|  |  |
| --- | --- |
| 3GPP TR 26.940 V0.2.0 (2025-05) | |
| Technical Report  In the document, delete either "TS" and "Specification" or "TR" and "Report" as applicable. These instances are shown with yellow highlighting. Also ensure the copyright date, version, spec number, title and Release (also highlighted) are correct. Below, replace <TSG name> by the [appropriate text](#tsgNames). | |
| 3rd Generation Partnership Project;  Technical Specification Group Services and System Aspects;  Study on Ultra Low Bit rate Speech Codecs  (Release 20) | |
|  | |
|  |  |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  https://www.3gpp.org |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2025, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 4

Introduction 4

1 Scope 5

2 References 6

3 Definitions of terms, symbols and abbreviations 6

3.1 Terms 6

3.2 Symbols 6

3.3 Abbreviations 7

4 Application scenarios for ultra-low bit rate communication services 7

5 Channel characteristics and service-related dependencies 7

6 Design constraints 7

7 Existing technologies and feasibility evidence 8

8 Performance requirements 8

9 Test methodologies 8

10 Considered work plan for potential normative work 8

11 Conclusions and recommendations 9

Proforma copyright release text block 10

X.1 The right to copy 10

Abstract Test Suite (ATS) text block 11

Y Abstract Test Suite (ATS) 11

Y.1 Introduction 11

Y.2 The TTCN Graphical form (TTCN.GR) 11

Y.3 The TTCN Machine Processable form (TTCN.MP) 11

Annex <A> (normative): <Normative annex for a Technical Specification> 12

Annex <B> (informative): <Informative annex for a Technical Specification> 13

B.1 Heading levels in an annex 13

Annex <F> (informative): Change history 15

For definitive guidance on drafting 3GPP TSs and TRs, see [3GPP TS 21.801](https://www.3gpp.org/DynaReport/21801.htm).

Ensure all blue guidance text is removed before submitting the TS/TR to the TSG for approval.

# Foreword

This clause is mandatory; do not alter the text in any way other than to choose between "Specification" and "Report".

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

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

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In drafting the TS/TR, pay particular attention to the use of modal auxiliary verbs! TRs shall not contain any normative provisions.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# Introduction

This clause is optional. If it exists, it shall be the second unnumbered clause.

# 1 Scope

This clause shall start on a new page.

[The present document develops recommendations for potential normative work on an ultra-low bit rate codec for the use case of IMS voice services over Geostationary Orbit (GEO) access.]

Editor’s Note:

3GPP SA1 has studied the use case IMS Voice Call Using GEO Access, and the results are documented in TR 22.887. Normative service requirements and KPIs on IMS voice call using GEO satellite access will be introduced in TS 22.261 at TSG#107. GEO satellites are on a 35,786 km distance from the earth, which noticeably impacts signal propagation delay (one way approx. 285ms), data rate, and channel conditions due to e.g. atmospheric attenuation. Compared to terrestrial links, this poses significant new challenges for the voice codecs and services:

The overall transmission data rate assumed for GEO satellite systems is very constrained due to e.g. high path loss, atmospheric attenuation, energy constraints for terminals etc.. In TR 22.887, a total transmission data rate of [1-3] kbit/s is assumed. This transmission data rate are lower than what current 3GPP protocol stacks and codecs can supports.

For GEO satellite access, the propagation delay (285ms) is much longer than for commonly used terrestrial links.

The GEO satellite link imposes different channel characteristics, e.g., due to atmospheric attenuation.

Currently, no 3GPP voice codec seems to support all the expected requirements for this use case. Considering bitrate alone, the lowest supported bitrate of any 3GPP codec is 4.75 kbit/s as provided by the narrow band AMR codec (TS 26.071). This makes it necessary to have a new feasibility study relating to ultra-low bitrate codecs suitable for voice using GEO access.

The primary focus of this study is to develop design constraints and performance requirements for a codec supporting use cases like IMS Voice Call over GEO and the resulting transmission parameters. The requirements can provide guidance on the evaluation of the candidate codecs during potential normative work.

**1. General considerations**

**- Bitrate:** TR 22.887 concludes that the transmission rates are lower than what current 3GPP protocol stacks and codecs can supports. Detailed analysis on available bitrate requires more study..

**- Quality:** Despite of the low bit rate, a good audio quality of the codec is of importance, to ensure a reasonable QoE. Detailed QoE requirements for such services are for study..

**- Complexity and memory demands:** Modern low bitrate codecs exhibit a large scale of complexity and memory demands. The codec is expected to be deployable on the processing capabilities as can be found in today’s smartphones. Exact complexity requirements are for study.

**- Robustness to network conditions**: the codec is expected to operate in typical network conditions (delay, loss, jitter, etc.). Details are for further study.

**2. Functional requirements**

- **Speech transcoding functions**: To achieve integration with the terrestrial voice communication system (4G/5G IMS architecture), it is necessary to consider tandeming with existing IMS voice codecs.

NOTE: Additional study areas or use cases, such as assessing the market potential and potential market-readiness of a new ULBC codec should be added with lower priority if time permits and once the exact requirements can be given.

It is expected that coordination with other working groups, e.g. SA2, CT1, RAN2 is needed in order to substantiate the design constraints of such a codec. However, it is not expected that this work creates any dependency for studies and normative in other working groups.

# 2 References

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

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

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

[22261] 3GPP TS 22.261: "Service requirements for the 5G system".

[22887] 3GPP TR 22.887: "Feasibility Study on satellite access - Phase 4".

[26071] 3GPP TS 26.071: "Mandatory speech CODEC speech processing functions; AMR speech Codec; General description".

[7-1] T. Muller et al., “Speech quality evaluation of neural audio codecs”, Interspeech 2024, https://www.isca-archive.org/interspeech\_2024/muller24c\_interspeech.pdfers, this corresponds to ca. 900 Mbyte. Other solution like SNAC operate with 19M parameters which correspond to 18 Mbyte. Implementations may also use 32bit floating point parameters which leads to an ROM increase by a factor four.

[7-2] Tdoc S4-250565, “Analysis of existing technologies for ULBC”

[7-3] <https://github.com/drowe67/codec2>

[7-4] <https://ffmpeg.org/>

[7-5] <https://melp.org/>

[7-6] <https://en.wikipedia.org/wiki/Harmonic_Vector_Excitation_Coding>

[7-7] <https://github.com/descriptinc/descript-audio-codec>

[7-8] <https://bellard.org/tsac/>

[7-9] ITU-T P.800 : Methods for subjective determination of transmission quality (08/1996)

[D.1] 3GPP [TS 26.131](https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1408), „Terminal acoustic characteristics for telephony; Requirements“

[D.2] 3GPP [TS 22.261](https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3107), “Service requirements for the 5G system”

[D.3] 3GPP TS.36.763, “Study on Narrow-Band Internet of Things (NB-IoT) / enhanced Machine Type Communication (eMTC) support for Non-Terrestrial Networks (NTN)”

[x] <doctype> <#>[ ([up to and including]{yyyy[-mm]|V<a[.b[.c]]>}[onwards])]: "<Title>".

It is preferred that the reference to TR 21.905 be the first in the list.

# 3 Definitions of terms, symbols and abbreviations

This clause and its three (sub) clauses are mandatory. The contents shall be shown as "void" if the TS/TR does not define any terms, symbols, or abbreviations.

## 3.1 Terms

For the purposes of the present document, the terms given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

Definition format (Normal)

**<defined term>:** <definition>.

**example:** text used to clarify abstract rules by applying them literally.

## 3.2 Symbols

For the purposes of the present document, the following symbols apply:

Symbol format (EW)

<symbol> <Explanation>

## 3.3 Abbreviations

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

GEO Geostationary Orbit

# 4 Application scenarios for ultra-low bit rate communication services

Editor’s Note:

From the WID:

- Document the application scenarios for ultra-low bit rate communication services taking into account the use cases and potential requirements documented in TR 22.887 related to IMS Voice Call Using GEO Access.

- Additional study areas or use cases, such as assessing the market potential and potential market-readiness of a new ULBC codec should be added with lower priority if time permits and once the exact requirements can be given.

Input on further application scenarios is invited.

## 4.1 Introduction

This clause introduces application scenarios that may be relevant for an Ultra-Low Bitrate speech Codec (ULBC). For each scenario, high-level prerequisites on the ULBC codec or the service operation are derived.

## 4.2 Scenario 1: IMS Voice Call over GEO

### 4.2.0 General

This is the primary application scenario in the context of voice communication via Geostationary Earth Orbit (GEO) satellites.

### 4.2.1 Background

Satellite communication plays an important role in extending terrestrial network coverage, ensuring seamless connectivity for users This technology unlocks new opportunities across various sectors, including smartphones and IoT.

Conversational two-party communication is foreseen as the most common and essential use case for IMS voice over GEO satellite access. This use case is documented in clause 5.1 of TR 22.887 [22887]. Due to the limitations of low transmission data rates (in the range of [1-3] kbps) as estimated in Table 7.4.2-1 in TS 22.261 [22261], an ultra-low bit rate codec is required. Users may rely on this service when they are beyond terrestrial network coverage but within satellite range, ensuring essential connectivity.

NOTE: The bit rate range of [1-3] kbps is a working assumption according to TS 22.261 and might be adjusted after coordination with RAN groups.

### 4.2.2 Scenario Description

#### 4.2.2.1 General

The typical UE is a handheld device supporting GEO satellite access with built-in microphones and loudspeakers or a monaural hands-free set, as outlined in Table 7.4.2-1 in TS 22.261 [22261]. This means there is no need for the user to extend the antenna or carry any extra devices to access the IMS voice call service.

The typical IMS voice service includes both regular call and emergency call as outlined in clause 6.46.11 in TS 22.261 [22261]. Such service can be provided as:

- Supplementary regular IMS voice service provided by the terrestrial operators, especially in areas without terrestrial coverage.

- Main regular IMS voice service provided by the satellite operators.

- An IMS voice call service using GEO satellite access in case of emergency situations.

Editor’s note: UEs under consideration are typical commercial devices, further details are TBD.

In the following, different scenarios establishing an end-to-end voice service are considered.

#### 4.2.2.2 [Main] Scenario: One UE connects via GEO-satellite access only

In a common scenario, one party in the conversation is assumed to be using a handheld mobile terminal over a GEO satellite network, while the other may be on a terrestrial mobile network (e.g., VoLTE, VoNR), a fixed-line connection, or another IMS-supported platform, as outlined in Figure 4.2.2.2-1, where a sketch of the bi-directional voice data flow for this main scenario is depicted.

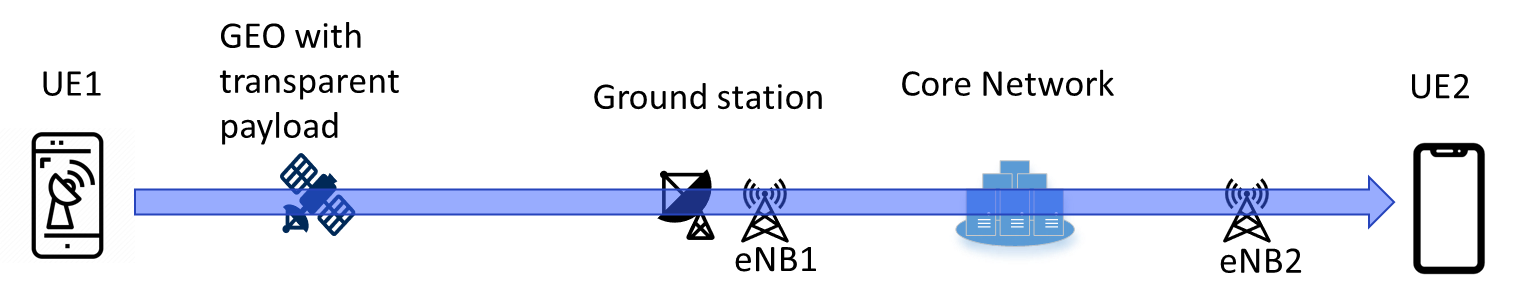


Figure 4.2.2.2-1: Bi-directional voice data flow for main scenario

NOTE: Core network typically stands for 3GPP core network and IMS core network.

In this scenario, UE1 is a phone supporting IMS voice call service over GEO satellite access. One may distinguish two cases depending on UE2:

* UE2 is a "regular" phone supporting IMS voice call service but not supporting ULBC nor GEO satellite access. In this case, UE2 may not be aware that UE1 is using a GEO satellite link during the communication and transcoding is performed in the core network.
* UE2 is an "upgraded" phone supporting IMS voice call service with ULBC, but using another access than satellite access (e.g., LTE, NR or WLAN). In this case UE2 may be able to communicate with UE1 using ULBC in a transcoder-free operation.

#### 4.2.2.3 [Sub-] Scenario: Both UEs connect via GEO-satellite access

In a less common scenario, both parties in the conversation are connected to a GEO satellite as outlined in Figure 4.2.2.3-1 using IMS-based communication services. This scenario may be less frequent than the [main] scenario but become relevant to support multiple contexts including disaster or cyberattack with potentially no terrestrial PLMN available which will be leading to the UEs communicating through NTN.

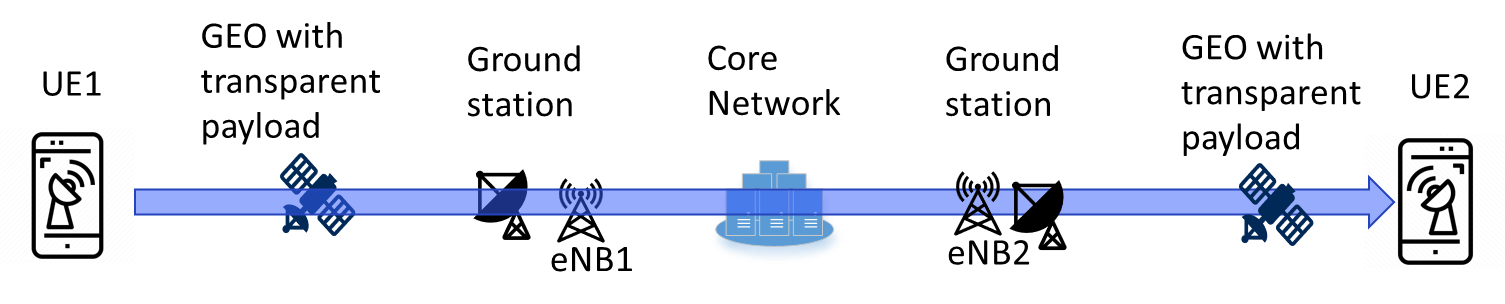


Figure 4.2.2.3-1: Bi-directional voice data flow for sub-scenario

NOTE: Core network typically stands for 3GPP core network and IMS core network.

When the GEO satellite operates in a transparent payload, the voice packets are transmitted to the ground before transmitted to the other UE, even if both UEs are connected to the same GEO satellite. In this case, the communication between UE1 and UE2 may use ULBC in a transcoder-free operation.

### 4.2.3 Derived high-level prerequisites

#### 4.2.3.1 General

The following general prerequisites for the ULBC apply based on application scenario 1.

Editor’s note: Some high-level prerequisites could be added such as:

"To serve application scenario 1, the ULBC codec is expected to meet the following high-level prerequisites:

- Very low bitrate support

- DTX support [to be confirmed]

- Error concealment

- Implementable in real-time (encoding and decoding) on at least a selective set of smartphones

- Good audio quality to ensure a reasonable QoE (Detailed QoE requirements for such services are for study).

Considerations on whether existing audio quality tests (TS 26.131 and TS 26.132) may be considered."

# 5 Channel characteristics and service-related dependencies

Editor’s Note:

2. Study GEO channel characteristics and derive service-related dependencies, e.g. bitrates, mouth- to-ear delay or loss/delay/jitter profiles.

NOTE: Any impact of ultra-low bitrate voice codec in NB-IoT services is outside of the scope of the study and is expected to be addressed by other working groups.

8. Coordinate work with other 3GPP groups e.g. SA2, RAN, CT1, and others as needed.

## 5.1 Estimation of mouth to ear delay for GEO scenarios

### 5.1.1 Overview

This clause estimates the mouth to ear (M2E) delay for IMS voice call over GEO satellites based on the application scenario introduced in clause 4.2. Two sub-scenarios are considered:

**- Main Scenario (see clause 4.2.2.2):** UE1 is connected via satellite while UE2 is connected via terrestrial network which corresponds to the signal flow UE1 àGEO satellite àGround stationàCore networkà eNodeB àUE2

* **Sub-Scenario 1 (see clause 4.2.2.3):** Both UEs are connected to a GEO satellite which corresponds to the signal flow UE1 àGEO satellite àGround stationàCore networkàGround stationàGEO satelliteàUE2

This approach aims to estimate the maximum and minimum delay components in the signal flow and finally to estimate a range of the. mouth-to-ear delay accordingly. The estimation assumes jitter free case and no network congestion.

NOTE: In practical deployments, various jitter and network conditions can arise.

Editor’s note: The scenarios and the terminology of this clause needs to be aligned with clause 4.) “Application Scenario” where a detailed description of the call scenarios is expected.

### 5.1.2 Delay components

#### 5.1.2.1 Overview

In this clause, the individual delay components that contribute to the mouth-to-ear delay are introduced and derived. The derived values are independent of the signal flow direction.

#### 5.1.2.2 UE Delay considering IMS codecs

[TS 26.131](https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1408) [D.1] defines the internal UE delay requirements and objectives depending on the components codec (frame size and algorithmic delay), air interface, jitter buffer depth and vendor specific delay budget. The UE delays in sending (UE1) and receiving directions (UE2) are not separated in TS 26.131, however the sum of the sending and receiving delays can be considered together.

The jitter buffer delay budget contains 40ms if the packet duration is 20ms and it includes the expected jitter profiles for terrestrial network transmission. In case of 40ms packet duration, the budget is doubled to 80ms, which is not further discussed. The value for the air interface in [D.1] just reflects the delay between UE and measurement equipment (2 ms) and needs be replaced by the expected delay for real air interface, i.e. air interface to GEO satellite or terrestrial network.

For MTSI-based speech only services with LTE and NR, the UE delay is outlined in the following table:

Table 5.1.2-1 UE delay components

|  |  |  |
| --- | --- | --- |
|  | UE delay in ms (Performance objective)  (Note 2) | UE delay in ms (Maximum requirement)  (Note 2) |
| Frame size (Note 1) | 20 | 20 |
| alg. Codec Delay (Note 1) | 5 | 12 |
| JBM (jitter free) (Note 3) | 40 | 40 |
| Vendor specific budget (Note 4) | 83 | 123 |
| UE delay Ts+Tr | 148 | 195 |
| Note 1: Values reflect the IMS codecs AMR/AMR-WB/EVS  Note 2: Requirements and Performance Objectives apply to the UE delay only (sum of send (Ts) and receive (Tr) delays) and only for MTSI-based speech-only with LTE, NR or WLAN access in error and jitter free conditions.  Note 3: JBM delay is considered as constant independent of the frame size.  Note 4: Vendor specific budget of TS 26.131 may change for GEO satellite connectivity | | |

Editor’s note: This table assumes LTE/NR air interface and needs to be updated for GEO satellite access air interface.

#### 5.1.2.3 Core network delay

The delay contribution of the core network consists of the packet transmission delay between two network entities, e.g. ground station to core network or core network to eNodeB. In case of the interop scenario GEO NTN to TN network, an additional delay component for transcoding needs to be considered. Assuming the frame size of both codecs is identical or a multiple of each other, only the algorithmic codec delay contributes to the transcoding delay, i.e. 5ms for AMR/AMR-WB or 12ms for EVS, and an additional delay margin for the processing of the transcoding (2 ms). This means, transcoding with AMR/AMR-WB adds 7ms and with EVS adds 14ms.

Table Y.5.1.2.3-1 Core network delay components

|  |  |  |
| --- | --- | --- |
|  | Minimum delay in ms | Maximum delay in ms |
| Network delay ground station to core network (Delay\_GSCN) | [ 5 Note1-1 ,  20 Note1-2] | [200 Note2] |
| Network delay eNodeB to core network (Delay\_eNBCN) | 5 | 20 |
| Transcoding | 7 | 14 |
| [Note1-1: In [D.2] 5 ms network latency is assumed]  [Note1-2: TS 23.501 assumes a static delay value for the CN PDB of 20ms between a UPF and 5G-AN. ]  [Note2: In some NTN deployments, the core network may need to be located far from the ground station due to factors like user distribution, geography, or other practical considerations. As a result, latency can increase, ping statistics between continents, for example, can reach up to 200ms.] | | |

#### 5.2.2.4 Transmission delay UE – GEO - Ground station

Clause 7.4.2 of [D.2] defines the KPI requirement for GEO based satellite access, i.e. 280ms. TR 36.763 clause 7.1.1 describes the max. and min. propagation delay contribution which depends on the location of the UE within the beam. As a result, the round-trip-delay can differ by 64ms which corresponds 32ms for one-way transmission. It is proposed to consider the 280ms as the max. transmission delay and consequently 248ms (280ms – 32ms) as the minimal transmission time. This assumes no retransmission over the GEO satellite link.

Table 5.1.2.4-1 Transmission delay GEO satellite

|  |  |  |
| --- | --- | --- |
|  | Minimum delay in ms | Maximum delay in ms |
| GEO transmission delay | 248 | 280 |
| Note: Transmission delay ground station to core network counted in 5.1.2.3-1. | | |

#### 5.2.2.5 ULBC Delay components

Table 5.1.2.2-1 lists the algorithmic delay for the IMS codecs AMR and EVS, i.e. in range of 5ms to 12ms. For ULBC, different delay values may result from codec processing delays as well as algorithmic delays. Exact numbers are for further study.

### 5.1.3 Estimation of Mouth-to-ear delay

Given the values in 5.1.2 the mouth-to-ear delay for scenario can be estimated for the two scenarios outlined in 5.1.1 by summing up the delay components according to the signal flow to derive a lower (minimum values as in tables 5.1.2.2-1, 5.1.2.3-1, 5.1.2.4-1) and an upper bound (maximum values as in tables 5.1.2.2-1, 5.1.2.3-1, 5.1.2.4-1).

As the bitrate for GEO satellite link is very restricted, options for minimizing the protocol overhead need to be considered. One option to reduce the protocol overhead are larger frame sizes or frame aggregation as the protocol stack is transmitted less often. Therefore, the following table outlines the delay values for codec frame sizes of 20ms, and in addition derived for 40ms, 80ms, 160ms and 320ms.

Editor’s Note: Current values assume algorithmic delay of AMR and EVS as given in 5.1.2.2-1. ULBC Delay components documented in 5.1.2 still need to be addressed. For the min. Delay\_GSCN, 20ms is assumed.

Table 5.1.3-1 Mouth-to-ear delay estimation depending on codec frame size

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Frame size in ms | Mouth to ear delay main scenario in ms  (GEO - TN) (Note 1) | | Mouth to ear delay sub-scenario 1 in ms  (GEO - GEO) (Note 2) | |
| **lower bound** | **upper bound** | **lower bound** | **upper bound** |
| 20 | 428 | 712 | 684 | 1155 |
| 40 | 448 | 732 | 704 | 1175 |
| 80 | 488 | 772 | 744 | 1215 |
| 160 | 568 | 852 | 824 | 1295 |
| 320 | 728 | 1012 | 984 | 1455 |
| Note 1: UE(frame size) + UE(alg. Delay) + UE(JBM) + UE(Vendor)+GEO transmission+Delay\_GSCN+Delay\_eNBCN  Note 2: UE(frame size) + UE(alg. Delay) + UE(JBM) + UE(Vendor)+2x GEO transmission+2x Delay\_GSCN | | | | |

Editor’s note: The scenarios and the terminology of this clause needs to be aligned with clause 4.) “Application Scenario” where a detailed description of the call scenarios is expected.

# 6 Design constraints

Editor’s Note:

3. Identify the relevant design constraints for such a codec, in coordination with other WGs, including:

- Bit rates

- Sample rate and audio bandwidth

- Frame length

- Complexity and memory demands

- Algorithmic delay

- Packet loss concealment (PLC)

- Potential use of noise suppression as part of the codec

- Discontinuous transmission including voice activity detection and comfort noise

- Speech quality

- Robustness to non-speech input

- Identify or develop objective measures to verify the design constraints as necessary (e.g., to measure complexity and memory demands)

6. Identify or develop objective measures to verify the design constraints as necessary (e.g., to measure complexity and memory demands)

8. Coordinate work with other 3GPP groups e.g. SA2, RAN, CT1, and others as needed.

## 6.1 General

The following clauses present the design constraints (DC) for an Ultra Low Bitrate Codec for the use in application scenarios as given in clause 4. Clause 6.2 outlines the DC parameter and clause 6.3 outlines objective verification methods of some DC parameter.

## 6.2 Design Constraint Parameter

Table 6.2-1 List of ULBC design constraint parameter

| Parameter | Design Constraint | Note |
| --- | --- | --- |
| Bit rates |  |  |
| Sample rate and audio bandwidth |  |  |
| Frame length |  |  |
| Complexity and memory demands |  |  |
| Algorithmic delay |  | The algorithmic delay is defined as the frame size buffering delay plus any other delays inherent in the codec algorithm (e.g., look-ahead, sample-rate conversion, and decoder post-processing) |
| Packet loss concealment (PLC) |  |  |
| Potential use of noise suppression as part of the codec |  |  |
| Discontinuous transmission including voice activity detection and comfort noise |  |  |
| Robustness to non-speech input |  | Editor’s note: May need to be in performance requirement |
|  |  |  |

Editor’s note: Speech quality to be addressed in the performance requirements.

## 6.2 Design Constraint Verification

Editor’s note: Algorithmic delay verification method for AI based codecs required.

# 7 Existing technologies and feasibility evidence

Editor’s Note:

1. Provide some evidence that the design criteria can be met, for example existing reference codecs.

## [7.1 Existing codec technologies for GEO scenarios

### 7.1.1 Overview

The present clause collects information on existing codec technologies that may be suitable for GEO application scenarios and categorizes those using the following definitions:

**- 3GPP IMS codecs**: Even though the operating bitrates are out of scope for the envisioned services, these codecs can be considered as reference condition regarding the performance requirements.

**- Conventional Ultra Low Bitrate Codecs**: These codecs are based on conventional signal and speech processing algorithms (like the 3GPP IMS codecs) and can operate in the envisioned bitrate range.

**- AI-based postprocessor**: The bitstream or decoder of a conventional ultra-low bitrate codec is decoded and enhanced by a AI based postprocessor

**- AI-based encoder and decoder**: Encoder and decoder implemented as Deep Neural Network (DNN) and can potentially operate in the envisioned bitrate range. Due to fundamental design differences regarding the algorithmic delay, two more sub-categories are defined:

**- Causal systems**: Codecs which can operate in real-time applications

**- Non-causal systems:** Codecs which can only operate in non-real-time applications due to large processing look-ahead or framing

The following codec properties are considered:

**- Audio bandwidth**: Support of NB, WB, SWB, FB and/or 12 kHz

**- Codec delay**: any delays inherent in the codec algorithm (e.g., look-ahead, sample-rate conversion, and decoder post-processing), excluding framing

**- Frame duration**: Block size or framing delay, may be bit rate dependent

**- Bitrates**: List of supported bitrates lower than [3] kbps or the lowest supported constant bitrate.

Editor’s note: Variable bitrate modes in combination with DTX are for FFS.

**- Specification / Access / Software**:

- Specification determines the level of development for a complete solution

- A: Fully specified and verified coding algorithm including all relevant system components such as channel resilience capability or discontinuous transmission

- B: Specification of codec algorithm only

- Access determines the availability of a reference implementation:

- A: Repository with open access

- B: No repository with open access known

- Software determines the deployment level of the reference software its level of optimization

- A: Software fully functional and optimized C-code

- B: Software fully functional and reference C-code

- C: Experimental software framework in Python only

- X: Software status unknown

NOTE: In the case of a DNN-based codec, both the implementation of the computational graph (i.e. the layer definitions and their order of execution) and the corresponding pretrained weights are needed for inference. For this reason, consider two sets of weights for the same computational graph as defining two different codecs.

Editor’s note: Documentation regarding complexity aspects (computational complexity, memory, ROM) is envisioned, however it requires a clear definition on the measurement first.

In the following a collection of speech codecs and the related parameter is presented in Table 7.1.1-1.

Table 7.1.1-1 List of existing codec technologies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Codec | Source | Audioband-width | Codec Delay  [ms] | Frame duration  [ms] | Bitrates  [kbps] | Specification / Access / Software |
| 3GPP IMS codecs | | | | | | |
| AMR | 3GPP TS [26.071](https://www.3gpp.org/DynaReport/26071.htm" \t "_blank) | NB | 5 | 20 | 4.75 | A / A / A |
| AMR-WB | 3GPP TS [26.171](https://www.3gpp.org/DynaReport/26171.htm" \t "_blank) | WB | 5.9375 | 20 | 6.6 | A / A / A |
| EVS | 3GPP TS [26.445](https://www.3gpp.org/DynaReport/26445.htm" \t "_blank) | NB | 12 | 20 | 7.2 | A / A / A |
| WB | 7.2 |
| SWB | 9.6 |
| Conventional Ultra Low Bitrate Codecs | | | | | | |
| MELP / MELPe  (Note1) | <https://www.compandent.com/about-melpe/> and melp\_faq | NB | 36.25 | 90 | 0.6 | A / B / A |
| 27.25 | 67.5 | 1.2 |
| 20.125 | 22.5 | 2.4 |
| AMBE-LR | https://www.dvsinc.com/software/technology.shtml#ambelr | NB |  |  | 1.6 – 1.8 | A / B / A |
| MPEG--HVXC | https://www.iso.org/obp/ui/en/#iso:std:iso-iec:14496:-3:ed-5:v1:en:sec:1.3 | NB |  |  | 2 – 4 | B / B / B  (Note2) |
| TWELP MR  (Note 1) | <https://dspini.com/vocoders/lowrate/twelp-lowrate/twelp300-3600-mr> | NB | 20 | 40 | 3.2 | A / B / A  (Note5) |
| 20 | 40 | 2.4 |
| 40 | 60 | 1.6 |
| 40 | 60 | 1.2 |
| 80 | 100 | 0.7 |
| 80 | 100 | 0.6 |
| 100 | 120 | 0.48 |
| 100 | 120 | 0.3 |
| Codec2 | https://github.com/drowe67/codec2 | NB | 40  (Note 3) | [TBD] | 0.45 | A / A / A |
| 40 | 0.7 |
| 40 | 1.2 |
| 40 | 1.3 |
| 40 | 1.4 |
| 40 | 1.6 |
| 20 | 2.4 |
| AI based decoders | | | | | | |
| WaveNet Codec2 | [Paper](https://arxiv.org/pdf/1712.01120) | WB | See Codec2 | 20 | 2.4 | B / B / X |
| CQNV  Codec2 | [Paper](https://arxiv.org/abs/2307.13295) | WB | 40 | 40-60 | 1.0  1.1 | B / B / X |
| AI based encoder and decoder (causal) | | | | | | |
| LPCnet | [Paper](https://jmvalin.ca/papers/lpcnet_codec.pdf), [Code](https://github.com/xiph/LPCNet), [Demo](https://jmvalin.ca/demo/lpcnet_codec/) | WB | 25 | 40 | 1.6 | B / A / C |
| LyraV2 (aka SoundStream)  (Note1) | [Blog](https://opensource.googleblog.com/2022/09/lyra-v2-a-better-faster-and-more-versatile-speech-codec.html),  [Paper (SoundStream)](https://arxiv.org/pdf/2107.03312), [Code](https://github.com/google/lyra), [Demo (SoundStream)](https://research.google/blog/soundstream-an-end-to-end-neural-audio-codec/) | WB | [TBD] | 20 | 3.2, 6, 9.2 | B / A / A |
| EnCodec | [Paper](https://arxiv.org/pdf/2210.13438), [Code](https://github.com/facebookresearch/encodec), [Demo](https://audiocraft.metademolab.com/encodec.html) | 24kHz | 0 | 13.3 | 1.5, 3, … | B / A / C |
| FB | 1000 | 13.3 | 6, 12, 24 |
| Mimi-Codec | [Paper(moshi)](https://kyutai.org/Moshi.pdf), [Code](https://huggingface.co/kyutai/mimi), [Demo (moshi)](https://moshi.chat/) | 24kHz | 0 | 80 | 0.55, 1.1 | B / A / C |
| TS3 | [Paper](https://arxiv.org/pdf/2411.18803), [Code](https://github.com/ga642381/speech-trident), Demo | WB | 0 | 20 | 0.64, 0.8 | B / B / X |
| TAAE | [Paper](https://arxiv.org/pdf/2411.19842v1), [Code](https://github.com/Stability-AI/stable-codec), [Demo](https://stability-ai.github.io/stable-codec-demo/) | WB | 0 | 20, 40 | 0.4, 0.7 | B / B / X |
| LMCodec2 | [Paper](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4915275), Code, Demo | [TBD] | [TBD] | [TBD] | [TBD] | [TBD] |
| AI based encoder and decoder (non-causal) | | | | | | |
| DAC | [Paper](https://arxiv.org/pdf/2306.06546), [Code](https://github.com/descriptinc/descript-audio-codec/tree/main), [Demo](https://descript.notion.site/Descript-Audio-Codec-11389fce0ce2419891d6591a68f814d5) | WB | 244 | 20 | 0.5, 1.0, 1.5, … | B / A / C |
| 24kHz | 366 | 13.3 | 0.75 1.5, 3, … |
| DAC-IBM | [Paper](https://arxiv.org/html/2410.08325v1) | 24kHz | 366 | 13.3 | 0.75, 1.5, 3 | B / A / C |
| SNAC | [Paper](https://arxiv.org/pdf/2410.14411), [Code](https://github.com/hubertsiuzdak/snac) | 24 kHz | 1000 | 80 | 0.98 | B / A / C |
| SpeechTokenizer | [Paper](https://arxiv.org/pdf/2308.16692), [Code](https://github.com/ZhangXInFD/SpeechTokenizer), [Demo](https://0nutation.github.io/SpeechTokenizer.github.io/) | WB | full-signal | 20 | 0.5, 1.0 | B / A / C |
| SemantiCodec | [Paper](https://arxiv.org/pdf/2405.00233), [Code](https://github.com/haoheliu/SemantiCodec-inference), [Demo](https://haoheliu.github.io/SemantiCodec/) | WB | full-signal | 10 | 1.25, …, 1.4 | B / A / C |
| 20 | 0.63, …, 0.70 |
| 40 | 0.31, …, 0.35 |
| FunCodec  (Note4) | [Paper](https://arxiv.org/pdf/2309.07405), [Code](https://github.com/modelscope/FunCodec), [Demo](https://funcodec.github.io/index.html) | WB | [TBD] | 20 | 0.5, 1.0, … | B / A / C |
| 40 | 0.25, 0.5, … |
| WavTokenizer (Note4) | [Paper](https://arxiv.org/pdf/2408.16532), [Code](https://github.com/jishengpeng/WavTokenizer), [Demo](https://wavtokenizer.github.io/) | 24kHz | [TBD] | 40 | 0.25, 0.5, … | B / A / C |
| 25 | 0.9 |
| BigCodec  (Note4) | [Paper](https://arxiv.org/html/2409.05377v1), [Code](https://github.com/Aria-K-Alethia/BigCodec), [Demo](https://aria-k-alethia.github.io/bigcodec-demo/) | WB | [TBD] | 12.5 | 1.04 | B / A / C |
| FocalCodec | [Paper](https://arxiv.org/pdf/2502.04465), [Code](https://github.com/lucadellalib/focalcodec), [Demo](https://lucadellalib.github.io/focalcodec-web/) | WB | [TBD] | 20 | 0.65 | B / A / C |
| 40 | 0.33 |
| 80 | 0.16 |

Note1: These codecs may include a noise

Note2: Decoder and reference encoder are available in the MPEG reference software and for MPEG members

Note3: Encoder with 20ms overlapping FFT and iFFT at decoder with same overlap-add size

Note4: Only non-causal version publicly available

Note5: Although TWELP does not have an open reference implementation, a complete quality assessment testbench is available at <https://dspini.com/files/downloads/speechsamples/C2_vs_TWELP.zip>.

Editor’s note: More codecs may be added to the table

### 7.1.2 Observations regarding codec parameter

The following observations are extracted from the Table 7.1.1-1

**- Audio bandwidth:** While conventional ultra-low bitrate codecs operate in the envisioned bitrate range on NB only, modern AI based codecs offer an audio bandwidth of WB or higher

**- Algorithmic codec delay:** frame size buffering delay plus any other delays inherent in the codec algorithm (e.g., look-ahead, sample-rate conversion, and decoder post-processing),

- IMS codecs operate in range between 25ms (AMR) and 32ms (EVS)

- Conventional audio codecs operate in the delay range of 60ms – 80ms (Codec2) up to 126.25ms (MELPe).

- Causal AI based codecs can operate between 20ms (TS3) and up to 80 ms (Mimi/Fun-Codec)

- Non-Causal AI based codecs require several 100ms (min. 500ms for DAC/DAC-IBM) or the entire input signal

**- Frame duration:** For conventional ultra-low bitrate codecs, the frame duration is increased compared to the regular 20ms VoIP framing. A similar tendency can be observed for AI based codecs (MimiCodec), even though some designs maintain the frame duration parameter but increase other parameter values such as bitrate (LyraV2).

**- Bitrate:** All listed codecs (except the IMS codecs and LyraV2) offer at least one bitrate mode lower than 3kbps.

### 7.1.3 Complexity Considerations:

The procedure on measuring complexity of AI based codecs and how to compare those to conventional codecs is an independent objective of the study item. Therefore, the source just wants to share some observations regarding computational complexity, RAM and ROM demand:

- **Complexity**: Almost all AI based codecs may have higher computational complexity than IMS codecs and conventional ultra-low bitrate codecs. One exception is LyraV2, which only requires 35% of an ARM A53 core according to measurements on RaspberryPi 3+. More real-time factor or RTF (ratio between the frame length of the input audio and the time needed for encoding/decoding) analysis for some AI based codecs has been done in [2]. The final method for complexity evaluation is for further study.

**- RAM**: RAM numbers may highly depend on the implementation esp. for AI based solutions. However, it should be noted that at least some of the AI based codecs in available implementation show a significant higher RAM demand compared to conventional codecs, e.g. LyraV2 show ca. 54 Mbyte while EVS was characterized with 294 kByte.

**- ROM** It is expected that ROM number for AI based solutions are significantly higher compared to legacy codecs, e.g. EVS was characterized with ca. 2MB while for TAAE 950M parameters are reported. In general, the number of parameters for AI based solution correlates to the ROM demand. Assuming 8bit integer values for the parameters, this corresponds to ca. 900 Mbyte. Other solution like SNAC operate with 19M parameters which correspond to 18 Mbyte. Implementations may also use 32bit floating point parameters which leads to an ROM increase by a factor four.

### 7.1.4 Performance Evaluation

NOTE: As this is a pre-evaluation, the test methodology gives no precedence for testing methodologies of ULBC candidates.

In order to get a first impression on the performance of the codecs listed in Table 1, an ITU-T P.808 ACR listening test has been conducted in English clean speech (4 talker x 6 samples). The direct signal was 32kHz sampled with SWB, normalized to -26 dBoV, to accommodate the SWB conditions. The following conditions were included:

- Conventional codecs:

- Codec2 at 0.7, 1.2, 2.4 kbps

- AMR at 4.75 kbps

- AMR-WB at 6.65 kbps

- EVS-WB at 7.2 kbps

- EVS-SWB at 9.6 kbps

- AI based codecs:

- LPCNET at 1.6 kbps

- LyraV2 at 3.2 kbps

- Mimi 0.55, 1.1 kbps

- SemantiCodec 0.34, 0.68, 1.35 kbps

- DAC\_ibm 1.5 kbps

- SNAC 0.98 kbps

- Excluded codecs

- EnCodec was not considered due the poor P.800 scoring at 1.5 kbps in [7-1]

- DAC was excluded in favour of the better optimized DAC-ibm for this bitrate range

- FunCodec will be added in a potential follow up test.

Figure 7.1.4-1 shows the Mean Opinion Scores and 95% confidence intervals of 24 subjects.

Ein Bild, das Text, Screenshot, Diagramm, Reihe enthält.

KI-generierte Inhalte können fehlerhaft sein.

Figure 7.1.4-1 P.808 Mean Opinion Scores and 95% confidence intervals

The following observations can be extracted from the results in Figure 7.1.4-1:

- Codec2 at any rate performs significantly worse than AMR at 4.75 kbps

- SemantiCodec, LyraV2, LPCnet and Mimi-Codec at 0.55 kbps score comparable to AMR-WB at 6.65 kbps

- Three conditions show a promising performance on par or slightly better than EVS at 9.6 kbps and therefore, might reach a level quality considerable for voice services. At least one is considered as causal codec

- Mimi-Codec at 1.1 kbps

and two more as non-causal codecs

- DAC-ibm at 1.5 kbps

- SNAC at 0.98 kbps

- Comparing conventional ultra-low bitrate codecs to AI based solutions, a significant quality gain of 2 MOS or more can be observed

### 7.1.5 Conclusion on existing technology

The following conclusions on existing technologies can be drawn based on this sub-clause 7.1:

- AI based solution may have the potential to provide substantial better audio quality compared to conventional solutions

- Whether AI based codecs fit into complexity constraints of current chipset is FFS.

Increasing the frame duration to larger blocks than the standard 20 ms VoIP size seems to be beneficial for ultra-low bitrate coding.

- None of the presented solutions can be considered as a complete candidate for ULBC.

- Besides an optimal match on the basic parameters such as frame length or bitrate, the ULBC candidate needs to be implementable on mobile devices. Furthermore, the candidate needs to support required system aspects such as channel resilience capability or discontinuous transmission, which is currently not supported by any AI-based solution.

]

## [7.1 Very low bitrate listening test results

### 7.1.1 Overview

There are several traditional (DSP based) speech codecs available for very low bit operation as described in [7-2]. They are commonly considered as vocoder-style codecs. They provide understandable voice quality with somewhat limited perceptual quality, often sounding a bit “synthetic”. Their bandwidth is usually limited to narrowband (4 kHz).

On the other hand there are several AI/ML based codecs coming from various universities, researchers, and companies working in the ML area. Some of these codecs target very low bitrates, e.g., 0.5 to 1 kbit/s and some aim at a bit higher operation points, e.g., 1-6 kbit/s. Some are optimized for voice, some are even trained only with certain language. Some codecs are optimized for general audio.

Different ML codecs use varying sampling rates and some of those are different to the common 3GPP sampling rates. Thus we included different bandwidth limits to the listening test. Bandwidths included are:

- Narrowband **(NB) 4 kHz**

**-** Mediumband **(MB) 6 kHz**

**-** Wideband **(WB) 8 kHz**

**-** 10 kHz

- Semi-Super Wideband **(SSWB) 12 kHz**

**-** Super Wideband (SWB) 16 kHz

- Fullband **(FB) 20 kHz**

For the listening test three DSP based vocoder style codecs were chosen:

- Codec2 0.7, **1.3**, **2.4,** and 3.2 kbit/s, tested using ffmpeg [7-3], 7-[4]

- MELP **2.4** kbit/s [7-5]

- MPEG4 HVXC **2.0** and 4.0 kbit/s [7-6]

3GPP solutions at slightly higher bitrates are represented by:

- AMR **4.75** and 7.95 kbit/s

- AMR-WB **6.6**, 8.85, and 12.65 kbit/s

- EVS-NB **5.9** and 7.2 kbit/s

- EVS-WB **5.9**, 7.2, and 9.6 kbit/s

- EVS-SWB **9.6** and 13.2 kbit/s

Finally tested ML/AI based codecs include:

- DAC (Descript Audio Codec) 44k 0.9, **1.7**, **2.6**, **3.4,** and 6.9 kbit/s [7-7]

- TSAC (Modified version of DAC) 44k 0.6, **1.2**, **2.5**, **3.2,** and 5.9 kbit/s [7-8]

**Bolded** bitrates and bandwidths were tested in both ACR and DCR listening tests.

The sizes of ML codec models vary wildly. Some of them are reasonably sized (e.g., few tens of megabytes), but sometimes they are several gigabytes in size.

For codecs in current tests:

- Descript-audio-codec uses “weight.pth”-file of size 292MB.

- TSAC codec apparently has for mono operation two files “dac\_mono\_q8.bin” and “tsac\_mono\_q8.bin” for a total of 128MB.

### 7.1.2 Listening test Results

Nokia conducted two listening tests using clean speech material. Clean speech consisted of Finnish language sentence pairs spoken by three males and three females, four sample pairs from each. The listening was conducted diotically using Sennheiser HD650 headphones in quiet listening booths. All listeners were experienced listeners, most of them were experts in the field of speech and audio processing. The first listening test used extended ACR5 test methodology, which is straightforward extension of standard P.800 ACR test [7-9]. The voting scale is extended as shown in Table 1 and Figure 1.

Table 1 Extended ACR5 voting scale

|  |  |
| --- | --- |
| Grade | Quality |
| 5.5 |  |
| 5 | Excellent |
| 4.5 |  |
| 4 | Good |
| 3.5 |  |
| 3 | Fair |
| 2.5 |  |
| 2 | Poor |
| 1.5 |  |
| 1 | Bad |
| 0.5 |  |

Kuva, joka sisältää kohteen teksti, kuvakaappaus, Fontti, viiva

Tekoälyn generoima sisältö voi olla virheellistä.

Figure 1 Example of extended ACR 5 voting scale in the listening test software.

### 7.1.3 Extended ACR5 listening test results

ACR listening results in Figure2 and Figure3 show that increased signal bandwidth increases perceptual quality up to a certain point. The high-frequency perception is very personal and also depends on hearing. Overall it seems that 12 kHz signal bandwidth is in the saturation region in this test. On the other hand it can be seen from the bandwidth experiment that 4 kHz signal bandwidth significantly limits the perceived quality of speech.

Regarding DSP based vocoders, MELP 2.4k and MPEG4 HVXC perform better than Codec2. From quality point of view, they could be considered as reasonable references representing DSP based vocoders.

3GPP standard codecs AMR, AMR-WB and EVS at their lowest supported bitrates perform as expected and provide good targets for performance.

Finally ML-based codecs DAC and TSAC have very good performance in clean speech, when considering their low bitrates. TSAC has somewhat better quality than DAC, which is in line with the web-page description, where they state that TSAC is an improved version of DAC. Both tested codecs have poor quality at their lowest tested bitrates below 1 kbit/s. However TSAC starting at 1.2 kbit/s and DAC starting at 1.7 kbit/s could be considered to represent ML-based codec references. Both of them were relatively easy to make run without issues.

Kuva, joka sisältää kohteen Tontti, diagrammi, viiva, teksti

Tekoälyn generoima sisältö voi olla virheellistä.

Figure 2 Extended ACR5 listening results of low bit rate voice codecs in clean speech.

A graph of different colored lines

AI-generated content may be incorrect.

Figure 3 Extended ACR clean speech listening results with 14 listeners in curve format and logarithmic bitrate.

### 7.1.4 CR listening test results

DCR listening test was constructed from a subset of conditions of the ACR listening test. The direct reference was the full-band signal. As can be seen listeners are more likely to notice degradations, when there is reference available. Overall the results are in line with the ACR test, with the exception that MELP is preferred over HVXC 2.0 in DCR test, when it is the other way around in the ACR test. The likely reason is that MELP has full 4 kHz bandwidth, but HVXC 2.0 has somewhat limited signal spectrum to approximately 3.7 kHz, and it therefore sounds a bit more muffled.

Kuva, joka sisältää kohteen teksti, diagrammi, kuvakaappaus, Tontti

Tekoälyn generoima sisältö voi olla virheellistä.

Figure 4 DCR listening test results of low bit rate voice codecs with clean speech

]

# 8 Performance requirements

Editor’s Note:

5a. Define performance requirements regarding speech quality, intelligibility, conversational quality, in particular taking into account:

- Clean speech and noisy speech

- Tandeming with existing IMS voice codecs

- Clean channel and GEO channel conditions

7. Identify relevant reference codecs for comparison and evaluation purposes.

8. Coordinate work with other 3GPP groups e.g. SA2, RAN, CT1, and others as needed.

# 9 Test methodologies

Editor’s Note:

5b. Identify appropriate test methodologies, regarding speech quality, intelligibility, conversational quality, in particular taking into account:

- Clean speech and noisy speech

- Tandeming with existing IMS voice codecs

- Clean channel and GEO channel conditions

# 10 Considered work plan for potential normative work

Editor’s Note:

9. Define potential normative work item objectives and timeline.

# 11 Conclusions and recommendations

# Proforma copyright release text block

(e.g. for PICS and PIXIT Proformas)

This text block shall immediately follow the heading of an element (i.e. clause or annex) containing a proforma or template which is intended to be copied by the user. Such an element shall always start on a new page.

## X.1 The right to copy

Notwithstanding the provisions of the copyright clause related to the text of the present document, the 3GPP Organizational Partners grant that users of the present document may freely reproduce the <proformatype> proforma in this clause|annex so that it can be used for its intended purposes and may further publish the completed <proformatype>.

# Abstract Test Suite (ATS) text block

This text should be used for ATS using TTCN. The subdivision is recommended.

# Y Abstract Test Suite (ATS)

## Y.1 Introduction

This ATS has been produced using the Tree and Tabular Combined Notation (TTCN) according to ISO/IEC 9646‑3 [x].

The ATS was developed on a separate TTCN software tool and therefore the TTCN tables are not completely referenced in the table of contents. The ATS itself contains a test suite overview part which provides additional information and references.

# Y.2 The TTCN Graphical form (TTCN.GR)

The TTCN.GR representation of this ATS is contained in an Adobe Portable Document Format™ file (<pdf\_file\_name>.PDF contained in archive <zip\_file\_name>.ZIP) which accompanies the present document.

# Y.3 The TTCN Machine Processable form (TTCN.MP)

The TTCN.MP representation corresponding to this ATS is contained in an ASCII file (<mp\_file\_name>.MP contained in archive <zip\_file\_name>.ZIP) which accompanies the present document.

Annex <A> (normative):  
<Normative annex for a Technical Specification>

Start each annex on a new page.

Annexes are labelled A, B, C, etc. and designated either "normative" or "informative" depending on their content.

Normative annexes only to appear in Technical Specifications. Use style "Heading 8".

Annex <B> (informative):  
<Informative annex for a Technical Specification>

Informative annexes may appear in both Technical Specifications and Technical Reports. Use style "Heading 8" for use in TSs.

Informative annexes shall not contain requirements for the implementation of the Technical Specification.

# B.1 Heading levels in an annex

Heading levels within an annex are used as in the main document, but for Heading level selection, the "A.", "B.", etc. are ignored. e.g. **B.1.2** is formatted using ***Heading 2*** style.

Annex <F> (informative):  
Change history

Use style "Heading 8" in TSs and "Heading 9" in TRs. Do not use "informative" in the title in TRs.

This is the last annex for TS/TSs which details the change history using the following table.  
This table is to be used for recording progress during the WG drafting process till TSG approval of this TS/TR.  
For TRs under change control, use one line per approved Change Request  
Date: use format YYYY-MM  
CR: four digits, leading zeros as necessary  
Rev: blank, or number (max two digits)  
Cat: use one of the letters A, B, C, D, F  
Subject/Comment: for TSs under change control, include full text of the subject field of the Change Request cover  
New vers: use format [n]n.[n]n.[n]n

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Change history | | | | | | | |
| Date | Meeting | TDoc | CR | Rev | Cat | Subject/Comment | New version |
| 2025-04 | SA4#131-bis-e | S4-250451 |  |  |  | Initial version submitted for SA4#131-bis-e | 0.0.1 |
| 2025-04 | SA4#131-bis-e | S4-250749 |  |  |  | Agreed version in SA4#131-bis-e | 0.1.0 |
| 2025-05 | SA4#132 | S4-251152 |  |  |  | Updated with agreed Tdoc: S4-250857,S4-251161, S4-250923, S4-251116 | 0.2.0 |