STANDARDIZATION OF THE SELECTABLE MODE VOCODER

S. Craig Greer[1], *Andrew DeJaco*[2]

[1]Chair, 3GPP2 TSG-C1.1 Voice Services, Nokia
[2]Qualcomm

ABSTRACT

The migration from the first generation of cellular telephony to the second has included a transition from an analog speech channel to a digital channel that employs digital speech codecs. As the deployment of these second-generation systems matures, system capacity concerns have increased the pressure for a more efficient encoding of speech. In addition, market pressures have contributed to the contradictory requirement of improved voice quality provided by these wireless systems. This paper presents the motivation for, and the execution of a program of standardizing a new variable-rate speech codec for the cdmaOne/cdma2000 wireless system. Several codecs have previously been standardized. This new codec, known as the selectable mode vocoder or SMV, offers a significant improvement in voice quality over that of existing codec standards as well as the flexibility of allowing the system operator to make tradeoffs between voice quality and system capacity.

1. INTRODUCTION

The primary service provided by second generation wireless telephony systems is speech communications. Service providers of wireless telephony have always been faced with the trade-off of providing high quality communications with that of maximizing the capacity of the air interface carrying the communication. The design of the speech codec employed by these second-generation systems is at the heart of this tradeoff, which manifests itself as a compromise between allowable bit rate and intrinsic voice quality. As the state of the art in speech coding advances, the improvements are realized and new codecs are standardized to provide either improved voice quality over the previous generation, or equivalent voice quality at a reduced bit rate, contributing to an increase in system capacity.

The cdmaOne/cdma2000 system, hereafter known as CDMA, is unique in that it is designed for a variable-rate encoding of speech. Therefore, new CDMA codec designs are necessarily variable rate and must operate within the constraints of the CDMA system. System capacity is directly tied to the average bit rate of these variable rate codecs. The goal therefore is to maximize the voice quality given a fixed average bit rate, or to minimize the average bit rate for a given level of voice quality.

SMV provides a number of operating points, or average bit rate set points, allowing operators to make their own voice quality/system capacity tradeoffs. These tradeoffs can be made on a static basis, such as call set-up, or on a dynamic basis, allowing for a reduction in voice quality to manage localized capacity problems. These tradeoffs can be independently made on the forward and reverse links.

In Section 2 the CDMA system is highlighted from the point of view of a speech codec designer to provide a basis for the constraints placed upon any new codec designed for such a system. A brief review is made of the existing CDMA speech codec standards. In Section 3 the motivation for a new CDMA speech codec is presented along with requirements for this new codec. Section 4 presents the standardization process followed, including the selection from the SMV candidates and the resulting performance of the selected codec. Section 5 discusses the expected capacity gains and deployment scenarios possible with the new CDMA standard.

2. BACKGROUND MATERIAL

2.1. CDMA system

In an effort to maximize system capacity, the CDMA air interface was designed with variable rate encoding of speech in mind. A variable rate codec designed for CDMA is limited to operation within one of two rate sets and consequentially to a set of four rates defined by each of the rate sets. The two rate sets and their associated source encoding rates are as follows:

Rate set 1: {0.8, 2.0, 4.0, 8.55} Kb/sec.
Rate set 2: {1.0, 2.7, 6.2, 13.3} Kb/sec.

In addition to the limitation on the number of rates, the CDMA system is designed with a 20 ms frame size, limiting the frame size choices available to the codec designer. Necessarily, codec rate changes are limited to once per frame.

Another unique aspect of the CDMA system is its separation of the source coding from the channel coding. The channel coding for the air interface is pre-designed and not considered part of the design of a new speech codec.

As the channel coding used to protect the encoded speech is part of the CDMA system, it is the CDMA system that informs the speech codec when a frame erasure takes place. Therefore, a CDMA speech codec will, with high probability either receive an entire frame of data correctly, or receive notification that an entire frame of data is bad.

Another aspect of the CDMA system that impacts the speech codec is the use of the speech channel for transmitting signaling information. This feature is known as "dim and burst" and forces the speech codec to operate at a maximum of half rate while the other half of the channel is used to transmit signaling information. The design of the MOS test conditions used to test the SMV assume a 1% dim and burst rate.

A final aspect of the CDMA system that must be taken into consideration is the lack of CRC protection for the quarter rate and eighth rate of the cdmaOne system. CRC protection for these rates was added to the cdma2000 system. From a speech codec designer's point of view, this represents the largest difference between these two systems. The impact of the lack of CRC protection in cdmaOne is the additional requirement that the quarter rate not be used on the reverse link of cdmaOne systems.

2.2. Existing CDMA speech codecs

The first speech codec designed for CDMA uses rate set 1 and is known by the TIA standard number IS-96C. The codec was standardized in 1993 but did not meet increasing voice quality expectations. Therefore, it not widely deployed in the US.

In an effort to improve voice quality, rate set 2 was created to allow more coding bandwidth. The codec designed for rate set 2, known as IS-733, was fielded in the 1995-1997 timeframe and provides a voice quality that is on par with competing systems in the industry. However, as the average bit rate for IS-733 is on the order of 6.6 Kb/sec, system capacity was reduced approximately 33% from IS-96C. This codec was widely deployed in the US.

A third speech codec development effort was undertaken to maintain the voice quality provided by the IS-733 standard, but to provide it using the more bandwidth efficient rate set 1. This successful development resulted in the 1996 standard known as IS-127 or EVRC, operating with an average bit rate of 4.2 Kb/sec. This codec is now transitioning into CDMA systems.

3. SMV PROGRAM AND REQUIREMENTS

Several years have passed since the completion of a codec standard for CDMA. Systems have been fielded and capacity concerns have increased. The state of the art for speech coding has advanced as well. The time was right for another look at the voice quality/system capacity tradeoff. This new effort, known as the Selectable Mode Vocoder (SMV) began in the TIA TR45.5 standards organization, but soon transitioned into the internationally represented 3GPP2 TSG-C standards body. An organization that represents the CDMA carriers known as the CDMA Development Group (CDG) provided a set of requirements for SMV [1]. This document formed the basis for the requirements and test plan developed for the SMV standardization process [2].

3.1. Compatibility and complexity requirements

Several constraints were placed on SMV due to the fact that its design is part of an existing fielded wireless system. The first requirement was for the codec to operate within the constraints of the existing rate set 1. Also required was a frame length compatible with the 20 ms frame structure of CDMA. Complexity limits were driven by the need to use SMV with existing infrastructure. As a result, SMV complexity limits were limited to those required of the EVRC codec; 800 instructions per 20 ms frame and RAM + ¼ ROM <= 30 Kbytes.

To accommodate the lack of CRC in the cdmaOne system, it was decided to restrict the SMV, as IS-127, from using the quarter rate, but only in mode 0. The use of quarter rate is essential for maintaining voice quality while reducing the average bit rate below that of IS-127.

3.2. Multi-mode operation

SMV is required to operate at a number of average bit rate setpoints, allowing for each service provider to perform their own voice quality/system capacity tradeoff. As a result, the SMV design includes multiple modes; each associated with an average bit rate limitation. Further, SMV is required to seamlessly transition from mode to mode as often as once per frame without producing artifacts. This and the tight requirements on complexity forced each SMV codec candidate designer to use the same set of codecs for each of the modes of operation.

3.3. Voice quality and average bit rate

Further requirements on SMV were, for each mode, to maximize the voice quality subject to the constraint of a limitation on the average bit rate associated with each mode. Voice quality goals were set in terms of existing CDMA reference codecs. The limitation in the average bit rate was set as a percentage of the IS-127 average bit rate and separate limits were defined for clean and noisy input speech. The voice quality goal and average bit rate (ABR) limitation as a percentage of IS-127 for the three original modes tested are as shown in Table 1.

Mode	Voice Quality Goal	Clean Speech ABR (% IS-127)	Noisy Speech ABR (% IS-127)
0	Max (IS-733, IS-127)	100.0%	100.0%
1	Min (IS-733, IS-127)	71.0%	71.0%
2	G.723.1 @ 6.3 Kb/sec (IS-96C in frame erasure conditions)	56.0%	60.0%

Table 1: SMV Mode Definitions and Quality Goals

Each mode was tested in 12 conditions. For each mode, the voice quality goals are the same for all conditions tested. Since the voice quality performance of IS-127 and IS-733 were deemed to be roughly equivalent, the performance of the better of the two in a particular condition was chosen as the goal for mode 0. The performance of the worse of the two in a particular condition was chosen as the goal for mode 1. A codec that performs as well as or better than the better of the two references in all conditions will be a significant improvement over either.

3.4. Delay requirements

The part of the end-to-end delay under the control of the codec designer is the algorithmic delay of the speech codec itself. This delay limit was set to the same original requirement for IS-127 of 40 ms or less.

3.5. Robustness requirements

As with any speech codec used in a wireless environment, SMV is required to operate in the presence of frame erasures. Since the

CDMA system notifies the speech decoder of frame erasures, an algorithm handling those erasures must be designed as part of the decoder. Noisy speech is another fact of operation in a wireless environment, therefore use of a noise suppressor with SMV was highly recommended and if used, required for all conditions tested. A reasonable performance with music was required for music-on-hold situations. Another requirement is equivalent performance with varying languages, as SMV will be deployed internationally. Maintained performance with high and low level speech and with a tandem of SMV with itself was also required.

4. SMV SELECTION AND PERFORMANCE

A total of eight companies developed codecs for submission as a candidate for the selectable mode vocoder. These companies are Conexant, DVSI, Ericsson, Lucent, Motorola, Nokia, Nortel and Qualcomm. Each candidate was required to abide by the terms agreed upon for participation in the program and to provide a codec that met all of the requirements highlighted in Section 3 [2].

Each candidate was required to submit a document demonstrating compliance with all but the voice quality and average bit rate requirements. Of the candidates showing compliance, the candidate codec with the best voice quality based upon mean opinion score (MOS) testing was declared the winner. The laboratory performing the MOS testing was also responsible for verifying compliance with the average bit rate requirements. As the results in the extensive MOS testing could easily have resulted in a difficult decision to make, detailed selection criteria was prepared in advance.

The MOS test consisted of 9 experiments, 3 for each mode of operation. Each mode was tested under clean speech conditions, with frame erasures and with background noise added to the input speech. All input speech was modified IRS filtered with the exception of experiments one and three for mode 0. Input speech for these experiments was flat weighted. Table 2 summarizes the MOS test conditions used in this selection test.

Exp.	Cond. #	Condition
1	1	Clear channel
1	2/3	Low/High input level, (-32/-12 dB)
1	4	Clear channel tandem
2	5	Clear channel
2	6/7/8	1%/2%/3% FER plus 1% signaling
3	9	Car noise @ 15 dB SNR
3	10	Street noise @ 15 dB SNR at 1% FER
3	11	Car noise @ 20 dB SNR and 1% FER
3	12	Office babble noise @ 20 dB SNR

Table 2: SMV Selection Test Conditions

For each condition, the voice quality performance of each of the eight SMV candidates was measured along with the appropriate reference codecs. Table 3 displays the reference codecs used in these tests. In Table 3, PCM(a) is the output of the G.711 codec with an input of speech weighted by car noise to 30.5 dB. PCM(b) is the output of G.711 with an input of speech weighted by car noise to 20 dB and then passed through the IS-127 noise suppressor.

Mode 0 Exp. 1	**Mode 1 Expt. 1**	**Mode 2 Expt. 1**
IS-733, IS-127, Source, G.711	IS-733, IS-127, Source, G.711	IS-127, G.723@6.3Kb/sec Source, G.711
Mode 0 Exp. 2	**Mode 1 Expt. 2**	**Mode 2 Expt. 2**
IS-733, IS-127, Source, G.711	IS-733, IS-127, Source, G.711	IS-127, IS-96C, Source, G.711
Mode 0 Exp. 3	**Mode 1 Expt. 3**	**Mode 2 Expt. 3**
IS-733, IS-127, PCM(a), PCM(b)	IS-733, IS-127, PCM(a), PCM(b)	IS-127, G.723 @ 6.3Kb/sec, PCM(a), PCM(b)

Table 3: SMV Selection Test Reference Codecs

For each candidate and for each mode, a metric was created from an average of the MOS scores of each of the 12 conditions tested. The selection criteria stipulated that if a single codec candidate scored the highest in each of the three modes, that candidate would be selected as the SMV codec. The additional selection criteria that were created would handle any scenario where no single codec received the highest score in all three modes.

The results of the SMV selection phase MOS test [3] made for an easy selection however, as the candidate codec from Conexant received the highest score in each of the three modes of operation. Averaged across the experiments, the Conexant codec also had the best scores in each of the three experiments. The new SMV candidate met or exceeded the voice quality goals in every condition. Using Table 2 as a reference, Table 4 displays in bold the conditions that the new SMV codec exceeded the voice quality goals defined by Table 1.

Meets **Exceeds**		**Mode 0 Ref. Max(IS-733, IS-127)**	**Mode 1 Ref. Min(IS-733, IS-127)**	**Mode 2 Ref.**
Exp.	**Cond**			
1	1	IS-127	**IS-733**	G.723.1
1	2	**IS-127**	**IS-733**	G.723.1
1	3	IS-733	IS-127	G.723.1
1	4	**IS-733**	IS-733	G.723.1
2	5	IS-127	**IS-733**	**IS-96C**
2	6	IS-127	IS-733	**IS-96C**
2	7	IS-127	IS-733	**IS-96C**
2	8	IS-127	IS-733	**IS-96C**
3	9	**IS-733**	**IS-127**	**G.723.1**
3	10	IS-733	IS-127	**G.723.1**
3	11	IS-733	**IS-127**	**G.723.1**
3	12	IS-733	**IS-127**	**G.723.1**

Table 4: SMV Performance across Conditions with Goals

5. CAPACITY GAINS AND DEPLOYMENT ASPECTS

5.1. Capacity Gains Over IS-127 Due to SMV

The reduced-rate operating modes of the SMV codec have a dramatic effect on system capacity in comparison to IS-127. For a CDMA system the average transmit power is directly proportional to the average bit rate generated by the speech codec. Therefore, statistics on the percentage of time a CDMA codec spends in each rate can be used to estimate the reduction in transmit power due to a more efficient encoding of speech. This reduction in average transmit power directly translates to improved system capacity. Table 5 below provides, for each mode and for the forward and reverse links, the percentage of time the SMV codec spends in each of the four rates. A new mode 3 that was not tested in the selection phase MOS test is included. Note that the active speech ADR is the average encoding rate for non-1/8 rate frames. This measure is not influenced by the voice activity factor of the speech file and thus describes the efficiency of the SMV codec operating modes for the source coding of speech.

	Mode	% Full Rate	%1/2 Rate	%1/4 Rate	%1/8 Rate	Active Speech ADR (Kbps)
Rev. Link	0	37.00	4.50	0.00	58.5	8.01
	1	20.33	9.39	9.78	60.48	5.82
	2	8.42	21.19	9.17	61.22	4.50
	3	3.65	26.04	9.09	61.22	3.95
Fwd. Link	0	42.50	5.50	0.00	52.00	7.98
	1	23.52	10.87	11.32	54.30	5.82
	2	9.74	24.51	10.60	55.14	4.50
	3	4.22	30.11	10.51	55.14	3.95

Table 5: SMV Use of the Encoding Rates by Mode and Link

The first two columns of Table 6 show the increase in cdma2000 air interface capacity over IS-127 due to the SMV codec for both the forward and reverse links [4]. This increase is due to the reduction in average transmit power in a particular sector for a particular SMV mode.

	cdma2000 Forward Link	cdma2000 Reverse Link	Erlang-B System Capacity
Mode 0	0%	0%	0%
Mode 1	27%	16%	34%
Mode 2	49%	29%	61%
Mode 3	60%	35%	75%

Table 6: Capacity Gains Due to SMV

Note that in Table 6 the capacity gains are greater on the forward link than on the reverse link. Since the cdma2000 system is significantly forward link-limited, [5] these forward link gains are realized as total system gains. Using the forward link increases and calculating the trunking efficiencies gains from this increase in server capacity, the Erlang-B system capacity is calculated in column 3 of Table 6 assuming a 2% call blocking probability. As can be seen from Table 6, the 3 reduced-rate operating points of the SMV codec offer significant capacity gains of 34%, 61% and 75% respectively.

5.2. System Deployment Aspects

Mode 0 of SMV is useful for the system operator who wants improved voice quality over IS-127 and is willing to use a codec with the same capacity requirements as IS-127. Mode 1 is designed for the system operator who wants to maintain the quality provided by EVRC in addition to realizing a capacity benefit. Mode 2 is for the system operator who is willing to sacrifice some voice quality robustness in order to realize the significant capacity gain provided by mode 2 of SMV. Similarly, mode 3 of SMV provides even more capacity gains.

We have seen that there is a limitation on the use of SMV to mode 0 for the reverse link of cdmaOne systems. Furthermore, the system is forward link-limited such that mode 0 can be run on the reverse link while mode 2 is run on the forward link and a 61% overall system capacity gain can be achieved [4]. It has also been demonstrated that the effects of higher voice quality are more likely to be realized in the quieter listening environment of the reverse link [6]. Therefore, a realistic deployment scenario is to deploy mode 0 of SMV on the reverse link and to use a reduced-rate mode of SMV on the forward link.

6. CONCLUSIONS

A new speech codec for cdmaOne/cdma2000 wireless systems has been developed and is currently undergoing standardization in the 3GPP2 standards organization. This codec is a realization of advances made in the field of speech coding and is used to provide improved speech quality and coding efficiency as well as the ability for the system operator to make tradeoffs between the two design goals.

SMV also takes advantage of the improved air interface of cdma2000 and makes extensive use of the quarter rate to provide IS-127-like voice quality at reduced average bit rates. The selected candidate for SMV meets or exceeds all speech quality goals in 36 separate conditions tested.

The air interface capacity gains over IS-127 that are attributable to the various modes of SMV have been reviewed and a deployment scenario that complements the characteristics of the CDMA system has been presented.

At this time, a collaboration phase is underway to further improve various characteristics of the SMV.

7. REFERENCES

[1] CDMA Development Group, "Carrier Requirements for SMV", *3GPP2-C11-19990615-008*, Jun. 1999.

[2] S. C. Greer, Chair - 3GPP2 TSG-C1.1, "SMV Test Plan, Version 11.1", *3GPP2-C11-20000425-003*, Apr. 2000.

[3] F. Corcoran, "SMV Selection Test – Final Host and Listening Lab Report", *3GPP2-C11-20000821-003R1*, Aug. 2000.

[4] A. DeJaco, "SMV Capacity Increases", *3GPP2-C11-20001016-010R1*, Oct. 2000.

[5] Lucent, Motorola, Nokia, Nortel, Qualcomm, Samsung, "CDG Evolution Study Report", *CDG Evolution Team Document,* Dec. 1999.

[6] L. Thorpe & R. Rabipour, "Changes in Voice Quality Judgements as a Function of Background Noise Level in the Listening Environment", *IEEE Workshop on Speech Coding*, Sep. 2000.