	C (X-chk)	99%	D	99%	D (X-chk)	99%	E	99%	E (X-chk)	99%
CuT cond.	Reference cond.	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 24k, Office Bkgr	G.722 at 48k, Office Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 32k, Office Bkgr	G.722 at 56k, Office Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

In this experiment, all Candidates pass all requirements.

9.1.8 Test Results – Experiment 3c

Table 47: Pairwise comparisons for Experiment 3c

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A	99%	B	99%	C	99%	D	99%	E	99%
CuT condition	Reference condition	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 24k, Babble Bkgr	G.722 at 48k, Babble Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
CuT at 32k, Babble Bkgr	G.722 at 56k, Babble Bkgr	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

In this experiment, all Candidates pass all requirements.

9.1.9 Test Results – Experiment 4

Table 48: Pairwise comparisons for Experiment 4

CuT	Reference	ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test		ToR Test	
		A	99%	A (X-chk)	99%	B	99%	B (X-chk)	99%	C	99%	C (X-chk)	99%	D	99%	D (X-chk)	99%	E	99%	E (X-chk)	99%
CuT cond.	Ref cond.	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CuT at 32k, Music	G.722 at 56k, Music	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Fail

In this experiment, Candidates A, B, C and D pass the requirement. Candidate E fails the requirement in crosschecking lab C.

9.1.10 Test Results Summary – Codec Comparison and MOS Analysis of the Candidates

Table 49: Candidate Codec Comparison

Exp.	Coder condition	Reference cond.	Lab A Test-Ref	Lab A-X Test-Ref	Lab B Test-Ref	Lab B-X Test-Ref	Lab C Test-Ref	Lab C-X Test-Ref	Lab D Test-Ref	Lab D-X Test-Ref	Lab E Test-Ref	Lab E-X Test-Ref
1a	CuT at 8k, -26dB	G.729A at 8k, -26dB	0.104	0.135	0.313	0.156	0.531	-0.115	0.188	0.490	0.188	0.260
1a	CuT at 12k, -26dB	G.729E, -26dB	-0.010	0.135	-0.240	-0.146	0.365	-0.240	-0.104	0.271	0.000	-0.208
1a	CuT at 8k, -16dB	G.729A at 8k, -16dB	0.083	0.115	0.073	0.156	0.302	0.073	0.260	0.417	0.208	0.135
1a	CuT at 12k, -16dB	G.729E, -16dB	-0.010	0.104	-0.104	-0.156	-0.010	-0.083	-0.104	0.063	0.104	0.010
1a	CuT at 8k, -36dB	G.729A at 8k, -36dB	0.552	0.417	0.396	0.740	0.833	0.344	0.698	0.677	0.417	0.740
1a	CuT at 12k, -36dB	G.729E, -36dB	0.125	0.458	-0.125	0.135	0.854	0.104	0.260	0.531	0.583	0.406
1a	CuT at 8k, 3% FER	G.729A at 8k, 3% FER	0.302	0.354	0.292	0.240	0.698	0.104	0.354	0.573	0.208	0.250
1a	CuT at 12k, 3%FER	G.729A at 8k, 3% FER	0.521	0.365	0.375	0.417	0.948	0.313	0.531	0.823	0.375	0.438
1b	CuT at 14k, -26dB	G.729A at 8k, -26dB	0.969	0.698	0.740	0.250	1.188	1.313	1.125	0.760	0.000	1.302
1b	CuT at 14k, -26dB	G.722.2 at 8.85k, -26dB	0.021	-0.073	0.344	0.677	0.698	0.167	0.281	0.104	0.365	0.750
1b	CuT at 24k, -26dB	G.722 at 48k, -26dB	0.365	0.115	0.365	1.073	0.802	0.490	0.760	0.510	1.250	1.094
1b	CuT at 32k, -26dB	G.722 at 56k, -26dB	0.354	0.073	0.229	0.885	0.885	0.385	0.292	0.271	0.927	1.052
1b	CuT at 14k, -16dB	G.729A at 8k, -16dB	0.948	0.823	0.708	0.146	1.083	1.208	1.042	0.792	0.021	1.042
1b	CuT at 14k, -16dB	G.722.2 at 8.85k, -16dB	0.021	-0.063	0.281	0.292	0.208	0.135	0.188	0.094	0.292	0.125
1b	CuT at 24k, -16dB	G.722 at 48k, -16dB	0.271	0.115	0.365	0.750	0.552	0.354	0.219	0.333	0.833	0.458
1b	CuT at 32k, -16dB	G.722 at 56k, -16dB	0.063	0.042	-0.156	0.406	0.156	0.115	-0.115	-0.042	0.469	0.240
1b	CuT at 14k, -36dB	G.729A at 8k, -36dB	1.125	0.781	0.688	0.427	1.469	1.344	1.115	0.854	0.458	1.469
1b	CuT at 14k, -36dB	G.722.2 at 8.85k, -36dB	0.344	0.260	0.635	1.198	1.344	0.688	0.604	0.469	1.281	1.313
1b	CuT at 24k, -36dB	G.722 at 48k, -36dB	0.750	0.781	0.906	1.698	1.427	1.281	0.917	0.875	1.708	1.490
1b	CuT at 32k, -36dB	G.722 at 56k, -36dB	0.552	0.615	0.833	1.438	1.375	1.208	0.677	0.729	1.396	1.427
1b	CuT at 24k, 1% FER	G.722 at 48k, 0% FER	0.240	0.094	0.302	0.885	0.813	0.406	0.458	0.302	1.156	0.885
1b	CuT at 32k, 1% FER	G.722 at 56k, 0% FER	0.156	0.021	-0.031	0.719	0.698	0.146	0.073	-0.083	0.719	0.833
2a	CuT at 8k, Music Bkgr	G.729A at 8k, Music Bkgr	0.250	-	0.021	-	0.240	-	0.063	-	0.104	-
2a	CuT at 12k, Music Bkgr	G.729A at 8k, Music Bkgr	0.542	-	0.083	-	0.906	-	0.521	-	0.219	-
2b	CuT at 8k, Office Bkgr	G.729A at 8k, Office Bkgr	0.135	-	0.115	-	0.271	-	0.010	-	-0.063	-
2b	CuT at 12k, Office Bkgr	G.729A at 8k, Office Bkgr	0.490	-	0.177	-	0.875	-	0.323	-	0.177	-
2c	CuT at 8k, Babble Bkgr	G.729A at 8k, Babble Bkgr	0.052	0.250	0.271	0.396	0.281	0.094	0.083	-0.042	0.094	-0.104
2c	CuT at 12k, Babble Bkgr	G.729A at 8k, Babble Bkgr	0.177	0.729	0.281	0.323	0.792	0.177	0.281	0.188	0.146	0.104
4	CuT at 32k, Music	G.722 at 56k, Music	0.260	-0.010	0.271	1.031	0.219	0.010	-0.146	0.281	0.781	-0.573
3a	CuT at 24k, Music Bkgr	G.722 at 48k, Music Bkgr	-14.6	-13.6	-9.6	-11.6	-10.6	-19.6	-12.6	-12.6	-10.6	-9.6
3a	CuT at 32k, Music Bkgr	G.722 at 56k, Music Bkgr	-9.6	-8.6	-13.6	-10.6	-10.6	-14.6	-8.6	-11.6	-15.6	-8.6
3b	CuT at 24k, Office Bkgr	G.722 at 48k, Office Bkgr	-9.6	-12.6	-11.6	-7.6	-8.6	-12.6	-11.6	-10.6	-12.6	-10.6
3b	CuT at 32k, Office Bkgr	G.722 at 56k, Office Bkgr	-9.6	-12.6	-9.6	-9.6	-8.6	-11.6	-9.6	-11.6	-14.6	-8.6
3c	CuT at 24k, Babble Bkgr	G.722 at 48k, Babble Bkgr	-9.6	-	-12.6	-	-10.6	-	-12.6	-	-13.6	-
3c	CuT at 32k, Babble Bkgr	G.722 at 56k, Babble Bkgr	-9.6	-	-9.6	-	-9.6	-	-9.6	-	-10.6	-

Table 50: Candidate A

Req.	Coder under Test Candidate A				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	4.094	0.682	G.729A at 8k, -26dB	c08	3.990	0.703	0.100	0.104	NWT	-0.165	Pass	-0.235	Pass
1a	CuT at 12k, -26dB	c17	4.354	0.665	G.729E, -26dB	c09	4.365	0.651	0.095	-0.010	NWT	-0.157	Pass	-0.223	Pass
1a	CuT at 8k, -16dB	c18	3.969	0.760	G.729A at 8k, -16dB	c10	3.885	0.832	0.115	0.083	NWT	-0.190	Pass	-0.270	Pass
1a	CuT at 12k, -16dB	c19	4.281	0.706	G.729E, -16dB	c11	4.292	0.724	0.103	-0.010	NWT	-0.171	Pass	-0.242	Pass
1a	CuT at 8k, -36dB	c20	4.031	0.774	G.729A at 8k, -36dB	c12	3.479	0.781	0.112	0.552	NWT	-0.186	Pass	-0.263	Pass
1a	CuT at 12k, -36dB	c21	4.146	0.740	G.729E, -36dB	c13	4.021	0.754	0.108	0.125	NWT	-0.178	Pass	-0.253	Pass
1a	CuT at 8k, 3% FER	c22	3.656	0.819	G.729A at 8k, 3% FER	c14	3.354	0.906	0.125	0.302	NWT	-0.206	Pass	-0.292	Pass
1a	CuT at 12k, 3%FER	c23	3.875	0.771	G.729A at 8k, 3% FER	c14	3.354	0.906	0.121	0.521	BT	0.201	Pass	0.285	Pass
1b	CuT at 14k, -26dB	c30	4.063	0.629	G.729A at 8k, -26dB	c18	3.094	0.782	0.102	0.969	BT	0.169	Pass	0.240	Pass
1b	CuT at 14k, -26dB	c30	4.063	0.629	G.722.2 at 8.85k, -26dB	c19	4.042	0.739	0.099	0.021	NWT	-0.164	Pass	-0.232	Pass
1b	CuT at 24k, -26dB	c31	4.354	0.680	G.722 at 48k, -26dB	c20	3.990	0.733	0.102	0.365	NWT	-0.169	Pass	-0.239	Pass
1b	CuT at 32k, -26dB	c32	4.417	0.592	G.722 at 56k, -26dB	c21	4.063	0.693	0.093	0.354	NWT	-0.154	Pass	-0.218	Pass
1b	CuT at 14k, -16dB	c33	4.156	0.730	G.729A at 8k, -16dB	c22	3.208	0.939	0.121	0.948	BT	0.201	Pass	0.285	Pass
1b	CuT at 14k, -16dB	c33	4.156	0.730	G.722.2 at 8.85k, -16dB	c23	4.135	0.776	0.109	0.021	NWT	-0.180	Pass	-0.255	Pass
1b	CuT at 24k, -16dB	c34	4.281	0.660	G.722 at 48k, -16dB	c24	4.010	0.657	0.095	0.271	NWT	-0.157	Pass	-0.223	Pass
1b	CuT at 32k, -16dB	c35	4.375	0.700	G.722 at 56k, -16dB	c25	4.313	0.621	0.095	0.063	NWT	-0.158	Pass	-0.224	Pass
1b	CuT at 14k, -36dB	c36	3.865	0.776	G.729A at 8k, -36dB	c26	2.740	0.714	0.108	1.125	BT	0.178	Pass	0.253	Pass
1b	CuT at 14k, -36dB	c36	3.865	0.776	G.722.2 at 8.85k, -36dB	c27	3.521	0.781	0.112	0.344	NWT	-0.186	Pass	-0.264	Pass
1b	CuT at 24k, -36dB	c37	4.042	0.724	G.722 at 48k, -36dB	c28	3.292	0.845	0.114	0.750	NWT	-0.188	Pass	-0.266	Pass
1b	CuT at 32k, -36dB	c38	4.031	0.852	G.722 at 56k, -36dB	c29	3.479	0.951	0.130	0.552	NWT	-0.215	Pass	-0.306	Pass
1b	CuT at 24k, 1% FER	c39	4.229	0.703	G.722 at 48k, 0% FER	c20	3.990	0.733	0.104	0.240	NWT	-0.171	Pass	-0.243	Pass
1b	CuT at 32k, 1% FER	c40	4.219	0.668	G.722 at 56k, 0% FER	c21	4.063	0.693	0.098	0.156	NWT	-0.162	Pass	-0.231	Pass
2a	CuT at 8k, Music Bkgr	c07	4.260	0.811	G.729A at 8k, Music Bkgr	c06	4.010	0.801	0.116	0.250	NWT	-0.192	Pass	-0.273	Pass
2a	CuT at 12k, Music Bkgr	c08	4.552	0.613	G.729A at 8k, Music Bkgr	c06	4.010	0.801	0.103	0.542	BT	0.170	Pass	0.242	Pass
2b	CuT at 8k, Office Bkgr	c07	4.281	0.736	G.729A at 8k, Office Bkgr	c06	4.146	0.906	0.119	0.135	NWT	-0.197	Pass	-0.279	Pass
2b	CuT at 12k, Office Bkgr	c08	4.635	0.600	G.729A at 8k, Office Bkgr	c06	4.146	0.906	0.111	0.490	BT	0.183	Pass	0.260	Pass
2c	CuT at 8k, Babble Bkgr	c07	4.542	0.695	G.729A at 8k, Babble Bkgr	c06	4.490	0.665	0.098	0.052	NWT	-0.162	Pass	-0.230	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.667	0.574	G.729A at 8k, Babble Bkgr	c06	4.490	0.665	0.090	0.177	BT	0.148	Pass	0.210	Fail
4	CuT at 32k, Music	c08	4.240	0.778	G.722 at 56k, Music	c07	3.979	0.754	0.111	0.260	NWT	-0.183	Pass	-0.259	Pass

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	0	-	G.722 at 48k, Music Bkgr	c05	5	14.60	15.802	-14.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	0	-	G.722 at 56k, Music Bkgr	c06	0	9.60	10.105	-9.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	1	-	G.722 at 48k, Office Bkgr	c05	1	10.60	8.456	-9.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	0	-	G.722 at 56k, Office Bkgr	c06	0	9.60	10.105	-9.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 24k, Babble Bkgr	c07	0	-	G.722 at 48k, Babble Bkgr	c05	0	9.60	10.105	-9.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 32k, Babble Bkgr	c08	0	-	G.722 at 56k, Babble Bkgr	c06	0	9.60	10.105	-9.6	NMT10%	2.706	Pass	5.412	Pass

Table 51: Candidate A - Crosscheck

Exp.	Coder under Test		Candidate A-Xchk			Reference				Terms of Reference Test					
	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	4.396	0.774	G.729A at 8k, -26dB	c08	4.260	0.743	0.110	0.135	NWT	-0.181	Pass	-0.257	Pass
1a	CuT at 12k, -26dB	c17	4.615	0.639	G.729E, -26dB	c09	4.479	0.696	0.096	0.135	NWT	-0.159	Pass	-0.226	Pass
1a	CuT at 8k, -16dB	c18	4.479	0.740	G.729A at 8k, -16dB	c10	4.365	0.727	0.106	0.115	NWT	-0.175	Pass	-0.248	Pass
1a	CuT at 12k, -16dB	c19	4.563	0.577	G.729E, -16dB	c11	4.458	0.695	0.092	0.104	NWT	-0.152	Pass	-0.216	Pass
1a	CuT at 8k, -36dB	c20	4.167	0.804	G.729A at 8k, -36dB	c12	3.750	0.740	0.111	0.417	NWT	-0.184	Pass	-0.262	Pass
1a	CuT at 12k, -36dB	c21	4.396	0.688	G.729E, -36dB	c13	3.938	0.792	0.107	0.458	NWT	-0.177	Pass	-0.251	Pass
1a	CuT at 8k, 3% FER	c22	4.021	0.929	G.729A at 8k, 3% FER	c14	3.667	0.914	0.133	0.354	NWT	-0.220	Pass	-0.312	Pass
1a	CuT at 12k, 3% FER	c23	4.031	0.864	G.729A at 8k, 3% FER	c14	3.667	0.914	0.128	0.365	BT	0.212	Pass	0.301	Pass
1b	CuT at 14k, -26dB	c30	3.615	0.813	G.729A at 8k, -26dB	c18	2.917	0.914	0.125	0.698	BT	0.206	Pass	0.293	Pass
1b	CuT at 14k, -26dB	c30	3.615	0.813	G.722.2 at 8.85k, -26dB	c19	3.688	0.874	0.122	-0.073	NWT	-0.201	Pass	-0.286	Pass
1b	CuT at 24k, -26dB	c31	3.927	0.798	G.722 at 48k, -26dB	c20	3.813	0.812	0.116	0.115	NWT	-0.192	Pass	-0.273	Pass
1b	CuT at 32k, -26dB	c32	4.208	0.794	G.722 at 56k, -26dB	c21	4.135	0.763	0.112	0.073	NWT	-0.186	Pass	-0.264	Pass
1b	CuT at 14k, -16dB	c33	3.635	0.872	G.729A at 8k, -16dB	c22	2.813	0.886	0.127	0.823	BT	0.210	Pass	0.298	Pass
1b	CuT at 14k, -16dB	c33	3.635	0.872	G.722.2 at 8.85k, -16dB	c23	3.698	0.860	0.125	-0.063	NWT	-0.207	Pass	-0.293	Pass
1b	CuT at 24k, -16dB	c34	3.917	0.804	G.722 at 48k, -16dB	c24	3.802	0.854	0.120	0.115	NWT	-0.198	Pass	-0.281	Pass
1b	CuT at 32k, -16dB	c35	4.167	0.854	G.722 at 56k, -16dB	c25	4.125	0.729	0.115	0.042	NWT	-0.189	Pass	-0.269	Pass
1b	CuT at 14k, -36dB	c36	3.615	0.875	G.729A at 8k, -36dB	c26	2.833	0.981	0.134	0.781	BT	0.222	Pass	0.315	Pass
1b	CuT at 14k, -36dB	c36	3.615	0.875	G.722.2 at 8.85k, -36dB	c27	3.354	0.833	0.123	0.260	NWT	-0.204	Pass	-0.289	Pass
1b	CuT at 24k, -36dB	c37	3.969	0.787	G.722 at 48k, -36dB	c28	3.188	0.898	0.122	0.781	NWT	-0.201	Pass	-0.286	Pass
1b	CuT at 32k, -36dB	c38	4.104	0.801	G.722 at 56k, -36dB	c29	3.490	0.995	0.130	0.615	NWT	-0.215	Pass	-0.306	Pass
1b	CuT at 24k, 1% FER	c39	3.906	0.769	G.722 at 48k, 0% FER	c20	3.813	0.812	0.114	0.094	NWT	-0.189	Pass	-0.268	Pass
1b	CuT at 32k, 1% FER	c40	4.156	0.812	G.722 at 56k, 0% FER	c21	4.135	0.763	0.114	0.021	NWT	-0.188	Pass	-0.267	Pass
2c	CuT at 8k, Babble Bkgr	c07	4.094	0.755	G.729A at 8k, Babble Bkgr	c06	3.844	0.838	0.115	0.250	NWT	-0.190	Pass	-0.270	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.573	0.611	G.729A at 8k, Babble Bkgr	c06	3.844	0.838	0.106	0.729	BT	0.175	Pass	0.248	Pass
4	CuT at 32k, Music	c08	3.781	0.897	G.722 at 56k, Music	c07	3.792	0.882	0.128	-0.010	NWT	-0.212	Pass	-0.301	Pass

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	1	-	G.722 at 48k, Music Bkgr	c05	5	14.6	12.905	-13.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	2	-	G.722 at 56k, Music Bkgr	c06	1	10.6	6.282	-8.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	2	-	G.722 at 48k, Office Bkgr	c05	5	14.6	10.469	-12.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	1	-	G.722 at 56k, Office Bkgr	c06	4	13.6	11.769	-12.6	NMT10%	2.706	Pass	5.412	Pass

Table 52: Candidate B

Req.	Coder under Test Candidate B				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	3.990	0.733	G.729A at 8k, -26dB	c08	3.677	0.761	0.108	0.313	NWT	-0.178	Pass	-0.253	Pass
1a	CuT at 12k, -26dB	c17	3.865	0.705	G.729E, -26dB	c09	4.104	0.747	0.105	-0.240	NWT	-0.173	Fail	-0.246	Pass
1a	CuT at 8k, -16dB	c18	3.760	0.843	G.729A at 8k, -16dB	c10	3.688	0.744	0.115	0.073	NWT	-0.190	Pass	-0.269	Pass
1a	CuT at 12k, -16dB	c19	3.833	0.790	G.729E, -16dB	c11	3.938	0.708	0.108	-0.104	NWT	-0.179	Pass	-0.254	Pass
1a	CuT at 8k, -36dB	c20	3.896	0.774	G.729A at 8k, -36dB	c12	3.500	0.883	0.120	0.396	NWT	-0.198	Pass	-0.281	Pass
1a	CuT at 12k, -36dB	c21	3.760	0.750	G.729E, -36dB	c13	3.885	0.819	0.113	-0.125	NWT	-0.187	Pass	-0.266	Pass
1a	CuT at 8k, 3% FER	c22	3.302	0.884	G.729A at 8k, 3% FER	c14	3.010	0.827	0.124	0.292	NWT	-0.204	Pass	-0.290	Pass
1a	CuT at 12k, 3%FER	c23	3.385	0.887	G.729A at 8k, 3% FER	c14	3.010	0.827	0.124	0.375	BT	0.205	Pass	0.290	Pass
1b	CuT at 14k, -26dB	c30	4.333	0.706	G.729A at 8k, -26dB	c18	3.594	0.924	0.119	0.740	BT	0.196	Pass	0.278	Pass
1b	CuT at 14k, -26dB	c30	4.333	0.706	G.722.2 at 8.85k, -26dB	c19	3.990	0.877	0.115	0.344	NWT	-0.190	Pass	-0.269	Pass
1b	CuT at 24k, -26dB	c31	4.250	0.940	G.722 at 48k, -26dB	c20	3.885	0.916	0.134	0.365	NWT	-0.221	Pass	-0.314	Pass
1b	CuT at 32k, -26dB	c32	4.365	0.783	G.722 at 56k, -26dB	c21	4.135	0.841	0.117	0.229	NWT	-0.194	Pass	-0.275	Pass
1b	CuT at 14k, -16dB	c33	4.313	0.758	G.729A at 8k, -16dB	c22	3.604	0.864	0.117	0.708	BT	0.194	Pass	0.275	Pass
1b	CuT at 14k, -16dB	c33	4.313	0.758	G.722.2 at 8.85k, -16dB	c23	4.031	0.774	0.111	0.281	NWT	-0.183	Pass	-0.259	Pass
1b	CuT at 24k, -16dB	c34	4.292	0.724	G.722 at 48k, -16dB	c24	3.927	0.861	0.115	0.365	NWT	-0.190	Pass	-0.269	Pass
1b	CuT at 32k, -16dB	c35	4.188	0.758	G.722 at 56k, -16dB	c25	4.344	0.708	0.106	-0.156	NWT	-0.175	Pass	-0.248	Pass
1b	CuT at 14k, -36dB	c36	4.281	0.764	G.729A at 8k, -36dB	c26	3.594	0.924	0.122	0.688	BT	0.202	Pass	0.287	Pass
1b	CuT at 14k, -36dB	c36	4.281	0.764	G.722.2 at 8.85k, -36dB	c27	3.646	0.917	0.122	0.635	NWT	-0.201	Pass	-0.286	Pass
1b	CuT at 24k, -36dB	c37	4.313	0.621	G.722 at 48k, -36dB	c28	3.406	0.878	0.110	0.906	NWT	-0.181	Pass	-0.257	Pass
1b	CuT at 32k, -36dB	c38	4.292	0.710	G.722 at 56k, -36dB	c29	3.458	0.962	0.122	0.833	NWT	-0.202	Pass	-0.286	Pass
1b	CuT at 24k, 1% FER	c39	4.188	0.825	G.722 at 48k, 0% FER	c20	3.885	0.916	0.126	0.302	NWT	-0.208	Pass	-0.295	Pass
1b	CuT at 32k, 1% FER	c40	4.104	0.888	G.722 at 56k, 0% FER	c21	4.135	0.841	0.125	-0.031	NWT	-0.206	Pass	-0.293	Pass
2a	CuT at 8k, Music Bkgr	c07	4.240	0.778	G.729A at 8k, Music Bkgr	c06	4.219	0.699	0.107	0.021	NWT	-0.176	Pass	-0.250	Pass
2a	CuT at 12k, Music Bkgr	c08	4.302	0.618	G.729A at 8k, Music Bkgr	c06	4.219	0.699	0.095	0.083	BT	0.157	Fail	0.223	Fail
2b	CuT at 8k, Office Bkgr	c07	4.198	0.705	G.729A at 8k, Office Bkgr	c06	4.083	0.706	0.102	0.115	NWT	-0.168	Pass	-0.239	Pass
2b	CuT at 12k, Office Bkgr	c08	4.260	0.637	G.729A at 8k, Office Bkgr	c06	4.083	0.706	0.097	0.177	BT	0.160	Pass	0.228	Fail
2c	CuT at 8k, Babble Bkgr	c07	4.625	0.548	G.729A at 8k, Babble Bkgr	c06	4.354	0.580	0.081	0.271	NWT	-0.135	Pass	-0.191	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.635	0.526	G.729A at 8k, Babble Bkgr	c06	4.354	0.580	0.080	0.281	BT	0.132	Pass	0.187	Pass
4	CuT at 32k, Music	c08	4.083	0.914	G.722 at 56k, Music	c07	3.813	1.029	0.140	0.271	NWT	-0.232	Pass	-0.330	Pass

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	2	-	G.722 at 48k, Music Bkgr	c05	2	11.60	7.293	-9.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	0	-	G.722 at 56k, Music Bkgr	c06	4	13.60	14.637	-13.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	1	-	G.722 at 48k, Office Bkgr	c05	3	12.60	10.648	-11.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	3	-	G.722 at 56k, Office Bkgr	c06	3	12.60	6.430	-9.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 24k, Babble Bkgr	c07	0	-	G.722 at 48k, Babble Bkgr	c05	3	12.60	13.485	-12.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 32k, Babble Bkgr	c08	1	-	G.722 at 56k, Babble Bkgr	c06	1	10.60	8.456	-9.6	NMT10%	2.706	Pass	5.412	Pass

Table 53: Candidate B - Crosscheck

Req.	Coder under Test Candidate B X-chk				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	4.156	0.701	G.729A at 8k, -26dB	c08	4.000	0.725	0.103	0.156	NWT	-0.170	Pass	-0.242	Pass
1a	CuT at 12k, -26dB	c17	4.167	0.749	G.729E, -26dB	c09	4.313	0.654	0.102	-0.146	NWT	-0.168	Pass	-0.238	Pass
1a	CuT at 8k, -16dB	c18	4.177	0.632	G.729A at 8k, -16dB	c10	4.021	0.767	0.101	0.156	NWT	-0.168	Pass	-0.238	Pass
1a	CuT at 12k, -16dB	c19	4.198	0.720	G.729E, -16dB	c11	4.354	0.562	0.093	-0.156	NWT	-0.154	Fail	-0.219	Pass
1a	CuT at 8k, -36dB	c20	4.104	0.640	G.729A at 8k, -36dB	c12	3.365	0.860	0.109	0.740	NWT	-0.181	Pass	-0.257	Pass
1a	CuT at 12k, -36dB	c21	4.156	0.730	G.729E, -36dB	c13	4.021	0.711	0.104	0.135	NWT	-0.172	Pass	-0.244	Pass
1a	CuT at 8k, 3% FER	c22	3.615	0.988	G.729A at 8k, 3% FER	c14	3.375	0.921	0.138	0.240	NWT	-0.228	Pass	-0.323	Pass
1a	CuT at 12k, 3%FER	c23	3.792	0.882	G.729A at 8k, 3% FER	c14	3.375	0.921	0.130	0.417	BT	0.215	Pass	0.305	Pass
1b	CuT at 14k, -26dB	c30	4.781	0.527	G.729A at 8k, -26dB	c18	4.531	0.648	0.085	0.250	BT	0.141	Pass	0.200	Pass
1b	CuT at 14k, -26dB	c30	4.781	0.527	G.722.2 at 8.85k, -26dB	c19	4.104	0.732	0.092	0.677	NWT	-0.152	Pass	-0.216	Pass
1b	CuT at 24k, -26dB	c31	4.656	0.559	G.722 at 48k, -26dB	c20	3.583	0.790	0.099	1.073	NWT	-0.163	Pass	-0.232	Pass
1b	CuT at 32k, -26dB	c32	4.583	0.627	G.722 at 56k, -26dB	c21	3.698	0.698	0.096	0.885	NWT	-0.158	Pass	-0.225	Pass
1b	CuT at 14k, -16dB	c33	4.656	0.577	G.729A at 8k, -16dB	c22	4.510	0.665	0.090	0.146	BT	0.149	Fail	0.211	Fail
1b	CuT at 14k, -16dB	c33	4.656	0.577	G.722.2 at 8.85k, -16dB	c23	4.365	0.698	0.092	0.292	NWT	-0.153	Pass	-0.217	Pass
1b	CuT at 24k, -16dB	c34	4.583	0.660	G.722 at 48k, -16dB	c24	3.833	0.777	0.104	0.750	NWT	-0.172	Pass	-0.244	Pass
1b	CuT at 32k, -16dB	c35	4.500	0.632	G.722 at 56k, -16dB	c25	4.094	0.682	0.095	0.406	NWT	-0.157	Pass	-0.223	Pass
1b	CuT at 14k, -36dB	c36	4.615	0.587	G.729A at 8k, -36dB	c26	4.188	0.812	0.102	0.427	BT	0.169	Pass	0.240	Pass
1b	CuT at 14k, -36dB	c36	4.615	0.587	G.722.2 at 8.85k, -36dB	c27	3.417	0.627	0.088	1.198	NWT	-0.145	Pass	-0.206	Pass
1b	CuT at 24k, -36dB	c37	4.625	0.585	G.722 at 48k, -36dB	c28	2.927	0.714	0.094	1.698	NWT	-0.156	Pass	-0.221	Pass
1b	CuT at 32k, -36dB	c38	4.552	0.578	G.722 at 56k, -36dB	c29	3.115	0.709	0.093	1.438	NWT	-0.154	Pass	-0.219	Pass
1b	CuT at 24k, 1% FER	c39	4.469	0.648	G.722 at 48k, 0% FER	c20	3.583	0.790	0.104	0.885	NWT	-0.172	Pass	-0.245	Pass
1b	CuT at 32k, 1% FER	c40	4.417	0.735	G.722 at 56k, 0% FER	c21	3.698	0.698	0.103	0.719	NWT	-0.171	Pass	-0.243	Pass
2c	CuT at 8k, Babble Bkgr	c07	4.396	0.747	G.729A at 8k, Babble Bkgr	c06	4.000	0.795	0.111	0.396	NWT	-0.184	Pass	-0.261	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.323	0.852	G.729A at 8k, Babble Bkgr	c06	4.000	0.795	0.119	0.323	BT	0.197	Pass	0.279	Pass
4	CuT at 32k, Music	c08	4.240	0.867	G.722 at 56k, Music	c07	3.208	1.004	0.135	1.031	NWT	-0.224	Pass	-0.318	Pass

Req.	Coder under Test				Reference			Terms of Reference Test							
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	3	-	G.722 at 48k, Music Bkgr	c05	5	14.6	8.417	-11.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	0	-	G.722 at 56k, Music Bkgr	c06	1	10.6	11.219	-10.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	2	-	G.722 at 48k, Office Bkgr	c05	0	9.6	5.299	-7.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	0	-	G.722 at 56k, Office Bkgr	c06	0	9.6	10.105	-9.6	NMT10%	2.706	Pass	5.412	Pass

Table 54: Candidate C

Req.	Coder under Test Candidate C				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	3.938	0.880	G.729A at 8k, -26dB	c08	3.406	0.815	0.122	0.531	NWT	-0.202	Pass	-0.287	Pass
1a	CuT at 12k, -26dB	c17	4.323	0.747	G.729E, -26dB	c09	3.958	0.710	0.105	0.365	NWT	-0.174	Pass	-0.247	Pass
1a	CuT at 8k, -16dB	c18	3.896	0.827	G.729A at 8k, -16dB	c10	3.594	0.878	0.123	0.302	NWT	-0.203	Pass	-0.289	Pass
1a	CuT at 12k, -16dB	c19	4.208	0.893	G.729E, -16dB	c11	4.219	0.784	0.121	-0.010	NWT	-0.201	Pass	-0.285	Pass
1a	CuT at 8k, -36dB	c20	3.635	0.835	G.729A at 8k, -36dB	c12	2.802	0.720	0.113	0.833	NWT	-0.186	Pass	-0.264	Pass
1a	CuT at 12k, -36dB	c21	4.083	0.829	G.729E, -36dB	c13	3.229	0.703	0.111	0.854	NWT	-0.183	Pass	-0.260	Pass
1a	CuT at 8k, 3% FER	c22	3.583	0.879	G.729A at 8k, 3% FER	c14	2.885	0.806	0.122	0.698	NWT	-0.201	Pass	-0.286	Pass
1a	CuT at 12k, 3%FER	c23	3.833	0.890	G.729A at 8k, 3% FER	c14	2.885	0.806	0.123	0.948	BT	0.203	Pass	0.288	Pass
1b	CuT at 14k, -26dB	c30	4.177	0.711	G.729A at 8k, -26dB	c18	2.990	0.747	0.105	1.188	BT	0.174	Pass	0.247	Pass
1b	CuT at 14k, -26dB	c30	4.177	0.711	G.722.2 at 8.85k, -26dB	c19	3.479	0.665	0.099	0.698	NWT	-0.164	Pass	-0.233	Pass
1b	CuT at 24k, -26dB	c31	4.292	0.739	G.722 at 48k, -26dB	c20	3.490	0.711	0.105	0.802	NWT	-0.173	Pass	-0.245	Pass
1b	CuT at 32k, -26dB	c32	4.438	0.646	G.722 at 56k, -26dB	c21	3.552	0.679	0.096	0.885	NWT	-0.158	Pass	-0.224	Pass
1b	CuT at 14k, -16dB	c33	4.146	0.696	G.729A at 8k, -16dB	c22	3.063	0.831	0.111	1.083	BT	0.183	Pass	0.260	Pass
1b	CuT at 14k, -16dB	c33	4.146	0.696	G.722.2 at 8.85k, -16dB	c23	3.938	0.779	0.107	0.208	NWT	-0.176	Pass	-0.250	Pass
1b	CuT at 24k, -16dB	c34	4.396	0.607	G.722 at 48k, -16dB	c24	3.844	0.654	0.091	0.552	NWT	-0.151	Pass	-0.214	Pass
1b	CuT at 32k, -16dB	c35	4.313	0.701	G.722 at 56k, -16dB	c25	4.156	0.701	0.101	0.156	NWT	-0.167	Pass	-0.237	Pass
1b	CuT at 14k, -36dB	c36	3.990	0.673	G.729A at 8k, -36dB	c26	2.521	0.598	0.092	1.469	BT	0.152	Pass	0.216	Pass
1b	CuT at 14k, -36dB	c36	3.990	0.673	G.722.2 at 8.85k, -36dB	c27	2.646	0.680	0.098	1.344	NWT	-0.161	Pass	-0.229	Pass
1b	CuT at 24k, -36dB	c37	4.146	0.711	G.722 at 48k, -36dB	c28	2.719	0.627	0.097	1.427	NWT	-0.160	Pass	-0.227	Pass
1b	CuT at 32k, -36dB	c38	4.135	0.720	G.722 at 56k, -36dB	c29	2.760	0.645	0.099	1.375	NWT	-0.163	Pass	-0.231	Pass
1b	CuT at 24k, 1% FER	c39	4.302	0.713	G.722 at 48k, 0% FER	c20	3.490	0.711	0.103	0.813	NWT	-0.170	Pass	-0.241	Pass
1b	CuT at 32k, 1% FER	c40	4.250	0.681	G.722 at 56k, 0% FER	c21	3.552	0.679	0.098	0.698	NWT	-0.162	Pass	-0.230	Pass
2a	CuT at 8k, Music Bkgr	c07	3.542	0.739	G.729A at 8k, Music Bkgr	c06	3.302	0.872	0.117	0.240	NWT	-0.193	Pass	-0.274	Pass
2a	CuT at 12k, Music Bkgr	c08	4.208	0.695	G.729A at 8k, Music Bkgr	c06	3.302	0.872	0.114	0.906	BT	0.188	Pass	0.267	Pass
2b	CuT at 8k, Office Bkgr	c07	3.760	0.891	G.729A at 8k, Office Bkgr	c06	3.490	0.871	0.127	0.271	NWT	-0.210	Pass	-0.298	Pass
2b	CuT at 12k, Office Bkgr	c08	4.365	0.667	G.729A at 8k, Office Bkgr	c06	3.490	0.871	0.112	0.875	BT	0.185	Pass	0.263	Pass
2c	CuT at 8k, Babble Bkgr	c07	4.031	0.703	G.729A at 8k, Babble Bkgr	c06	3.750	0.834	0.111	0.281	NWT	-0.184	Pass	-0.261	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.542	0.631	G.729A at 8k, Babble Bkgr	c06	3.750	0.834	0.107	0.792	BT	0.176	Pass	0.250	Pass
4	CuT at 32k, Music	c08	3.958	0.780	G.722 at 56k, Music	c07	3.740	1.008	0.130	0.219	NWT	-0.215	Pass	-0.305	Pass

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	4	-	G.722 at 48k, Music Bkgr	c05	5	14.60	6.689	-10.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	1	-	G.722 at 56k, Music Bkgr	c06	2	11.60	9.544	-10.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	3	-	G.722 at 48k, Office Bkgr	c05	2	11.60	5.483	-8.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	1	-	G.722 at 56k, Office Bkgr	c06	0	9.60	7.385	-8.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 24k, Babble Bkgr	c07	0	-	G.722 at 48k, Babble Bkgr	c05	1	10.60	11.219	-10.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 32k, Babble Bkgr	c08	0	-	G.722 at 56k, Babble Bkgr	c06	0	9.60	10.105	-9.6	NMT10%	2.706	Pass	5.412	Pass

Table 55: Candidate C - Crosscheck

Req.	Coder under Test Candidate C-Xchk				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	3.719	0.817	G.729A at 8k, -26dB	c08	3.833	0.675	0.108	-0.115	NWT	-0.179	Pass	-0.254	Pass
1a	CuT at 12k, -26dB	c17	3.906	0.697	G.729E, -26dB	c09	4.146	0.696	0.101	-0.240	NWT	-0.166	Fail	-0.236	Fail
1a	CuT at 8k, -16dB	c18	3.792	0.794	G.729A at 8k, -16dB	c10	3.719	0.706	0.108	0.073	NWT	-0.179	Pass	-0.254	Pass
1a	CuT at 12k, -16dB	c19	3.896	0.732	G.729E, -16dB	c11	3.979	0.781	0.109	-0.083	NWT	-0.181	Pass	-0.256	Pass
1a	CuT at 8k, -36dB	c20	3.781	0.771	G.729A at 8k, -36dB	c12	3.438	0.844	0.117	0.344	NWT	-0.193	Pass	-0.274	Pass
1a	CuT at 12k, -36dB	c21	4.042	0.832	G.729E, -36dB	c13	3.938	0.792	0.117	0.104	NWT	-0.194	Pass	-0.275	Pass
1a	CuT at 8k, 3% FER	c22	3.219	0.797	G.729A at 8k, 3% FER	c14	3.115	0.881	0.121	0.104	NWT	-0.200	Pass	-0.285	Pass
1a	CuT at 12k, 3%FER	c23	3.427	0.855	G.729A at 8k, 3% FER	c14	3.115	0.881	0.125	0.313	BT	0.207	Pass	0.294	Pass
1b	CuT at 14k, -26dB	c30	4.365	0.634	G.729A at 8k, -26dB	c18	3.052	0.786	0.103	1.313	BT	0.170	Pass	0.242	Pass
1b	CuT at 14k, -26dB	c30	4.365	0.634	G.722.2 at 8.85k, -26dB	c19	4.198	0.734	0.099	0.167	NWT	-0.164	Pass	-0.232	Pass
1b	CuT at 24k, -26dB	c31	4.500	0.580	G.722 at 48k, -26dB	c20	4.010	0.747	0.097	0.490	NWT	-0.160	Pass	-0.226	Pass
1b	CuT at 32k, -26dB	c32	4.646	0.542	G.722 at 56k, -26dB	c21	4.260	0.637	0.085	0.385	NWT	-0.141	Pass	-0.200	Pass
1b	CuT at 14k, -16dB	c33	4.271	0.672	G.729A at 8k, -16dB	c22	3.063	0.737	0.102	1.208	BT	0.168	Pass	0.239	Pass
1b	CuT at 14k, -16dB	c33	4.271	0.672	G.722.2 at 8.85k, -16dB	c23	4.135	0.749	0.103	0.135	NWT	-0.170	Pass	-0.241	Pass
1b	CuT at 24k, -16dB	c34	4.458	0.648	G.722 at 48k, -16dB	c24	4.104	0.747	0.101	0.354	NWT	-0.167	Pass	-0.237	Pass
1b	CuT at 32k, -16dB	c35	4.448	0.647	G.722 at 56k, -16dB	c25	4.333	0.627	0.092	0.115	NWT	-0.152	Pass	-0.216	Pass
1b	CuT at 14k, -36dB	c36	4.292	0.614	G.729A at 8k, -36dB	c26	2.948	0.838	0.106	1.344	BT	0.175	Pass	0.249	Pass
1b	CuT at 14k, -36dB	c36	4.292	0.614	G.722.2 at 8.85k, -36dB	c27	3.604	0.801	0.103	0.688	NWT	-0.170	Pass	-0.242	Pass
1b	CuT at 24k, -36dB	c37	4.552	0.578	G.722 at 48k, -36dB	c28	3.271	0.814	0.102	1.281	NWT	-0.168	Pass	-0.239	Pass
1b	CuT at 32k, -36dB	c38	4.490	0.649	G.722 at 56k, -36dB	c29	3.281	0.804	0.105	1.208	NWT	-0.174	Pass	-0.247	Pass
1b	CuT at 24k, 1% FER	c39	4.417	0.675	G.722 at 48k, 0% FER	c20	4.010	0.747	0.103	0.406	NWT	-0.170	Pass	-0.241	Pass
1b	CuT at 32k, 1% FER	c40	4.406	0.625	G.722 at 56k, 0% FER	c21	4.260	0.637	0.091	0.146	NWT	-0.151	Pass	-0.214	Pass
2c	CuT at 8k, Babble Bkgr	c07	4.729	0.470	G.729A at 8k, Babble Bkgr	c06	4.635	0.545	0.073	0.094	NWT	-0.121	Pass	-0.172	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.813	0.418	G.729A at 8k, Babble Bkgr	c06	4.635	0.545	0.070	0.177	BT	0.116	Pass	0.165	Pass
4	CuT at 32k, Music	c08	3.854	0.962	G.722 at 56k, Music	c07	3.844	0.862	0.132	0.010	NWT	-0.218	Pass	-0.309	Pass

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	0	-	G.722 at 48k, Music Bkgr	c05	10	19.6	21.828	-19.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	0	-	G.722 at 56k, Music Bkgr	c06	5	14.6	15.802	-14.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	0	-	G.722 at 48k, Office Bkgr	c05	3	12.6	13.485	-12.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	0	-	G.722 at 56k, Office Bkgr	c06	2	11.6	12.346	-11.6	NMT10%	2.706	Pass	5.412	Pass

Table 56: Candidate D

Req.	Coder under Test Candidate D				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	3.594	0.889	G.729A at 8k, -26dB	c08	3.406	0.958	0.133	0.188	NWT	-0.221	Pass	-0.313	Pass
1a	CuT at 12k, -26dB	c17	3.813	0.874	G.729E, -26dB	c09	3.917	0.735	0.117	-0.104	NWT	-0.193	Pass	-0.274	Pass
1a	CuT at 8k, -16dB	c18	3.667	0.842	G.729A at 8k, -16dB	c10	3.406	0.841	0.121	0.260	NWT	-0.201	Pass	-0.285	Pass
1a	CuT at 12k, -16dB	c19	3.708	0.807	G.729E, -16dB	c11	3.813	0.799	0.116	-0.104	NWT	-0.192	Pass	-0.272	Pass
1a	CuT at 8k, -36dB	c20	3.396	0.814	G.729A at 8k, -36dB	c12	2.698	0.809	0.117	0.698	NWT	-0.194	Pass	-0.275	Pass
1a	CuT at 12k, -36dB	c21	3.615	0.933	G.729E, -36dB	c13	3.354	0.995	0.139	0.260	NWT	-0.230	Pass	-0.327	Pass
1a	CuT at 8k, 3% FER	c22	3.031	1.010	G.729A at 8k, 3% FER	c14	2.677	0.912	0.139	0.354	NWT	-0.230	Pass	-0.326	Pass
1a	CuT at 12k, 3%FER	c23	3.208	1.035	G.729A at 8k, 3% FER	c14	2.677	0.912	0.141	0.531	BT	0.233	Pass	0.330	Pass
1b	CuT at 14k, -26dB	c30	3.771	0.900	G.729A at 8k, -26dB	c18	2.646	0.781	0.122	1.125	BT	0.201	Pass	0.285	Pass
1b	CuT at 14k, -26dB	c30	3.771	0.900	G.722.2 at 8.85k, -26dB	c19	3.490	0.894	0.130	0.281	NWT	-0.214	Pass	-0.304	Pass
1b	CuT at 24k, -26dB	c31	4.063	0.805	G.722 at 48k, -26dB	c20	3.302	0.919	0.125	0.760	NWT	-0.206	Pass	-0.293	Pass
1b	CuT at 32k, -26dB	c32	4.063	0.723	G.722 at 56k, -26dB	c21	3.771	0.852	0.114	0.292	NWT	-0.188	Pass	-0.268	Pass
1b	CuT at 14k, -16dB	c33	3.698	0.908	G.729A at 8k, -16dB	c22	2.656	0.856	0.127	1.042	BT	0.210	Pass	0.299	Pass
1b	CuT at 14k, -16dB	c33	3.698	0.908	G.722.2 at 8.85k, -16dB	c23	3.510	0.929	0.133	0.188	NWT	-0.219	Pass	-0.311	Pass
1b	CuT at 24k, -16dB	c34	3.698	0.872	G.722 at 48k, -16dB	c24	3.479	0.917	0.129	0.219	NWT	-0.214	Pass	-0.303	Pass
1b	CuT at 32k, -16dB	c35	3.906	0.809	G.722 at 56k, -16dB	c25	4.021	0.906	0.124	-0.115	NWT	-0.205	Pass	-0.291	Pass
1b	CuT at 14k, -36dB	c36	3.750	0.858	G.729A at 8k, -36dB	c26	2.635	0.713	0.114	1.115	BT	0.188	Pass	0.267	Pass
1b	CuT at 14k, -36dB	c36	3.750	0.858	G.722.2 at 8.85k, -36dB	c27	3.146	0.882	0.126	0.604	NWT	-0.208	Pass	-0.295	Pass
1b	CuT at 24k, -36dB	c37	3.865	0.790	G.722 at 48k, -36dB	c28	2.948	0.933	0.125	0.917	NWT	-0.206	Pass	-0.293	Pass
1b	CuT at 32k, -36dB	c38	3.927	0.771	G.722 at 56k, -36dB	c29	3.250	0.871	0.119	0.677	NWT	-0.196	Pass	-0.278	Pass
1b	CuT at 24k, 1% FER	c39	3.760	0.830	G.722 at 48k, 0% FER	c20	3.302	0.919	0.126	0.458	NWT	-0.209	Pass	-0.297	Pass
1b	CuT at 32k, 1% FER	c40	3.844	0.862	G.722 at 56k, 0% FER	c21	3.771	0.852	0.124	0.073	NWT	-0.205	Pass	-0.290	Pass
2a	CuT at 8k, Music Bkgr	c07	3.781	0.931	G.729A at 8k, Music Bkgr	c06	3.719	0.867	0.130	0.063	NWT	-0.215	Pass	-0.305	Pass
2a	CuT at 12k, Music Bkgr	c08	4.240	0.750	G.729A at 8k, Music Bkgr	c06	3.719	0.867	0.117	0.521	BT	0.193	Pass	0.275	Pass
2b	CuT at 8k, Office Bkgr	c07	3.708	1.025	G.729A at 8k, Office Bkgr	c06	3.698	0.860	0.137	0.010	NWT	-0.226	Pass	-0.320	Pass
2b	CuT at 12k, Office Bkgr	c08	4.021	0.821	G.729A at 8k, Office Bkgr	c06	3.698	0.860	0.121	0.323	BT	0.201	Pass	0.285	Pass
2c	CuT at 8k, Babble Bkgr	c07	4.063	0.805	G.729A at 8k, Babble Bkgr	c06	3.979	0.808	0.116	0.083	NWT	-0.192	Pass	-0.273	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.260	0.771	G.729A at 8k, Babble Bkgr	c06	3.979	0.808	0.114	0.281	BT	0.188	Pass	0.267	Pass
4	CuT at 32k, Music	c08	3.406	1.011	G.722 at 56k, Music	c07	3.552	1.113	0.154	-0.146	NWT	-0.254	Pass	-0.360	Pass

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	0	-	G.722 at 48k, Music Bkgr	c05	3	12.60	13.485	-12.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	1	-	G.722 at 56k, Music Bkgr	c06	0	9.60	7.385	-8.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	1	-	G.722 at 48k, Office Bkgr	c05	3	12.60	10.648	-11.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	0	-	G.722 at 56k, Office Bkgr	c06	0	9.60	10.105	-9.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 24k, Babble Bkgr	c07	0	-	G.722 at 48k, Babble Bkgr	c05	3	12.60	13.485	-12.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 32k, Babble Bkgr	c08	0	-	G.722 at 56k, Babble Bkgr	c06	0	9.60	10.105	-9.6	NMT10%	2.706	Pass	5.412	Pass

Table 57: Candidate D - Crosscheck

Req.	Coder under Test Candidate D-Xchk				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	3.927	0.714	G.729A at 8k, -26dB	c08	3.438	0.723	0.104	0.490	NWT	-0.171	Pass	-0.243	Pass
1a	CuT at 12k, -26dB	c17	4.323	0.673	G.729E, -26dB	c09	4.052	0.655	0.096	0.271	NWT	-0.158	Pass	-0.225	Pass
1a	CuT at 8k, -16dB	c18	4.115	0.832	G.729A at 8k, -16dB	c10	3.698	0.822	0.119	0.417	NWT	-0.197	Pass	-0.280	Pass
1a	CuT at 12k, -16dB	c19	4.354	0.846	G.729E, -16dB	c11	4.292	0.664	0.110	0.063	NWT	-0.181	Pass	-0.257	Pass
1a	CuT at 8k, -36dB	c20	3.427	0.764	G.729A at 8k, -36dB	c12	2.750	0.616	0.100	0.677	NWT	-0.166	Pass	-0.235	Pass
1a	CuT at 12k, -36dB	c21	3.729	0.774	G.729E, -36dB	c13	3.198	0.734	0.109	0.531	NWT	-0.180	Pass	-0.256	Pass
1a	CuT at 8k, 3% FER	c22	3.406	0.901	G.729A at 8k, 3% FER	c14	2.833	0.902	0.130	0.573	NWT	-0.215	Pass	-0.305	Pass
1a	CuT at 12k, 3%FER	c23	3.656	0.856	G.729A at 8k, 3% FER	c14	2.833	0.902	0.127	0.823	BT	0.210	Pass	0.298	Pass
1b	CuT at 14k, -26dB	c30	4.042	0.739	G.729A at 8k, -26dB	c18	3.281	0.804	0.111	0.760	BT	0.184	Pass	0.261	Pass
1b	CuT at 14k, -26dB	c30	4.042	0.739	G.722.2 at 8.85k, -26dB	c19	3.938	0.856	0.115	0.104	NWT	-0.191	Pass	-0.271	Pass
1b	CuT at 24k, -26dB	c31	4.271	0.788	G.722 at 48k, -26dB	c20	3.760	0.818	0.116	0.510	NWT	-0.192	Pass	-0.272	Pass
1b	CuT at 32k, -26dB	c32	4.344	0.708	G.722 at 56k, -26dB	c21	4.073	0.798	0.109	0.271	NWT	-0.180	Pass	-0.255	Pass
1b	CuT at 14k, -16dB	c33	4.135	0.643	G.729A at 8k, -16dB	c22	3.344	0.693	0.096	0.792	BT	0.159	Pass	0.226	Pass
1b	CuT at 14k, -16dB	c33	4.135	0.643	G.722.2 at 8.85k, -16dB	c23	4.042	0.710	0.098	0.094	NWT	-0.162	Pass	-0.229	Pass
1b	CuT at 24k, -16dB	c34	4.177	0.754	G.722 at 48k, -16dB	c24	3.844	0.910	0.121	0.333	NWT	-0.199	Pass	-0.283	Pass
1b	CuT at 32k, -16dB	c35	4.177	0.858	G.722 at 56k, -16dB	c25	4.219	0.757	0.117	-0.042	NWT	-0.193	Pass	-0.274	Pass
1b	CuT at 14k, -36dB	c36	4.063	0.805	G.729A at 8k, -36dB	c26	3.208	0.951	0.127	0.854	BT	0.210	Pass	0.298	Pass
1b	CuT at 14k, -36dB	c36	4.063	0.805	G.722.2 at 8.85k, -36dB	c27	3.594	0.802	0.116	0.469	NWT	-0.192	Pass	-0.272	Pass
1b	CuT at 24k, -36dB	c37	4.229	0.827	G.722 at 48k, -36dB	c28	3.354	0.962	0.129	0.875	NWT	-0.214	Pass	-0.304	Pass
1b	CuT at 32k, -36dB	c38	4.260	0.757	G.722 at 56k, -36dB	c29	3.531	0.994	0.128	0.729	NWT	-0.211	Pass	-0.299	Pass
1b	CuT at 24k, 1% FER	c39	4.063	0.844	G.722 at 48k, 0% FER	c20	3.760	0.818	0.120	0.302	NWT	-0.198	Pass	-0.281	Pass
1b	CuT at 32k, 1% FER	c40	3.990	0.814	G.722 at 56k, 0% FER	c21	4.073	0.798	0.116	-0.083	NWT	-0.192	Pass	-0.273	Pass
2c	CuT at 8k, Babble Bkgr	c07	4.427	0.750	G.729A at 8k, Babble Bkgr	c06	4.469	0.664	0.102	-0.042	NWT	-0.169	Pass	-0.240	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.656	0.540	G.729A at 8k, Babble Bkgr	c06	4.469	0.664	0.087	0.188	BT	0.144	Pass	0.205	Fail
4	CuT at 32k, Music	c08	3.875	0.954	G.722 at 56k, Music	c07	3.594	1.101	0.149	0.281	NWT	-0.246	Pass	-0.349	Pass

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	0	-	G.722 at 48k, Music Bkgr	c05	3	12.6	13.485	-12.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	0	-	G.722 at 56k, Music Bkgr	c06	2	11.6	12.346	-11.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	0	-	G.722 at 48k, Office Bkgr	c05	1	10.6	11.219	-10.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	0	-	G.722 at 56k, Office Bkgr	c06	2	11.6	12.346	-11.6	NMT10%	2.706	Pass	5.412	Pass

Table 58: Candidate E

Req.	Coder under Test Candidate E				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MP}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	4.531	0.695	G.729A at 8k, -26dB	c08	4.344	0.737	0.103	0.188	NWT	-0.171	Pass	-0.243	Pass
1a	CuT at 12k, -26dB	c17	4.583	0.592	G.729E, -26dB	c09	4.583	0.627	0.088	0.000	NWT	-0.145	Pass	-0.207	Pass
1a	CuT at 8k, -16dB	c18	4.563	0.662	G.729A at 8k, -16dB	c10	4.354	0.767	0.103	0.208	NWT	-0.171	Pass	-0.243	Pass
1a	CuT at 12k, -16dB	c19	4.563	0.558	G.729E, -16dB	c11	4.458	0.710	0.092	0.104	NWT	-0.152	Pass	-0.216	Pass
1a	CuT at 8k, -36dB	c20	4.177	0.768	G.729A at 8k, -36dB	c12	3.760	0.830	0.115	0.417	NWT	-0.191	Pass	-0.271	Pass
1a	CuT at 12k, -36dB	c21	4.542	0.695	G.729E, -36dB	c13	3.958	0.893	0.116	0.583	NWT	-0.191	Pass	-0.271	Pass
1a	CuT at 8k, 3% FER	c22	3.896	0.900	G.729A at 8k, 3% FER	c14	3.688	0.966	0.135	0.208	NWT	-0.223	Pass	-0.316	Pass
1a	CuT at 12k, 3%FER	c23	4.063	0.868	G.729A at 8k, 3% FER	c14	3.688	0.966	0.133	0.375	BT	0.219	Pass	0.311	Pass
1b	CuT at 14k, -26dB	c30	4.448	0.630	G.729A at 8k, -26dB	c18	4.448	0.663	0.093	0.000	BT	0.154	Fail	0.219	Fail
1b	CuT at 14k, -26dB	c30	4.448	0.630	G.722.2 at 8.85k, -26dB	c19	4.083	0.721	0.098	0.365	NWT	-0.162	Pass	-0.229	Pass
1b	CuT at 24k, -26dB	c31	4.573	0.611	G.722 at 48k, -26dB	c20	3.323	0.733	0.097	1.250	NWT	-0.161	Pass	-0.228	Pass
1b	CuT at 32k, -26dB	c32	4.479	0.680	G.722 at 56k, -26dB	c21	3.552	0.663	0.097	0.927	NWT	-0.160	Pass	-0.227	Pass
1b	CuT at 14k, -16dB	c33	4.479	0.632	G.729A at 8k, -16dB	c22	4.458	0.695	0.096	0.021	BT	0.158	Fail	0.225	Fail
1b	CuT at 14k, -16dB	c33	4.479	0.632	G.722.2 at 8.85k, -16dB	c23	4.188	0.744	0.100	0.292	NWT	-0.165	Pass	-0.234	Pass
1b	CuT at 24k, -16dB	c34	4.510	0.615	G.722 at 48k, -16dB	c24	3.677	0.747	0.099	0.833	NWT	-0.163	Pass	-0.232	Pass
1b	CuT at 32k, -16dB	c35	4.479	0.632	G.722 at 56k, -16dB	c25	4.010	0.775	0.102	0.469	NWT	-0.169	Pass	-0.239	Pass
1b	CuT at 14k, -36dB	c36	4.521	0.665	G.729A at 8k, -36dB	c26	4.063	0.779	0.104	0.458	BT	0.173	Pass	0.245	Pass
1b	CuT at 14k, -36dB	c36	4.521	0.665	G.722.2 at 8.85k, -36dB	c27	3.240	0.677	0.097	1.281	NWT	-0.160	Pass	-0.227	Pass
1b	CuT at 24k, -36dB	c37	4.552	0.630	G.722 at 48k, -36dB	c28	2.844	0.686	0.095	1.708	NWT	-0.157	Pass	-0.223	Pass
1b	CuT at 32k, -36dB	c38	4.385	0.655	G.722 at 56k, -36dB	c29	2.990	0.688	0.097	1.396	NWT	-0.160	Pass	-0.227	Pass
1b	CuT at 24k, 1% FER	c39	4.479	0.680	G.722 at 48k, 0% FER	c20	3.323	0.733	0.102	1.156	NWT	-0.169	Pass	-0.239	Pass
1b	CuT at 32k, 1% FER	c40	4.271	0.732	G.722 at 56k, 0% FER	c21	3.552	0.663	0.101	0.719	NWT	-0.167	Pass	-0.237	Pass
2a	CuT at 8k, Music Bkgr	c07	4.656	0.577	G.729A at 8k, Music Bkgr	c06	4.552	0.630	0.087	0.104	NWT	-0.144	Pass	-0.205	Pass
2a	CuT at 12k, Music Bkgr	c08	4.771	0.470	G.729A at 8k, Music Bkgr	c06	4.552	0.630	0.080	0.219	BT	0.133	Pass	0.188	Pass
2b	CuT at 8k, Office Bkgr	c07	4.521	0.680	G.729A at 8k, Office Bkgr	c06	4.583	0.556	0.090	-0.063	NWT	-0.148	Pass	-0.210	Pass
2b	CuT at 12k, Office Bkgr	c08	4.760	0.497	G.729A at 8k, Office Bkgr	c06	4.583	0.556	0.076	0.177	BT	0.126	Pass	0.179	Fail
2c	CuT at 8k, Babble Bkgr	c07	4.771	0.470	G.729A at 8k, Babble Bkgr	c06	4.677	0.533	0.073	0.094	NWT	-0.120	Pass	-0.170	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.823	0.384	G.729A at 8k, Babble Bkgr	c06	4.677	0.533	0.067	0.146	BT	0.111	Pass	0.157	Fail
4	CuT at 32k, Music	c08	4.021	1.056	G.722 at 56k, Music	c07	3.240	0.992	0.148	0.781	NWT	-0.244	Pass	-0.347	Pass

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	2	-	G.722 at 48k, Music Bkgr	c05	3	12.60	8.329	-10.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	0	-	G.722 at 56k, Music Bkgr	c06	6	15.60	16.980	-15.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	5	-	G.722 at 48k, Office Bkgr	c05	8	17.60	7.962	-12.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	0	-	G.722 at 56k, Office Bkgr	c06	5	14.60	15.802	-14.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 24k, Babble Bkgr	c07	2	-	G.722 at 48k, Babble Bkgr	c05	6	15.60	11.570	-13.6	NMT10%	2.706	Pass	5.412	Pass
3c	CuT at 32k, Babble Bkgr	c08	0	-	G.722 at 56k, Babble Bkgr	c06	1	10.60	11.219	-10.6	NMT10%	2.706	Pass	5.412	Pass

Table 59: Candidate E - Crosscheck

Req.	Coder under Test Candidate E-Xchk				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	Mean	StdDev	Reference cond.	c #	Mean	StdDev	SE _{MD}	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
1a	CuT at 8k, -26dB	c16	3.719	0.903	G.729A at 8k, -26dB	c08	3.458	0.893	0.130	0.260	NWT	-0.214	Pass	-0.304	Pass
1a	CuT at 12k, -26dB	c17	3.708	0.807	G.729E, -26dB	c09	3.917	0.790	0.115	-0.208	NWT	-0.191	Fail	-0.270	Pass
1a	CuT at 8k, -16dB	c18	3.781	0.861	G.729A at 8k, -16dB	c10	3.646	0.808	0.120	0.135	NWT	-0.199	Pass	-0.283	Pass
1a	CuT at 12k, -16dB	c19	3.844	0.825	G.729E, -16dB	c11	3.833	0.902	0.125	0.010	NWT	-0.206	Pass	-0.293	Pass
1a	CuT at 8k, -36dB	c20	3.552	0.766	G.729A at 8k, -36dB	c12	2.813	0.898	0.120	0.740	NWT	-0.199	Pass	-0.283	Pass
1a	CuT at 12k, -36dB	c21	3.844	0.786	G.729E, -36dB	c13	3.438	0.904	0.122	0.406	NWT	-0.202	Pass	-0.287	Pass
1a	CuT at 8k, 3% FER	c22	3.125	0.897	G.729A at 8k, 3% FER	c14	2.875	0.909	0.130	0.250	NWT	-0.215	Pass	-0.306	Pass
1a	CuT at 12k, 3%FER	c23	3.313	0.933	G.729A at 8k, 3% FER	c14	2.875	0.909	0.133	0.438	BT	0.220	Pass	0.312	Pass
1b	CuT at 14k, -26dB	c30	4.198	0.763	G.729A at 8k, -26dB	c18	2.896	0.761	0.110	1.302	BT	0.182	Pass	0.258	Pass
1b	CuT at 14k, -26dB	c30	4.198	0.763	G.722.2 at 8.85k, -26dB	c19	3.448	0.724	0.107	0.750	NWT	-0.177	Pass	-0.252	Pass
1b	CuT at 24k, -26dB	c31	4.281	0.610	G.722 at 48k, -26dB	c20	3.188	0.715	0.096	1.094	NWT	-0.159	Pass	-0.225	Pass
1b	CuT at 32k, -26dB	c32	4.375	0.653	G.722 at 56k, -26dB	c21	3.323	0.688	0.097	1.052	NWT	-0.160	Pass	-0.227	Pass
1b	CuT at 14k, -16dB	c33	4.063	0.737	G.729A at 8k, -16dB	c22	3.021	0.882	0.117	1.042	BT	0.194	Pass	0.275	Pass
1b	CuT at 14k, -16dB	c33	4.063	0.737	G.722.2 at 8.85k, -16dB	c23	3.938	0.693	0.103	0.125	NWT	-0.171	Pass	-0.242	Pass
1b	CuT at 24k, -16dB	c34	4.198	0.690	G.722 at 48k, -16dB	c24	3.740	0.729	0.102	0.458	NWT	-0.169	Pass	-0.240	Pass
1b	CuT at 32k, -16dB	c35	4.260	0.669	G.722 at 56k, -16dB	c25	4.021	0.725	0.101	0.240	NWT	-0.166	Pass	-0.236	Pass
1b	CuT at 14k, -36dB	c36	3.948	0.686	G.729A at 8k, -36dB	c26	2.479	0.649	0.096	1.469	BT	0.159	Pass	0.226	Pass
1b	CuT at 14k, -36dB	c36	3.948	0.686	G.722.2 at 8.85k, -36dB	c27	2.635	0.713	0.101	1.313	NWT	-0.167	Pass	-0.237	Pass
1b	CuT at 24k, -36dB	c37	4.052	0.686	G.722 at 48k, -36dB	c28	2.563	0.612	0.094	1.490	NWT	-0.155	Pass	-0.220	Pass
1b	CuT at 32k, -36dB	c38	4.031	0.672	G.722 at 56k, -36dB	c29	2.604	0.624	0.094	1.427	NWT	-0.155	Pass	-0.220	Pass
1b	CuT at 24k, 1% FER	c39	4.073	0.743	G.722 at 48k, 0% FER	c20	3.188	0.715	0.105	0.885	NWT	-0.174	Pass	-0.247	Pass
1b	CuT at 32k, 1% FER	c40	4.156	0.701	G.722 at 56k, 0% FER	c21	3.323	0.688	0.100	0.833	NWT	-0.166	Pass	-0.235	Pass
2c	CuT at 8k, Babble Bkgr	c07	4.458	0.614	G.729A at 8k, Babble Bkgr	c06	4.563	0.499	0.081	-0.104	NWT	-0.133	Pass	-0.189	Pass
2c	CuT at 12k, Babble Bkgr	c08	4.667	0.496	G.729A at 8k, Babble Bkgr	c06	4.563	0.499	0.072	0.104	BT	0.119	Fail	0.168	Fail
4	CuT at 32k, Music	c08	3.292	1.045	G.722 at 56k, Music	c07	3.865	0.890	0.140	-0.573	NWT	-0.232	Fail	-0.329	Fail

Req.	Coder under Test				Reference				Terms of Reference Test						
Exp.	Coder condition	c #	#(1,2)	-	Reference cond.	c #	#(1,2)	+10%	Chi.Sq.	Test-Ref	Test	95% cri	95% CI	99% cri	99% CI
3a	CuT at 24k, Music Bkgr	c07	1	-	G.722 at 48k, Music Bkgr	c05	1	10.6	8.456	-9.6	NMT10%	2.706	Pass	5.412	Pass
3a	CuT at 32k, Music Bkgr	c08	1	-	G.722 at 56k, Music Bkgr	c06	0	9.6	7.385	-8.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 24k, Office Bkgr	c07	0	-	G.722 at 48k, Office Bkgr	c05	1	10.6	11.219	-10.6	NMT10%	2.706	Pass	5.412	Pass
3b	CuT at 32k, Office Bkgr	c08	1	-	G.722 at 56k, Office Bkgr	c06	0	9.6	7.385	-8.6	NMT10%	2.706	Pass	5.412	Pass

9.2 Experiment 5 (Clean Speech; wide band case, Bit rate granularity)

The purpose of Experiment 5 was to evaluate whether the 2 kbit/s fine granularity of G.729EV candidate codecs brings graceful quality improvement from 14 kbit/s up to 32 kbit/s. It assesses the quality for wideband clean speech signals at 10 bit rates from 14 to 32 kbit/s with 2 kbit/s steps. SG12 agreed with the use of wideband extension to P.862 (WB-PESQ) for this experiment.

9.2.1 Test Organization

The individual WB-PESQ scores for 80 speech files (4 samples x 4 talkers x 5 languages) were computed, and then averaged to produce a single WB-PESQ score for a given bit-rate. The database is common to all candidates. The 5 languages are: Korean, Japanese, French, English, and German - with 2 male and 2 female talkers for each language and 4 samples per talker-.

The following companies kindly volunteered to process 16 clean speech files in their native languages: Samsung (Korean), Matsushita (Japanese), VoiceAge (French), Mindspeed (North American English), Siemens (German). Each company has performed the processing of its native language long file (concatenation of 16 files) with the 5 executables at bit rates from 14 to 32 kbit/s and then computed the WB-PESQ scores after individual files extraction. These results were provided with the corresponding database to the Q10/16 Rapporteur for cross-checking.

The filenames of speech samples are built on the following way: 5aLCGySz.cnn where

- L stands for the language (E: English, F: French, G: German, J: Japanese; K: Korean)
- C is the candidate designator: A, B, C, D or E
- G is gender of talker (i.e. F for female and M for male)
- y is the number 1 or 2
- S stands for sample and z is the sample number: 1, 2, 3 or 4

The factors for Experiment 5 can be found in Table 60 and the list of conditions in Table 61.

Table 60: Factors for Experiment 5

Main Codec Conditions		
Candidate Codecs	1	
Rates	10	14, 16, 18, 20, 22, 24, 26, 28, 30, & 32 kbit/s
RBERs	-	All conditions are error free
FERs	-	All conditions have no frame erasures
Input level	-	all at the nominal level: -26dB relative to OVL
Tandeming	-	no tandemed conditions
Noise		Clean speech
Input Characteristic	1	Band limited signals 50-7000 Hz
References		(flat input, nominal level, with no noise)
Direct	1	
Common Conditions		
Number of languages	5	English, French, German, Japanese, Korean
Number of talkers per language	4	2 male and 2 female
Speech samples per language	16	4 samples per talker
Quality assessment method	1	WB-PESQ

Table 61: Conditions for Experiment 5

Number	Test Condition	Input signal
1	Direct	Band limited to 50-7000Hz
2	CuT-14 k	Band limited to 50-7000Hz
3	CuT-16 k	Band limited to 50-7000Hz
4	CuT-18 k	Band limited to 50-7000Hz
5	CuT-20 k	Band limited to 50-7000Hz
6	CuT-22 k	Band limited to 50-7000Hz
7	CuT-24 k	Band limited to 50-7000Hz
8	CuT-26 k	Band limited to 50-7000Hz
9	CuT-28 k	Band limited to 50-7000Hz
10	CuT-30 k	Band limited to 50-7000Hz
11	CuT-32 k	Band limited to 50-7000Hz

9.2.2 Test Results – Experiment 5

The mean values (obtained by averaging the 80 individual scores for each bit rate) are plotted as a function of bit-rate in Figure 59 for the five candidate coders. The results for each individual language (obtained by averaging the 16 individual scores for each bit rate) are also shown in Figure 60, Figure 61, Figure 62, Figure 63, Figure 64, and Figure 64. The average WB-PESQ scores for the five databases and also the whole database (80 files) are given in Table 62, Table 63, Table 64, Table 65, and Table 66 for the five candidates respectively.

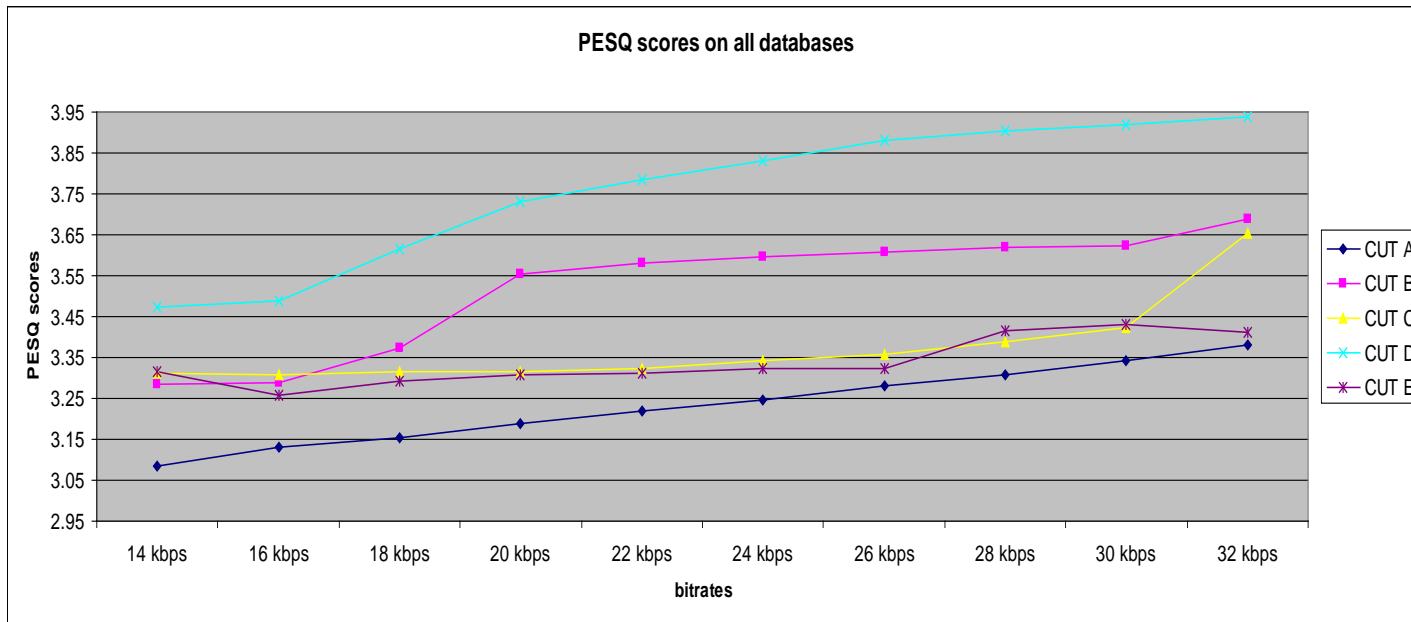


Figure 59 – WB-PESQ scores on all databases

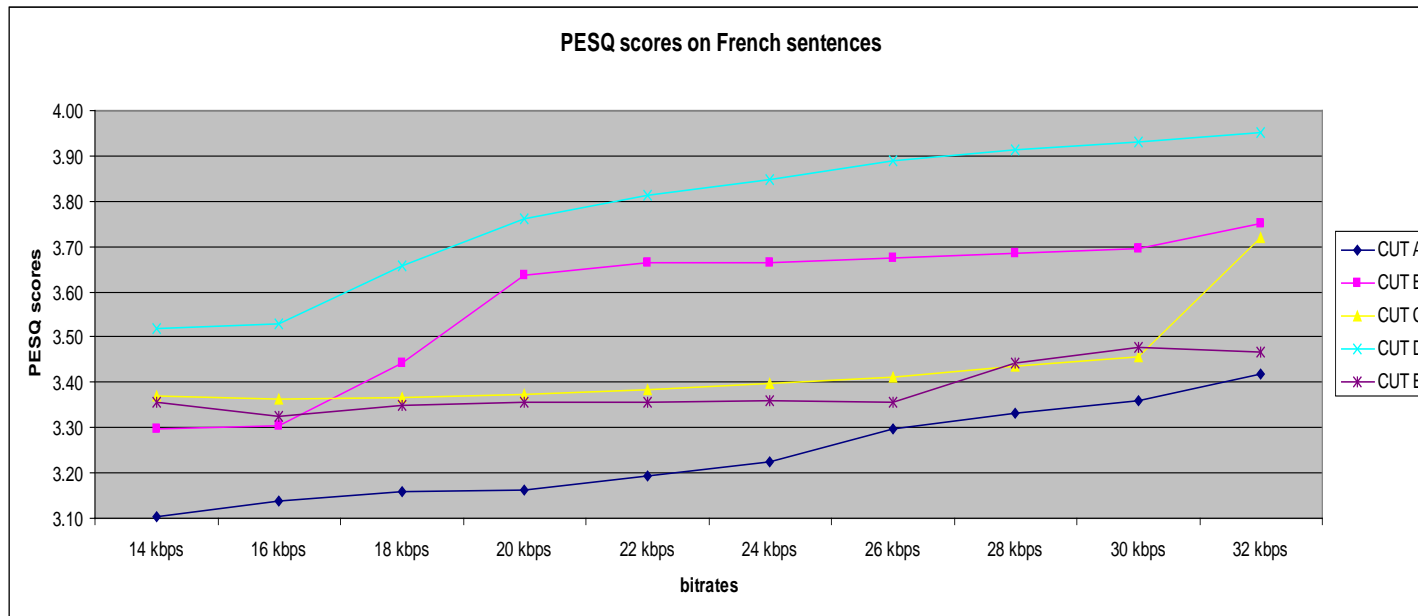


Figure 60 – WB-PESQ scores on French database

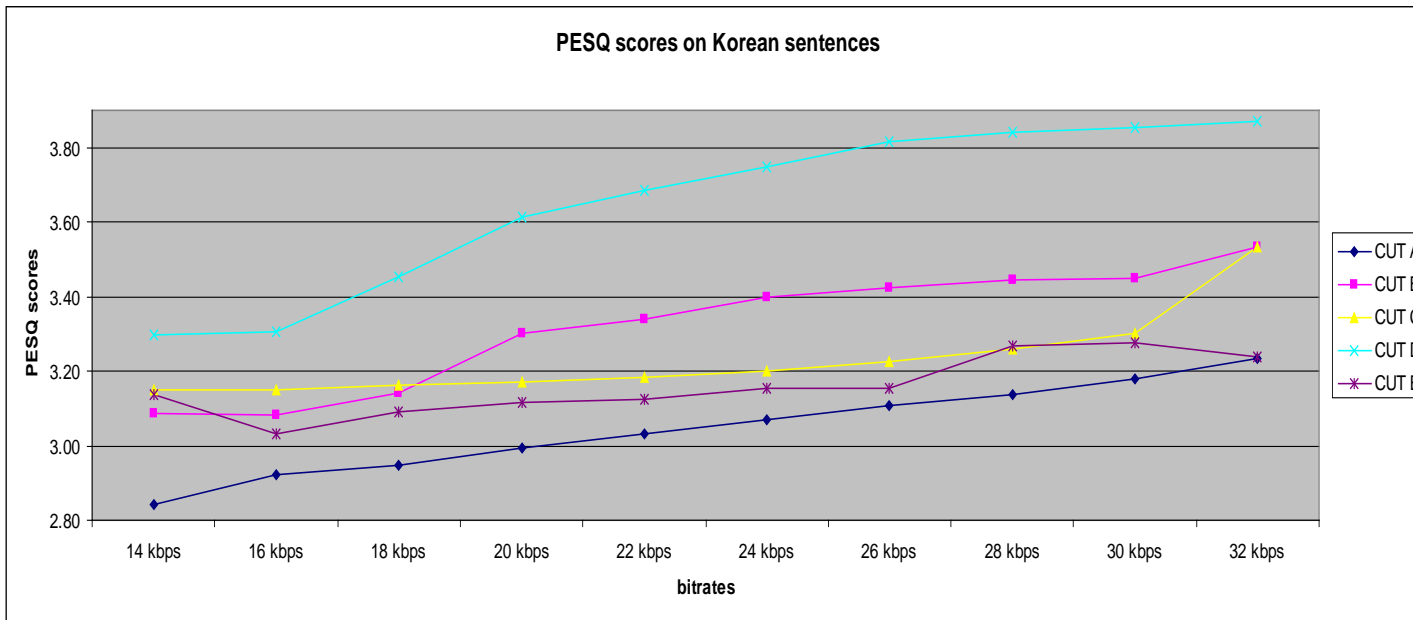


Figure 61 – WB-PESQ scores on Korean database

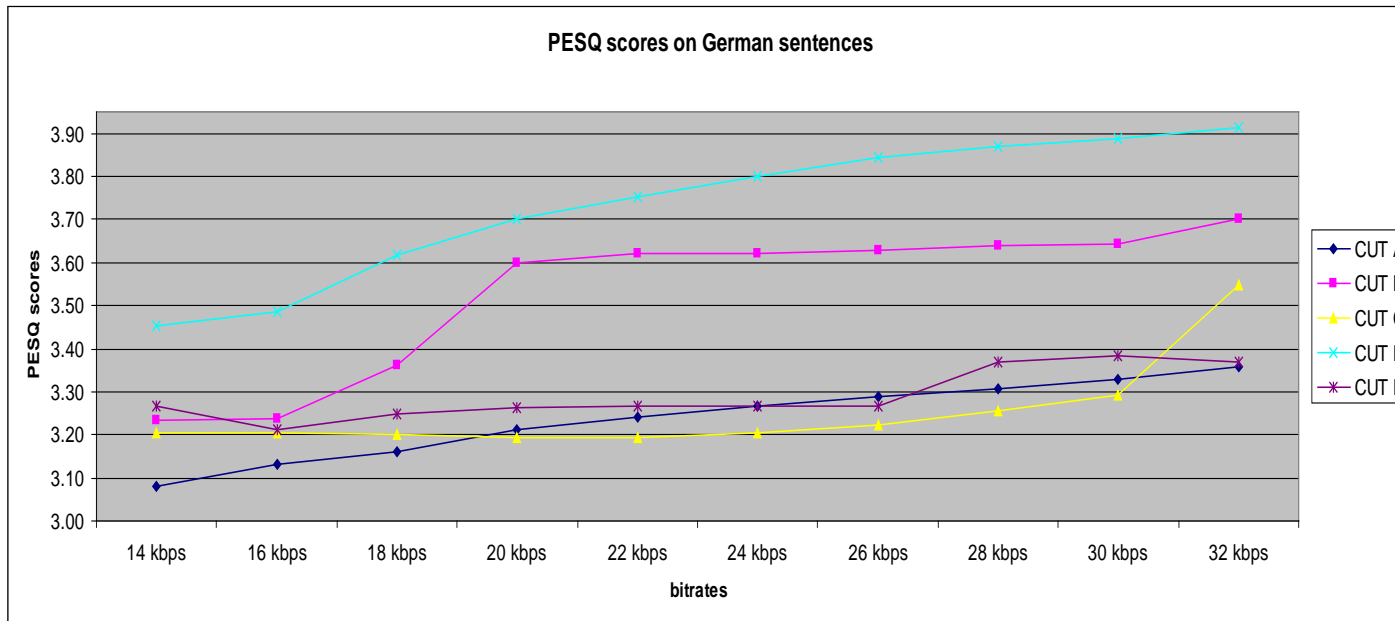


Figure 62 – WB-PESQ scores on German database

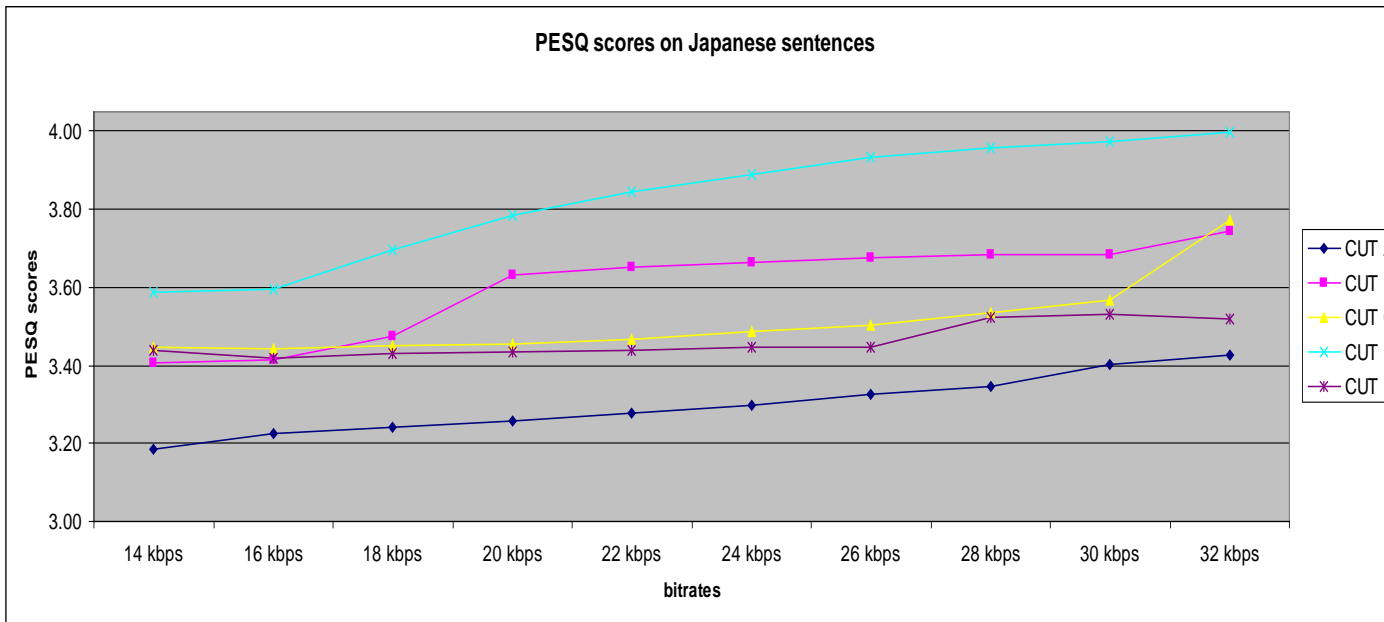


Figure 63 – WB-PESQ scores on Japanese database

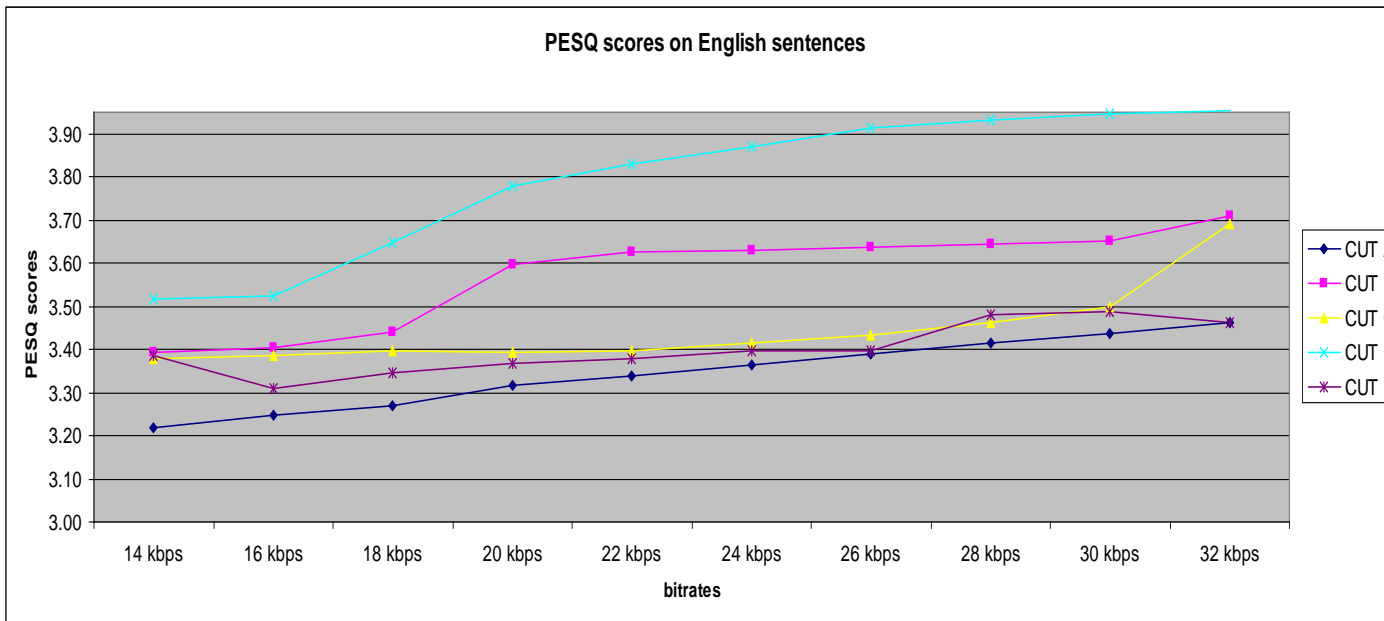


Figure 64 – WB-PESQ scores on North American English database

Table 62: Experiment 5 results for coder A (WB-PESQ scores)

Bit rates	French	Korean	German	Japanese	N-A English	All
14 kbit/s	3.10	2.84	3.08	3.18	3.22	3.09
16 kbit/s	3.14	2.92	3.13	3.22	3.25	3.13
18 kbit/s	3.16	2.95	3.16	3.24	3.27	3.15
20 kbit/s	3.16	2.99	3.21	3.26	3.32	3.19
22 kbit/s	3.19	3.03	3.24	3.28	3.34	3.22
24 kbit/s	3.22	3.07	3.27	3.30	3.36	3.24
26 kbit/s	3.30	3.11	3.29	3.33	3.39	3.28
28 kbit/s	3.33	3.14	3.31	3.34	3.42	3.31
30 kbit/s	3.36	3.18	3.33	3.40	3.44	3.34
32 kbit/s	3.42	3.23	3.36	3.43	3.46	3.38

Table 63: Experiment 5 results for coder B (WB-PESQ scores)

Bit rates	French	Korean	German	Japanese	N-A English	All
14 kbit/s	3.30	3.09	3.23	3.41	3.39	3.28
16 kbit/s	3.30	3.08	3.24	3.42	3.40	3.29
18 kbit/s	3.44	3.14	3.36	3.47	3.44	3.37
20 kbit/s	3.63	3.30	3.60	3.63	3.60	3.55
22 kbit/s	3.66	3.34	3.62	3.65	3.63	3.58
24 kbit/s	3.66	3.40	3.62	3.66	3.63	3.60
26 kbit/s	3.67	3.42	3.63	3.67	3.64	3.61
28 kbit/s	3.68	3.44	3.64	3.68	3.64	3.62
30 kbit/s	3.70	3.45	3.64	3.69	3.65	3.62
32 kbit/s	3.75	3.53	3.70	3.74	3.71	3.69

Table 64: Experiment 5 results for coder C (WB-PESQ scores)

Bit rates	French	Korean	German	Japanese	N-A English	All
14 kbit/s	3.37	3.15	3.21	3.45	3.38	3.31
16 kbit/s	3.36	3.15	3.21	3.44	3.39	3.31
18 kbit/s	3.37	3.16	3.20	3.45	3.40	3.32
20 kbit/s	3.37	3.17	3.19	3.45	3.39	3.32
22 kbit/s	3.38	3.18	3.19	3.47	3.40	3.32
24 kbit/s	3.40	3.20	3.20	3.49	3.41	3.34
26 kbit/s	3.41	3.22	3.22	3.50	3.43	3.36
28 kbit/s	3.43	3.26	3.26	3.53	3.46	3.39
30 kbit/s	3.45	3.30	3.29	3.57	3.50	3.42
32 kbit/s	3.72	3.53	3.55	3.77	3.69	3.65

Table 65: Experiment 5 results for coder D (WB-PESQ scores)

Bit rates	French	Korean	German	Japanese	N-A English	All s
14 kbit/s	3.52	3.30	3.45	3.59	3.52	3.47
16 kbit/s	3.53	3.31	3.49	3.60	3.53	3.49
18 kbit/s	3.66	3.45	3.62	3.69	3.65	3.61
20 kbit/s	3.76	3.62	3.70	3.79	3.78	3.73
22 kbit/s	3.81	3.68	3.75	3.85	3.83	3.79
24 kbit/s	3.85	3.75	3.80	3.89	3.87	3.83
26 kbit/s	3.89	3.82	3.84	3.93	3.91	3.88
28 kbit/s	3.91	3.84	3.87	3.96	3.93	3.90
30 kbit/s	3.93	3.86	3.89	3.97	3.95	3.92
32 kbit/s	3.95	3.87	3.91	4.00	3.95	3.94

Table 66: Experiment 5 results for coder E (WB-PESQ scores)

Bit rates	French	Korean	German	Japanese	N-A English	All
14 kbit/s	3.36	3.14	3.27	3.44	3.39	3.32
16 kbit/s	3.32	3.03	3.21	3.42	3.31	3.26
18 kbit/s	3.35	3.09	3.25	3.43	3.35	3.29
20 kbit/s	3.36	3.12	3.26	3.44	3.37	3.31
22 kbit/s	3.36	3.12	3.27	3.44	3.38	3.31
24 kbit/s	3.36	3.15	3.27	3.45	3.40	3.32
26 kbit/s	3.36	3.15	3.27	3.45	3.40	3.32
28 kbit/s	3.44	3.27	3.37	3.52	3.48	3.42
30 kbit/s	3.48	3.28	3.38	3.53	3.49	3.43
32 kbit/s	3.47	3.24	3.37	3.52	3.46	3.41

9.3 Frequency Responses of Candidates

At the January 2005 SG 12 plenary meeting, a method was agreed to measure the frequency response of a codec and to evaluate its effective bandwidth. The frequency responses were computed using STL2005 tool *freqresp*. This tool computes and outputs the average amplitude spectra in ASCII and also produces a bitmap file. As recommended by SG12, P.50 test signals which are representative of speech signals were used to compute the frequency response: P50m.16k for male speech and P50f.16k for female speech.

Blinding of executable was as follows:

- A: France Telecom
- B: ETRI
- C: VoiceAge
- D: Siemens Matsushita Mindspeed
- E: Samsung

The frequency response is in Figure 65 and Figure 66 for male and female speech, respectively.

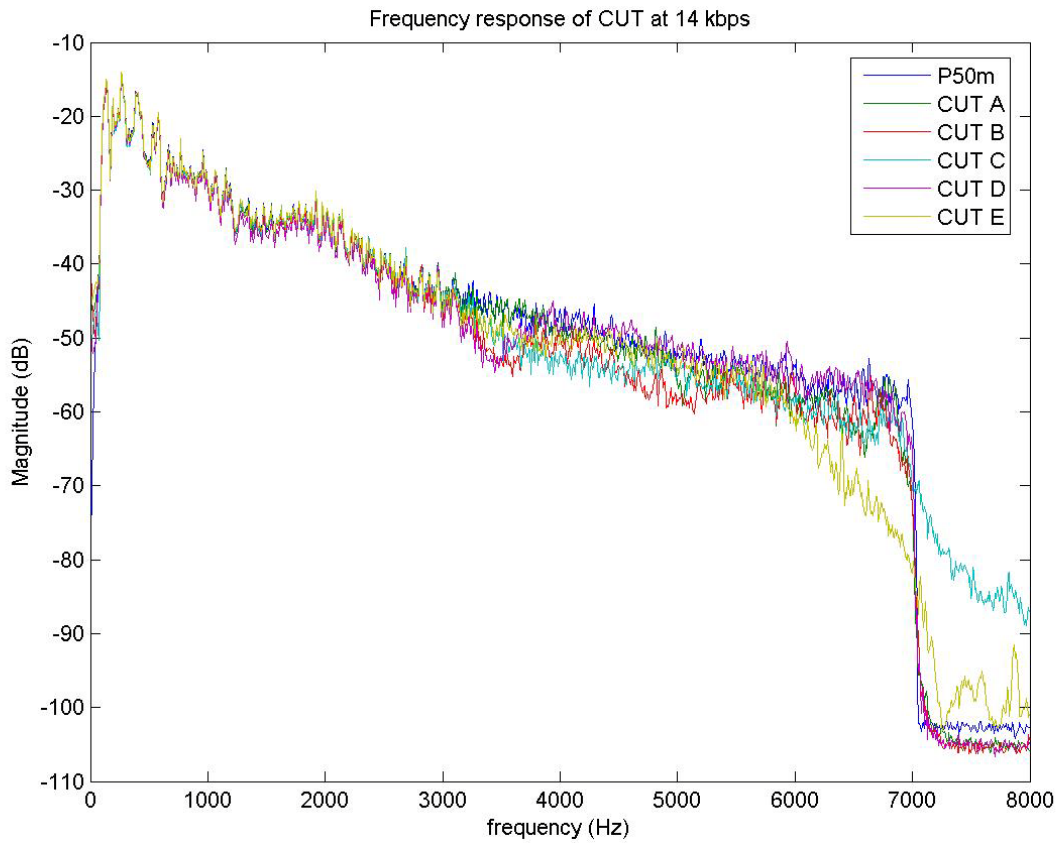


Figure 65(a) – Male speech (P50m.16k), 14 kbit/s

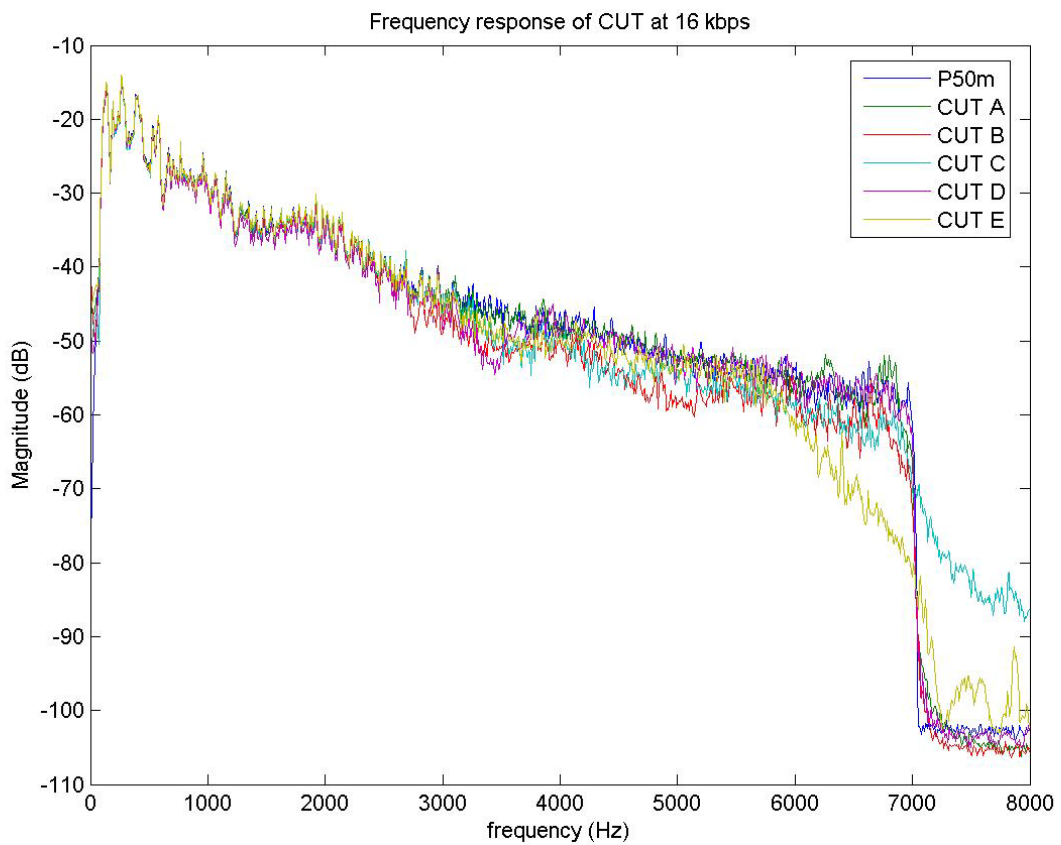


Figure 65(b) – Male speech (P50m.16k), 16 kbit/s

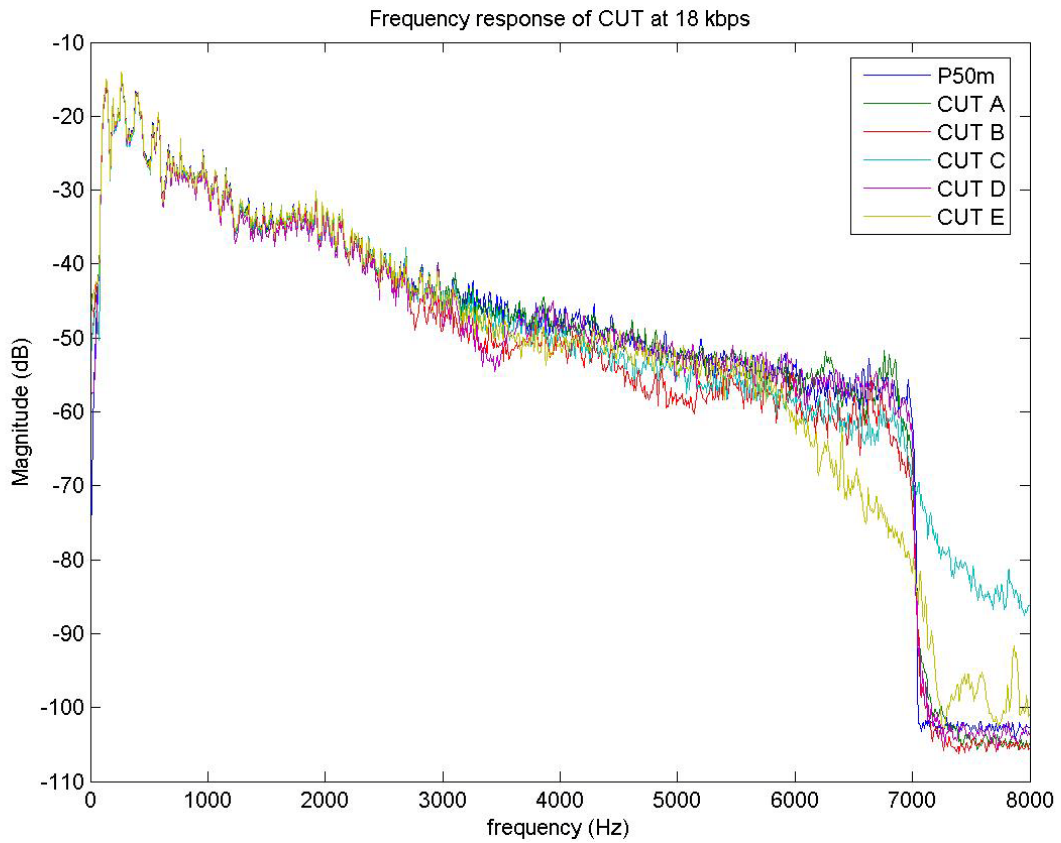


Figure 65(c) – Male speech (P50m.16k), 18 kbit/s

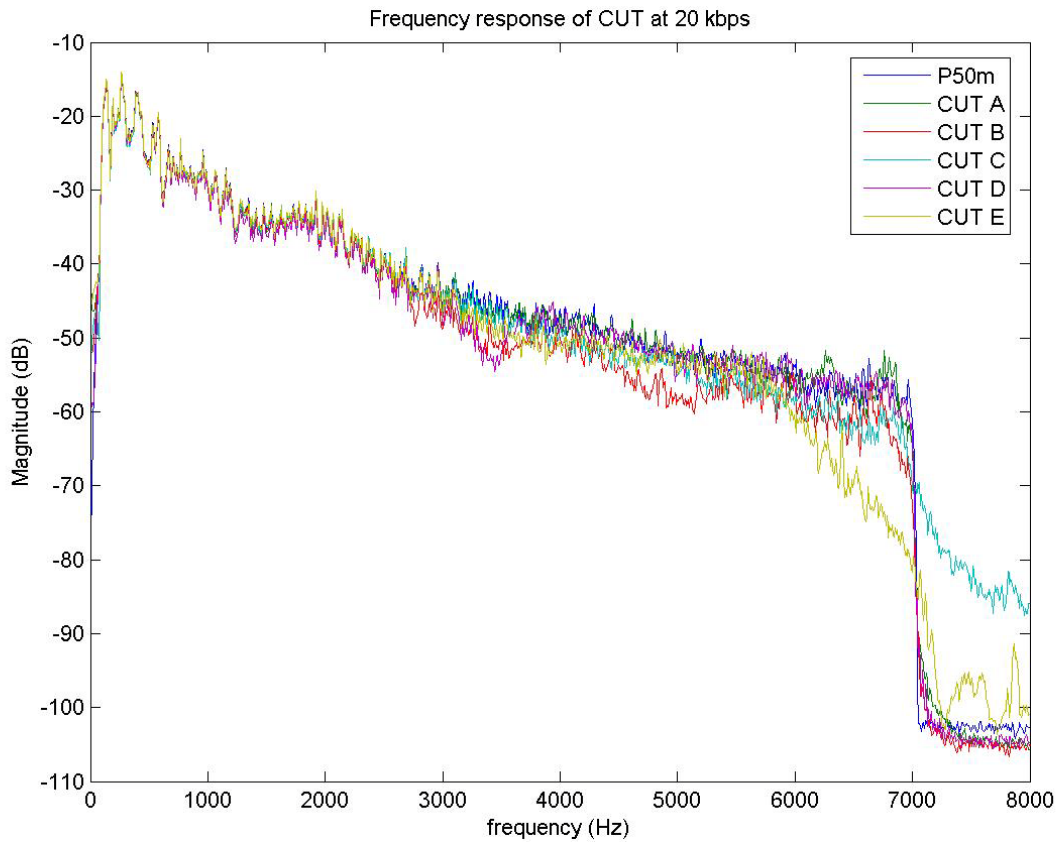


Figure 65(d) – Male speech (P50m.16k), 20 kbit/s

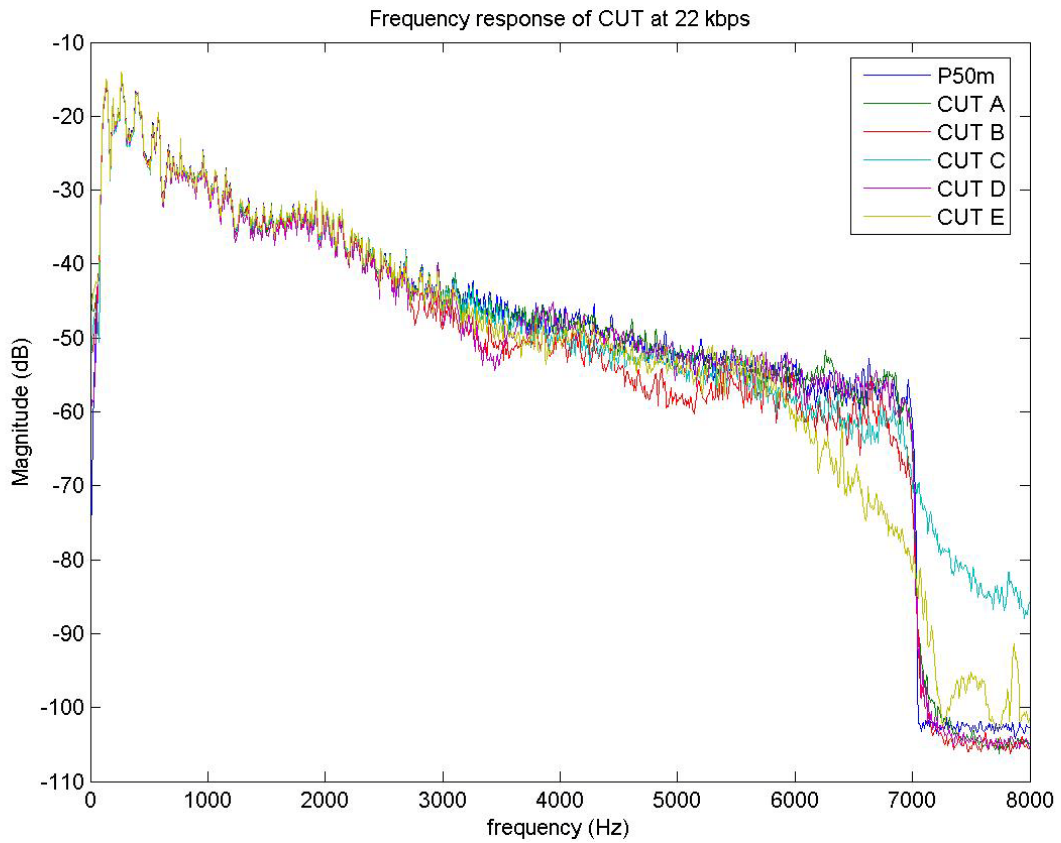


Figure 65(e) – Male speech (P50m.16k), 22 kbit/s

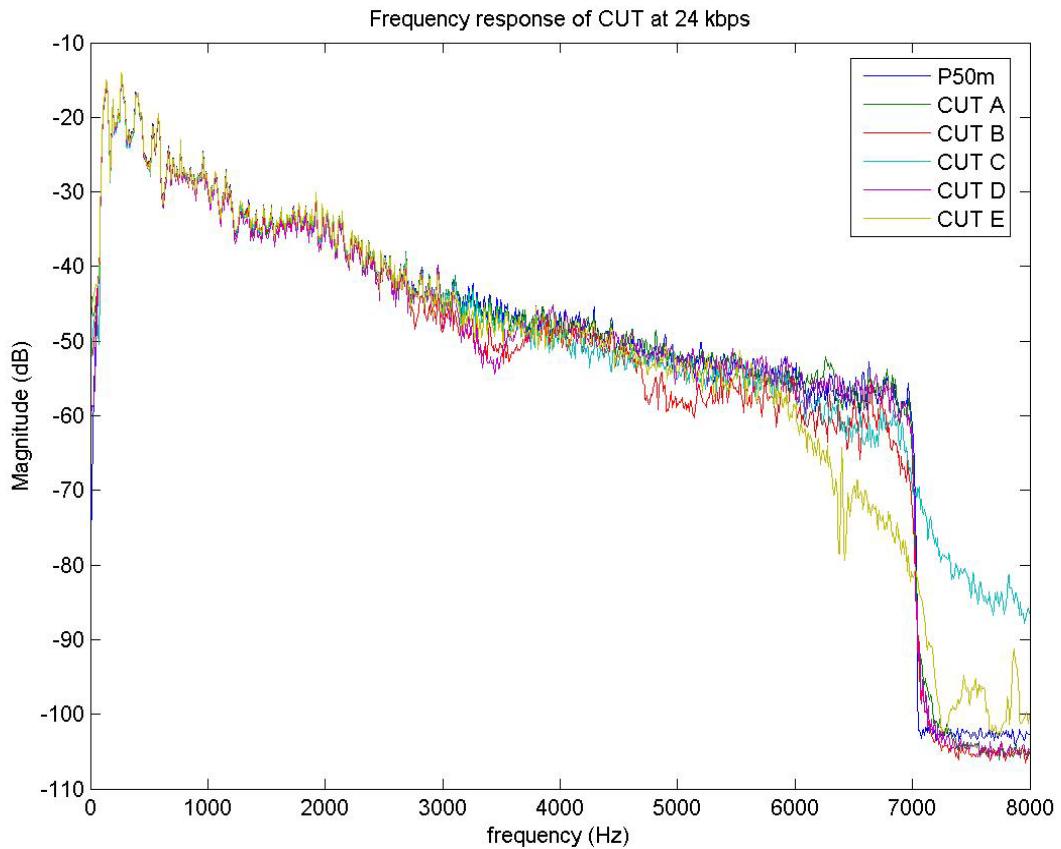


Figure 65(f) – Male speech (P50m.16k), 24 kbit/s

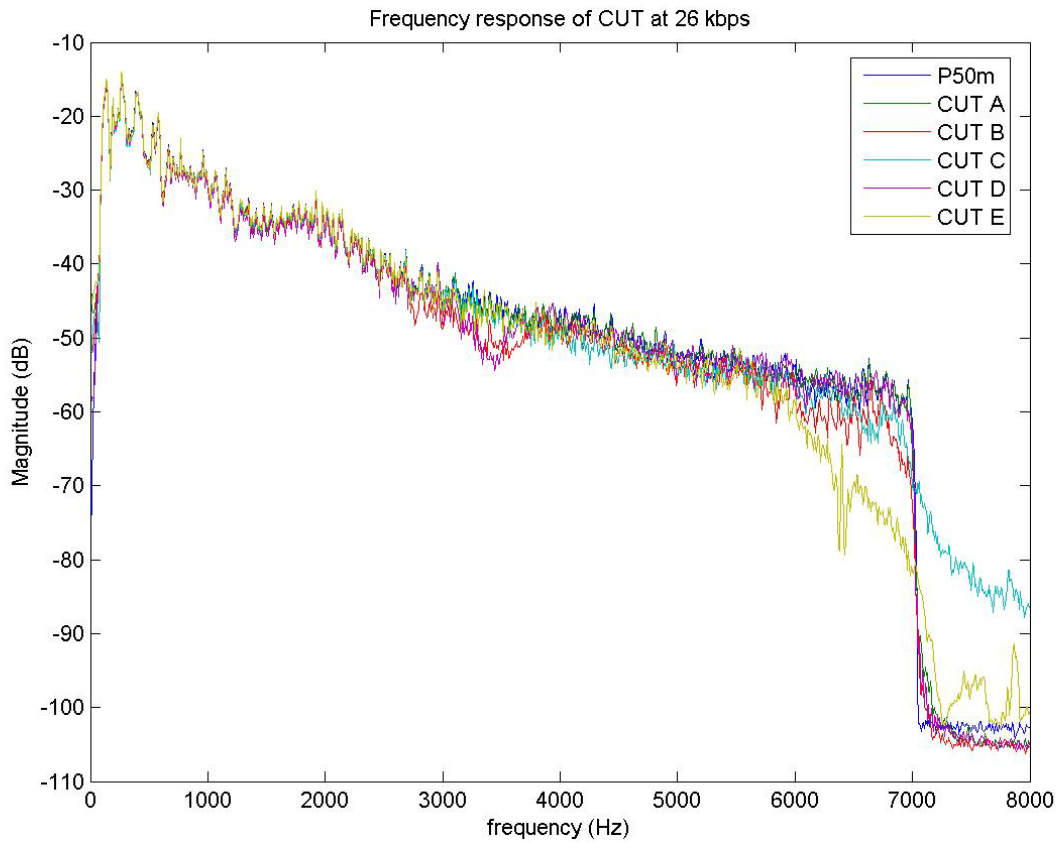


Figure 65(g) – Male speech (P50m.16k), 26 kbit/s

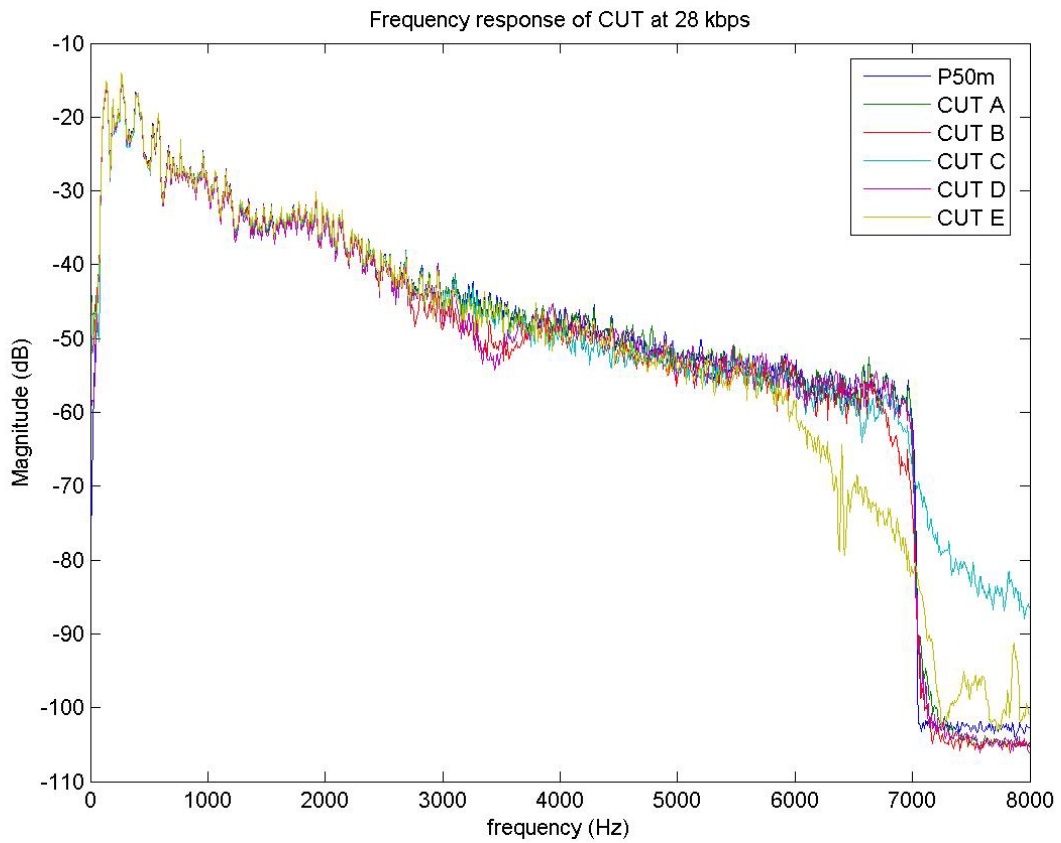


Figure 65(h) – Male speech (P50m.16k), 28 kbit/s

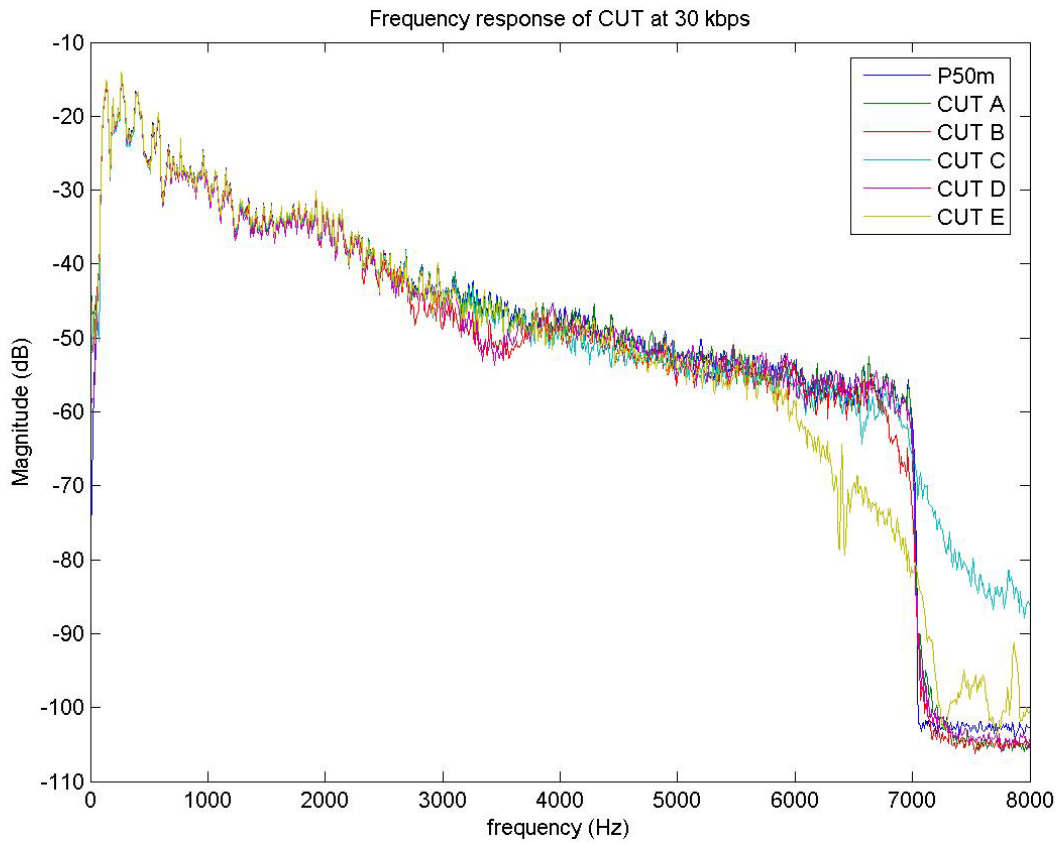


Figure 65(i) – Male speech (P50m.16k), 30 kbit/s

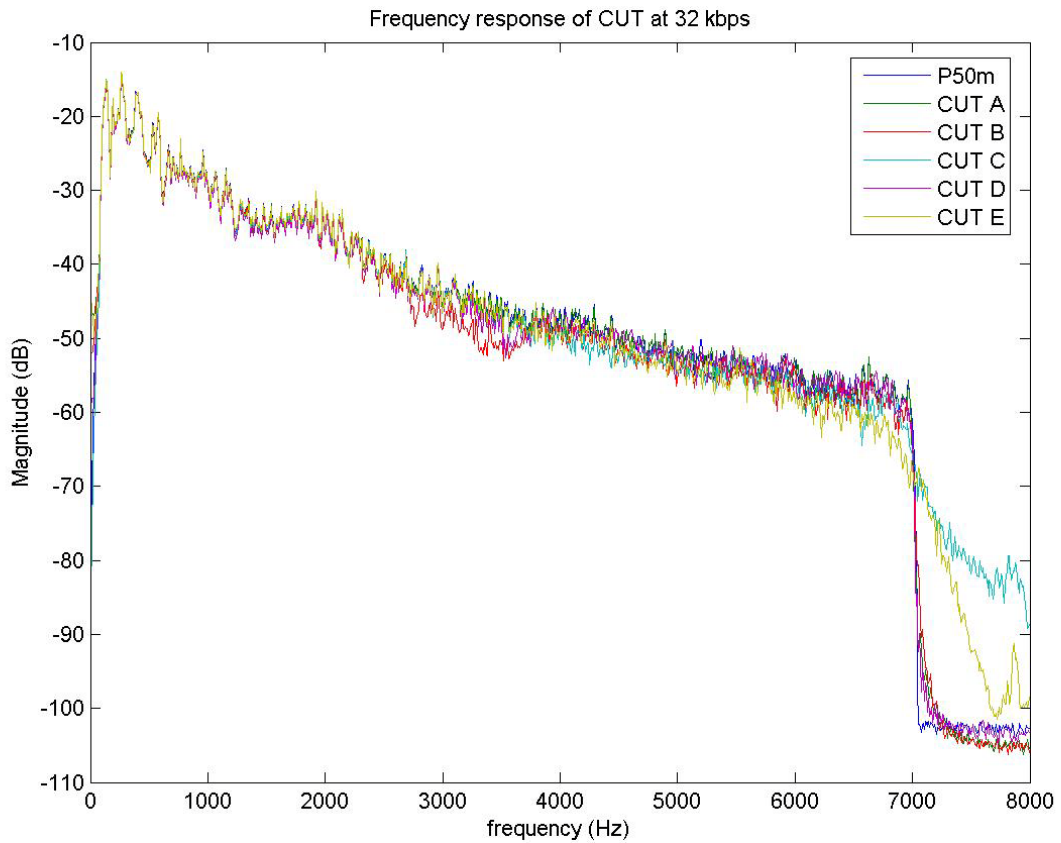


Figure 65(j) – Male speech (P50m.16k), 32 kbit/s

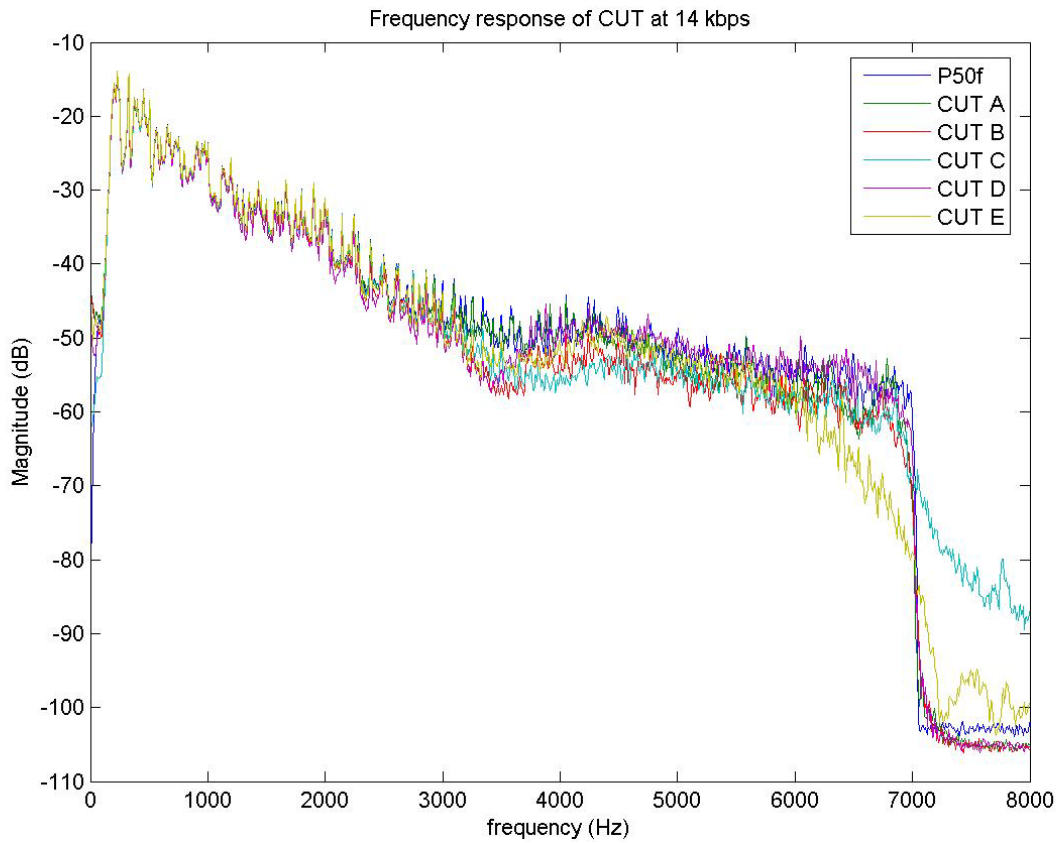


Figure 66(a) – Female speech, 14 kbit/s

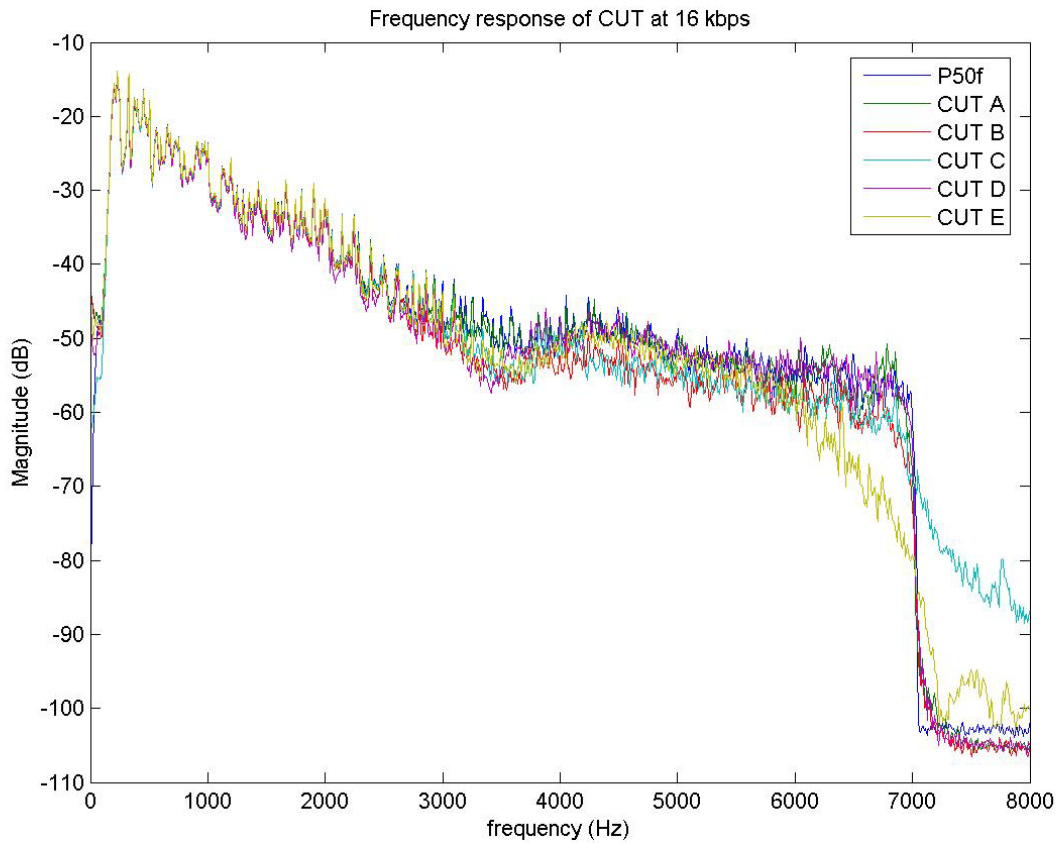


Figure 66(b) – Female speech, 16 kbit/s

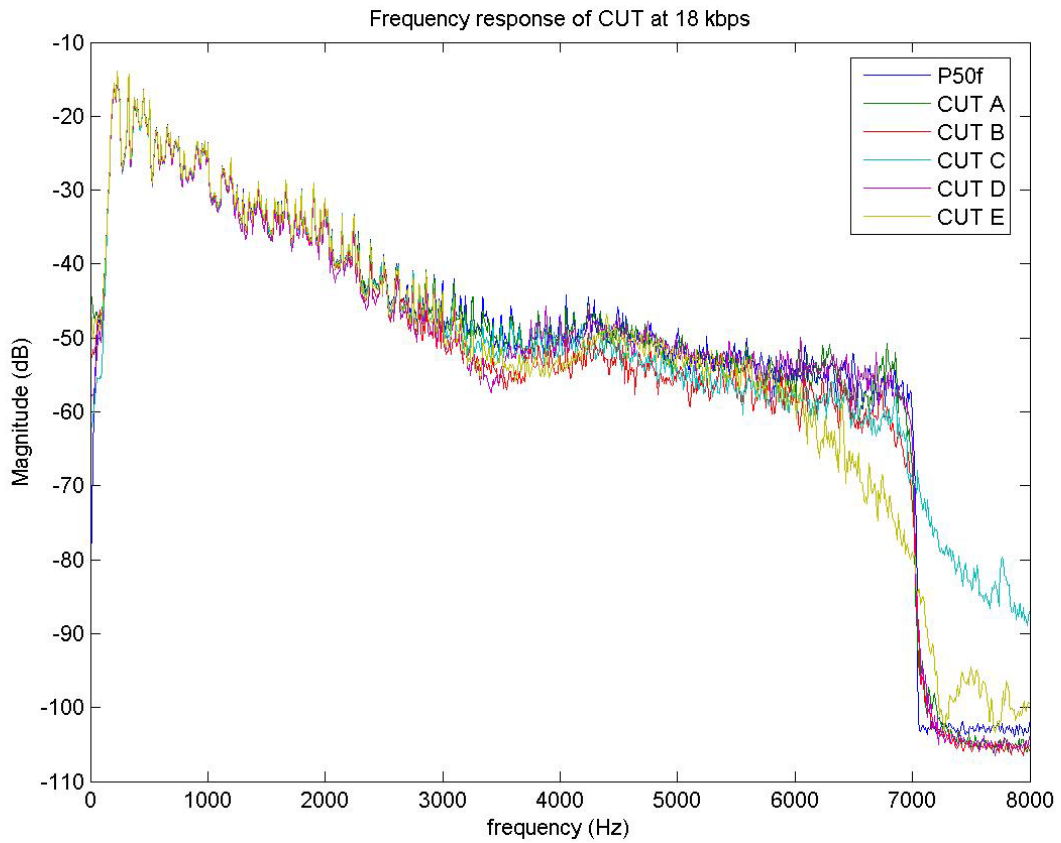


Figure 66(c) – Female speech, 18 kbit/s

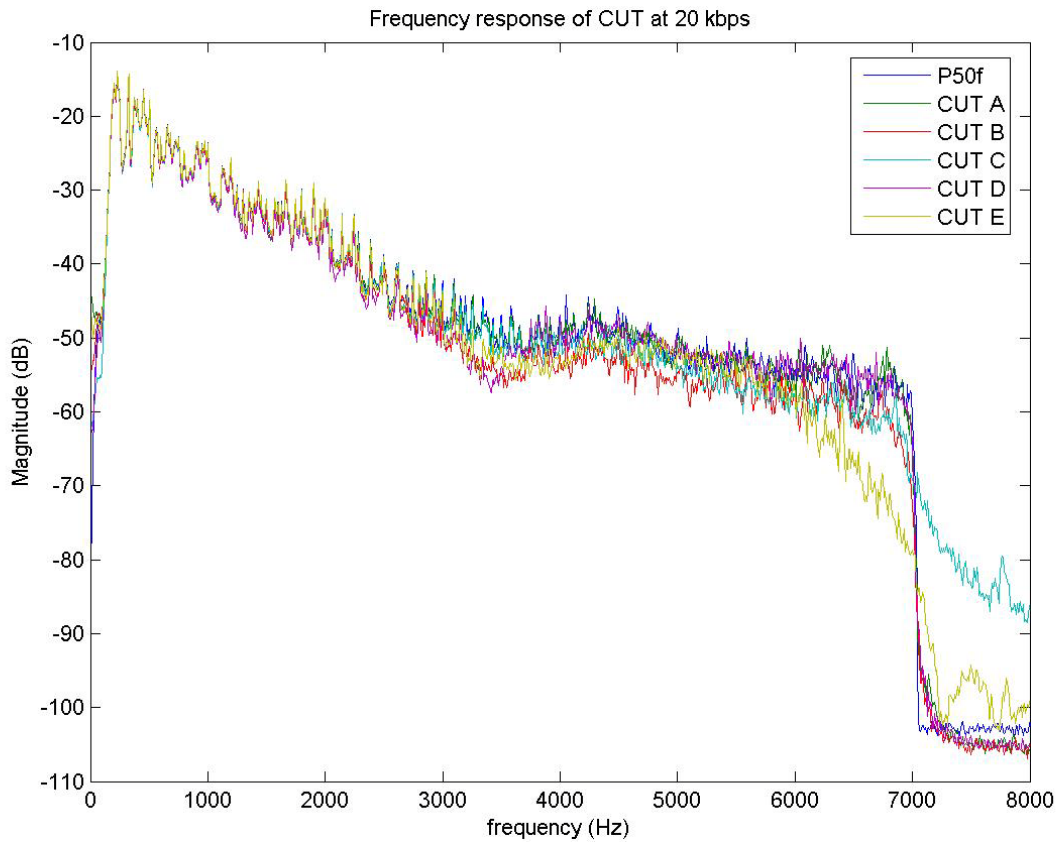


Figure 66(d) – Female speech, 20 kbit/s

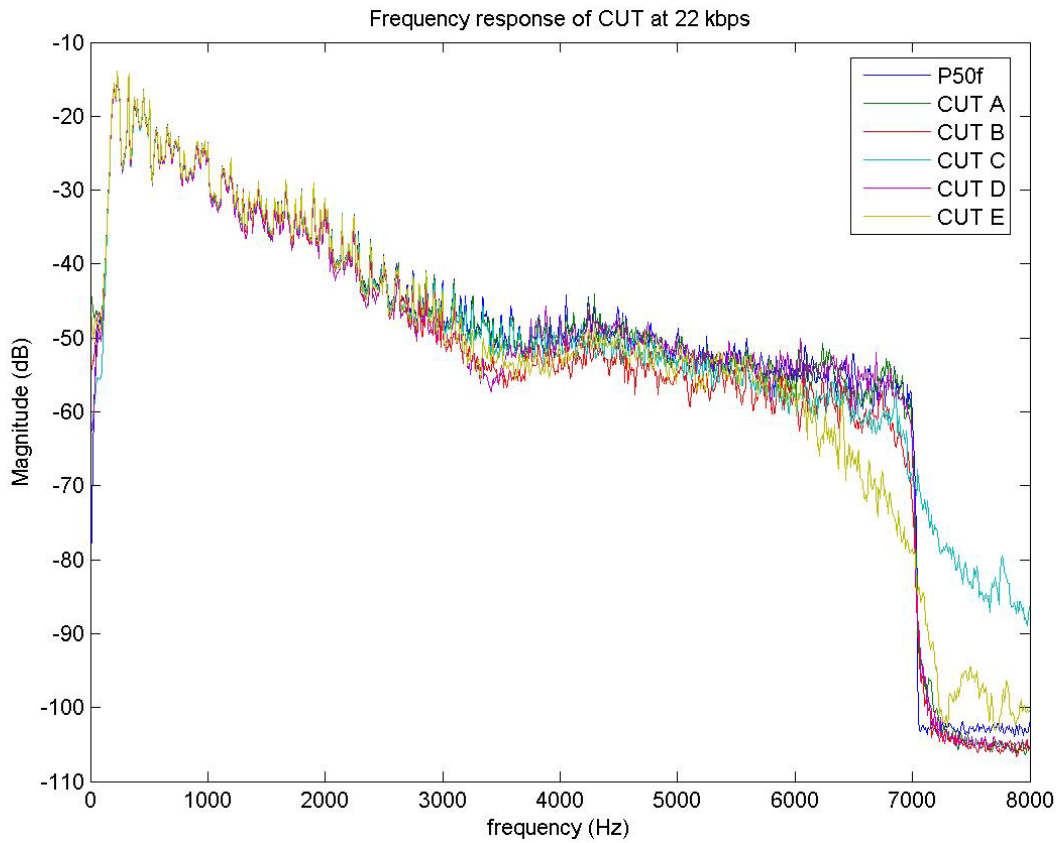


Figure 66(e) – Female speech, 22 kbit/s

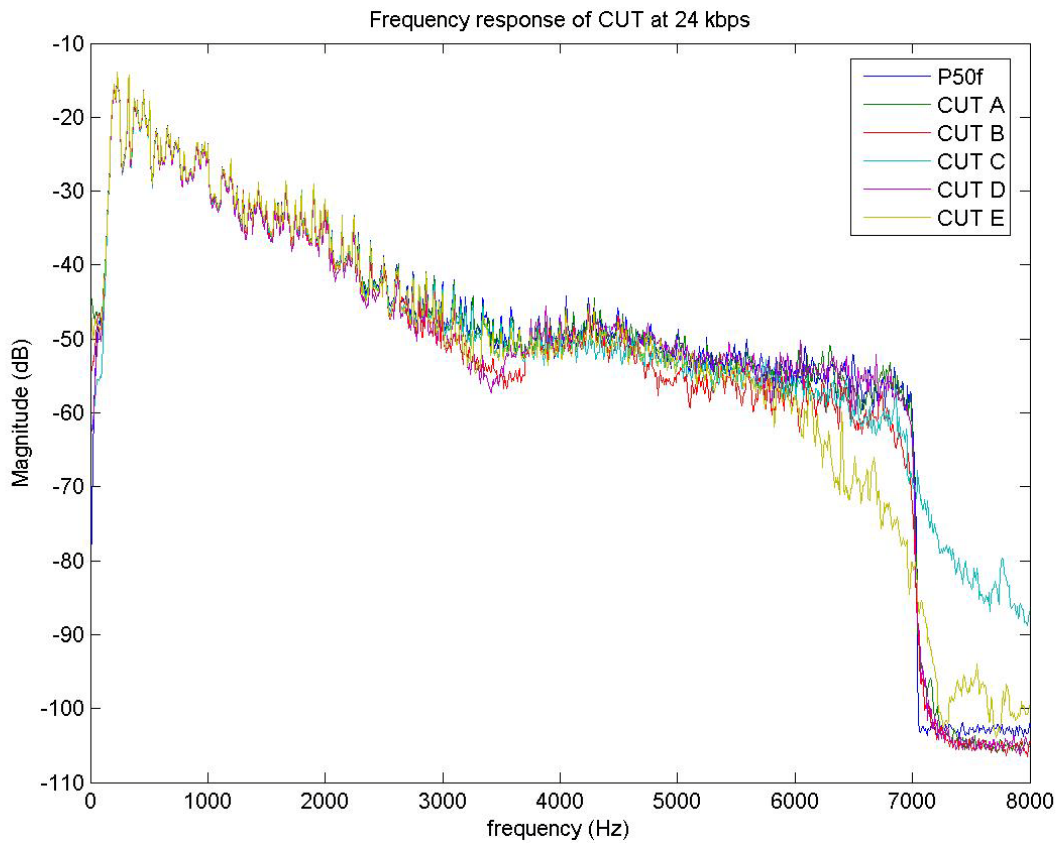


Figure 66(f) – Female speech, 24 kbit/s

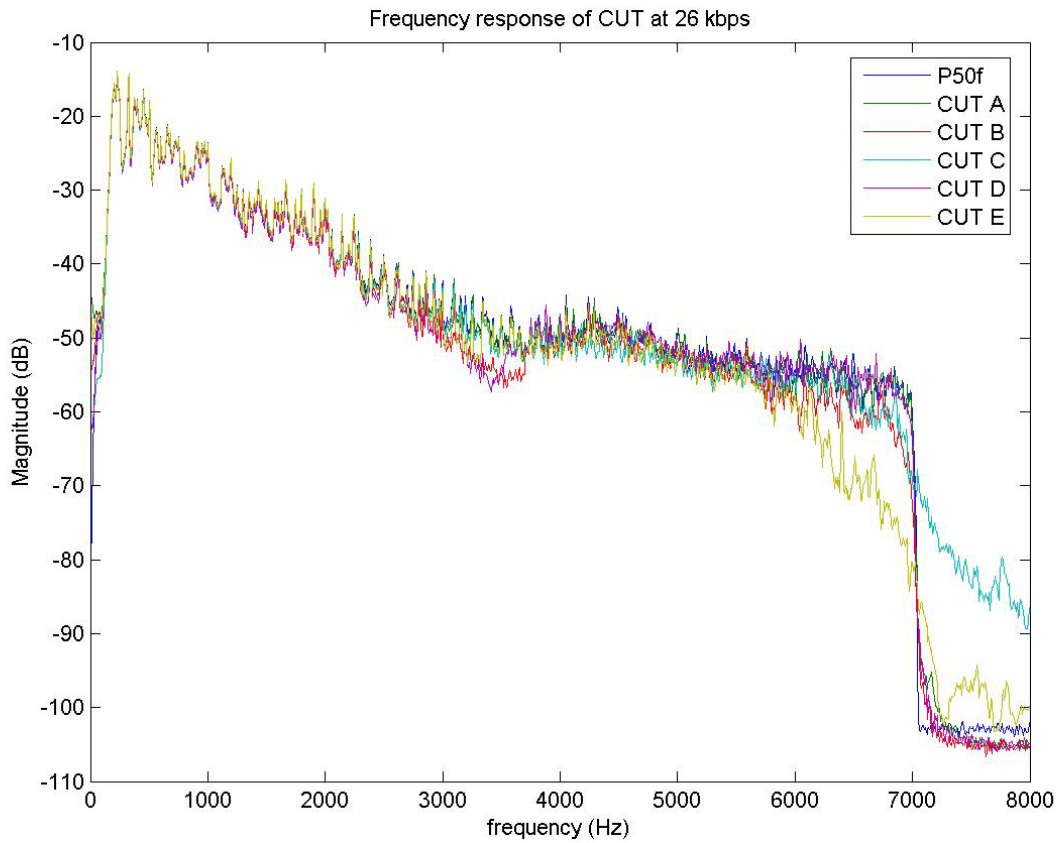


Figure 66(g) – Female speech, 26 kbit/s

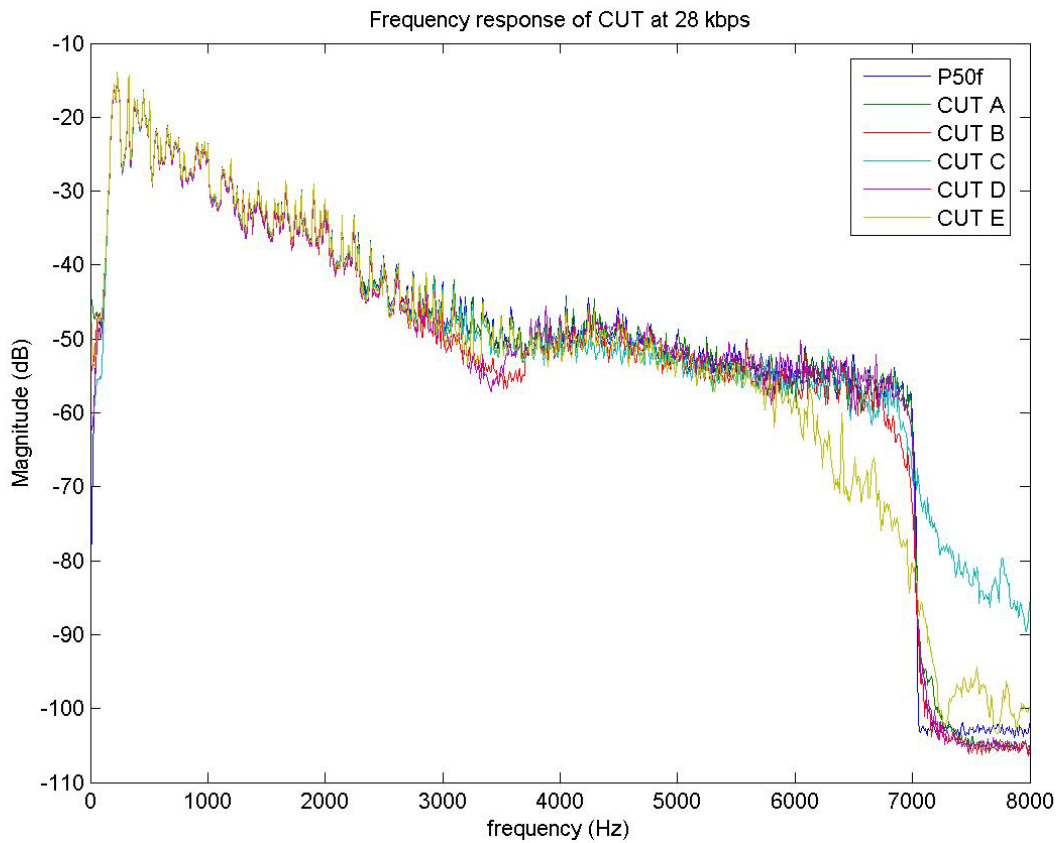


Figure 66(h) – Female speech, 28 kbit/s

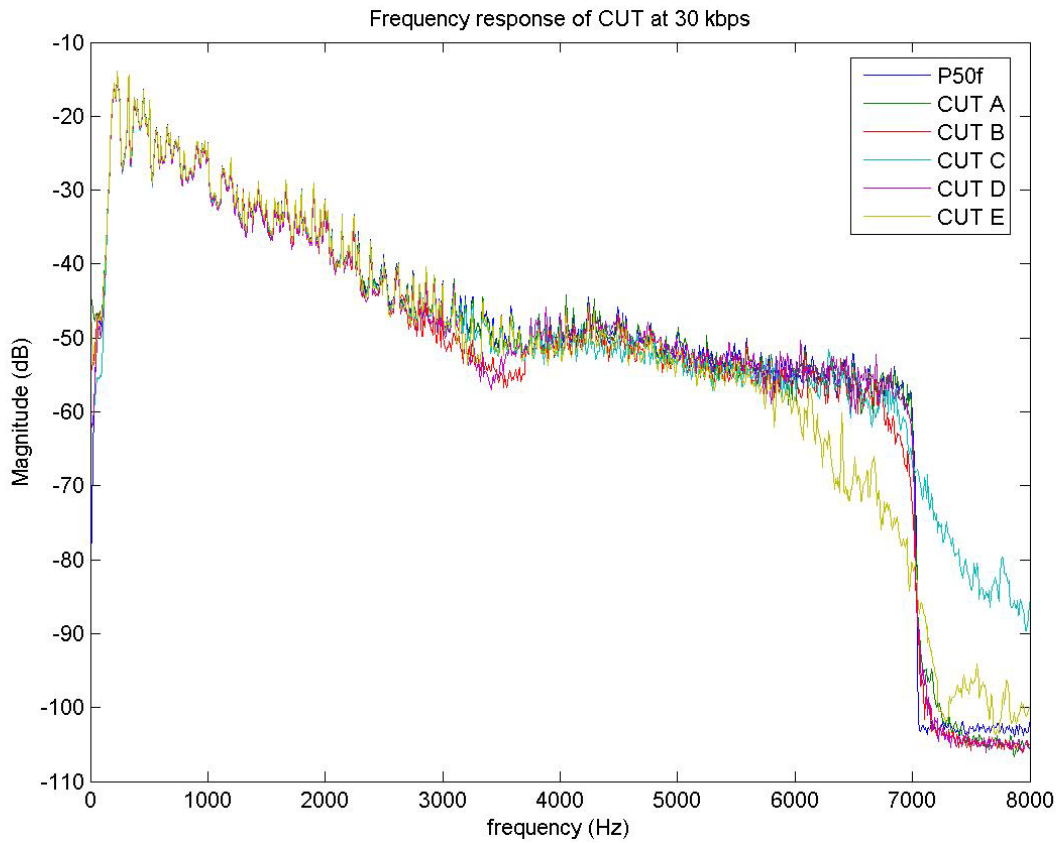


Figure 66(i) – Female speech, 30 kbit/s

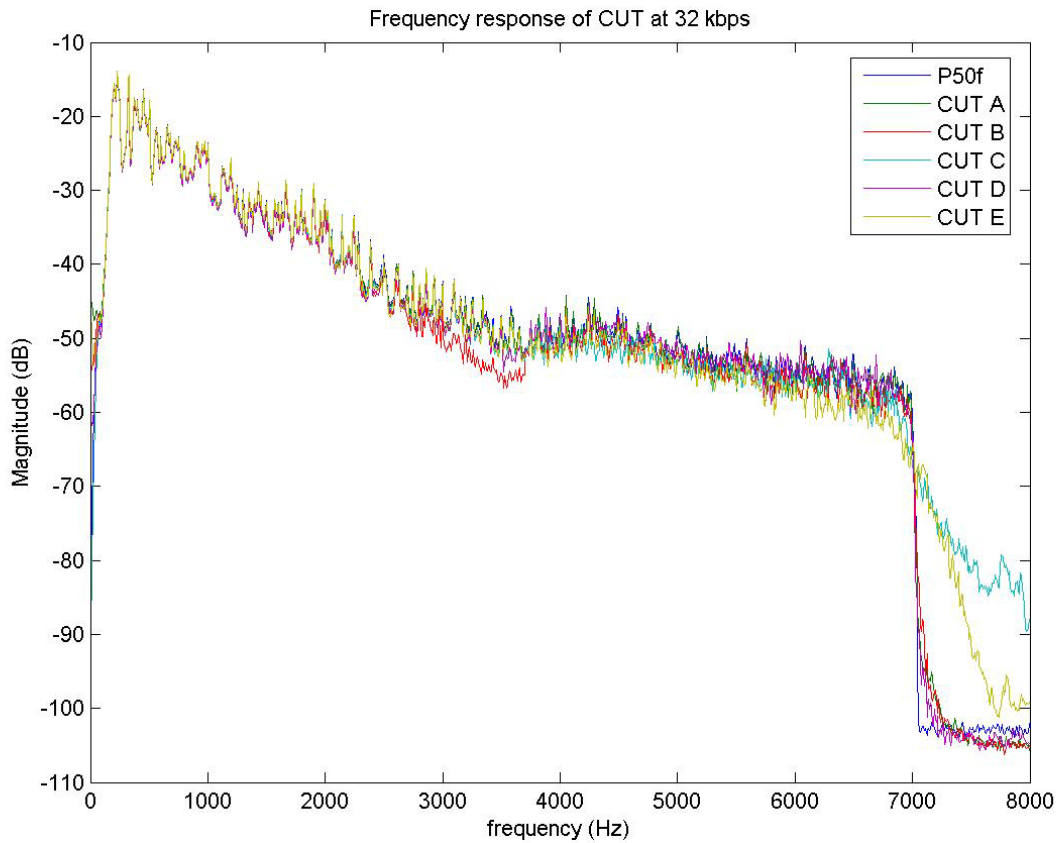


Figure 66(j) – Female speech, 32 kbit/s