# Summary

This document describes the results of Experiments 1 and 2 of the 14kHz Low-Complexity Audio Coding Algorithm at 24, 32 and 48 kbps Extension to ITU-T G.722.1 Subjective Characterisation Tests performed by France Telecom. In the following, this codec will be called, for convenience, "G.722.1C". The test was performed according to the ITU-T Characterisation Test Plan using the processed speech material provided by France Telecom.

In experiment 1, G.722.1C met all performance requirements for a 99% and 95% confidence interval (those requirements were relative to the reference codec MPEG4 AAC-LD) and moreover it is even scored better than AAC-LD at 24 and 32 kbps.

For the characterisation test, in experiment 1, G.722.1C was also compared to other reference codecs, with the following results:

G.722.1C is scored worse than the 3GPP extended AMRWB codec and worse than the 3GPP enhanced aacPlus codec for a 95% confidence interval at 24 kbps.

G.722.1C is scored not worse than the 3GPP extended AMRWB codec and not worse than the 3GPP enhanced aacPlus codec for a 95% confidence interval at 32 kbps.

In experiment 2, G.722.1C met all performance requirements for a 99% and 95% confidence interval (those requirements were relative to the reference codec MPEG4 AAC-LD) and moreover it is even scored better than AAC-LD in most of the cases. For the characterisation test, in experiment 2, G.722.1C was also compared to other reference codecs, with the following results:

G.722.1C is scored worse than the 3GPP extended AMRWB codec for a 99% and 95% confidence interval in most of the cases.

G.722.1C is scored not worse than the 3GPP enhanced aacPlus codec for a 99% and 95% confidence interval in most of the cases.

# Contributors

This document has been reviewed by the ITU-T Q7/12 experts group under the chairmanship of the two Q7/12 Rapporteurs in liaison with ITU-T Q10/16.

**Q7/12 Rapporteurs**:

|  |  |
| --- | --- |
| Catherine Quinquis  France Telecom  France | Tel: +33 2 96 05 14 93  Fax: +33 2 96 05 3530  Email: [catherine.quinquis@francetelecom.com](mailto:catherine.quinquis@francetelecom.com) |
| Paolo Usai  ETSI  France | Tel: +33 4 92 94 42 36  Fax: +33 4 92 94 52 06  Email: [paolo.usai@etsi.org](mailto:paolo.usai@etsi.org) |

# Experiment 1

France Telecom tested G.722.1C for Experiment 1 of the Characterisation Test according to the specifications in the Characterisation Test Plan. France Telecom provided the processed speech material used in this experiment.

The codec performance was assessed for clean speech. The Experiment was performed using the 5-point scale ACR method with P.341extension-weighted speech.

Source speech for Experiment 1 was a sub-set of the French France Telecom listening lab speech database and consisted of 5 different sentence-pairs for each of the three male and three female talkers. France Telecom was also responsible for the processing of the source speech through all the Experiment 1 conditions, which are indicated in Table 1.

**Table 1** - Conditions for Experiment 1

|  |  |
| --- | --- |
| Number | Condition |
| C01 | Direct |
| C02 | MNRU Q=58dB |
| C03 | MNRU Q=50dB |
| C04 | MNRU Q=42dB |
| C05 | MNRU Q=34dB |
| C06 | MNRU Q=26dB |
| C07 | MNRU Q=18dB |
| C08 | Candidate codec at 24 kbps |
| C09 | MPEG-4 AAC-LD at 24 kbps |
| C10 | 3GPP extended AMRWB at 24 kbps |
| C11 | 3GPP enhanced aacPlus at 24 kbps |
| C12 | Candidate codec at 32kbps |
| C13 | MPEG-4 AAC-LD at 32 kbps |
| C14 | 3GPP extended AMRWB at 32 kbps |
| C15 | 3GPP enhanced aacPlus at 32 kbps |
| C16 | Candidate codec at 48 kbps |
| C17 | MPEG-4 AAC-LD at 48 kbps |
| C18 | MPEG-4 AAC-LD at 64 kbps |

The 32 non-expert French-speaking listeners (nominally balanced for gender) were arranged in four groups of 8 listeners who used headphones. The speech material was presented binaurally. The listening level was calibrated to generate a nominal listening level of -73dB SPL for a file at -26 dBov played on the computer.

The randomization sequences were used for Experiment 1. Votes were collected through a computer terminal and saved to a central computer, where the data were analyzed at 99% and 95% confidence interval.

The results of Experiment 1 are in table 2

**Table 2** – Mos scores for Experiment 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Averaged  MOS score on male speakers | SD male | Averaged  MOS score on female speakers | SD female | Averaged  MOS score | SD |
| 1 | 4.52 | 0.62 | 4.35 | 0.81 | 4.44 | 0.72 |
| 2 | 4.21 | 0.72 | 4.04 | 0.81 | 4.13 | 0.77 |
| 3 | 4.34 | 0.65 | 3.72 | 0.93 | 4.03 | 0.86 |
| 4 | 3.96 | 0.81 | 3.69 | 0.84 | 3.82 | 0.83 |
| 5 | 2.95 | 0.98 | 2.69 | 0.90 | 2.82 | 0.95 |
| 6 | 1.89 | 0.82 | 1.70 | 0.68 | 1.79 | 0.76 |
| 7 | 1.25 | 0.56 | 1.19 | 0.51 | 1.22 | 0.54 |
| 8 | 3.66 | 0.83 | 3.39 | 0.98 | 3.52 | 0.92 |
| 9 | 2.85 | 0.91 | 2.86 | 0.98 | 2.86 | 0.94 |
| 10 | 4.24 | 0.68 | 3.99 | 0.85 | 4.11 | 0.78 |
| 11 | 3.71 | 0.81 | 3.64 | 0.90 | 3.67 | 0.85 |
| 12 | 3.95 | 0.88 | 3.66 | 0.97 | 3.80 | 0.93 |
| 13 | 3.41 | 0.98 | 3.42 | 0.94 | 3.41 | 0.96 |
| 14 | 4.00 | 0.78 | 3.81 | 0.84 | 3.91 | 0.81 |
| 15 | 3.95 | 0.76 | 3.74 | 0.82 | 3.84 | 0.80 |
| 16 | 4.31 | 0.72 | 3.89 | 0.99 | 4.10 | 0.89 |
| 17 | 4.26 | 0.73 | 3.98 | 0.86 | 4.12 | 0.81 |
| 18 | 4.22 | 0.78 | 4.14 | 0.84 | 4.18 | 0.81 |







Verification of requirements for experiment 1:

**Table 3** – Requirements for Experiment 1

|  |  |
| --- | --- |
| Quality with clean speech | 1) at 24 kbit/s  Not worse than AAC-LD at 24 kbit/s ( 99% confidence level)  2) at 32kbit/s  Not worse than AAC-LD at 32 kbit/s ( 99% confidence level)  3) at 48 kbit/s  Not worse than AAC-LD at 48 kbit/s ( 99% confidence level) |

The comparisons are made between pairs of mean opinion scores Yrefand Ytest, having standard deviations sref and stest.

Therefore, the test codec is not worse than the reference codec if:

where nref and ntest are the numbers of observations for each mean, (1-α) is the confidence level, ν = nref+ntest-2 is the number of degrees of freedom, and so2 is the pooled estimator of the common variance s2 given by:

For a confidence level of 99%, , the test codec is not worse than the reference codec if: :

For a confidence level of 95%, , the test codec is not worse than the reference codec if: :

G.722.1C passes all requirements for a 99% confidence interval as can be shown in table 4. It is also scored not worse than the reference codec MPEG AAC-LD for a 95% confidence interval

Compared with the other reference codecs, G.722.1C is worse than 3GPP enhanced aacPlus at 24 kbps for a 95% confidence interval. G.722.1C is worse than 3GPP extended AMRWB at 24 kbps for a for a 95% confidence interval. G.722.1C is not worse than 3GPP enhanced aacPlus at 32 kbps for a 95% confidence interval. G.722.1C is not worse than 3GPP extended AMRWB at 32 kbps for a for a 95% confidence interval.

**Table 4** – Comparison with "not worse than" criterion for Experiment 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | Mean |  | Not worse than  At 99% CI |  | Not worse than  95% CI |  |
| Cut at 24kbps | 3.52 | 0.92 |  |  |  |  |
| AAC-LD at 24kbps | 2.86 | 0.94 | 2.67 | true | 2.72 | true |
| AMRWB+ at 24kbps | 4.11 | 0.78 | 3.94 | false | 3.99 | false |
| AacPlus at 24kbps | 3.67 | 0.85 | 3.49 | true | 3.54 | false |
| Cut at 32kbps | 3.80 | 0.93 |  |  |  |  |
| AAC-LD at 32kbps | 3.41 | 0.96 | 3.22 | true | 3.27 | true |
| AMRWB+ at 32kbps | 3.91 | 0.81 | 3.72 | true | 3.78 | true |
| AacPlus at 32kbps | 3.84 | 0.80 | 3.66 | true | 3.72 | true |
| Cut at 48kbps | 4.10 | 0.89 |  |  |  |  |
| AAC-LD at 48kbps | 4.12 | 0.81 | 3.94 | true | 4.00 | true |
| AAC-LD at 64kbps | 4.18 | 0.81 | 4.00 | true | 4.05 | true |

The comparisons are made between pairs of mean opinion scores Yrefand Ytest, having standard deviations sref and stest.

Therefore, the test codec is better than the reference codec if:

where nref and ntest are the numbers of observations for each mean, (1-α) is the confidence level, ν = nref+ntest-2 is the number of degrees of freedom, and so2 is the pooled estimator of the common variance s2 given by:

For a confidence level of 99%, , the test codec is better than the reference codec if: :

For a confidence level of 95%, , the test codec is better than the reference codec if: :

In table 5, it is shown that G.722.1C is better than the reference codec MPEG AAC-LD for a 95% confidence interval at the same bit rate for 24kbps and 32 kbps.

**Table 5** – Comparison with "better than" criterion for Experiment 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | Mean |  | Better than at  99% CI |  | Better than at  95% CI |  |
| Cut at 24kbps | 3.52 | 0.92 |  |  |  |  |
| AAC-LD at 24kbps | 2.86 | 0.94 | 3.05 | true | 3.00 | true |
| AMRWB+ at 24kbps | 4.11 | 0.78 | 4.29 | false | 4.24 | false |
| AacPlus at 24kbps | 3.67 | 0.85 | 3.86 | false | 3.80 | false |
| Cut at 32kbps | 3.80 | 0.93 |  |  |  |  |
| AAC-LD at 32kbps | 3.41 | 0.96 | 3.61 | true | 3.55 | true |
| AMRWB+ at 32kbps | 3.91 | 0.81 | 4.09 | false | 4.03 | false |
| AacPlus at 32kbps | 3.84 | 0.80 | 4.02 | false | 3.97 | false |
| Cut at 48kbps | 4.10 | 0.89 |  |  |  |  |
| AAC-LD at 48kbps | 4.12 | 0.81 | 4.30 | false | 4.24 | false |
| AAC-LD at 64kbps | 4.18 | 0.81 | 4.35 | false | 4.30 | false |

# Experiment 2

France Telecom tested G.722.1C for Experiment 2 of the Characterisation Test according to the specifications in the Characterisation Test Plan. France Telecom provided the processed speech material used in this experiment.

The codec performance was assessed for reverberant noisy speech. The Experiment was performed using the 5-point scale DCR method with P.341extension-weighted speech.

Source speech for Experiment 2 was a sub-set of the French France Telecom listening lab speech database and consisted of 5 different sentence-pairs for each of the three male and three female talkers. France Telecom was also responsible for the processing of the source speech through all the Experiment 2 conditions, which are indicated in Table 6.

Office noise as well as interfering talker condition were also generated by France Telecom.

**Table 6** - Conditions for Experiment 2

|  |  |
| --- | --- |
| Number | Condition |
| C01 | Direct |
| C02 | MNRU Q=58dB |
| C03 | MNRU Q=50dB |
| C04 | MNRU Q=42dB |
| C05 | MNRU Q=34dB |
| C06 | MNRU Q=26dB |
| C07 | MNRU Q=18dB |
| C08 | Candidate codec at 24 kbps |
| C09 | MPEG-4 AAC-LD at 24 kbps |
| C10 | 3GPP extended AMRWB at 24 kbps |
| C11 | 3GPP enhanced aacPlus at 24 kbps |
| C12 | Candidate codec at 32kbps |
| C13 | MPEG-4 AAC-LD at 32 kbps |
| C14 | 3GPP extended AMRWB at 32 kbps |
| C15 | 3GPP enhanced aacPlus at 32 kbps |
| C16 | Candidate codec at 48 kbps |
| C17 | MPEG-4 AAC-LD at 48 kbps |
| C18 | MPEG-4 AAC-LD at 64 kbps |

The 32 non-expert French-speaking listeners (nominally balanced for gender) were arranged in four groups of 8 listeners who used headphones. The speech material was presented binaurally. The listening level was calibrated to generate a nominal listening level of -73dB SPL for a file at -26 dBov played on the computer.

The randomization sequences were used for Experiment 2. Votes were collected through a computer terminal and saved to a central computer, where the data were analyzed at 99% and 95% confidence interval.

Experiment 2a:

The results of Experiment 2a are in table 7

**Table 7** – Mos scores for Experiment 2a

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Averaged  MOS score on male speakers | SD male | Averaged  MOS score on female speakers | SD female | Averaged  MOS score | SD |
| 1 | 4.63 | 0.58 | 4.48 | 0.78 | 4.55 | 0.69 |
| 2 | 4.48 | 0.68 | 4.51 | 0.71 | 4.49 | 0.69 |
| 3 | 4.35 | 0.74 | 4.40 | 0.76 | 4.38 | 0.75 |
| 4 | 3.40 | 1.01 | 3.59 | 0.89 | 3.49 | 0.95 |
| 5 | 2.30 | 1.10 | 2.50 | 1.09 | 2.40 | 1.09 |
| 6 | 1.59 | 0.80 | 1.59 | 0.78 | 1.59 | 0.79 |
| 7 | 1.18 | 0.46 | 1.13 | 0.36 | 1.15 | 0.41 |
| 8 | 3.69 | 1.04 | 4.09 | 0.88 | 3.89 | 0.98 |
| 9 | 2.44 | 1.05 | 2.83 | 1.10 | 2.64 | 1.09 |
| 10 | 4.14 | 0.82 | 4.20 | 0.82 | 4.17 | 0.81 |
| 11 | 3.34 | 0.93 | 3.61 | 0.98 | 3.48 | 0.96 |
| 12 | 4.08 | 0.91 | 4.28 | 0.80 | 4.18 | 0.86 |
| 13 | 3.39 | 0.91 | 3.67 | 1.00 | 3.53 | 0.97 |
| 14 | 4.32 | 0.75 | 4.41 | 0.79 | 4.36 | 0.77 |
| 15 | 3.83 | 0.80 | 4.10 | 0.86 | 3.97 | 0.84 |
| 16 | 4.36 | 0.77 | 4.44 | 0.69 | 4.40 | 0.73 |
| 17 | 4.31 | 0.74 | 4.43 | 0.78 | 4.37 | 0.76 |
| 18 | 4.43 | 0.68 | 4.49 | 0.66 | 4.46 | 0.67 |







Verification of requirements for experiment 2a:

**Table 8** – Requirements for Experiment 2a

|  |  |
| --- | --- |
| Quality with Reverberant Speech + office noise (SNR 15 dB) | 1) at 24 kbit/s  Not worse than AAC-LD at 24 kbit/s ( 99% confidence level)  2) at 32kbit/s  Not worse than AAC-LD at 32 kbit/s ( 99% confidence level)  3) at 48 kbit/s  Not worse than AAC-LD at 48 kbit/s ( 99% confidence level) |

The comparisons are made between pairs of mean opinion scores Yrefand Ytest, having standard deviations sref and stest.

Therefore, the test codec is not worse than the reference codec if:

where nref and ntest are the numbers of observations for each mean, (1-α) is the confidence level, ν = nref+ntest-2 is the number of degrees of freedom, and so2 is the pooled estimator of the common variance s2 given by:

For a confidence level of 99%, , the test codec is not worse than the reference codec if: :

For a confidence level of 95%, , the test codec is not worse than the reference codec if: :

G.722.1C passes all requirements for a 99% confidence interval as can be shown in table 9. It is also scored not worse than the reference codec MPEG AAC-LD for a 95% confidence interval

Compared with the other reference codecs, G.722.1C is not worse than 3GPP enhanced aacPlus at the same bit rate for 24 and 32 kbps for a 95% confidence interval. G.722.1C is worse than 3GPP extended AMRWB for a 95% confidence interval.

**Table 9** – Comparison with "not worse than" criterion for Experiment 2a

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | Mean |  | Not worse than  At 99% CI |  | Not worse than  95% CI |  |
| Cut at 24kbps | 3.89 | 0.98 |  |  |  |  |
| AAC-LD at 24kbps | 2.64 | 1.09 | 2.42 | true | 2.48 | true |
| AMRWB+ at 24kbps | 4.17 | 0.81 | 3.98 | false | 4.03 | false |
| AacPlus at 24kbps | 3.48 | 0.96 | 3.28 | true | 3.34 | true |
| Cut at 32kbps | 4.18 | 0.86 |  |  |  |  |
| AAC-LD at 32kbps | 3.53 | 0.97 | 3.34 | true | 3.39 | true |
| AMRWB+ at 32kbps | 4.36 | 0.77 | 4.19 | false | 4.24 | false |
| AacPlus at 32kbps | 3.97 | 0.84 | 3.79 | true | 3.84 | true |
| Cut at 48kbps | 4.40 | 0.73 |  |  |  |  |
| AAC-LD at 48kbps | 4.37 | 0.76 | 4.21 | true | 4.26 | true |
| AAC-LD at 64kbps | 4.46 | 0.67 | 4.31 | true | 4.36 | true |

The comparisons are made between pairs of mean opinion scores Yrefand Ytest, having standard deviations sref and stest.

Therefore, the test codec is better than the reference codec if:

where nref and ntest are the numbers of observations for each mean, (1-α) is the confidence level, ν = nref+ntest-2 is the number of degrees of freedom, and so2 is the pooled estimator of the common variance s2 given by:

For a confidence level of 99%, , the test codec is better than the reference codec if: :

For a confidence level of 95%, , the test codec is better than the reference codec if: :

In table 10, it is shown that G.722.1C is better than the reference codec MPEG AAC-LD for a 95% confidence interval at the same bit rate for 24kbps and 32 kbps.

Compared with the other reference codecs, G.722.1C is better than 3GPP enhanced aacPlus for a 95% confidence interval.

**Table 10** – Comparison with "better than" criterion for Experiment 2a

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | Mean |  | Better than at  99% CI |  | Better than at  95% CI |  |
| Cut at 24kbps | 3.89 | 0.98 |  |  |  |  |
| AAC-LD at 24kbps | 2.64 | 1.09 | 2.85 | true | 2.79 | true |
| AMRWB+ at 24kbps | 4.17 | 0.81 | 4.35 | false | 4.30 | false |
| AacPlus at 24kbps | 3.48 | 0.96 | 3.68 | true | 3.62 | true |
| Cut at 32kbps | 4.18 | 0.86 |  |  |  |  |
| AAC-LD at 32kbps | 3.53 | 0.97 | 3.72 | true | 3.66 | true |
| AMRWB+ at 32kbps | 4.36 | 0.77 | 4.53 | false | 4.48 | false |
| AacPlus at 32kbps | 3.97 | 0.84 | 4.15 | true | 4.09 | true |
| Cut at 48kbps | 4.40 | 0.73 |  |  |  |  |
| AAC-LD at 48kbps | 4.37 | 0.76 | 4.52 | false | 4.48 | false |
| AAC-LD at 64kbps | 4.46 | 0.67 | 4.60 | false | 4.56 | false |

Experiment 2b:

The results of Experiment 2b are in table 11

**Table 11** – Mos scores for Experiment 2b

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Averaged  MOS score on male speakers | SD male | Averaged  MOS score on female speakers | SD female | Averaged  MOS score | SD |
| 1 | 4.74 | 0.53 | 4.81 | 0.42 | 4.78 | 0.48 |
| 2 | 4.61 | 0.65 | 4.67 | 0.57 | 4.64 | 0.61 |
| 3 | 4.63 | 0.67 | 4.70 | 0.53 | 4.66 | 0.60 |
| 4 | 4.17 | 0.82 | 4.33 | 0.71 | 4.25 | 0.77 |
| 5 | 3.18 | 1.03 | 3.23 | 1.01 | 3.20 | 1.02 |
| 6 | 1.81 | 0.72 | 1.89 | 0.74 | 1.85 | 0.73 |
| 7 | 1.23 | 0.47 | 1.21 | 0.48 | 1.22 | 0.47 |
| 8 | 3.79 | 0.96 | 4.22 | 0.88 | 4.01 | 0.95 |
| 9 | 2.63 | 1.02 | 2.99 | 1.09 | 2.81 | 1.07 |
| 10 | 4.36 | 0.74 | 4.43 | 0.64 | 4.40 | 0.69 |
| 11 | 3.81 | 1.01 | 4.05 | 0.85 | 3.93 | 0.94 |
| 12 | 4.24 | 0.90 | 4.52 | 0.77 | 4.38 | 0.85 |
| 13 | 3.32 | 1.01 | 3.82 | 1.10 | 3.57 | 1.09 |
| 14 | 4.42 | 0.84 | 4.42 | 0.83 | 4.42 | 0.83 |
| 15 | 4.28 | 0.76 | 4.40 | 0.73 | 4.34 | 0.75 |
| 16 | 4.66 | 0.58 | 4.79 | 0.50 | 4.72 | 0.54 |
| 17 | 4.50 | 0.65 | 4.50 | 0.81 | 4.50 | 0.73 |
| 18 | 4.69 | 0.55 | 4.68 | 0.62 | 4.68 | 0.59 |







Verification of requirements for experiment 2b:

**Table 12** – Requirements for Experiment 2b

|  |  |
| --- | --- |
| Quality with Reverberant Speech + interfering talker noise (SNR 15 dB) | 1) at 24 kbit/s  Not worse than AAC-LD at 24 kbit/s ( 99% confidence level)  2) at 32kbit/s  Not worse than AAC-LD at 32 kbit/s ( 99% confidence level)  3) at 48 kbit/s  Not worse than AAC-LD at 48 kbit/s ( 99% confidence level) |

The comparisons are made between pairs of mean opinion scores Yrefand Ytest, having standard deviations sref and stest.

Therefore, the test codec is not worse than the reference codec if:

Where nref and ntest are the numbers of observations for each mean, (1-α) is the confidence level, ν = nref+ntest-2 is the number of degrees of freedom, and so2 is the pooled estimator of the common variance s2 given by:

For a confidence level of 99%, , the test codec is not worse than the reference codec if: :

For a confidence level of 95%, , the test codec is not worse than the reference codec if: :

G.722.1C passes all requirements for a 99% confidence interval as can be shown in table 13. It is also scored not worse than the reference codec MPEG AAC-LD for a 95% confidence interval

Compared with the other reference codecs, G.722.1C is not worse than 3GPP enhanced aacPlus at the same bit rate for 24 and 32 kbps for a 95% confidence interval. G.722.1C is not worse than 3GPP extended AMRWB for a 99% confidence interval at 32kbps and worse for the other cases

**Table 13** – Comparison with "not worse than" criterion for Experiment 2b

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | Mean |  | Not worse than  At 99% CI |  | Not worse than  95% CI |  |
| Cut at 24kbps | 3.79 | 0.96 |  |  |  |  |
| AAC-LD at 24kbps | 2.63 | 1.02 | 2.60 | true | 2.66 | true |
| AMRWB+ at 24kbps | 4.36 | 0.74 | 4.22 | false | 4.27 | false |
| AacPlus at 24kbps | 3.81 | 1.01 | 3.74 | true | 3.79 | true |
| Cut at 32kbps | 4.24 | 0.90 |  |  |  |  |
| AAC-LD at 32kbps | 3.32 | 1.01 | 3.37 | true | 3.43 | true |
| AMRWB+ at 32kbps | 4.42 | 0.84 | 4.24 | true | 4.29 | false |
| AacPlus at 32kbps | 4.28 | 0.76 | 4.17 | true | 4.22 | true |
| Cut at 48kbps | 4.66 | 0.58 |  |  |  |  |
| AAC-LD at 48kbps | 4.50 | 0.65 | 4.37 | true | 4.41 | true |
| AAC-LD at 64kbps | 4.69 | 0.55 | 4.56 | true | 4.60 | true |

The comparisons are made between pairs of mean opinion scores Yrefand Ytest, having standard deviations sref and stest.

Therefore, the test codec is better than the reference codec if:

where nref and ntest are the numbers of observations for each mean, (1-α) is the confidence level, ν = nref+ntest-2 is the number of degrees of freedom, and so2 is the pooled estimator of the common variance s2 given by:

For a confidence level of 99%, , the test codec is better than the reference codec if: :

For a confidence level of 95%, , the test codec is better than the reference codec if: :

In table 14, it is shown that G.722.1C is better than the reference codec MPEG AAC-LD for a 95% confidence interval at the same bit rate for 24kbps, 32 kbps and 48 kbps.

**Table 14** – Comparison with "better than" criterion for Experiment 2b

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | Mean |  | Better than at  99% CI |  | Better than at  95% CI |  |
| Cut at 24kbps | 3.79 | 0.96 |  |  |  |  |
| AAC-LD at 24kbps | 2.63 | 1.02 | 3.02 | true | 2.95 | true |
| AMRWB+ at 24kbps | 4.36 | 0.74 | 4.57 | false | 4.52 | false |
| AacPlus at 24kbps | 3.81 | 1.01 | 4.13 | false | 4.07 | false |
| Cut at 32kbps | 4.24 | 0.90 |  |  |  |  |
| AAC-LD at 32kbps | 3.32 | 1.01 | 3.78 | true | 3.72 | true |
| AMRWB+ at 32kbps | 4.42 | 0.84 | 4.59 | false | 4.54 | false |
| AacPlus at 32kbps | 4.28 | 0.76 | 4.50 | false | 4.46 | false |
| Cut at 48kbps | 4.66 | 0.58 |  |  |  |  |
| AAC-LD at 48kbps | 4.50 | 0.65 | 4.63 | true | 4.59 | true |
| AAC-LD at 64kbps | 4.69 | 0.55 | 4.80 | false | 4.77 | false |

Experiment 2c:

The results of Experiment 2c are in table 15

**Table 15** – Mos scores for Experiment 2c

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Averaged  MOS score on male speakers | SD male | Averaged  MOS score on female speakers | SD female | Averaged  MOS score | SD |
| 1 | 4.52 | 0.70 | 4.66 | 0.48 | 4.59 | 0.60 |
| 2 | 4.41 | 0.78 | 4.49 | 0.71 | 4.45 | 0.74 |
| 3 | 4.34 | 0.86 | 4.51 | 0.66 | 4.43 | 0.77 |
| 4 | 3.97 | 0.93 | 4.10 | 0.86 | 4.04 | 0.90 |
| 5 | 2.81 | 1.08 | 3.07 | 1.10 | 2.94 | 1.09 |
| 6 | 1.75 | 0.68 | 1.90 | 0.83 | 1.82 | 0.76 |
| 7 | 1.11 | 0.32 | 1.25 | 0.52 | 1.18 | 0.44 |
| 8 | 3.64 | 1.12 | 3.92 | 1.02 | 3.78 | 1.08 |
| 9 | 2.41 | 1.17 | 2.73 | 1.24 | 2.57 | 1.21 |
| 10 | 4.09 | 0.87 | 4.25 | 0.82 | 4.17 | 0.85 |
| 11 | 3.60 | 0.99 | 3.92 | 1.04 | 3.76 | 1.03 |
| 12 | 4.10 | 0.95 | 4.22 | 0.84 | 4.16 | 0.89 |
| 13 | 3.10 | 1.11 | 3.60 | 1.01 | 3.35 | 1.09 |
| 14 | 4.18 | 0.87 | 4.17 | 0.95 | 4.17 | 0.91 |
| 15 | 4.17 | 0.96 | 4.15 | 0.83 | 4.16 | 0.90 |
| 16 | 4.42 | 0.72 | 4.27 | 0.92 | 4.34 | 0.83 |
| 17 | 4.25 | 0.86 | 4.40 | 0.75 | 4.32 | 0.81 |
| 18 | 4.40 | 0.81 | 4.45 | 0.74 | 4.42 | 0.78 |







Verification of requirements for experiment 2c:

**Table 16** – Requirements for Experiment 2c

|  |  |
| --- | --- |
| Quality with Reverberant Speech + interfering talker noise (SNR 15 dB) | 1) at 24 kbit/s  Not worse than AAC-LD at 24 kbit/s ( 99% confidence level)  2) at 32kbit/s  Not worse than AAC-LD at 32 kbit/s ( 99% confidence level)  3) at 48 kbit/s  Not worse than AAC-LD at 48 kbit/s ( 99% confidence level) |

The comparisons are made between pairs of mean opinion scores Yrefand Ytest, having standard deviations sref and stest.

Therefore, the test codec is not worse than the reference codec if:

where nref and ntest are the numbers of observations for each mean, (1-α) is the confidence level, ν = nref+ntest-2 is the number of degrees of freedom, and so2 is the pooled estimator of the common variance s2 given by:

For a confidence level of 99%, , the test codec is not worse than the reference codec if: :

For a confidence level of 95%, , the test codec is not worse than the reference codec if: :

G.722.1C passes all requirements for a 99% confidence interval as can be shown in table 17. It is also scored not worse than the reference codec MPEG AAC-LD for a 95% confidence interval

Compared with the other reference codecs, G.722.1C is not worse than 3GPP enhanced aacPlus for a 95% confidence interval. G.722.1C is not worse than 3GPP extended AMRWB for a 99% confidence interval at the bit rate of 32kbps and worse for the other cases

**Table 17** – Comparison with "not worse than" criterion for Experiment 2c

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | Mean |  | Not worse than  At 99% CI |  | Not worse than  95% CI |  |
| Cut at 24kbps | 3.78 | 1.08 |  |  |  |  |
| AAC-LD at 24kbps | 2.57 | 1.21 | 2.33 | true | 2.40 | true |
| AMRWB+ at 24kbps | 4.17 | 0.85 | 3.97 | false | 4.03 | false |
| AacPlus at 24kbps | 3.76 | 1.03 | 3.54 | true | 3.61 | true |
| Cut at 32kbps | 4.16 | 0.89 |  |  |  |  |
| AAC-LD at 32kbps | 3.35 | 1.09 | 3.15 | true | 3.21 | true |
| AMRWB+ at 32kbps | 4.17 | 0.91 | 3.98 | true | 4.04 | true |
| AacPlus at 32kbps | 4.16 | 0.90 | 3.97 | true | 4.03 | true |
| Cut at 48kbps | 4.34 | 0.83 |  |  |  |  |
| AAC-LD at 48kbps | 4.32 | 0.81 | 4.15 | true | 4.20 | true |
| AAC-LD at 64kbps | 4.42 | 0.78 | 4.25 | true | 4.30 | true |

The comparisons are made between pairs of mean opinion scores Yrefand Ytest, having standard deviations sref and stest.

Therefore, the test codec is better than the reference codec if:

Where nref and ntest are the numbers of observations for each mean, (1-α) is the confidence level, ν = nref+ntest-2 is the number of degrees of freedom, and so2 is the pooled estimator of the common variance s2 given by:

For a confidence level of 99%, , the test codec is better than the reference codec if: :

For a confidence level of 95%, , the test codec is better than the reference codec if: :

In table 18, it is shown that G.722.1C is better than the reference codec MPEG AAC-LD for a 95% confidence interval at the same bit rate for 24kbps and 32 kbps.

**Table 18** – Comparison with "better than" criterion for Experiment 2c

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | Mean |  | Better than at  99% CI |  | Better than at  95% CI |  |
| Cut at 24kbps | 3.78 | 1.08 |  |  |  |  |
| AAC-LD at 24kbps | 2.57 | 1.21 | 2.81 | true | 2.74 | true |
| AMRWB+ at 24kbps | 4.17 | 0.85 | 4.37 | false | 4.31 | false |
| AacPlus at 24kbps | 3.76 | 1.03 | 3.98 | false | 3.91 | false |
| Cut at 32kbps | 4.16 | 0.89 |  |  |  |  |
| AAC-LD at 32kbps | 3.35 | 1.09 | 3.56 | true | 3.50 | true |
| AMRWB+ at 32kbps | 4.17 | 0.91 | 4.36 | false | 4.30 | false |
| AacPlus at 32kbps | 4.16 | 0.90 | 4.34 | false | 4.29 | false |
| Cut at 48kbps | 4.34 | 0.83 |  |  |  |  |
| AAC-LD at 48kbps | 4.32 | 0.81 | 4.49 | false | 4.44 | false |
| AAC-LD at 64kbps | 4.42 | 0.78 | 4.59 | false | 4.54 | false |