INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2001-2004

COM 16 – LS 21 – E

English only

Original: English

Question(s):	9/16	Geneva, 20-30 May 2003

LIAISON STATEMENT

Source:	ITU-T SG 16
Title:	LS to 3GPP, 3GPP2 and IETF on Transport Layer for MSC-VBR and EV

LIAISON STATEMENT

To:	**3GPP, 3GPP2 and IETF**
Approval:	**Agrred to at ITU-T SG 16 meeting (Geneva, 20-30 May 2003)**
For:	**Action**
Deadline:	**10 November 2003**

Contact:	Mr Yushi Naito Mitsubishi Electric Corporation 5-1-1 Ofuna Kamakura 247-8501 Japan	Tel: +81 467 41 2449 Fax: +81 467 41 2019 Email: naitoy @ isl.melco.co.jp
	Mr Jonas Svedberg Ericsson Research Sweden	Tel: +46-920-202854 Fax: +46-920-202099 Email: Jonas.Svedberg @ ericsson.com
	Mr Eyal Shlomot Conexant Systems 4000 MacArthur Blvd Newport Beach, CA 92660 USA	Tel: +1-949-579-4988 Fax: +1-949-579-6361 Email: eyal.shlomot @ mindspeed.com

ITU-T Q.9/16 continued its work toward the standardisation of a Multi-mode Source Controlled Variable-Bit Rate speech codec, called MSC-VBR, and for an Embedded Variable rate codec, called EV. One of the main applications for both variable rate codecs is 3G cellular telephony, and another important application is VoIP, possibly transported by 3G cellular networks. Both coding systems incorporate elaborated schemes for trade off between bandwidth, bit-rate, and quality. In order to properly progress in the standardisation process of the MSC-VBR and of the EV, Q.9/16 wishes to ask 3GPP and 3GPP2 for further clarifications regarding the requirements and limitations of the transport layer for speech communications applications in their current and future networks. Such clarifications are required to ensure that the building blocks of the MSC-VBR and the EV will fit into the transport framework of these networks.

Q.9/16 kindly requests both organizations to inform Q.9/16 about:

A) Specific transport characteristics of the respective air interface transport layer and/or network transport layer (of applications you identify to benefit from an MSC-VBR codec), such as:

- Bit rates limitations and requirements for sets of building blocks for MSC-VBR codec.
- Typical frame error patterns and ranges and/or bit error patterns and ranges for MSC-VBR codec transport.
- Maximum number of concurrent frame rates, (frame wise switching is assumed to be possible between concurrent frame rates).
- Any other air interface and/or network interface characteristics that may affect the transport of voice using an MSC-VBR type of codec, e.g. uplink and downlink specifics that may affect the requirements in the MSC-VBR ToR.

B) Specific transport characteristics of the respective air interface transport layer and/or network transport layer (of applications that you identify to benefit from an EV codec), such as:

- The availability of a transport layer that may use the EV rate flexibility.
- Bit rates limitations and requirements for sets of building blocks for EV codec.
- Typical frame error patterns and ranges and/or bit error patterns and ranges for EV codec transport.
- Any other air interface and/or network characteristic that may affect the transport of voice using an EV type of codec (e.g. QoS variability possibilities within a transport layer).

C) Specific transport characteristics of the VoIP(RTP/UDP/IP) channels that have been specified for the respective system, e.g. requirements imposed by physical channels pre-selected for system transport of VoIP, and possible requirements imposed on the codec by the use of selected link level Header Compression schemes in combination with VoIP.

Please find the current Terms of Reference (ToR) for both approaches attached to this document, (note that both ToR should be considered preliminary). Detailed feedback from both organizations will be valuable in finalizing these ToR.

Attachments:
1) Annex A- Proposed Terms of Reference of MSC-VBR (Draft 4.0)
2) Annex B - Proposed Terms of Reference for Embedded VBR (EV) speech codec standardization

ANNEX A

Proposed Terms of Reference of MSC-VBR (Draft 4.0)

The ad-hoc group for MSC-VBR reviewed the ToR, as provided by the Rapporteur in TD.23. Since the text of the ToR went through numerous changes, the ad-hoc group decided to generate a clean version of the ToR, which can be used for further work. The ad-hoc group also elaborated on the applications for the MSC-VBR, in order to be able to tighten the requirements and the objectives of the ToR.

1. Applications

Several possible applications were previously identified for the MSC-VBR, and are listed below, under Primary Applications and Secondary Applications.

Primary Applications:

- 3G wireless applications, e.g. IMT-2000
- VoIP
- Very low-rate PSTN visual telephony;
- Personal communications;
- Simultaneous voice and data systems; and
- Mobile-telephony satellite systems.

Secondary Applications:

- Circuit multiplication equipment (CME);
- Packet circuit multiplication equipment;
- Low-rate mobile visual telephony;
- Message retrieval systems; and
- Private networks.

The ad-hoc group wishes to elaborate further on the primary applications for MSC-VBR.

A significant shift in speech communication is taking place lately, where the traditional routing of speech through switched networks is replaced by speech communications through packet networks. The proliferation of wireless speech communication in 2G was a major driving factor in speech communication by packets, but the flexibility and the robustness of the packet networks make them very attractive for speech communications, either for enterprise applications or for wide area networks. Speech communication via packet networks is sometimes called “voice over packets”, or VoP. Currently, VoP applications are still tightly connected to the switch networks (via gateways), since the switched networks are still used for a significant portion of speech traffic.

The MSC-VBR will play a central role in providing a unified high-quality speech for each packet network (3G wireless, WAN, Enterprise), for the connection between the switched network and the packet networks via gateways, and for TFO and TrFO between the networks. The MSC-VBR will be in particular suitable for networks with “soft capacity” property, such as 3G wireless networks and IP networks.

For example, for VoIP, the current situation is that the encoded speech bitstream will be encapsulated using IP/UDP/RTP. According to a document IETF RFC 1889 for Real Time protocol standard the decoder will send to the encoder some RTCP information on the status of the transmission (e.g. packet loss ratio). As defined in the standard, this information can only be sent with a minimum time interval of 5 seconds. Then, it will not be possible to have a good control of the source bitrate according to the network state. Nevertheless a good way to avoid network congesting would be to use source controlled coders which are going to reduce the bitrate at a minimum but still keeps the high encoding quality.

ITU-T already standardized many codecs for narrow band speech coding (e.g. G.729 and its Annexes E, D for bitrates at (6.4, 8, 11.8 kbit/s), G.723.1 at bitrates of 5.3-6.3 kbit/s). The utilization of DTX and CNG could give a flavour of source controlled but more elaborated strategies are yet available.

Besides these codecs were not designed for networks which can have high packet error rate. The MSC VBR should address these points. The adoption of an advanced source controlled variable rate speech coding standard for all of these said networks will not only improve the performance of each network, but will enable seamless speech interoperability between all of these networks.

2. Open Questions in ToR and Recommendation by Ad-Hoc Group

In this section we list the questions that are open on the ToR and the recommendation of the ad-hoc group for each of the questions.

Table 1 - Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm (Note 1)

Parameter	**Requirement(s)**	**Objective(s)**
1. Building blocks of the MSC-VBR (Note 2) ***Details will be re-visited after receiving comments from other bodies.***	Several different building blocks operating at certain total source bit-rates. 0 kbit/s building block shall be one of the building blocks.	
2. Frame size	The frame size should be set to an integer multiple of 5, 10 or 20 ms	
3. Operating modes (Note 3), (Note 4)	Multiple modes resulting in different targeted average bit-rates (ABRs). 4 modes will be considered: WB(H): highest ABR mode for wideband speech WB(L): lowerABR mode for wideband speech NB(H): higher ABR mode for narrowband speech NB(L): Lower ABR mode for narrowband speech	Continuous ABR ranging between that of WB(H) and NB(L)
4. Maximum bit-rate building block allowed	WB(H): 13.3 kbit/s WB(L): TBD kbit/s NB(H): 8.5 kbit/s NB(L): 6.4 kbit/s	MVBR(H): 13 kbit/s MVBR(M1): for further study MVBR(M2): for further study MVBR(L): for further study

Table 1 (Part 2/9)

Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm

Parameter	**Requirement(s)**	**Objective(s)**
5. Average bit-rate (ABR) (as a function of voice activity factor, VAF, measured on clean speech) (Note)	WB(H): around that of VAF×[13.2 kbit/s for clean speech condition + 1.2 kbit/s for background noise conditions] WB(L) TBD NB(H): around that of VAF×[6.6 kbit/s for clean speech condition + 0.6 kbit/s for background noise conditions] NB(L): around that of VAF×[4.4 kbit/s for clean speech condition + 0.4 kbit/s for background noise conditions]	As low as possible
6. Speech quality in error-free condition at nominal input level of –26.15 dB with respect to the OVL point (-20 dBm0) ***(Chech values for WB cases)*** ***Target quality shall be re-visited after receiving comments from other bodies.Contributions are solicited*.**	WB(H): TBD WB(L) TBD NB(H): TBD NB(L): TBD	TBD TBD TBD TBD

Table F.1 (Part 3/9)

Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm

Parameter	**Requirement(s)**	**Objective(s)**
7. Operating mode switching (Note6)	No annoying effect. Resulting quality shall not be worse than that of lower layer involved	
8. Speech quality dependency on the input signal level between –36.15dB and -16.15 dB with respect to the overload point ***(Chech values for WB cases)*** ***Target quality shall be re-visited after finalizing the nominal levels.***	WB(H): TBD WB(L): TBD NB(H): TBD NB(L): TBD	TBD TBD TBD TBD

Table 1 (Part 4/9)

Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm

9. Quality dependency on speakers	*for further study* *(TD 39(GEN) is the replying LS from SG12. Details shall be re-visited after receiving responses from other bodies)*	*WB(H): TBD* *WB(L): TBD* *NB(H): TBD* *NB(L): TBD*
10. Tandeming capability of each mode for speech at nominal input level (Note7) ***Target quality shall be re-visited visited after receiving comments from other bodies. Contributions are invited for all requirements.***	2 tandem of WB(H): TBD 2 tandem of WB(L): TBD 2 tandem of NB(H): TBD 2 tandem of NB(L): TBD	
Cross tandeming capability of different modes for speech at nominal input level (Note7) ***Target quality shall be re-visited visited after receiving comments from other bodies. Contributions are invited to revisit this aspect.***		Tandem of WB(H) (first) into WB(L) (second) [WB(H)→WB(L)]: Not worse than that of single WB(L) at nominal input level Tandem of WB(L) (first) into NB(H) (second) [WB(L)→NB(H)]: Not worse than that of single NB(H) at nominal input level Tandem of NB(H) (first) into NB(L) (second) [NB(H)→NB(L)]: Not worse than that NB(L) at nominal input level Tandem of NB(L) (first) into TBD (second) [NB(L)→TBD]: Not worse than that of G.729D at nominal input level
Tandeming with other ITU-T speech encoding standards (Note 7) Tandeming with regional digital mobile radio (DMR) standards (Note 7)	For Further Study	

Table 1 (Part 5/9)

Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm

11. Speech performance with bit errors for an input signal nominal level of -26.15 dB with respect to the overload point:		
- Random BER = 10^{-3} (Note8)	for further study	WB(H):Not worse than WB(H) at error free conditions WB(L):Not worse than WB(L) at error free conditions NB(H):Not worse than NB(H) at error free conditions NB(L):Not worse than NB(L) at error free conditions
Detected Frame Erasures (Note 9) - 3% Bursty (air interface pattern) (Note10) -5% Bursty (VoIP pattern) (Note 11) ***Target quality shall be re-visited visited after receiving comments from other bodies.*** ***Contributions are invited.***	WB(H):TBD WB(L): TBD NB(H): TBD NB(L): TBD	WB(H):TBD WB(L):TBD NB(L): TBD NB(L): TBD
Undetected Burst Errors: ***Contributions are invited.***		For further study

Table 1 (Part 6/9)

Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm

12. Capability to transmit music (e.g. music on hold) ***Target quality shall be re-visited visited after receiving comments from other bodies.***	WB(H):TBD WB(L): TBD NB(H): TBD NB(L): TBD	
Performance in the presence of background music		For further study

Table 1 (Part 7/9)

Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm

13. Performance in the presence of background noise (Note 10) ***Introduction of noise suppression: contributions are invited*** ***Testing methodology: Replies can be found in TD 57(GEN).*** ***Target quality shall be re-visited visited after receiving comments from other bodies.*** ***Selection of noise type to be tested shall be discussed in test planing.***		
– Car Noise at a SNR of 15 dB	WB(H): TBD WB(L): TBD. NB(H): TBD NB(L): TBD	WB(H): TBD WB(L): TBD NB(L): TBD NB(L): Not worse than that of ITU-T Rec. G.729 at 8 kbit/s at the input SNR of 22 dB.TBD
– Babble Noise at a SNR of 20 dB	WB(H): TBD WB(L): TBD NB(H): TBD NB(L): TBD	WB(H): TBD WB(L): TBD NB(H): TBD NB(L): TBD
– Street Noise at a SNR of 20 dB	WB(H): TBD WB(L): TBD NB(H): TBD NB(L): TBD	WB(H): TBD WB(L): TBD NB(H): TBD NB(L): TBD
– Interfering Talker at a SNR of 20 dB	WB(H): TBD WB(L): TBD NB(H): TBD NB(L): TBD	WB(H): TBD WB(L): TBD NB(H): TBD NB(L): TBD

Table 1 (Part 8/9)

Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm

Parameter	**Requirement(s)**	**Objective(s)**
14. Capability to operate at wider bandwidths		10-15 kHz speech band-width, when operating at higher bit-rates (Note 13)
15. One way coder/decoder delay (Note 14) - total CODEC delay ***Contributions are invited***	≤ 60 ms	as low as possible
16. Capability to transmit voiceband data (for NB modes) (Note 15)	V.21 300 bit/s V.22 600/ 1200 bit/s V.23 1200 bit/s V.26 bis 1200 bit/s V.26 ter 1200 bit/s	V.22bis 2400 bit/s V.26 bis 2400 bit/s V.26 ter 2400 bit/s V.27 ter 2400/4800 bit/s V.29 up to 4800 bit/s
17. Capability to transmit signaling and information tones (for NB modes) ***Contributions are invited on channel impairment conditions***	DTMF, TTY/TDD Modulation schemes necessary to support ITU.T Rec. V.18. Tones have to be transmitted with as little distortion as possible.	MF
18. Capability to support speech recognition ***Contributions are invited***		for further study)
19. Capability to support speaker recognition and verification ***Contributions are invited***		for further study
20. Encoder/Decoder synchronisation ***Contributions are invited***		for further study
21. Effect of switching signal source to the codec (Note		
21. Idle channel noise (weighted) Idle channel noise (single frequency) (for NB modes) ***further sudy for WB modes***		(Note17)

Table F.1 (Part 9/9)

Performance Requirements and Objectives for MSC-VBR Speech-Coding Algorithm

Parameter	Requirement(s)	Objective(s)
22. Complexity ***Contributions are invited***	*Combined encoder and decoder to be implementable on a commercially available (single CPU) fixed-point DSP device (Note18)*	As low as possible (Input for ITU-T members regarding RAM limitation is needed.)
23. Implementation	Bit exact 16-bit fixed-point using ITU-T library tools	Interoperable floating-point implementation to follow. Interoperable with the implementation using 40 bit accumulator
24. Specification description	Bit-exact fixed-point modular ANSI-C code using ITU-T library tools in Electronic format (Note19)	
25. Interoperability with existing standards (Note20)		desirable

Notes to Table 1

1. The requirements and objectives refer to the distortion introduced between the input PCM interface of the coder and the output PCM interface of the decoder. Because the principal intended application of this coder is for terminal equipment and because in future digital networks it may be possible to transport this codec's variable bit-rate stream without conversion to other ITU-T standards, the testing and specification will be for linear PCM input and output with at least 13 bits of accuracy. Since the codec must interface to present standards and because there are network applications (such as DCME) which may use MSC-VBR speech coding, a specification for interfacing 8 bit PCM (A/μ law) will also be recommended.
2. Building blocks refer to speech coding sub-systems operating at particular total source bit-rates. For example, a building block might operate at a total source rate of 6 kbit/s. The 0 kbit/s gross rate building block denotes DTX operation.
3. The operating mode of a MSC-VBR coding system has the following properties:
   - A particular operating mode with a targeted average bit rate (ABR) can be realized by using an appropriate mix of different building blocks, with a certain percentage for each of them. For example, 70% of active speech might be coded with 8 kbit/s, while the other 30% coded by 6 kbit/s. And 90% of inactive speech area (silence or background noise) might be coded at 1 kbit/s and the other 10% coded at 2 kbit/s.
   - A particular operating mode can use all, or a sub-set, of the available building blocks.
   - Multiple operating modes at different target ABRs can be designed to provide flexibility for application control, such as network control of the desired ABR for congestion control, trade-off between quality and ABR (or system capacity), i.e. higher quality at higher ABR (therefore lower system capacity) or lower quality at lower ABR (therefore greater system capacity).
4. Bandwidth for “WB” is assumed to be 50 Hz to 7 kHz, but it can be extended to more than 7 kHz.
5. It is understood that the average bit-rate (ABR) is typically a changing figure depending upon the characteristics of the input speech signal, such as voice activity factor (VAF, measured on clean speech), the level of the background noise, the characteristics of the background noise, etc. It is therefore appropriate to define the ABR requirement as a function of the VAF, with increased margin for background noise conditions. The host lab will provide a supplemental database for the measuring of the ABR by the proponents during the development, and final tuning capability of the ABR (on the test materials by the host lab)
6. Operating mode switching refers to the on-the-fly change of operating mode, for example, from WB(H) to NB(H) during a phone call. Frequent/non-frequent switching across the different bandwidth should be taken into account.
7. Since tandeming is such that a circuit switched network connects between two Packet switched networks, tandeming will be tested with coding frame boundary aligned and the ITU-T Rec. G.711at 64 kbit/s as an intermediate coder.
8. The bit error rate is defined at the speech decoder input after channel decoding.

9. Detected frame erasures require that the decoder is informed that the incoming frame is in error. Frame erasures will assume 20ms frame length. Therefore if the frame length of a codec is 10ms, the marked erasure will result in two codec frame erasure, and if the frame length is 5 ms the marked erasure will result in four codec frame erasure.
10. Since MSC VBR will be mostly used in packet switched networks (such as wireless or wireline networks), frame error patterns of such networks should be provided.
11. Frame erasure pattern for VoIP should be provided.
12. Noise suppression technique is allowed to be applied. The input SNR for reference codec shall be xdB higher than that for candidate codec.
13. Wider bandwidths(10-15 kHz) may be desirable for multimedia, news broadcasting, news gathering, VoIP and visual-telephone applications. Capability to operate at wider bandwidths may be separately specified in the future Annex to this draft Recommendation.
14. **Algorithmic delay** includes the frame size delay plus any other delays inherent in the algorithm (look-ahead, noise suppression and error correcting codes for algorithm purposes). **Processing delay** is the additional delay caused by implementation with a finite speed processor. The minimal processing delay is assumed to be the frame size. **Total codec delay** is the sum of algorithmic delay and processing delay. The **serial channel delay** is the delay caused by transmitting the channel signal over a serial channel matched to its bit rate as customarily done in tested hardware. (For block coders this delay is usually equal to the block size of the coder.) The **total system delay** consists of the algorithmic delay, the processing delay, the serial channel delay and such other delays caused by the test equipment and interfaces connected to these. The total system delay can be measured. The **test system delay** consists of the delay caused by the test equipment and the interface between the speech encoder/decoder and the test equipment. The test system delay can be measured by passing PCM data directly through the system, bypassing only speech encoder and decoder. The total codec delay can be calculated by subtracting the test system delay and serial channel delay from the total system delay
15. These modulation schemes are picked up from the Table 7 of ITU Recommendation G.763 (ADPCM DCME).
16.
17. Considering that for ITU-T Recommendation G.726 at 32 kbit/s the transfer characteristic for very low signal levels is highly non-linear (due to the mid-tread quantizer) and is likely to differ from that of a selected MSC-VBR algorithm. It may be necessary to set an objective performance requirement in accord with network and subscriber terminal idle channel noise requirement.
18. Early evaluation of the complexity (comprising encoder, decoder, VAD, noise suppression, post filtering etc) will be performed on the basis of the detailed level description of the algorithms (e.g. number of multiplication's, number of shifts, etc.). The final complexity evaluation will be based on ITU fixed-point library.
19. Modular means a software implementation made in accordance to the guidelines given in the ITU-T Software Tools Library user’s manual.
20. The interoperability means tandem free operation. It is possible that certain operating modes might be interoperable to some other operating modes of existing standards.

ANNEX B

Proposed Terms of Reference for Embedded VBR (EV) speech codec standardization

The following ToR proposal is the output of an ad-hoc group meeting.
Before terms of reference can be defined, it is necessary to consider the applications that the next generation audio coding algorithm is assumed to address. As an objective, it is assumed that a unique "toll-quality" audio embedded algorithm ought to be selected, that is, an algorithm capable of addressing all applications listed below. This approach is consistent with recent efforts within the ITU-T and is the only approach likely to permit the ITU-T to select algorithms with wider scope of applications than those selected by regional standards bodies.

In the following the applications foreseen for an Embedded Variable Bit Rate audio coder are listed. These applications are partitioned into two groups: a primary group and a secondary group. The primary group comprises those applications that should benefit from embedded schemes while having a great potential use i.e. applications that are most likely to employ EV audio coding early and in large numbers. As a result, primary applications are expected to "drive" the development of the standard, at least as regards schedule. The secondary group comprises those applications likely to benefit from the availability of a EV audio coding standard, but which are either unlikely to employ large numbers of EV audio coding devices or, at least on an interim basis, can also utilise some other audio coding standards without adversely impacting the economics of their application.
The following applications are proposed as primary applications:

- packetised voice (VoIP, VoATM, IP phone, private networks)
- high quality audio/video conferencing
- Applications that benefit from
- congestion control
- differentiated QoS,
- 3G wireless (Packet switched conversational multimedia, multimedia content distribution)
- Multimedia streaming (e.g. video + audio involving bit-rate tradeoff)

The following applications are proposed as secondary applications:

- multicast content distribution (offline/online)
- Message retrieval systems
- CME/Trunking equipment
- Applications that require music on hold (e.g. VoIP)

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

- Primary signals of interest are speech but in high quality audio conferencing, background signals shall be considered not as the noise any more, but as a part of signals that is carrying information
- To cope with heterogeneous accesses and terminals, it is important to consider bit-rate scalability but also bandwidth scalability and complexity scalability
- Narrowband/Wideband speech capability with HiFi bandwidth capability as objective (up to 20 kHz)
- In order to smoothen bandwidth switching effects, intermediate bandwidth or other means of providing a gradual transition will be needed.

- In spite of the development of broadband access (xDSL), a lot of users would only have service access via PSTN modems, or mobile links. The bit range shall allow both types of users to benefit from EV scheme: the bit range will have to cover low bit rate (<8 kbit/s) to higher bit rate (≅ 32 kbit/s). Also for mobile users, it is highly desirable to introduce bitrates compatible with mobile links.
- Bit-rate scalability of the codec that will enable the trade-off between video quality and speech quality will be appreciated in designing video conferencing system. Contributions are invited providing information on the granularity of bitrate scalability.
- To maintain a good quality of services requiring interactivity, it is necessary to maintain the overall delay as low as possible. But the delay requirement tends to have less importance in applications involving packetized voice, possibly combined with other media and/or in heterogeneous network environment. A trade-off must be found between low delays and flexibility (scalability, ability to operate in various conditions with many types of signals etc.).

Performance Requirements & Objectives

Performance requirements and objectives for an Embedded Variable Bit Rate audio coding algorithm are proposed in Table 1.

Table 1 Performance requirements and objectives for an embedded audio coding algorithm (Ver. 2.0)

3.3.1.1 Parameter	**Requirement**	**Objective**
1. Embedded bitstream • Number of layers • Bit rates in kbit/s • Bandwidths in kHz • Sampling rate in kHz	At least 5 layers R1≤6,4/R2≤12/R3≤14/R4≤24/R5≤32 R1,R2:[0,3-3,4]/R3-R5:[0,05-7] 8/16	Finer increments of bit-rate. 4/10/12/16/24+
2. Embedde bitstream for wider bandwidth extension (Note 1) • Number of layers • Bit rates in kbit/s • Bandwidths in kHz • Sampling rate in kHz	TBD TBD TBD TBD ***Contributions are invited***	2 layers 48 kbit/s[12kHz], 64 kbit/s[16kHz] R6:[0,02-12]/R7:[0,02-16] 24, 32 kbit/s
3. Speech quality in error-free condition at nominal input level of -26 dB with respect to the OVL point	Equivalent to R1: G.729 at 8 kbit/s R2: G.729 E at 11,8 kbit/s R3: G.722 at 48 kbit/s R4: G.722 at 56 kbit/s R5: G.722 at 64 kbit/s (Note 2)	Equivalent to R1: G729 E at 11,8 kbit/s R2: G711 at 64 kbit /s R3: G722 at 56 kbit/s R4: G.722 at 64 kbit/s R5: direct for Super Wideband extension, R6 (24 kHz sampling) equivalent to Direct in the corresponding bandwidth R7 (32 kHz sampling) equivalent to Direct in the corresponding bandwidth
4. Speech quality in error-free condition at nominal level of -16dB with respect to the OVL point	Equivalent to their respective references at the same input levels	Equivalent to their respective references at the same input levels

5. Speech quality in error-free condition at nominal level of -36 dB with respect to the OVL point	Equivalent to their respective references at the same input levels	Equivalent to their respective references at the same input levels
6. Effect of switching between layers at the decoder side. Special attention needs to be paid to bandwidth switching (Note 3)	No annoying effect. Resulting quality shall not be worse than that of lower layer involved	
7. Speech quality in error conditions for an input signal nominal level of -26 dB (Note 4) • BER = 10-3 • 3% FER Random • 3 % FER Bursty • Undetected Burst Errors	t.b.d	
8. Quality dependency on speakers	Not worse than their respective references at the same input levels	
9. Music in error-free condition at input signal nominal level –26 dB with respect to the overload point	R1: Recognisable as music Equivalent to R2: G.729 E at 11.8 kbit/s R3: G.722.2 at 12.65 kbit/s R4: G.722 at 48 kbit/s R5: G.722 at 56 kbit/s (Note 2)	Equivalent to R1: G.729 E at 11.8 kbit/s R2: G.711 at 64 kbit /s R3: G.722 at 48 kbit/s R4: G.722 at 56 kbit/s R5: G.722 at 64 kbit/s for Super Wideband extension, R6 (24 kHz sampling) equivalent to Direct at corresponding bandwidth R7 (32 kHz sampling) equivalent to Direct at corresponding bandwidth
10. Interoperability for speech • Interoperability with other ITU-T speech encoding standards • Interoperability with 2G and 3G mobile radio systems	For further study	Bit-stream interoperability is desirable.
11. Performance in the presence of background noises • Background music at a SNR of x dB • Office noise at a SNR of x dB • Car Noise at a SNR of x dB • Babble Noise at a SNR of x dB • Interfering Talker at a SNR of x dB • Street Noise at a SNR of x dB	Equivalent to R1: G729 at 8 kbit/s R2: G729 E at 11.8 kbit/s R3: G722.2 at 12.65 kbit/s R4: G.722.2 at 19.85 kbit/s R5: G.722.2 at 23.85 kbit/s (Note 2)	Equivalent to R1: G729 E at 11.8 kbit/s R2: G711 R3: G722.2 at 14.25 kbit/s R4: G.722.2 at 23.85 kbit/s R5: G.722 at 64 kbit/s (Note 5)
12. Algorithmic delay (Note 6) •	$\leq$ 60 ms	as low as possible
13. Frame size	multiple of 10 ms	

14. Capability to transmit voiceband data	Tbd	
15. Capability to transmit signaling and information tones	Tbd	
16. Capability to support speech recognition	For further study	
17. Capability to work with VAD algorithm for discontinuous transmission	Needed	
18. Encoder/Decoder synchronisation	For further study	
19. Multipoint Control Unit operation		mixing at lower complexity than decoding + encoding
20.Effect of switching signal sources to the codec (Note 7)	Tbd	For further study
21. Convergence time	Tbd	
22. Idle channel noise • Unweighted • Weighted • Single frequency	Tbd	
23. Complexity	Combined encoder and decoder to be implementable on a commercially available (single CPU) fixed point DSP device	As low as possible. Fine-grained complexity control both for encoder and decoder
24. Memory		As low as possible
25. Specification description and implementation	Bit-exact fixed-point modular ANSI-C code using basic operators set provided in the ITU-T Software Tool Library to follow.	Interoperable floating-point implementation (Electronic format)
26. A/D and D/A converter accuracy	Tbd	
27. Attenuation/frequency response of encoder and decoder analog circuitry	Tbd	

Notes to Table 1

1. This capability will be included as a future Annex.
2. Requirements for the Super-wideband extension (10-15 kHz bandwidth) should be provided.
3. Operating mode switching refers to the on-the-fly change of operating mode, for example, from R5 to R3 during a phone call. Frequent/non-frequent switching across the different bandwidth should be taken into account.
4. Packet network characteristics need to be taken into account. Core layers are expected to be less subjected to packet losses than enhancement layers
5. Objectives for the Super-wideband extension (10-15 kHz bandwidth) should be provided.
6. **Algorithmic delay** includes the frame size delay plus any other delays inherent in the algorithm (look-ahead, noise suppression and error correcting codes for algorithm purposes). **Processing delay** is the additional delay caused by implementation with a finite speed processor. The minimal processing delay is assumed to be the frame size. **Total codec delay** is the sum of algorithmic delay and processing delay. The **serial channel delay** is the delay caused by transmitting the channel signal over a serial channel matched to its bit rate as customarily done in tested hardware. (For block coders this delay is usually equal to the block size of the coder.) The **total system delay** consists of the algorithmic delay, the processing delay, the serial channel delay and such other delays caused by the test equipment and interfaces connected to these. The total system delay can be measured. The **test system delay** consists of the delay caused by the test equipment and the interface between the speech encoder/decoder and the test equipment. The test system delay can be measured by passing PCM data directly through the system, bypassing only speech encoder and decoder. The total codec delay can be calculated by subtracting the test system delay and serial channel delay from the total system delay
7. Switching signal source to the codec may occur when the pooled-codec configuration should be adopted by the system (e.g. DCMS).