**3GPP TSG-SA WG4 Meeting #Audio SWG AH S4aA250204**

**Erlangen, 23 – 25 Sept. 2025**

**Source: vivo, Samsung, MediaTek Inc., Bytedance, Nokia, Xiaomi, Spreadtrum**

**Title: [FS\_ULBC] Discussion on Audio Bandwidth for ULBC**

**Agenda item: 4.4**

**Document for: DISCUSSION and AGREEMENT**

**1. Introduction**

This document provides a technical discussion on the audio bandwidth design constraints for the Ultra-Low Bitrate Codec (ULBC) [1]. As 3GPP is to define in this work the requirements for a codec intended for extremely constrained environments, mainly on satellite (GEO) communications, it is crucial to prioritize design choices that maximize robustness and efficiency.

Recent discussions have considered possible mandating Wideband (WB) and Super-Wideband (SWB) support for ULBC [2]. However, we argue that for the target use cases, mandatory support for Narrowband (NB) is essential, while whether wider bandwidths should be considered further enhancements of this baseline. This paper synthesizes key technical arguments and presents new experimental data to support a design philosophy focusing on performance in the most challenging conditions.

**2. Reasons for Change**

Establishing the appropriate audio bandwidth requirements is fundamental to the success of the ULBC standard. Per SID descriptions the voice over GEO scenario is highest priority goal of the ULBC study.

"*The objective of this study is to develop recommendations for potential normative work on an ultra-low bit rate codec with* ***primary*** *application for* ***voice over GEO****. …….*

*NOTE:* ***Other application scenarios*** *and related objectives, such as IMS voice call over 3GPP NGSO/GSO Satellites and Terrestrial Networks, are not excluded from the scope of the study but* ***addressed with lower priority*** *and shall not impact the completion timeline of the study while completing all objectives for the primary application.*".

A design that over-prioritizes wide audio bandwidths may lead to significant drawbacks in terms of performance at the target bitrates, computational complexity, power consumption, and interoperability with legacy systems. This contribution provides evidence that NB is a basic foundation for the primary objectives of ULBC, ensuring the highest possible reliability and intelligibility when network resources are scarcest. Furthermore, support of both NB and WB, where applicable, can ensure meeting of higher quality expectations.

**3. Discussion**

A robust design for ULBC requires a careful balance of trade-offs. The following points highlight key considerations for the audio bandwidth requirement.

**3.1. The Global Reality of NB Usage and the Cost of Inefficiency**

Although 4G and 5G networks are expanding, Narrowband (NB) voice services remain a significant and critical part of the global mobile ecosystem. According to the GSMA report [12] [13], 2G and 3G connections—which primarily use the AMR-NB codec for voice—still accounted for a combined 20% of the global technology mix at the end of 2023. This share is substantially higher in developing regions, reaching as high as 81% in Sub-Saharan Africa and 46% in the Middle East and North Africa. This represents hundreds of millions of users for whom NB is their primary voice service.

Furthermore, NB serves as the universal fallback for interoperability. Even users with advanced VoLTE devices frequently experience NB quality when making calls to users on different networks or in areas with limited coverage like Circuit Switched (CS) fallback [3] [4]. This means a substantial portion of global voice minutes are still carried over NB.

If ULBC is lack of NB mode, it would create a fundamentally inefficient system for connecting to this massive global user base. Every call between a WB ULBC user and an NB user would have the following drawback, when a WB ULBC user calls an NB user, the entire upper half of the audio band (from ~4kHz to ~8kHz) is transmitted over the expensive and scarce satellite link, only to be discarded at the gateway. This means a significant portion of the bitrate is wasted transmitting data the recipient can never hear.

For a service defined by its ultra-low bitrate, this inefficiency is unacceptable. From the overall system view, a native NB mode in ULBC is the most direct and efficient solution, ensuring that network resource required is minimal and no bandwidth is wasted and that connections to these critical, widespread legacy networks are as reliable as possible.

**3.2. Meeting User Expectations in a 'Last Resort' Scenario**

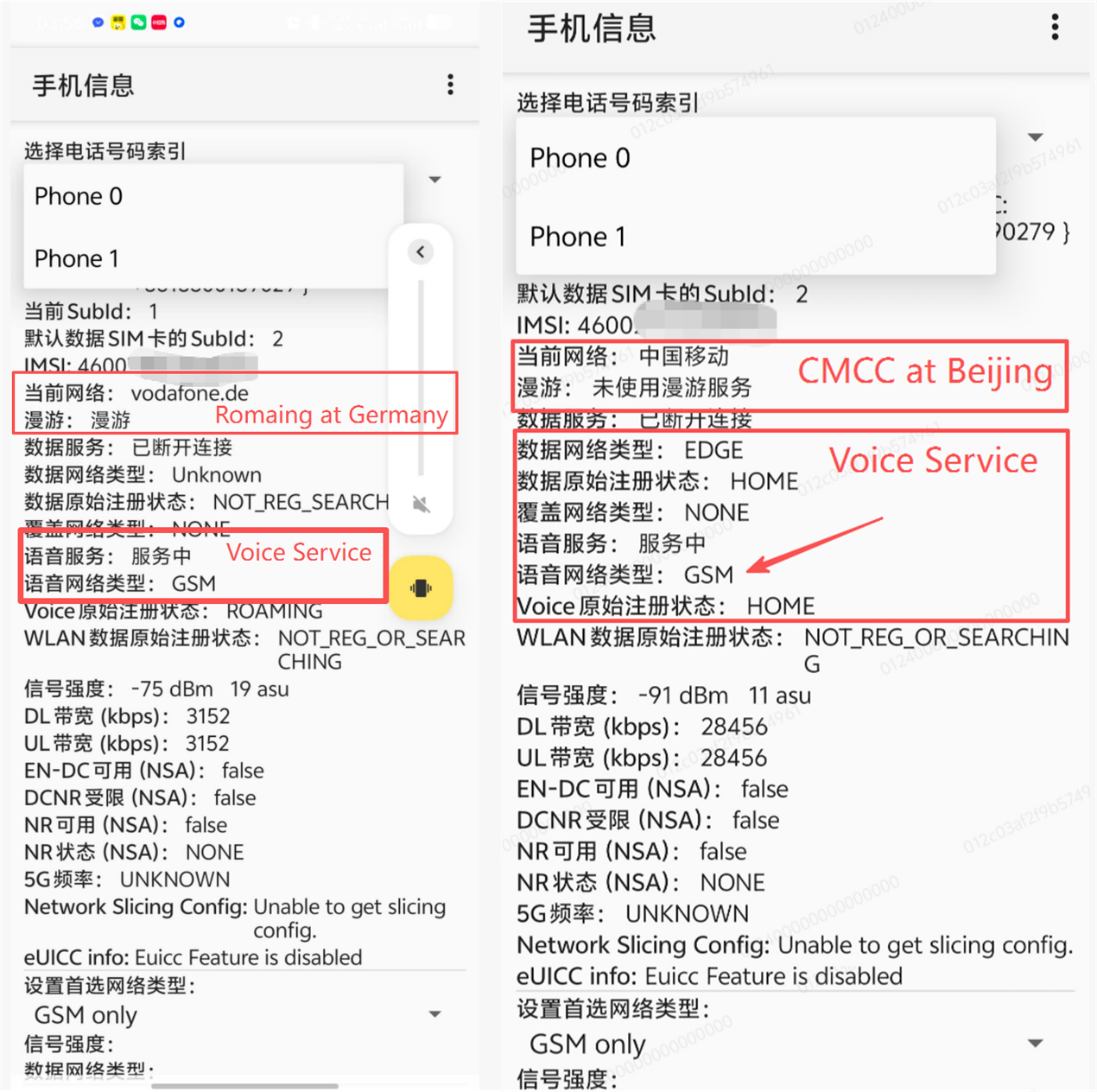
The scenario for a ULBC call over GEO is critical: it is the last resort when all other communication methods have failed. Before turning to a GEO device, a user has likely already tried their regular mobile phone. In poor conditions, that phone would have automatically fallen back to its most robust mode, which is often AMR-NB.

This experience sets the user's baseline expectation. The GEO call, being the final option, must be at least as reliable as the NB fallback the user just experienced. If a WB-only ULBC fails or provides poor quality in conditions where an NB connection would have worked, the "last resort" service has failed to meet its most basic requirement.

Therefore, NB support in ULBC is not just for connecting to other NB phones; it is essential for ensuring the service is a truly robust and reliable lifeline that meets the user's expectation of a final, working connection. Where applicable, higher bandwidths, e.g., WB, are expected to provide even higher quality.

**3.3. Typical Use Case for ULBC in GEO Scenarios: Emergency Communications**

A primary application for GEO voice calls is in emergency communications, such as a rescue team operating in a remote wilderness like the Himalayan mountains. To understand the requirements for this scenario, we must first consider the user’s expectation for a ‘last resort’ communication service. Even in well-developed areas with comprehensive network construction, such as Erlangen, Germany, or Beijing, China, GSM coverage acts as the ultimate fallback, as shown in Figure 1. Users experience a graceful degradation of audio quality from NR (e.g., Band 78) to LTE, and finally to GSM. Because GSM often utilizes favorable frequency bands for better coverage, it becomes the most reliable terrestrial connection, establishing the NB call as the user's baseline for a 'last resort' service.



**Figure 1: GSM connections in both Erlangen, Germany and Beijing, China**

Building on this, the rescue mission may consist of several squads, each equipped with a ULBC-enabled GEO phone, coordinating with a central base camp that is often connected to the TN via a PSTN gateway. This scenario presents a mixed-connectivity environment:

**Squad A** may be deep within the search area, completely outside TN coverage and relying solely on GEO voice.

**Squad B** might be operating at the fringe of the coverage area, where their terrestrial service has fallen back to a GSM network.

The **Base Camp** and other emergency response agencies are typically accessed via PSTN, which is an NB service.



**Figure 2: Emergency Communications: A Mixed-Connectivity Rescue Mission in the Himalayas**

In such critical situations, the terminating audio endpoints are almost exclusively NB. Emergency communication systems prioritize robustness, employing traditional NB audio codecs (e.g., Codec2, MELP) designed for reliability under extreme conditions. This makes the inefficiency described in section 3.1 even more severe. Transmitting wideband audio over a scarce satellite link, only for the upper frequencies to be discarded by an NB endpoint like a PSTN-based command center, is an unacceptable waste of resources in a life-or-death situation. As this emergency rescue scenario represents a key and practical use case for GEO voice (already in use in real-world rescue missions, e.g., in China, see the link to the [video](https://www.bilibili.com/video/BV1NvAJeaEcs/?share_source=copy_web&vd_source=b4a857c132244a90796c387749f76672&t=318) <https://www.bilibili.com/video/BV1NvAJeaEcs/?share_source=copy_web&vd_source=b4a857c132244a90796c387749f76672&t=318>, as shown the screen-snap in Figure 3, which the satellite voice call is the same technology that source also demonstrated on the Geneva meetings), it underscores the necessity of NB support in ULBC.



**Figure 3: A real example that GEO voice call used for rescue and Emergency Communications**

Consequently, the evaluation of any ULBC candidate shall prioritize testing for intelligibility and robustness in NB mode to ensure it is fit for its primary purpose.

**3.4. Performance at Very Low Bitrates**

A primary concern for any low-bitrate codec is that forcing a wider bandwidth spreads the available data too thinly, which can degrade core speech clarity rather than improve it. Research on modern neural codecs has shown that at very low bitrates, a lower sampling rate can achieve higher perceptual quality. Forcing a WB codec to operate at ~1 kbps may compromise intelligibility, especially in the presence of packet loss where a simpler, band-limited NB signal could be more robustly reconstructed.

**3.5. Complexity and Power Consumption**

Modern AI-based codec architectures do not scale in complexity as gracefully as older technologies. Doubling the input sampling rate from NB to WB can lead to a 2x to 4x increase in complexity for common CNN or Transformer-based models [5]. If WB-only is mandated, it would impose an unnecessarily high computational burden, which is a critical issue for the power-constrained mobile devices expected in the ULBC ecosystem. A native NB mode offers the opportunity to deliver high-quality voice at a significantly lower complexity and power budget.

**4. Experimental Analysis for Higher Bandwidth**

While the preceding sections argue for the necessity of NB, this section uses experimental data to demonstrate that mandating support for Super-Wideband (SWB) or Full band (FB) is an inefficient use of resources for a bitrate-constrained codec like ULBC. The data shows that Wideband (WB) already achieves excellent quality at low bitrates, and the marginal gains from SWB come at a significant bitrate cost. To provide concrete data, an experiment was conducted using the publicly available Descript Audio Codec (DAC) [6].

**4.1. Experiment Setup**

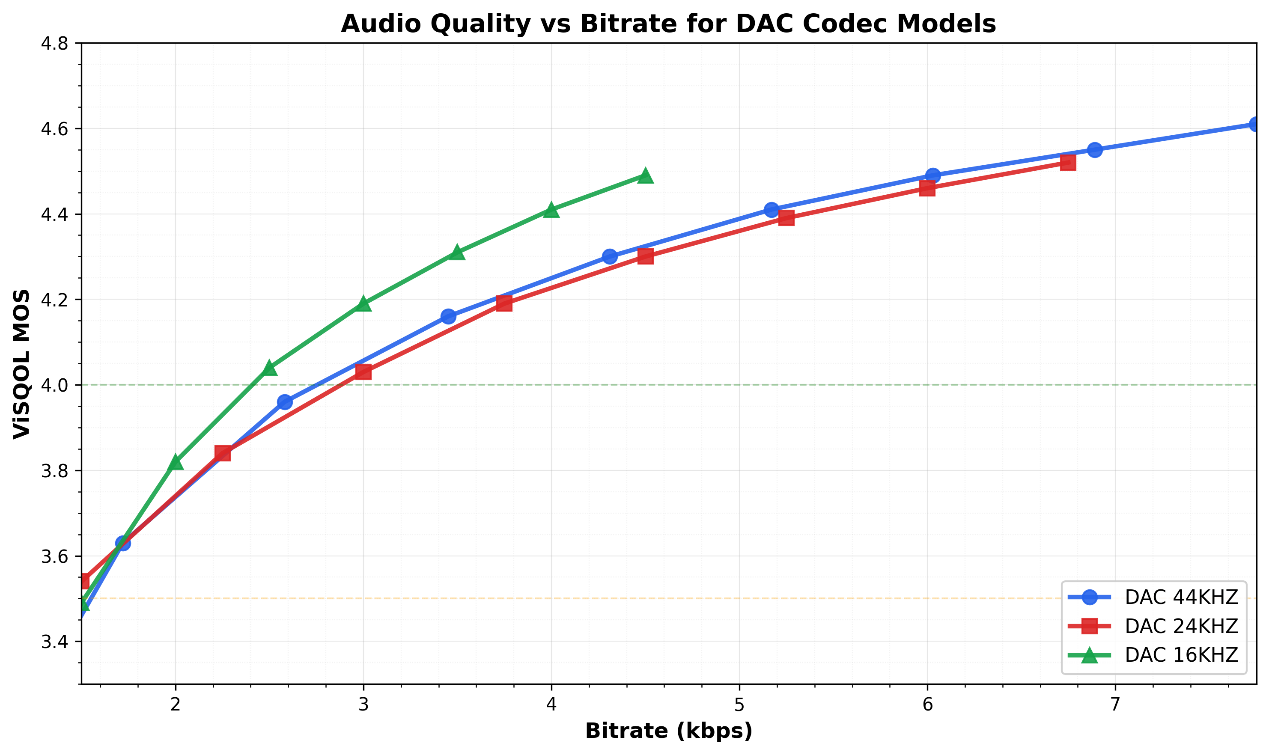
The evaluation was performed using pre-trained DAC models operating at three different sampling rates: 44.1 kHz [7], 24 kHz (SWB) [8], and 16 kHz (WB) [9]. A set of 100 clean speech samples from the MS-SNSD [10] dataset was used as the test corpus. For each model, the effective bitrate was varied by changing the number of active quantization codebooks from one to nine. The output quality was measured objectively using the ViSQOL [11] algorithm in speech mode, which provides a Mean Opinion Score (MOS) estimate.

The key specifications for the tested DAC models are as follows:

|  |
| --- |
| **Model specifications:**  - 44.1kHz: stride [2,4,8,8] (512x), 9 codebooks, 10-bit each  - 24kHz: stride [2,4,5,8] (320x), 32 codebooks, 10-bit each  - 16kHz: stride [2,4,5,8] (320x), 12 codebooks, 10-bit each  **16 kHz model (WB):**  Compression: 320x (via stride [2,4,5,8])  Frame Rate: 50 Hz  Codebooks: 12 available (10-bit each)  Bitrate per codebook: 0.50 kbps  **24 kHz model (SWB):**  Compression: 320x (via stride [2,4,5,8])  Frame Rate: 75 Hz  Codebooks: 32 available (10-bit each)  Bitrate per codebook: 0.75 kbps  **44.1 kHz model (SWB):**  Compression: 512x (via stride [2,4,8,8])  Frame Rate: ~86.1 Hz  Codebooks: 9 available (10-bit each)  Bitrate per codebook: ~0.86 kbps |

**4.2. Key Findings**

The results in Figure 4 show that Wideband (WB) provides a highly efficient path to excellent quality, while Super-Wideband (SWB) offers diminishing returns.



**Figure 4: Audio Quality vs Bitrate for DAC Codec Models (16 kHz, 24 kHz and 44.1 kHz analysis;)**

* The WB model achieves excellent perceptual quality (ViSQOL MOS > 4.0) at a competitive bitrate of around 2.5 kbps.
* The 24 kHz SWB model requires a higher bitrate to achieve the same quality as the WB model. More importantly, increasing the sampling rate further to 44.1 kHz provides barely any perceptible improvement over the 24 kHz model, making it a highly inefficient use of bandwidth for voice content .
* While the exact MOS scores may differ with other audio samples or objective metrics, this experiment reveals a clear and valuable trend: for a service like ULBC, the bitrate cost of selecting WB to SWB is not justified by the corresponding in quality.

**5. Proposal**

Based on the discussion and the experimental evidence, we propose the following design constraints for the ULBC audio bandwidth:

1. The codec shall be able to operate with an 8 kHz sampling rate, supporting an audio bandwidth of 50 – 4000 Hz.
2. The codec shall be able to operate with a 16 kHz sampling rate (50 – 8000 Hz audio bandwidth) to offer enhanced quality where channel conditions and device capabilities permit. For example, WB support can be limited to higher bitrates than NB operation.
3. The necessity and feasibility of including Super-Wideband (SWB) and Fullband (FB) support remains for further study.

We propose to update the design constraints table in TR 26.940 as follows:

\* \* \* First Change \* \* \* \*

Table 6.2-1 List of ULBC design constraint parameter

| Parameter | Design Constraint | Note |
| --- | --- | --- |
| Bit rates |  |  |
| Sample rate and audio bandwidth | The ultra low bitrate codec shall support sampling rates of 8kHz (NB) and 16kHz (WB). | The supported audio bandwidth for:  - NB ranges from 50 – 4000Hz  - WB ranges from 50 – 8000Hz |
| 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 |
|  |  |  |

\* \* \* End of Changes \* \* \* \*

**Reference**

[1] SP-250378 "New SID on Ultra Low Bitrate Speech Codec".

[2] S4-251394 "Pseudo-CR on Design Constraints Audio Bandwidth".

[3] Public switched telephone network [Internet]. Wikipedia, The Free Encyclopedia; Available from: https://en.wikipedia.org/wiki/Public\_switched\_telephone\_network

[4] 4G/LTE - Call Processing: CS Fallback [Internet]. ShareTechNote; [cited 2025 Aug 15] Available from: https://www.sharetechnote.com/html/Handbook\_LTE\_CS\_Fallback.html

[5] S4-251333 "Considerations on measuring ULBC complexity ".

[6] R. Kumar, P. Seetharaman, A. Luebs, I. Kumar, and K. Kumar, "High-Fidelity Audio Compression with Improved RVQGAN," arXiv preprint arXiv:2306.06546, 2023.

[7] Descript, Inc., "Descript Audio Codec (DAC) Pretrained Model 44.1 kHz," Version 0.0.1, Feb. 2023. [Online]. Available: https://github.com/descriptinc/descript-audio-codec/releases/download/1.0.0/weights\_44khz\_16kbps.pth

[8] Descript, Inc., "Descript Audio Codec (DAC) Pretrained Model 24 kHz," Version 0.0.4, Jun. 2023. [Online]. Available: https://github.com/descriptinc/descript-audio-codec/releases/download/0.0.4/weights\_24khz.pth

[9] Descript, Inc., "Descript Audio Codec (DAC) Pretrained Model 16 kHz," Version 0.0.5, Jul. 2023. [Online]. Available: https://github.com/descriptinc/descript-audio-codec/releases/download/0.0.5/weights\_16khz.pth

[10] Reddy, Chandan KA, et al. "A scalable noisy speech dataset and online subjective test framework." arXiv preprint arXiv:1909.08050 (2019).

[11] Chinen, Michael, et al. "ViSQOL v3: An open source production ready objective speech and audio metric." 2020 twelfth international conference on quality of multimedia experience (QoMEX). IEEE, 2020.

[12] GSMA. The State of Mobile Internet Connectivity 2024 [Internet]. London: GSMA; 2024 Oct [cited 2025 Aug 15]. Available from: www.gsma.com/somic

[13] GSMA. The Mobile Economy 2024 [Internet]. London: GSMA; 2024 [cited 2025 Aug 15]. Available from: www.gsma.com