International Telecommunication Union

ITU-T Technical Paper

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(25 March 2011)

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

GSTP-G.711AppIII
Performance of ITU-T G.711 Appendix III (Audio quality enhancement toolbox)

ITU-T

International Telecommunication Union

Summary

This document addresses the performance assessment of the Audio quality enhancement toolbox defined in Appendix III of Recommendation ITU-T G.711. It collects performance evaluation results provided during the standardization of ITU-T G.711.1 and ITU-T G.711 Appendix III.

Change Log

This document contains Version 1 of the ITU-T Technical Paper entitled "*Performance of ITU-T G.711 Appendix III (Audio quality enhancement toolbox)*" approved at the ITU-T Study Group 16 meeting held in Geneva, 14-25 March 2011.

Editors:

David Virette
Huawei Technologies
China
Tel: +49 89 15 88 34 41 48
Email: david.virette@huawei.com

Claude Lamblin
France Telecom Orange
France
Tel: +33 2 96 05 13 03
Fax: +33 2 96 05 35 30
Email: claude.lamblin@orange-ftgroup.com

Contents

Page

1 SCOPE 1

2 REFERENCES 1

3 DEFINITIONS 2

4 ABBREVIATIONS, ACRONYMS AND CONVENTIONS 2

5 G.711 APPENDIX III HISTORY 2

6 SCOPE OF THE AUDIO QUALITY ENHANCEMENT TOOLBOX 3

7 TOOLBOX OVERVIEW 4

7.1 NOISE SHAPING TOOL 4
7.2 NOISE GATE TOOL 4
7.3 POSTFILTER TOOL 4
7.4 FRAME ERASURE CONCEALMENT TOOL 4

8 TOOLBOX COMPLEXITY, MEMORY AND DELAY 4

9 G.711.1 CHARACTERIZATION PHASE STEP 1 5

9.1 QUALITY PERFORMANCE ASSESSMENT 5
9.2 QUALITY TEST RESULTS 6
9.3 QUALITY PERFORMANCE COMPARISON 8

10 G.711.1 CHARACTERIZATION PHASE STEP 2 10

10.1 QUALITY PERFORMANCE ASSESSMENT 10
10.2 QUALITY TEST RESULTS 11
10.3 QUALITY PERFORMANCE COMPARISON 13

11 G.711 APPENDIX III VERIFICATION PHASE 16

11.1 QUALITY PERFORMANCE ASSESSMENT 16
11.2 QUALITY TEST RESULTS 17
11.3 QUALITY PERFORMANCE COMPARISON 18

12 OVERVIEW OF G.711 APPENDIX III QUALITY PERFORMANCE 19

12.1 CLEAN SPEECH SIGNALS 19
12.2 MUSIC SIGNALS 19
12.3 NOISY SIGNALS 19

List of Tables

Page

TABLE 1 - G.711 APPENDIX III COMPLEXITY (WORST CASE) [1] 4

TABLE 2 - G.711 APPENDIX III MEMORY CONSUMPTION [1] 4

TABLE 3 - ALGORITHMIC DELAY OF G.711 APPENDIX III TOOLS [1] 4

TABLE 4 - OVERVIEW OF THE QUALITY ASSESSMENT TESTS IN G.711.1 CHARACTERIZATION PHASE STEP 1 (NARROWBAND EXPERIMENTS) 5

TABLE 5 - LANGUAGES AND LABORATORY ALLOCATION IN G.711.1 CHARACTERIZATION PHASE STEP 1 5

TABLE 6 - TOOL COMBINATIONS TESTED IN G.711.1 CHARACTERIZATION PHASE STEP 1 6

TABLE 7 – G.711.1 CHARACTERIZATION PHASE STEP 1 – EXPERIMENT 1A RESULTS (CLEAN SPEECH) 6

TABLE 8 – G.711.1 CHARACTERIZATION PHASE STEP 1 – EXPERIMENT 2A RESULTS (MUSIC) 7

TABLE 9 – G.711.1 CHARACTERIZATION PHASE STEP 1 - EXPERIMENT 3A RESULTS (BACKGROUND MUSIC SNR=25DB) .. 7

TABLE 10 – G.711.1 CHARACTERIZATION PHASE STEP 1 - EXPERIMENT 3B RESULTS (OFFICE NOISE SNR=20 DB) 7

TABLE 11 – G.711.1 CHARACTERIZATION PHASE STEP 1 - EXPERIMENT 3C RESULTS (BABBLE NOISE SNR=30 DB) 7

TABLE 12 – G.711.1 CHARACTERIZATION PHASE STEP 1 - EXPERIMENT 3D RESULTS (INTERFERING TALKER SNR=15 DB)8

TABLE 13 – G.711.1 CHARACTERIZATION PHASE STEP 1 - COMPARISONS IN EXPERIMENT 1 (CLEAN SPEECH) 8

TABLE 14 – G.711.1 CHARACTERIZATION PHASE STEP 1 - COMPARISONS IN EXPERIMENT 2A (MUSIC) 9

TABLE 15 – G.711.1 CHARACTERIZATION PHASE STEP 1 - COMPARISONS IN EXPERIMENT 3A (BACKGROUND MUSIC SNR=25DB) 9

TABLE 16 – G.711.1 CHARACTERIZATION PHASE STEP 1 - COMPARISONS IN EXPERIMENT 3B (OFFICE NOISE SNR=20DB) 9

TABLE 17 – G.711.1 CHARACTERIZATION PHASE STEP 1 - COMPARISONS IN EXPERIMENT 3C (BABBLE NOISE SNR=30DB) 9

TABLE 18 – G.711.1 CHARACTERIZATION PHASE STEP 1 - COMPARISONS IN EXPERIMENT 3D (INTERFERING TALKER SNR=15DB) 9

TABLE 19: OVERVIEW OF THE QUALITY ASSESSMENT TESTS IN G.711.1 CHARACTERIZATION PHASE STEP 2 10

TABLE 20 - LANGUAGES AND LABORATORY ALLOCATION IN G.711.1 CHARACTERIZATION PHASE STEP 2 10

TABLE 21: TOOL COMBINATIONS TESTED IN G.711.1 CHARACTERIZATION PHASE STEP 2 10

TABLE 22 – G.711.1 CHARACTERIZATION PHASE STEP 2 - EXPERIMENT 1A RESULTS (CLEAN SPEECH) 11

TABLE 23 – G.711.1 CHARACTERIZATION PHASE STEP 2 - EXPERIMENT 2A RESULTS (MUSIC) 12

TABLE 24 – G.711.1 CHARACTERIZATION PHASE STEP 2 - EXPERIMENT 3A RESULTS (BACKGROUND MUSIC SNR=25 DB) 12

TABLE 25 – G.711.1 CHARACTERIZATION PHASE STEP 2 - EXPERIMENT 3B RESULTS (OFFICE NOISE SNR=20 DB) 13

TABLE 26 – G.711.1 CHARACTERIZATION PHASE STEP 2 - EXPERIMENT 3C RESULTS (BABBLE NOISE SNR=30 DB) 13

TABLE 27 – G.711.1 CHARACTERIZATION PHASE STEP 2 - EXPERIMENT 3D RESULTS (INTERFERING TALKER SNR=15 DB) 13

TABLE 28 – G.711.1 CHARACTERIZATION PHASE STEP 2 – COMPARISONS IN EXPERIMENT 1A (CLEAN SPEECH) 14

TABLE 29 – G.711.1 CHARACTERIZATION PHASE STEP 2 - COMPARISONS IN EXPERIMENT 2A (MUSIC) 14

TABLE 30 – G.711.1 CHARACTERIZATION PHASE STEP 2 - COMPARISONS IN EXPERIMENT 3A (BACKGROUND MUSIC SNR=25DB) 15

TABLE 31 – G.711.1 CHARACTERIZATION PHASE STEP 2 - COMPARISONS IN EXPERIMENT 3B (OFFICE NOISE SNR=20DB) 15

TABLE 32 – G.711.1 CHARACTERIZATION PHASE STEP 2 - COMPARISONS IN EXPERIMENT 3C (BABBLE NOISE SNR=30DB) 15

TABLE 33 – G.711.1 CHARACTERIZATION PHASE STEP 2 - COMPARISONS IN EXPERIMENT 3D (INTERFERING TALKER

Page

SNR=15dB) 15
Table 34: Overview of the quality assessment tests in G.711 Appendix III verification phase 16
Table 35 - Languages and laboratory allocation in G.711 Appendix III verification phase 16
Table 36: Tool combinations tested in G.711 Appendix III verification phase 16
Table 37 – G.711 Appendix III verification phase - Experiment 1 results (clean speech) 17
Table 38 – G.711 Appendix III verification phase - Experiment 2 results (music) 17
Table 39 – G.711 Appendix III verification phase - Experiment 3 results (various background noise types) 17
Table 40 – G.711 Appendix III verification phase - Comparisons in Experiment 1 (clean speech) 18
Table 41 – G.711 Appendix III verification phase - Comparisons in Experiment 2 (music) 18
Table 42 – G.711 Appendix III verification phase - Comparisons in Experiment 3 (various office noise types) 18

List of Figures

Page

Figure 1 – Timeline of G.711.1 and G.711 Appendix III development 3

ITU-T Technical Paper GSTP-G711AppIII

Performance of ITU-T G.711 Appendix III (Audio quality enhancement toolbox)

1 Scope

This technical paper addresses the performance assessment of the audio quality enhancement toolbox defined in Appendix III of Recommendation ITU-T G.711 [1]. It collects performance evaluation results provided during the standardization of ITU-T G.711.1 and ITU-T G.711 Appendix III.

2 References

This Technical Paper is open to the general public whereas TDs are not accessible by the general public. Therefore, for easy access of all TDs referenced below have been made available at http://ftp3.itu.int/av-arch/audio-site/tpref/.

[1] ITU-T G.711 Amendment 2 (2009), "*New Appendix III on audio quality enhancement toolbox*"

[2] TD 197/Gen-16 Attachment 3, "*Reply LS to ITU-T SG 16 on audio coding matters (COM16-LS 79, COM16-LS 36; COM16-LS 356)*", source: ITU-T SG 12, Geneva, 26 October- 6 November 2009

[3] ITU-T Recommendation G.711.1 (2008), "*Wideband embedded extension for G.711 pulse code modulation*"

[4] TD 225R1/WP3-16, "*Q10/16 Rapporteur meeting report (Geneva, 16-19 January 2007)*", source: Rapporteur Q10/16, Geneva, 30 March 2007

[5] TD 228R1/WP3-16, "*Q10/16 Rapporteur meeting report (Geneva, 22-30 March 2007)*", source: Rapporteur Q10/16, Geneva, 30 March 2007

[6] TD 283/WP3-16, "*Report of Question 10/16 'Software tools for signal processing standardization activities and maintenance and extension of existing voice coding standards'*", source: Rapporteur Q10/16; Geneva, 26 June - 6 July 2007

[7] TD 297R1/WP3-16, "*Report of Q10/16 Rapporteur's meeting report (Geneva, 28 January - 1 February 2008)*", source: Rapporteur Q10/16; Geneva, 1 February 2008

[8] Recommendation ITU-T G.711 (1988), "*Pulse Code Modulation (PCM) of voice frequencies*"

[9] TD 479/Gen-16 Attachment 1, "*G.711 WB extension optimisation/characterization Quality Assessment Test Plan*", source: Rapporteurs Q7/12, Geneva, 22 April - 2 May 2008, (also in attachment 1 of AC-0801-Q10-36, Geneva, 28 January – 1 February 2008)

[10] AC-0801-Q10-37, "*Revised Processing Plan for G.711WB Optimisation/Characterization Phase*", source: Rapporteur Q10/16, Geneva, 28 January – 1 February 2008

[11] TD 479/Gen-16, Annex, "*Summary of Results of optimisation/characterization tests for G.711 wideband extension codec*", source: Rapporteurs Q7/12, Geneva, 22 April - 2 May 2008, (also in annex of AC-0801-Q10-36, Geneva, 28 January – 1 February 2008)

[12] AC-0809-Q10-32, "*G.711.1 characterization phase step2 quality assessment test plan*", source: Rapporteur Q10/16, Geneva, 25 September– 3 October 2008

[13] AC-0809-Q10-33, "*Processing Plan for G.711.1 Characterization Phase Step 2*", source: Rapporteur Q10/16, Geneva, 25 September– 3 October 2008

[14] TD 27/Gen-16 attachment 3, "*Summary of characterization step 2 of G.711.1*", source: Rapporteurs Q7/12, Geneva, 27 January - 6 February 2009

[15] AC-0907-Q10-29, "*Quality Assessment Test Plan of for pre- or post processing module for ITU-T G.711 legacy codec*", source: Rapporteurs Q10/16, Geneva, 6 – 9 July 2009

[16] TD 79/WP3-16, "*Processing plan for G.711-Plus (final version)*", source: Rapporteurs Q10/16, Geneva, 26 October - 6 November 2009

3 Definitions

3.1 **codec:** encoding and decoding algorithm.

3.2 **narrowband audio:** audio signals sampled at 8 kHz.

3.3 **wideband audio:** audio signals sampled at 16 kHz within 50-7000 Hz bandwidth.

4 Abbreviations, acronyms and conventions

The following abbreviations, acronyms, and conventions are used in this document.

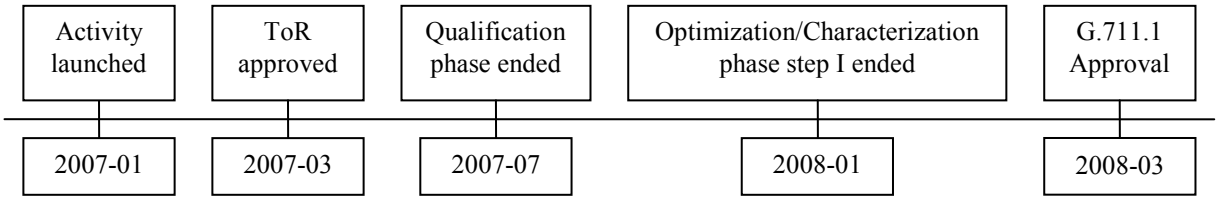
ACR	Absolute Category Rating
BFER	Burst frame erasure
BT	Better than
CCR	Comparison Category Rating
DCR	Degradation Category Rating
FERC	Frame erasure concealment
labX	Listening Laboratory X with X= A or B (each experiment was performed by two listening laboratories)
Legacy	Legacy G.711 encoder and/or decoder denotes the G.711 standardized in 1972
meanX	Overall mean opinion score (MOS) or degradation MOS (DMOS) for confounded talkers/music items in listening laboratory X (X= A or B)
NB	Narrowband
NG	Noise gate
NS*	Noise shaping
	NOTE: This abbreviation is used in G.711.1 and should not be mistaken with the abbreviation NS used in recommendation ITU-T G.718 for noise reduction.
nwt	Not worse than
PCM	Pulse code modulation
PF	Postfilter
RFER	Random frame erasure
stdX	standard deviation for a condition in listening laboratory X (X= A or B)
WMOPS	Weighted million operations per second

5 G.711 Appendix III History

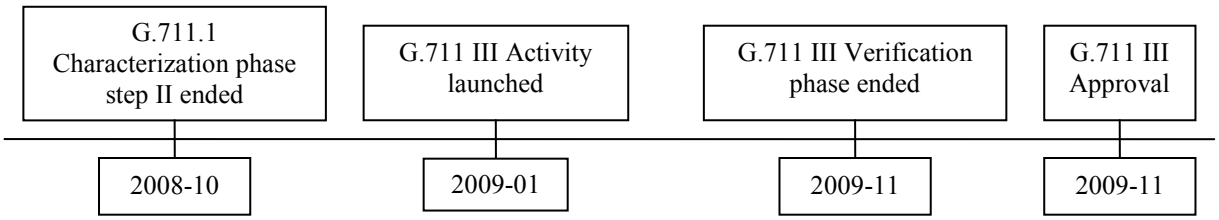
The timeline for the development of G.711.1 and G.711 Appendix III is found in Figure 1. Appendix III to G.711 [1] describes a toolbox that comprises four tools with algorithms initially

developed in the context of G.711.1 wideband speech and audio codec [3]. Therefore, G.711 Appendix III history is closely related to G.711.1 history.

ITU-T G.711.1 standardization was launched in Q10/16 January 2007 meeting [4]. Its objective was to develop a low-complexity, low-delay, wideband speech codec, with an embedded scalable structure on top of G.711. After the approval of terms of reference in March 2007 (see in Annex Q10.H of [5]), the qualification phase was launched. Five candidate coders (in floating point) participated in the qualification phase (March–July 2007). In summer 2007, the qualification meeting took place [6]. Then a characterization phase was organized where all five candidate companies worked together in order to provide a single candidate. The characterization phase was conducted in two steps: step 1 to check requirements and some objectives and step 2 to verify further the quality of G.711.1, and especially of its Appendix I. At WP3/16 February 2008 plenary meeting, G.711WB Characterization Phase Step 1 was successfully completed and G.711WB was moved for Consent under AAP [7]. Approved in March 2008, the Recommendation has been published as G.711.1. During G.711.1 characterization phase, it was noted that, G.711.1 and its Appendix I brought quality improvement for narrowband voice inter-working with the "legacy" G.711 [8]. Therefore in January 2009, it was proposed to standardize G.711.1 pre and post processing functions as extensions to G.711 legacy codec. A verification phase took place to characterize the quality enhancements in emulated real usage conditions. This verification phase ended in November 2009 with the approval of Appendix III of G.711.



(a) Timeline of G.711.1 development



(b) Timeline of G.711 Appendix III development

Figure 1 – Timeline of G.711.1 and G.711 Appendix III development

6 Scope of the audio quality enhancement toolbox

Standardized for PSTN use in 1972, ITU-T G.711 is now widely used in more recent VoIP/ packetized networks and often specified as mandatory or recommended codec in various network applications for inter-working purposes. Although the G.711 quality is classified as good – usually called "toll" quality – the G.711 codec suffers from audible white quantization noise [3]. Furthermore, as terminal acoustic characteristics have evolved to accommodate wider audio bandwidths, this quantization noise becomes more audible and annoying for voice service customers. The toolbox defined in G.711 Appendix III comprises pre- or post- processing modules providing quality enhancements to the legacy ITU-T G.711 codec for inter-working purposes.

7 Toolbox overview

The four tools, extracted from G.711.1, are: a noise shaping tool (NS), a noise gate tool (NG), a postfilter tool (PF), and frame erasure concealment tool (FERC). They are briefly described below. All these tools can be used separately or in combination [1].

7.1 Noise shaping tool

Applied at the encoder side, the noise shaping tool is a pre-processing module which is used in combination with a modified G.711 encoder to perceptually shape the coding noise of the PCM encoder and produces a compatible bitstream.

7.2 Noise gate tool

Applied at the decoder side, the noise gate tool is a post-processing module which increases the clearness of the audio signal during quasi-silent periods.

7.3 Postfilter tool

Applied at the decoder side, the postfilter tool is a post-processing module which reduces the PCM quantization noise of legacy G.711.

7.4 Frame erasure concealment tool

Applied at the decoder side, the frame erasure concealment tool is used to extrapolate the signal in case of erased frames.

8 Toolbox complexity, memory and delay

The complexity of G.711 Appendix III was assessed during its verification phase. Table 1 gives the worst-case complexity of the audio quality enhancement toolbox and Table 2 gives its memory usage. Table 3 details the delay of G.711 Appendix III tools.

Table 1 - G.711 Appendix III complexity (worst case) [1]

NS	NG	PF	FERC	FERC+PF+NG
0.87	0.23	2.02	2.05	3.31

Table 2 - G.711 Appendix III memory consumption [1]

Memory type	NS	NG	PF	FERC
Static RAM (kWords)	0.093	0.003	0.353	0.984
Scratch RAM (kWords)	0.107	0.012	0.529	0.314
Data ROM (kWords)	0.088	0	0.191	0.121
Program ROM (number of basic operators)	191	37	593	728

Table 3 - Algorithmic delay of G.711 Appendix III tools [1]

NS	NG	PF	FERC	FERC+PF+NG
0 ms	0 ms	2 ms	5 ms	5 ms

9 G.711.1 Characterization Phase Step 1

Launched in July 2007, the G.711.1 Characterization Phase Step 1 involved optimization and characterization of the algorithm, and it was concluded in February 2008. During this phase, the quality performance was assessed in various conditions.

9.1 Quality performance assessment

In Characterization Phase Step 1, the quality of G.711.1 was assessed in five main formal subjective experiments. In the narrowband experiments, the quality of some tools or some combination of tools of G.711 Appendix III was also assessed. Table 4 gives an overview of these narrowband experiments. The quality assessment test plan for the formal subjective experiments was prepared by Q7/12 during SG 12 October 2007 Plenary meeting, the final version is found in [9].

Table 4 - Overview of the quality assessment tests in G.711.1 Characterization Phase Step 1 (narrowband experiments)

Exp	Methodology	Conditions
1a	ACR	Clean speech for narrowband speech signals: different input levels, frame erasure and bitstream interoperability with G.711 legacy system coder
2a	ACR	Music signal. Interoperability with G.711 legacy system coder
3a,b,c,d	DCR	Speech quality with background noise: – Exp.3a: background music SNR=25 dB – Exp.3b: office noise SNR=20 dB – Exp.3c: babble noise SNR=30 dB – Exp.3d: interfering talker SNR=15 dB

The processing test plan can be found in [10]. Processing batch files were distributed to the processing laboratories. Common FER pattern files and background noise files (except interfering talkers) were also provided to these laboratories. These experiments were performed in two different languages and their processing was performed twice by two different laboratories (for cross-checking purposes). The listening sessions were conducted in fall 2007 by five listening laboratories. Table 5 indicates, for each experiment, the two languages and the two listening/ processing laboratories.

The quality of two tools (NS and NG) and of two tool combinations (NS+NG and NS+NG+FERC) was assessed in G.711.1 Characterization Phase Step 1. Table 6 lists in which experiments and conditions (see condition numbers in Table 7 to Table 12) these tools and combinations were tested.

Table 5 - Languages and laboratory allocation in G.711.1 Characterization Phase Step 1

Exp	Lab A	Language A	Lab B	Language B
1a	VoiceAge	North-American English	ETRI	Korean
2a	Huawei	Music	NTT	Music
3	NTT	Japanese	ETRI	Korean

Table 6 - Tool combinations tested in G.711.1 Characterization Phase Step 1

Combinations	Exp.	Conditions
NS	1a	Clean Speech: -26 dBov
NS	2a	Music
NG	1a	Clean Speech: -26 dBov
NG	2a	Music
NS+NG	1a	Clean Speech: -26, -16, -36 dBov
NS+NG	2a	Music
NS+NG	3a, b, c, d	Background noise types: – Exp.3a: background music SNR=25 dB – Exp.3b: office noise SNR=20 dB – Exp.3c: babble noise SNR=30 dB – Exp.3d: interfering talker SNR=15 dB
NS+NG+FERC	1a	Clean Speech: 3% RFER & 3 %BFER at -26dBov

9.2 Quality test results

The quality test results were reviewed by Q7/12 experts during their January 2008 Rapporteurs meeting. The test results analysis is reported in section 2 and in the annex of their January 2008 Liaison Statement [11].

Table 7 and Table 8 give the results of Experiment 1a and 2a respectively, while Table 9 to Table 12 give the results of Experiment 3 for the four types of background noise. In Table 7, column FER, (B) denotes bursty FER, while (R) denotes random FER.

Table 7 – G.711.1 Characterization Phase Step 1 – Experiment 1a results (clean speech)

Combinations	Number	Conditions	Level	FER	meanA	stdA	meanB	stdB
-	8	Direct	-26 dBov	0%	4.11	0.72	4.38	0.69
Legacy G.711	9	G.711 @ 64k A-law	-26 dBov	0%	3.16	0.72	2.91	0.65
Legacy G.711	10	G.711 @ 64k A-law	-16 dBov	0%	3.88	0.69	3.86	0.72
Legacy G.711	11	G.711 @ 64k A-law	-36 dBov	0%	2.33	0.62	2.06	0.56
Legacy G.711 + App. I	12	G.711 @ 64k A-law w/ App. I	-26 dBov	3% (R)	3.07	0.73	2.78	0.71
Legacy G.711 + App. I	13	G.711 @ 64k A-law w/ App. I	-26 dBov	3% (B)	3.04	0.70	2.70	0.77
Legacy G.711	14	G.711 @ 64k μ-law	-26 dBov	0%	3.65	0.69	3.57	0.75
Legacy G.711	15	G.711 @ 64k μ-law	-16 dBov	0%	3.96	0.72	3.90	0.80
Legacy G.711	16	G.711 @ 64k μ-law	-36 dBov	0%	3.80	0.78	3.01	0.71
Legacy G.711 + App. I	17	G.711 @ 64k μ-law w/ App. I	-26 dBov	3% (R)	3.60	0.69	3.47	0.72
Legacy G.711 + App. I	18	G.711 @ 64k μ-law w/ App. I	-26 dBov	3% (B)	3.54	0.74	3.38	0.80
NS+NG	19	CuT @ 64k (R1) A-law	-26 dBov	0%	4.41	0.64	4.05	0.75
NS+NG	21	CuT @ 64k (R1) A-law	-16 dBov	0%	4.08	0.75	3.98	0.81
NS+NG	22	CuT @ 64k (R1) A-law	-36 dBov	0%	4.18	0.77	3.62	0.74
NS+NG+FERC	23	CuT @ 64k (R1) A-law	-26 dBov	3% (R)	4.26	0.73	3.94	0.75
NS+NG+FERC	25	CuT @ 64k (R1) A-law	-26 dBov	3% (B)	4.06	0.77	3.67	0.86
NS+NG	26	CuT @ 64k (R1) μ-law	-26 dBov	0%	4.41	0.70	4.07	0.8
NS+NG	28	CuT @ 64k (R1) μ-law	-16 dBov	0%	4.05	0.70	4.05	0.79
NS+NG	29	CuT @ 64k (R1) μ-law	-36 dBov	0%	4.39	0.65	3.76	0.80

Combinations	Number	Conditions	Level	FER	meanA	stdA	meanB	stdB
NS+NG+FERC	30	CuT @ 64k (R1) µ-law	-26 dBov	3% (R)	4.2	0.78	4.00	0.74
NS+NG+FERC	32	CuT @ 64k (R1) µ-law	-26 dBov	3% (B)	4.21	0.83	3.78	0.88
NG	33	G.711Enc-CuTDec A-law	-26 dBov	0%	3.12	0.72	3.05	0.75
NG	34	G.711Enc-CuTDec µ-law	-26 dBov	0%	4.15	0.75	3.82	0.85
NS	35	CuTEnc-G.711Dec A-law	-26 dBov	0%	4.35	0.67	3.87	0.81
NS	36	CutEnc-G.711Dec µ-law	-26 dBov	0%	4.25	0.74	4.03	0.79

Table 8 – G.711.1 Characterization Phase Step 1 – Experiment 2a results (music)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	6	Direct	3.90	0.80	3.43	0.97
Legacy G.711	7	G.711 @ 64k A-law	3.77	0.83	3.30	1.00
Legacy G.711	8	G.711 @ 64k µ-law	3.77	0.86	3.36	0.96
NS+NG	9	CuT @ 64k (R1) A-law	3.86	0.83	3.47	0.95
NS+NG	11	CuT @ 64k (R1) µ-law	3.86	0.82	3.52	1.01
NG	13	G.711Enc-CuTDec A-law	3.72	0.84	3.34	0.90
NG	14	G.711Enc-CuTDec µ-law	3.79	0.83	3.44	0.99
NS	15	CuTEnc-G.711Dec A-law	3.86	0.85	3.45	1.01
NS	16	CutEnc-G.711Dec µ-law	3.84	0.83	3.49	0.95

Table 9 – G.711.1 Characterization Phase Step 1 - Experiment 3a results (background music SNR=25dB)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	6	Direct	4.82	0.44	4.81	0.43
Legacy G.711	7	G.711 @ 64k A-law	4.56	0.59	4.35	0.59
Legacy G.711	8	G.711 @ 64k µ-law	4.60	0.61	4.39	0.60
NS+NG	9	CuT @ 64k (R1) A-law	4.77	0.47	4.58	0.54
NS+NG	11	CuT @ 64k (R1) µ-law	4.79	0.45	4.60	0.51

Table 10 – G.711.1 Characterization Phase Step 1 - Experiment 3b results (office noise SNR=20 dB)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	6	Direct	4.83	0.40	4.73	0.46
Legacy G.711	7	G.711 @ 64k A-law	4.77	0.46	4.68	0.50
Legacy G.711	8	G.711 @ 64k µ-law	4.76	0.45	4.57	0.56
NS+NG	9	CuT @ 64k (R1) A-law	4.82	0.41	4.68	0.51
NS+NG	11	CuT @ 64k (R1) µ-law	4.80	0.41	4.73	0.45

Table 11 – G.711.1 Characterization Phase Step 1 - Experiment 3c results (babble noise SNR=30 dB)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	6	Direct	4.78	0.47	4.80	0.43
Legacy G.711	7	G.711 @ 64k A-law	4.68	0.55	4.48	0.60

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
Legacy G.711	8	G.711 @ 64k µ-law	4.60	0.59	4.51	0.64
NS+NG	9	CuT @ 64k (R1) A-law	4.74	0.49	4.61	0.54
NS+NG	11	CuT @ 64k (R1) µ-law	4.77	0.52	4.76	0.44

Table 12 – G.711.1 Characterization Phase Step 1 - Experiment 3d results (interfering talker SNR=15 dB)

Comb.	Number	Conditions	meanA	stdA	meanB	stdB
Legacy G.711	6	Direct	4.73	0.54	4.85	0.38
Legacy G.711	7	G.711 @ 64k A-law	4.52	0.70	4.43	0.63
Legacy G.711	8	G.711 @ 64k µ-law	4.58	0.65	4.48	0.60
NS+NG	9	CuT @ 64k (R1) A-law	4.64	0.66	4.62	0.52
NS+NG	11	CuT @ 64k (R1) µ-law	4.64	0.62	4.65	0.53

9.3 Quality performance comparison

The comparisons, based on a statistical analysis, are computed with a confidence interval of 95%. The results of comparison are given by "nwt" for "not worse than" and "BT" for "better than". The results are provided in Table 13 and Table 14 for experiments 1 and 2 respectively and in Table 15 to Table 18 for the experiment with Background noise types.

Table 13 – G.711.1 Characterization Phase Step 1 - Comparisons in Experiment 1 (clean speech)

Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
A-law -26 dBov	19	NS+NG	9	Legacy G.711	BT	BT
A-law -16 dBov	21	NS+NG	10	Legacy G.711	BT	BT
A-law -36 dBov	22	NS+NG	11	Legacy G.711	BT	BT
A-law -26 dBov 3% RFER	23	NS+NG+FERC	12	Legacy G.711 + App. I	BT	BT
A-law -26 dBov 3% BFER	25	NS+NG+FERC	13	Legacy G.711 + App. I	BT	BT
µ-law -26 dBov	26	NS+NG	14	Legacy G.711	BT	BT
µ-law -16 dBov	28	NS+NG	15	Legacy G.711	BT	BT
µ-law -36 dBov	29	NS+NG	16	Legacy G.711	BT	BT
µ-law -26 dBov 3% RFER	30	NS+NG+FERC	17	Legacy G.711 + App. I	BT	BT
µ-law -26 dBov 3% BFER	32	NS+NG+FERC	18	Legacy G.711 + App. I	BT	BT
A-law -26 dBov	33	NG	9	Legacy G.711	nwt	BT
µ-law -26 dBov	34	NG	14	Legacy G.711	BT	BT
A-law -26 dBov	35	NS	9	Legacy G.711	BT	BT
µ-law -26 dBov	36	NS	14	Legacy G.711	BT	BT
A-law -26 dBov	19	NS+NG	33	NG	BT	BT
µ-law -26 dBov	26	NS+NG	34	NG	BT	BT
A-law -26 dBov	19	NS+NG	35	NS	nwt	BT
µ-law -26 dBov	26	NS+NG	36	NS	BT	nwt
A-law -26 dBov	35	NS	33	NG	BT	BT
µ-law -26 dBov	36	NS	34	NG	nwt	BT

Table 14 – G.711.1 Characterization Phase Step 1 - Comparisons in Experiment 2a (music)

Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
A-law	9	NS+NG	7	Legacy G.711	BT	BT
µ-law	11	NS+NG	8	Legacy G.711	BT	BT
A-law	13	NG	7	Legacy G.711	nwt	nwt
µ-law	14	NG	8	Legacy G.711	nwt	nwt
A-law	15	NS	7	Legacy G.711	nwt	BT
µ-law	16	NS	8	Legacy G.711	nwt	BT
A-law	9	NS+NG	13	NG	BT	BT
µ-law	11	NS+NG	14	NG	BT	nwt
A-law	9	NS+NG	15	NS	nwt	nwt
µ-law	11	NS+NG	16	NS	nwt	nwt

Table 15 – G.711.1 Characterization Phase Step 1 - Comparisons in Experiment 3a (background music SNR=25dB)

Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
A-law	9	NS+NG	7	Legacy G.711	BT	BT
µ-law	11	NS+NG	8	Legacy G.711	BT	BT

Table 16 – G.711.1 Characterization Phase Step 1 - Comparisons in Experiment 3b (office noise SNR=20dB)

Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
A-law	9	NS+NG	7	Legacy G.711	nwt	nwt
µ-law	11	NS+NG	8	Legacy G.711	nwt	BT

Table 17 – G.711.1 Characterization Phase Step 1 - Comparisons in Experiment 3c (babble noise SNR=30dB)

Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
A-law	9	NS+NG	7	Legacy G.711	nwt	BT
µ-law	11	NS+NG	8	Legacy G.711	BT	BT

Table 18 – G.711.1 Characterization Phase Step 1 - Comparisons in Experiment 3d (interfering talker SNR=15dB)

Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
A-law	9	NS+NG	7	Legacy G.711	BT	BT
µ-law	11	NS+NG	8	Legacy G.711	nwt	BT

10 G.711.1 Characterization Phase Step 2

10.1 Quality performance assessment

After the Consent meeting in February 2008, the G.711.1 Characterization Phase Step 2 was launched. It consisted of characterising further the quality of G.711.1 and especially the quality of its Appendix I with more formal subjective tests. Step2 was performed in summer 2008. The quality assessment test plan designed by Q7/12 is found in [12]. Table 19 gives an overview of the experiments.

Processing batch files, designed according the processing plan [13], were distributed to the processing laboratories. It should be noted that to characterize the coder in other VoIP use cases, the output filtering (RXIRS16) used in step 1, was removed in step 2. Common FER pattern files and background noise files (except interfering talkers) were provided to the processing laboratories. The experiments were performed in two different languages and their processing performed twice by two different laboratories for cross-checking. The listening sessions were conducted in summer 2008 by three listening laboratories. Table 20 indicates for each experiment the two languages and the two listening/processing laboratories.

In G.711.1 Characterization Phase Step 2, the quality of six combinations - NG, NS+NG, NG+PF, NS+NG+PF, NS+NG+FERC, and NS+NG+PF+FERC - was assessed. Table 21 lists in which experiments and conditions (condition numbers are given in Table 22 to Table 27) these tools and combinations were tested.

Table 19 - Overview of the quality assessment tests in G.711.1 Characterization Phase Step 2

Exp.	Methodology	Conditions
1	ACR	Clean speech for narrowband speech signals: different input levels, frame erasure and bitstream interoperability with G.711 legacy system coder
2	ACR	Music signal. Interoperability with G.711 legacy system coder
3a, 3b, 3c, 3d	DCR	Speech quality with background noise: – Exp.3a: background music SNR=25 dB – Exp.3b: office noise SNR=20 dB – Exp.3c: babble noise SNR=30 dB – Exp.3d: interfering talker SNR=15 dB

Table 20 – Languages and laboratory allocation in G.711.1 Characterization Phase Step 2

Exp	Lab A	Language A	Lab B	Language B
1a	France Telecom	French	NTT	Japanese
2a	France Telecom	Music	NTT	Music
3a,b,c,d	France Telecom	French	VoiceAge	North-American English

Table 21 - Tool combinations tested in G.711.1 Characterization Phase Step 2

Combinations	Exp.	Conditions
NG	1a	Clean Speech: -26 dBov
NG	2a	Music
NG	3a, 3b, 3c, 3d	Background noise types: – Exp.3a: background music SNR=25 dB – Exp.3b: office noise SNR=20 dB – Exp.3c: babble noise SNR=30 dB – Exp.3d: interfering talker SNR=15 dB
NS+NG	1a	Clean Speech: -26, -16, -36 dBov
NS+NG	2a	Music
NG+PF	1a	Clean Speech: -26 dBov
NG+ PF	2a	Music
NG+PF	3a, 3b, 3c, 3d	Background noise types: – Exp.3a: background music SNR=25 dB – Exp.3b: office noise SNR=20 dB – Exp.3c: babble noise SNR=30 dB – Exp.3d: interfering talker SNR=15 dB
NS+NG+PF	1a	Clean Speech: -26, -16, -36 dBov
NS+NG+PF	2a	Music
NS+NG+FERC	1a	Clean Speech: 3% RFER & 3 %BFER at -26dBov
NS+NG+PF+FERC	1a	Clean Speech: 3% RFER & 3 %BFER at -26dBov

10.2 Quality test results

The quality test results, analyzed by Q7/12 at their September 2008 meeting, were reviewed by Q10/16 at SG 16 January 2009 plenary meeting [14]. Table 22 and Table 23 give the results of Experiment 1a and 2a respectively while Table 24 to Table 27 give the results of Experiment 3 for the four types of background noise. In Table 22, column FER, (B) denotes bursty FER, while (R) denotes random FER.

Table 22 – G.711.1 Characterization Phase Step 2 - Experiment 1a results (clean speech)

Comb.	Number	Conditions	Level	FER	meanA	stdA	meanB	stdB
-	8	Direct	-26 dBov	0%	4.47	0.56	3.97	0.84
Legacy G.711	9	G.711 @ 64k A-law	-26 dBov	0%	3.29	0.82	3.42	0.90
Legacy G.711	10	G.711 @ 64k A-law	-16 dBov	0%	3.88	0.76	3.69	0.85
Legacy G.711	11	G.711 @ 64k A-law	-36 dBov	0%	2.38	0.79	2.66	0.83
Legacy G.711 + App. I	12	G.711 @ 64k A-law w/ App. I	-26 dBov	3% (R)	3.24	0.78	2.79	0.89
Legacy G.711 + App. I	13	G.711 @ 64k A-law w/ App. I	-26 dBov	3% (B)	3.10	0.77	2.82	0.89
Legacy G.711	14	G.711 @ 64k µ-law	-26 dBov	0%	3.64	0.74	3.49	0.81
Legacy G.711	15	G.711 @ 64k µ-law	-16 dBov	0%	3.86	0.81	3.65	0.86
Legacy G.711	16	G.711 @ 64k µ-law	-36 dBov	0%	3.11	0.76	3.53	0.84
Legacy G.711 + App. I	17	G.711 @ 64k µ-law w/ App. I	-26 dBov	3% (R)	3.66	0.68	2.95	0.85
Legacy G.711 + App. I	18	G.711 @ 64k µ-law w/ App. I	-26 dBov	3% (B)	3.61	0.71	2.91	0.92

Comb.	Number	Conditions	Level	FER	meanA	stdA	meanB	stdB
NS+NG	19	G.711.1 @ 64k (R1) A-law	-26 dBov	0%	4.19	0.74	3.90	0.85
NS+NG	20	G.711.1 @ 64k (R1) A-law	-16 dBov	0%	4.01	0.78	3.94	0.81
NS+NG	21	G.711.1 @ 64k (R1) A-law	-36 dBov	0%	3.72	0.78	3.70	0.84
NS+NG+FERC	22	G.711.1 @ 64k (R1) A-law	-26 dBov	3% (R)	4.09	0.74	3.44	0.93
NS+NG+FERC	23	G.711.1 @ 64k (R1) A-law	-26 dBov	3% (B)	3.93	0.87	3.12	1.03
NS+NG	24	G.711.1 @ 64k (R1) μ-law	-26 dBov	0%	4.29	0.72	3.83	0.81
NS+NG	25	G.711.1 @ 64k (R1) μ-law	-16 dBov	0%	4.11	0.76	3.82	0.86
NS+NG	26	G.711.1 @ 64k (R1) μ-law	-36 dBov	0%	4.14	0.70	3.71	0.84
NS+NG+FERC	27	G.711.1 @ 64k (R1) μ-law	-26 dBov	3% (R)	4.11	0.74	3.39	0.88
NS+NG+FERC	28	G.711.1 @ 64k (R1) μ-law	-26 dBov	3% (B)	3.98	0.82	3.03	0.99
NG	29	G.711Enc-G.711.1Dec A-law	-26 dBov	0%	3.37	0.70	3.43	0.86
NG	30	G.711Enc-G.711.1Dec μ-law	-26 dBov	0%	3.91	0.80	3.74	0.84
NS+NG+PF	31	CuT @ 64k (R1) A-law	-26 dBov	0%	4.47	0.69	4.00	0.77
NS+NG+PF	32	CuT @ 64k (R1) A-law	-16 dBov	0%	4.41	0.67	3.93	0.86
NS+NG+PF	33	CuT @ 64k (R1) A-law	-36 dBov	0%	4.28	0.71	3.79	0.79
NS+NG+PF+FERC	34	CuT @ 64k (R1) A-law	-26 dBov	3% (R)	4.40	0.66	3.50	0.95
NS+NG+PF+FERC	35	CuT @ 64k (R1) A-law	-26 dBov	3% (B)	4.20	0.87	3.15	1.00
NS+NG+PF	36	CuT @ 64k (R1) μ-law	-26 dBov	0%	4.47	0.65	3.91	0.83
NS+NG+PF	37	CuT @ 64k (R1) μ-law	-16 dBov	0%	4.49	0.64	3.87	0.85
NS+NG+PF	38	CuT @ 64k (R1) μ-law	-36 dBov	0%	4.38	0.67	3.81	0.82
NS+NG+PF+FERC	39	CuT @ 64k (R1) μ-law	-26 dBov	3% (R)	4.38	0.71	3.49	0.92
NS+NG+PF+FERC	40	CuT @ 64k (R1) μ-law	-26 dBov	3% (B)	4.17	0.87	3.13	1.06
NG+PF	41	G.711Enc-CuTDec A-law	-26 dBov	0%	3.73	0.80	3.67	0.81
NG+PF	42	G.711Enc-CuTDec μ-law	-26 dBov	0%	4.18	0.72	3.78	0.82

Table 23 – G.711.1 Characterization Phase Step 2 - Experiment 2a results (music)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	5	Direct 1(1)	4.19	0.90	3.41	0.92
-	6	Direct 2(2)	4.30	0.78	3.50	0.97
Legacy G.711	7	G.711 @ 64k A-law	3.76	1.02	3.41	0.93
Legacy G.711	8	G.711 @ 64k μ-law	3.80	1.01	3.40	0.95
NS+NG	9	G.711.1 @ 64k (R1) A-law	3.98	0.94	3.48	0.89
NS+NG	10	G.711.1 @ 64k (R1) μ-law	4.04	0.92	3.51	0.94
NG	11	G.711Enc-G.711.1Dec A-law	3.75	1.01	3.47	0.91
NG	12	G.711Enc-G.711.1Dec μ-law	3.83	1.07	3.47	0.95
NS+NG+PF	13	CuT @ 64k (R1) A-law	3.93	1.00	3.47	0.93
NS+NG+PF	14	CuT @ 64k (R1) μ-law	3.97	1.01	3.42	0.96
NG+PF	15	G.711Enc-CuTDec A-law	3.85	0.98	3.43	0.88
NG+PF	16	G.711Enc-CuTDec μ-law	3.82	0.99	3.41	0.92

Notes:

(1) Direct 1: 8 kHz sampled input processed with FLAT1 filter in ITU-T STL 2005

(2) Direct 2: 16 kHz sampled input processed with P.341 filter in ITU-T STL 2005 down-sampled with HQ2.

Table 24 – G.711.1 Characterization Phase Step 2 - Experiment 3a results (background music SNR=25 dB)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	6	Direct	4.68	0.53	4.81	0.42
Legacy G.711	7	G.711 @ 64k A-law	4.19	0.72	3.79	0.75
Legacy G.711	8	G.711 @ 64k µ-law	4.23	0.72	4.38	0.71
NG	9	G.711Enc-G.711.1Dec A-law	4.20	0.74	3.73	0.76
NG	10	G.711Enc-G.711.1Dec µ-law	4.27	0.71	4.38	0.71
NG+PF	11	G.711Enc-CuTDec A-law	4.47	0.65	4.08	0.78
NG+PF	12	G.711Enc-CuTDec µ-law	4.53	0.60	4.53	0.67

Table 25 – G.711.1 Characterization Phase Step 2 - Experiment 3b results (office noise SNR=20 dB)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	6	Direct	4.72	0.52	4.56	0.59
Legacy G.711	7	G.711 @ 64k A-law	4.45	0.72	4.48	0.69
Legacy G.711	8	G.711 @ 64k µ-law	4.49	0.66	4.46	0.73
NG	9	G.711Enc-G.711.1Dec A-law	4.49	0.68	4.36	0.72
NG	10	G.711Enc-G.711.1Dec µ-law	4.44	0.70	4.41	0.64
NG+PF	11	G.711Enc-CuTDec A-law	4.42	0.76	4.48	0.70
NG+PF	12	G.711Enc-CuTDec µ-law	4.56	0.63	4.41	0.67

Table 26 – G.711.1 Characterization Phase Step 2 - Experiment 3c results (babble noise SNR=30 dB)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	6	Direct	4.74	0.47	4.85	0.41
Legacy G.711	7	G.711 @ 64k A-law	4.32	0.68	3.78	0.71
Legacy G.711	8	G.711 @ 64k µ-law	4.50	0.60	4.33	0.66
NG	9	G.711Enc-G.711.1Dec A-law	4.28	0.73	3.72	0.67
NG	10	G.711Enc-G.711.1Dec µ-law	4.31	0.68	4.27	0.71
NG+PF	11	G.711Enc-CuTDec A-law	4.53	0.63	4.15	0.66
NG+PF	12	G.711Enc-CuTDec µ-law	4.63	0.55	4.28	0.74

Table 27 – G.711.1 Characterization Phase Step 2 - Experiment 3d results (interfering talker SNR=15 dB)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	6	Direct	4.77	0.45	4.63	0.63
Legacy G.711	7	G.711 @ 64k A-law	4.21	0.78	4.17	0.80
Legacy G.711	8	G.711 @ 64k µ-law	4.26	0.76	4.39	0.69
NG	9	G.711Enc-G.711.1Dec A-law	4.17	0.71	4.10	0.84
NG	10	G.711Enc-G.711.1Dec µ-law	4.20	0.76	4.36	0.73
NG+PF	11	G.711Enc-CuTDec A-law	4.55	0.59	4.60	0.58
NG+PF	12	G.711Enc-CuTDec µ-law	4.54	0.66	4.63	0.57

10.3 Quality performance comparison

The comparisons, based on a statistical analysis, are computed with a confidence interval of 95%. The results of comparison are given by "nwt" for "not worse than" and "BT" for "better than". The results are provided in Table 28 and Table 29 for experiments 1a and 2a respectively and in Table 30 to Table 33 for the experiments with Background noise types.

Table 28 – G.711.1 Characterization Phase Step 2 – Comparisons in Experiment 1a (clean speech)

Conditions	Test N°	Test tool(s)	Ref N°	Ref tool(s)	Lab A	Lab B
A-law -26 dBov	31	NS+NG+PF	9	Legacy G.711	BT	BT
A-law -16 dBov	32	NS+NG+PF	10	Legacy G.711	BT	BT
A-law -36 dBov	33	NS+NG+PF	11	Legacy G.711	BT	BT
A-law -26 dBov 3% RFER	34	NS+NG+PF+FERC	12	Legacy G.711 + App. I	BT	BT
A-law -26 dBov 3% BFER	35	NS+NG+PF+FERC	13	Legacy G.711 + App. I	BT	BT
µ-law -26 dBov	36	NS+NG+PF	14	Legacy G.711	BT	BT
µ-law -16 dBov	37	NS+NG+PF	15	Legacy G.711	BT	BT
µ-law -36 dBov	38	NS+NG+PF	16	Legacy G.711	BT	BT
µ-law -26 dBov 3% RFER	39	NS+NG+PF+FERC	17	Legacy G.711 + App. I	BT	BT
µ-law -26 dBov 3% BFER	40	NS+NG+PF+FERC	18	Legacy G.711 + App. I	BT	BT
A-law -26 dBov	31	NS+NG+PF	19	NS+NG	BT	BT
A-law -16 dBov	32	NS+NG+PF	20	NS+NG	BT	nwt
A-law -36 dBov	32	NS+NG+PF	21	NS+NG	BT	nwt
A-law -26 dBov 3% RFER	34	NS+NG+PF+FERC	22	NS+NG+FERC	BT	nwt
A-law -26 dBov 3% BFER	35	NS+NG+PF+FERC	23	NS+NG+FERC	BT	nwt
µ-law -26 dBov	36	NS+NG+PF	24	NS+NG	BT	nwt
µ-law -16 dBov	37	NS+NG+PF	25	NS+NG	BT	nwt
µ-law -36 dBov	38	NS+NG+PF	26	NS+NG	BT	BT
µ-law -26 dBov 3% RFER	39	NS+NG+PF+FERC	27	NS+NG+FERC	BT	BT
µ-law -26 dBov 3% BFER	40	NS+NG+PF+FERC	28	NS+NG+FERC	BT	BT
A-law -26 dBov	41	NG+PF	29	NG	BT	BT
µ-law -26 dBov	42	NG+PF	30	NG	BT	nwt

Table 29 – G.711.1 Characterization Phase Step 2 - Comparisons in Experiment 2a (music)

Conditions	Test N°	Test tool(s)	Ref N°	Ref. tool(s)	Lab A	Lab B
A-law	13	NS+NG+PF	7	Legacy G.711	BT	nwt
µ-law	14	NS+NG+PF	8	Legacy G.711	BT	nwt
A-law	13	NS+NG+PF	9	NS+NG	nwt	nwt
µ-law	14	NS+NG+PF	10	NS+NG	nwt	fail
A-law	15	NG+PF	11	NG	nwt	nwt
µ-law	16	NG+PF	12	NG	nwt	nwt
A-law	11	NG	7	Legacy G.711	nwt	nwt
µ-law	12	NG	8	Legacy G.711	nwt	nwt
A-law	15	NG+PF	7	Legacy G.711	nwt	nwt
µ-law	16	NG+PF	8	Legacy G.711	nwt	nwt

Table 30 – G.711.1 Characterization Phase Step 2 - Comparisons in Experiment 3a (background music SNR=25dB)

Conditions	Test N°	Test tool(s)	Ref N°	Ref tool(s)	Lab A	Lab B
A-law	11	NG+PF	7	Legacy G.711	BT	BT
µ-law	12	NG+PF	8	Legacy G.711	BT	BT
A-law	11	NG+PF	9	NG	BT	BT
µ-law	12	NG+PF	10	NG	BT	BT
A-law	9	NG	7	Legacy G.711	nwt	nwt
µ-law	10	NG	8	Legacy G.711	nwt	nwt

Table 31 – G.711.1 Characterization Phase Step 2 - Comparisons in Experiment 3b (office noise SNR=20dB)

Conditions	Test N°	Test tool(s)	Ref N°	Ref. tool(s)	Lab A	Lab B
A-law	11	NG+PF	7	Legacy G.711	nwt	nwt
µ-law	12	NG+PF	8	Legacy G.711	nwt	nwt
A-law	11	NG+PF	9	NG	nwt	BT
µ-law	12	NG+PF	10	NG	BT	nwt
A-law	9	NG	7	Legacy G.711	nwt	fail
µ-law	10	NG	8	Legacy G.711	nwt	nwt

Table 32 – G.711.1 Characterization Phase Step 2 - Comparisons in Experiment 3c (babble noise SNR=30dB)

Conditions	Test N°	Test tool(s)	Ref N°	Ref. tool(s)	Lab A	Lab B
A-law	11	NG+PF	7	Legacy G.711	BT	BT
µ-law	12	NG+PF	8	Legacy G.711	BT	nwt
A-law	11	NG+PF	9	NG	BT	BT
µ-law	12	NG+PF	10	NG	BT	nwt
A-law	9	NG	7	Legacy G.711	nwt	nwt
µ-law	10	NG	8	Legacy G.711	fail	nwt

Table 33 – G.711.1 Characterization Phase Step 2 - Comparisons in Experiment 3d (interfering talker SNR=15dB)

Conditions	Test N°	Test tool(s)	Ref N°	Ref. tool(s)	Lab A	Lab B
A-law	11	NG+PF	7	Legacy G.711	BT	BT
µ-law	12	NG+PF	8	Legacy G.711	BT	BT
A-law	11	NG+PF	9	NG	BT	BT
µ-law	12	NG+PF	10	NG	BT	BT
A-law	9	NG	7	Legacy G.711	nwt	nwt
µ-law	10	NG	8	Legacy G.711	nwt	nwt

11 G.711 Appendix III verification phase

11.1 Quality performance assessment

The verification phase of the Appendix III to G.711 (also dubbed step 3) was performed in summer 2009. The quality assessment test plan designed by Q7/12 experts is found in [15]. Table 34gives an overview of the experiments.

The processing test plan is in [16]. Processing batch files were distributed to the processing laboratories. Common FER pattern files and background noise files (except interfering talkers) were also provided to these laboratories. These experiments were performed in two different languages and their processing performed twice by two different laboratories for cross-checking. The listening sessions were conducted in summer 2009 by two listening laboratories. Table 35 indicates for each experiment the two languages and the two listening/processing laboratories.

In G.711 Appendix III verification phase, the quality of one tool – PF – was assessed. Table 36 lists in which experiments and conditions (condition numbers are given in Table 37 to Table 39) this tool was tested.

Table 34 - Overview of the quality assessment tests in G.711 Appendix III verification phase

Exp	Methodology	Conditions
1	ACR	Clean speech for narrowband speech signals: different input levels, frame erasure and bitstream interoperability with G.711 legacy system coder
2	ACR	Music signal. Interoperability with G.711 legacy system coder
3	DCR	Speech quality with background noise types: – Exp. 3a: office noise SNR=20dB – Exp. 3b: babble noise SNR=30dB – Exp. 3c: interfering talker SNR=15dB

Table 35 - Languages and laboratory allocation in G.711 Appendix III verification phase

Exp	Lab A	Language A	Lab B	Language B
1	Huawei	Chinese	France Telecom	French
2	Huawei	Music	France Telecom	Music
3	Huawei	Chinese	France Telecom	French

Table 36: Tool combinations tested in G.711 Appendix III verification phase

Tool	Exp	Conditions
PF	1	Clean Speech: -26, -16, -36 dBov
PF + G.711 App. I	1	Clean Speech: 3% BFER at -26dBov (with G.711 Appendix I as PLC)
PF	2	Music
PF	3	Background noise types: – Exp. 3a: office noise SNR=20dB – Exp. 3b: babble noise SNR=30dB – Exp. 3c: interfering talker SNR=15dB

11.2 Quality test results

The results analysed by Q7/12 experts at their Rapporteurs' October 2009 meeting were reviewed by Q10/16 at SG 16 October 2009 plenary meeting [2]. Table 37 and Table 38 give the results of Experiment 1 and 2 respectively while Table 39 gives the results of Experiment 3 for the four types of background noise.

Table 37 – G.711 Appendix III verification phase - Experiment 1 results (clean speech)

Combinations	Number	Conditions	Level	BFER	meanA	stdA	meanB	stdB
-	8	Direct	-26 dBov	0%	4.50	0.67	4.43	0.68
Legacy G.711	9	G.711 @ 64k A-law	-26 dBov	0%	3.49	0.80	3.71	0.83
Legacy G.711	10	G.711 @ 64k A-law	-16 dBov	0%	4.12	0.72	4.10	0.83
Legacy G.711	11	G.711 @ 64k A-law	-36 dBov	0%	2.48	0.81	2.89	0.89
Legacy G.711+ App. I	12	G.711 @ 64k A-law w/ App. I	-26 dBov	3%	3.32	0.86	3.47	1.01
Legacy G.711	13	G.711 @ 64k µ-law	-26 dBov	0%	3.94	0.69	3.90	0.81
Legacy G.711	14	G.711 @ 64k µ-law	-16 dBov	0%	4.17	0.73	4.06	0.81
Legacy G.711	15	G.711 @ 64k µ-law	-36 dBov	0%	3.25	0.80	3.67	0.89
Legacy G.711+ App. I	16	G.711 @ 64k µ-law w/ App. I	-26 dBov	3%	3.76	0.76	3.73	0.98
PF	17	CuT A-law	-26 dBov	0%	4.18	0.73	4.25	0.78
PF	18	CuT A-law	-16 dBov	0%	4.45	0.72	4.27	0.72
PF	19	CuT A-law	-36 dBov	0%	4.07	0.72	3.78	0.92
PF+ G.711 App. I	20	CuT A-law w/ G.711 App. I	-26 dBov	3%	4.07	0.80	3.87	0.92
PF	21	CuT µ-law	-26 dBov	0%	4.42	0.67	4.27	0.77
PF	22	CuT µ-law	-16 dBov	0%	4.46	0.63	4.24	0.72
PF	23	CuT µ-law	-36 dBov	0%	4.34	0.74	4.17	0.78
PF+ G.711 App. I	24	CuT µ-law w/ G.711 App. I	-26 dBov	3%	4.17	0.78	3.95	0.90

Table 38 – G.711 Appendix III verification phase - Experiment 2 results (music)

Combinations	Number	Conditions	meanA	stdA	meanB	stdB
-	4	Direct	4.03	0.88	3.92	1.02
Legacy G.711	5	G.711 @ 64k A-law	3.99	0.85	3.85	0.91
Legacy G.711	6	G.711 @ 64k µ-law	3.92	0.89	3.91	0.96
PF	7	CuT A-law	3.97	0.87	3.89	0.99
PF	8	CuT µ-law	3.91	0.89	3.85	0.99

Table 39 – G.711 Appendix III verification phase - Experiment 3 results (various background noise types)

Combinations	Number	Conditions	Noise	meanA	stdA	meanB	stdB
-	4	Direct	Office noise SNR=20 dB	4.45	0.59	4.70	0.47
Legacy G.711	5	G.711 @ 64k A-law	Office noise SNR=20 dB	4.46	0.65	4.53	0.56
Legacy G.711	6	G.711 @ 64k µ-law	Office noise SNR=20 dB	4.40	0.61	4.50	0.59
PF	7	CuT A-law	Office noise SNR=20 dB	4.50	0.66	4.59	0.60
PF	8	CuT µ-law	Office noise SNR=20 dB	4.43	0.64	4.72	0.49
-	12	Direct	Babble noise SNR=30 dB	4.57	0.62	4.83	0.39

Combinations	Number	Conditions	Noise	meanA	stdA	meanB	stdB
Legacy G.711	13	G.711 @ 64k A-law	Babble noise SNR=30 dB	4.10	0.84	4.43	0.64
Legacy G.711	14	G.711 @ 64k µ-law	Babble noise SNR=30 dB	4.29	0.77	4.56	0.57
PF	15	CuT A-law	Babble noise SNR=30 dB	4.16	0.80	4.64	0.52
PF	16	CuT µ-law	Babble noise SNR=30 dB	4.25	0.77	4.61	0.58
-	20	Direct	Interfering talker SNR=15 dB	4.66	0.54	4.79	0.41
Legacy G.711	21	G.711 @ 64k A-law	Interfering talker SNR=15 dB	3.98	0.85	4.25	0.74
Legacy G.711	22	G.711 @ 64k µ-law	Interfering talker SNR=15 dB	4.26	0.76	4.34	0.60
PF	23	CuT A-law	Interfering talker SNR=15 dB	4.48	0.70	4.63	0.56
PF	24	CuT µ-law	Interfering talker SNR=15 dB	4.53	0.68	4.50	0.66

11.3 Quality performance comparison

The comparisons, based on a statistical analysis, are computed with a confidence interval of 95%. The results of comparison are given by "nwt" for "not worse than" and "BT" for "better than". The results are provided in Table 40 for Experiment 1, in Table 41 for Experiment 2 and in Table 42 for the experiment with various background noise types.

Table 40 – G.711 Appendix III verification phase - Comparisons in Experiment 1 (clean speech)

Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
A-law, -26 dBov	17	PF	9	Legacy G.711	BT	BT
A-law, -16 dBov	18	PF	10	Legacy G.711	BT	BT
A-law, -36 dBov	19	PF	11	Legacy G.711	BT	BT
A-law, -26 dBov, 3% BFER	20	PF+ G.711 App. I	12	Legacy G.711+ App. I	BT	BT
µ-law, -26 dBov	21	PF	13	Legacy G.711	BT	BT
µ-law, -16 dBov	22	PF	14	Legacy G.711	BT	BT
µ-law, -36 dBov	23	PF	15	Legacy G.711	BT	BT
µ -law, -26 dBov, 3% BFER	24	PF+ G.711 App. I	16	Legacy G.711 + App. I	BT	BT

Table 41 – G.711 Appendix III verification phase - Comparisons in Experiment 2 (music)

Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
A-law	7	PF	5	Legacy G.711	nwt	nwt
µ-law	8	PF	6	Legacy G.711	nwt	nwt

Table 42 – G.711 Appendix III verification phase - Comparisons in Experiment 3 (various office noise types)

Noise type	Conditions	Test N°	Test Tool(s)	Ref N°	Ref Tool(s)	Lab A	Lab B
Office noise SNR=20dB	A-law	7	PF	5	Legacy G.711	nwt	nwt
	µ-law	8	PF	6	Legacy G.711	nwt	BT
Babble noise SNR=30dB	A-law	15	PF	13	Legacy G.711	nwt	BT
	µ-law	16	PF	14	Legacy G.711	nwt	nwt
Interfering talker SNR=15dB	A-law	23	PF	21	Legacy G.711	BT	BT
	µ-law	24	PF	22	Legacy G.711	BT	BT

12 Overview of G.711 Appendix III quality performance

12.1 Clean speech signals

In clean speech, any combination of tools is better than legacy G.711 whatever the law, input level or frame erasure conditions. When G711 is used with either the noise shaping tool or the postfilter tool, it is also better than legacy G.711 for both laws. Whereas G.711, used only with the noise gate tool, is better than legacy G.711 in μ-law and not worse in A-law. The combination of the noise gate tool with the noise shaping tool is better than the noise gate tool alone for both laws. The combination of noise gate with the postfilter tool is better than the noise gate tool alone in A-law and not worse in μ-law.

12.2 Music signals

In presence of music, G.711 used with the combination of the noise shaping tool and the noise gate tool is better than legacy G.711 whichever the compression law. G.711 used with any other combinations or any tools used alone is not worse than legacy G.711.

12.3 Noisy signals

In presence of background music, the combination of the noise gate tool with either the noise shaping tool or the postfilter tool is better than legacy G.711. The combination of the noise gate tool with the postfilter tool is better than the noise gate tool used alone for both laws.

In presence of office noise, the combination of the noise gate tool with either the noise shaping tool or the postfilter tool is not worse than legacy G.711.

In presence of babble noise, the combination of the noise gate tool with the postfilter tool is better than legacy G.711 or the use of noise gate tool alone in A-law. In μ-law the combination of the noise gate tool with the noise shaping tool is better than legacy G.711.

In presence of interfering talker, in both laws, the combination of the noise gate tool with the postfilter tool is better than legacy G.711 or the noise gate tool used alone. The postfilter tool alone is also better than legacy G.711. The combination of the noise gate tool with the postfilter tool is better than the noise gate tool used alone for both laws. The combination of the noise gate tool with the noise shaping tool is better than legacy G.711 in A-law.