3GPP2 S.R0080-0
Version 1.0
Version Date: February 20, 2003

3RD GENERATION PARTNERSHIP PROJECT 2 "3GPP2"

CDMA2000 Wideband Speech Codec

Stage 1 Requirements

EDITOR
Sassan Ahmadi, Ph.D.
+1 (858) 831-5916
E-mail: sassan.ahmadi@nokia.com

REVISION HISTORY

REVISION HISTORY		
Revision number	***Description***	***Date***
1.0	*Initial publication*	*20 February 2003*

Table of Contents

List of Tables

List of Figures

This section is not applicable

1. INTRODUCTION

Wideband speech and audio coding has been gaining popularity in recent years. The emergence of the 3rd generation cellular systems as well as an increasing public demand for applications such as Wireless multimedia, Voice over IP, Videophone, ISDN teleconferencing systems, Multi-point interactive audiovisual communication, and Audio streaming require improved voice quality and intelligibility. Multimedia services are among the main deliverables of 3G wireless communications. This implies the use of high quality audio and speech in multimedia content. Even in 3G voice applications the introduction of wideband speech would be an important step for CDMA service providers to deliver speech quality beyond the traditional limits of wireline communication systems.

This document describes the cdma2000 Wideband Speech Codec Stage-1 requirements.

2. REFERENCES

[1] TIA/EIA/IS-95: Mobile Station – Base Station Compatibility Standard for Dual-Mode Spread Spectrum Systems

[2] 3GPP2 C.S0001-C.S0006: A family of standards, which comprise the CDMA2000 Mobile Station-Base Station compatibility specification

[3] 3GPP2 C.S0014: Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems

[4] 3GPP2 C.S0020: High Rate Speech Service Option 17 for Wideband Spread Spectrum Communication Systems

[5] 3GPP2 C.S0030: Selectable Mode Vocoder Service Option 56 for Wideband Spread Spectrum Communication Systems

[6] ITU-T/P.341 Filter: Wideband (50Hz – 7 kHz) audio spectral shaping filter

[7] ITU-T/G.722.2: ITU-T Recommendation for AMR-WB Codec

3. DEFINITIONS AND ABBREVIATIONS

Codec – Combined-name mnemonic for Coder/Decoder. In this document codec and coder terms are used interchangeably.
RAM – Random Access Memory
ROM – Read Only Memory
VAD – Voice Activity Detector
WMOPS – Weighted Million Operations per Second
ADR: Average Data Rate
TrFO: Transcoder Free Operation
VoIP: Voice over Internet Protocol

4. GENERAL FEATURE DESCRIPTION

This section contains the requirements for the CDMA2000 wideband coder. The proposed requirements are printed in *italic font*, with accompanying discussion in normal font.

4.1. Compatibility

A number of requirements are imposed on the wideband codec because of its intended application in rate-set II of TIA/EIA/IS-95/CDMA2000 systems and in radio configuration 4 (reverse link)/5 (forward link) of TIA/EIA/IS-2000 air-interface.

4.1.1. Frame Size

The wideband codec shall be compatible with a frame size of 20 ms.

The wideband codec will operate in TIA/EIA/IS-95/CDMA2000 systems that use 20 ms frames.

The wideband codec shall be a variable rate codec that is limited to the following four packet sizes and source coding bit rates:

Full Rate (FR):	*266 bits/frame corresponding to 13.3 kbps*
Half Rate (HR):	*124 bits/frame corresponding to 6.2 kbps*
Quarter Rate (QR):	*54 bits/frame corresponding to 2.7 kbps*
Eighth Rate (ER):	*20 bits/frame corresponding to 1.0 kbps*

IS95/CDMA2000 systems allow only the above four data rates for variable rate codecs operating in radio configuration 4/5.

4.1.2. Audio Bandwidth

The wideband codec shall operate on input speech with audio bandwidth of 50Hz-7000Hz. The input speech is pre-processed with ITU-T/P.341 filter.

4.1.3. Input/Output Sampling Rate

The input/output sampling rate for the wideband codec shall be 16 kHz. Additionally the wideband codec shall be able to operate with 8 kHz sampled input speech with appropriate upsampling/downsampling filters that do not increase the overall algorithmic delay beyond the maximum allowable delay.

4.1.4. Required Modes of Operation

The wideband coder shall operate in three separate modes, each associated with a specific average data rate (ADR) limit. All of the operational modes require the end-to-end transport of the compressed speech packet.

These three modes will be known as mode 0, the premium mode, mode 1, the standard mode, and mode 2, the economy mode.

The wideband coder shall be designed such that it operates at a specified maximum frame rate and a specified minimum frame rate regardless of the mode of operation.

The wideband decoder shall be designed such that it does not require the mode of operation as out-of-band information.

4.1.5. Interoperability with 3GPP/AMR-WB (ITU-T/G.722.2)

While the original work item required an additional mode of operation that is interoperable with 3GPP/AMR-WB (ITU-T/G.722.2) at 12.65 kbps, the voice services sub-working group agreed that the interoperability with AMR-WB is a desirable feature for the CDMA2000 wideband codec. The sub-working group further decided not to include interoperability as a requirement, rather an optional mode of operation.

The sub-working group further agreed to include the interoperability as part of the selection criteria for the wideband codec. In order to select a wideband coder candidate as the standard for CDMA2000, a test plan shall be developed and approved by TSG-C1.1 and all candidate codecs shall participate in subjective evaluation tests to be performed by the Host and Listening laboratories (to be designated by TSG-C) according to the development schedule. The results of the subjective tests shall be statistically analyzed using standard statistical analysis techniques such as analysis of variance (ANOVA) with appropriate confidence intervals, where the candidate codecs are tested against the terms of reference in each condition and in all modes to ensure that they meet the requirements. Further, the highest-ranking candidates that meet the requirements shall be tested against each other to select a candidate wideband speech coder as the standard. If the results of the standard statistical analyses indicate that the highest-ranking codecs are statistically equivalent, then the candidate enabled with the interoperability feature shall be selected as the standard. Therefore, the interoperability feature shall decide between statistically equivalent wideband codecs and to break a tie.

To establish interoperability with 3GPP/ITU-T wideband coder, the output packets of the wideband coder in this mode shall be decodable by AMR-WB/G.722.2 at 12.65 kbps and vice versa under all conditions.

The ADR of the interoperable mode (mode i) shall be slightly higher than that of mode 0 and a quality equivalent to or better than 3GPP/AMR-WB at 12.65 kbps is targeted for this mode of operation under all conditions.

This mode shall only use the full-rate for active speech. Silence Descriptor (SID) frames shall be transmitted at quarter-rate. Eighth-rate frames shall substitute the Discontinuous Transmission (DTX) frames to maintain frame compatibility with 3GPP/AMR-WB at 12.65 kbps as well as continuity of transmission required in CDMA2000 systems.

The algorithmic delay of this mode shall not exceed 35 ms. The memory and complexity requirements of the wideband coder shall be the same whether or not the interoperable mode is included.

4.2. Delay

The algorithmic delay of all modes of the wideband coder shall not exceed 35 ms, however, an algorithmic delay less than or equal to the minimum delay of narrowband SMV (TIA/EIA/IS-893) is preferable; i.e., 33 ms.

Algorithmic delay includes frame size and look-ahead delays in addition to any other delays inherent to the design of the encoder. Processing delay is not included.

4.3. Average Data Rate and Quality Targets for Clean Condition

For each mode and for clean speech, the ADR should not exceed the average data rate limits specified in Table 4.3-1. The reference AMR-WB coder shall operate in VAD/DTX mode under clean, channel error, and noise conditions.

Mode	Average Data Rate	FER Condition	Requirement	Objective
0	Same as TIA/EIA/IS-733	No Errors	No worse than AMR-WB 14.25k	No worse than AMR-WB 15.85k
		1% FER	No worse than AMR-WB 14.25k @ 1% FER	No worse than AMR-WB 15.85k @ 1% FER
		3% FER	Better than AMR-WB 14.25k @ 3% FER	Better than AMR-WB 15.85k @ 3% FER
		6% FER	Better than AMR-WB 14.25k @ 6% FER	Better than AMR-WB 15.85k @ 6% FER
1	Same as average of TIA/EIA/IS-733 (QCELP 13K) and TIA/EIA/IS-127 (EVRC)	No Errors	No worse than AMR-WB 12.65k	
		1% FER	No worse than AMR-WB 12.65k @ 1% FER	
		3% FER	Better than AMR-WB 12.65k @ 3% FER	
		6% FER	Better than AMR-WB 12.65k @ 6% FER	
2	Same as TIA/EIA/IS-127 (EVRC)	No Errors	No worse than AMR-WB 8.85k	
		1% FER	No worse than AMR-WB 8.85k @ 1% FER	Better than AMR-WB 8.85k @ 1% FER
		3% FER	Better than AMR-WB 8.85k @ 3% FER	
		6% FER	Better than AMR-WB 8.85k @ 6% FER	
Mode i	Slightly higher than that of mode 0	No Errors	No worse than AMR-WB 12.65k	Optional Mode
		1% FER	No worse than AMR-WB 12.65k @ 1% FER	
		3% FER	No worse than AMR-WB 12.65k @ 3% FER	
		6% FER	No worse than AMR-WB 12.65k @ 6% FER	

Table 4.3-1: ADR and quality targets for each mode in clean condition

*Note: All FER conditions are inclusive of a 1% Dim and Burst Signaling, for Test and Reference Codecs.

Due to the variable rate nature of the wideband coder, the bit rate limitation is specified as an average data rate. This ADR is specified for each mode of operation, and for both clean and noisy input speech. This is under the assumption that narrowband database is directly obtained from the corresponding wideband

database after appropriate filtering and decimation. Furthermore, it must be ensured that the process of wideband to narrowband conversion does not change the speech activity factor.

4.4. Average Data Rates and Quality Targets for Noisy Conditions

For each mode and for clean speech, the ADR should not exceed the average data rate limits specified in Table 4.4-1. *The packet-level signaling indicated in Table 4.4-1 refers to introducing dim and burst signaling after encoder and before the decoder on the output frame, which may involve a full-rate to a reduced-rate conversion. The reference AMR-WB coder shall operate in VAD/DTX mode under clean, channel error, and noise conditions.*

Mode	Average Data Rate	Noise Condition	Requirement	Objective
0	Same as TIA/EIA/IS-733 (QCELP 13K)	Car noise @ 10 dB SNR	No worse than AMR-WB @ 14.25 kbps with 10 dB SNR car noise input	No worse than AMR-WB @ 15.85 kbps with 10 dB SNR car noise input
		Car Noise @ 20 dB SNR + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 14.25 kbps with 20 dB SNR car noise input + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 15.85 kbps with 20 dB SNR car noise + 2% FER + 2% Packet-Level Signaling input
		Street Noise @ 15 dB SNR	No worse than AMR-WB @ 14.25 kbps with 15 dB SNR street noise input	No worse than AMR-WB @ 15.85 kbps with 15 dB SNR street noise input
		Office Noise @ 20 dB SNR + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 14.25 kbps with 20 dB SNR office noise input + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 15.85 kbps with 20 dB SNR office noise input + 2% FER + 2% Packet-Level Signaling
1	Same as average of TIA/EIA/IS-733 (QCELP 13K) and TIA/EIA/IS-127 (EVRC)	Car noise @ 10 dB SNR	No worse than AMR-WB @ 12.65 kbps with 10 dB SNR car noise input	
		Car Noise @ 20 dB SNR + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 12.65 kbps with 20 dB SNR car noise input + 2% FER + 2% Packet-Level Signaling	

		Street Noise @ 15 dB SNR	No worse than AMR-WB @ 12.65 kbps with 15 dB SNR street noise input	
		Office Noise @ 20 dB SNR + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 12.65 kbps with 20 dB SNR office noise input + 2% FER + 2% Packet-Level Signaling	
2	Same as TIA/EIA/IS-127 (EVRC)	Car noise @ 10 dB SNR	No worse than AMR-WB @ 8.85 kbps with 10 dB SNR car noise input	
		Car Noise @ 20 dB SNR + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 8.85 kbps with 20 dB SNR car noise input + 2% FER + 2% Packet-Level Signaling	
		Street Noise @ 15 dB SNR	No worse than AMR-WB @ 8.85 kbps with 15 dB SNR street noise input	
		Office Noise @ 20 dB SNR + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 8.85 kbps with 20 dB SNR office noise input + 2% FER + 2% Packet-Level Signaling	
Mode i	Slightly higher than that of mode 0	Car noise @ 10 dB SNR	No worse than AMR-WB @ 12.65 kbps with 10 dB SNR car noise input	Optional Mode
		Car Noise @ 20 dB SNR + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 12.65 kbps with 20 dB SNR car noise input + 2% FER + 2% Packet-Level Signaling	

		Street Noise @ 15 dB SNR	No worse than AMR-WB @ 12.65 kbps with 15 dB SNR street noise input	
		Office Noise @ 20 dB SNR + 2% FER + 2% Packet-Level Signaling	No worse than AMR-WB @ 12.65 kbps with 20 dB SNR office noise input + 2% FER + 2% Packet-Level Signaling	

Table 4.4-1: ADR and quality targets for each mode in noisy condition

4.5. Complexity

The complexity of the wideband codec is measured by the number of instructions required to process a frame of speech and by the amount of ROM and RAM required for the codec. Due to implementation considerations, the complexity of the wideband codec is strictly limited.

The complexity of the floating-point encoder and floating-point decoder combined shall not exceed 40 WMOPS.

The RAM and Table ROM requirements of the wideband coder shall be on the order of TIA/EIA/IS-893 (SMV) floating-point standard. The Program ROM for the wideband coder shall be lower than the corresponding figure in SMV standard floating-point C code. More specifically, Data RAM on the order of 10 Kwords, Data ROM (Table ROM) on the order of 12 Kwords, and Program ROM on the order of 18 Kwords are desirable subject to the definition of counting method to be adopted by TSG-C1.1. *(Note: 1 word=16 bits).*

4.6. Robustness

There are a number of requirements imposed upon the wideband coder due to its global application in a wireless telephony environment.

4.6.1. Music on Hold

The wideband coder shall reproduce music with reasonably good quality (recognizable as music) and without any objectionable audible distortion.

4.6.2. Mode Switching

One potential application of the wideband coder might include switching between the modes during the course of a single phone conversation. Therefore, the following is required:

The wideband coder shall be able to transition through modes 0, 1, and 2 without explicit out-of-band signaling into the decoder, and with no perceptible degradation in voice quality that is attributable to the mode switching itself (i.e., seamless mode switching).

In the applications where the interoperable mode, mode I, is utilized (i.e., mobile-to-mobile calls between 3GPP and 3GPP2 wireless systems) the wideband speech coder shall operate only in mode i and mode switching is not allowed.

4.6.3. Robustness to Frames with Corrupted Information

The TIA/EIA/IS-95/CDMA2000 systems notify the speech codec, including the wideband coder, when a known frame error occurs. In fact, the wideband coder must be designed with the reality of a 0-6% frame error rate (FER) in mind. The TIA/EIA/IS-95/CDMA2000 systems occasionally label as good a frame that has been corrupted due to channel errors. In this case the wideband coder receives a frame of random bits that has been marked as good and with a certain frame rate.

All modes shall be robust to frame errors up to 6% FER. The performance of the wideband coder in all modes shall gracefully degrade with increasing frame error rate ranging from 0% to 6%.

The wideband coder shall be further robust to frames with corrupted rate information that have been marked good with no objectionable artifacts produced.

4.6.4. Multiple Languages

The wideband coder will be used in many different countries, and therefore by people speaking many different languages.

The speech quality of the wideband coder shall not have language dependency.

4.6.5. In-Band Signaling Support

Since there is no dedicated signaling channel for voice services in CDMA2000 systems, the traffic channel is occasionally used to transport signaling information. For dim-and-burst signaling, the encoder is not allowed to transmit full-rate packets and the rate usage is constrained to half-rate maximum. For blank-and-burst signaling, the encoder is not allowed to send information packets and a frame erasure (a blank packet type that contains no information bits is transmitted to the receiver which is dealt with similar to a frame erasure condition) is declared at the receiver.

The wideband coder shall be capable of handling dim and burst and blank and burst signaling in all modes of operation.

4.6.6. Forward Link Signaling Inter-working Function

Generating Rate ½ frames in a timely manner is critical for signaling during power control and handoffs. In a packet-core network, the delays in messaging between network elements may exceed system requirements for signaling. In transcoder free operation (TrFO; i.e., configuration of a speech or multimedia call for which transcoders are not present in the communication path) and voice-over-IP (VoIP) calls, where the vocoders are at the end points, external to the network, make the problem even worse. TrFO and VoIP calls are important transport mechanisms for any wideband coder.

The wideband coder shall either supply/support Rate 1 to Rate ½ translation, or treat forward link Dim & Burst signaling as Blank & Burst as needed.

The process of Rate 1 to Rate ½ translation shall be a memoryless transformation and shall have minimal complexity.

4.6.7. Format of the Standard

Most, if not all implementations of the wideband coder will be using fixed-point arithmetic. Therefore, the performance of the fixed point wideband coder must not be degraded from the floating-point version that is developed.

A fixed-point version of the wideband coder shall also be developed that has speech quality performance and robustness equivalent to that of the floating-point version of the wideband coder.

4.6.8. All-Ones 1/8 Rate Frame Handling

The wideband encoder shall not transmit 1/8 rate frames containing all ones. The wideband decoder shall treat all ones 1/8 rate frames as erased frames.

4.6.9. All-Zeros Frame Handling

The wideband encoder shall not generate all zeros frames. The wideband decoder is required to treat all-zeros frames in Rate 1, Rate ½, Rate ¼, and Rate 1/8 as erased frames.

4.6.10. Noise Suppression

While a noise pre-processing module is an optional part of the standard, an example solution shall be included with the standard implementation and for the reference. Each candidate should use noise suppression as an integrated part of the wideband coding algorithm. If noise suppression is used, it shall be used in all conditions, including both links of the tandem condition.

A reference stand-alone noise-suppression will be provided to be included with the reference conditions. Although noise suppression is algorithmically optional its functionality is mandatory.

Normal Procedures with Successful Outcome

This section is not applicable

- **Authorization**
- **De-Authorization**
- **Registration**
- **De-Registration**
- **Activation**

- **De-Activation**

- **Invocation**

- **Normal Operation with Successful Outcome**

- **Call Detail Record**

Exception Procedures or Unsuccessful Outcome

This section is not applicable

- **Registration**

- **De-Registration**

- **Activation**

- **De-Activation**

- **Invocation**

- **Exceptions While Roaming**

- **Exceptions During Intersystem Handoff**

Alternative Procedures

This section is not applicable

Interactions with Other Wireless Services

This section is not applicable