AN ADAPTIVE MULTI-RATE SPEECH CODER FOR DIGITAL CELLULAR TELEPHONY

Erdal Paksoy, Juan Carlos De Martin†, Alan McCree, Christian G. Gerlach, Anand Anandakumar, Wai-Ming Lai and Vishu Viswanathan

DSP Solutions R&D Center, Texas Instruments, Dallas, TX

ABSTRACT

We have developed an adaptive multi-rate (AMR) speech coder designed to operate under the GSM digital cellular full rate (22.8 kb/s) and half rate (11.4 kb/s) channels and to maintain high quality in the presence of highly varying background noise and channel conditions. Within each total rate, several codec modes with different source/channel bit rate allocations are used. The speech coders in each codec mode are based on the CELP algorithm operating at rates ranging from 11.85 kb/s down to 5.15 kb/s, where the lowest rate coder is a source controlled multi-modal speech coder. The decoders monitor channel quality at both ends of the wireless link using the soft values for the received bits and assist the base station in selecting the codec mode that is appropriate for a given channel condition. The coder was submitted to the GSM AMR standardization competition and met the qualification requirements in an independent formal MOS test.

1. INTRODUCTION

In digital cellular communication systems, one of the major challenges is that of designing a coder that is able to provide high quality speech throughout a wide variety of channel conditions. Ideally, a good solution must provide the highest possible quality in the clean channel conditions while maintaining good quality in very heavily disturbed channels. Traditionally, digital cellular applications use a single coding mode where a fixed source/channel bit allocation provides a compromise solution between clean and degraded channel performance. Clearly, a solution which is well suited for clean channels would use most of the available bits for source coding with only minimal error protection, while a solution designed for poor channels would use a lower rate speech coder protected with a large amount of forward error correction (FEC).

One way to obtain good performance across a wide range of conditions is to allow the network to monitor the state of the communication channel and direct the coders to adjust the allocation of bits between source and channel coding accordingly. This can be implemented via an adaptation algorithm whereby the network selects one of a number of available speech coders, called **codec modes**, each with a predetermined source/channel bit allocation. This concept is called **adaptive multi-rate (AMR) coding** and is a form of network-controlled multimodal coding of speech [1].

The AMR concept is the centerpiece of ETSI's GSM AMR standardization activity, which aims to define a new European cellular communication system designed to support an AMR mechanism in both the half rate and full rate channels. This paper describes an AMR coder we have developed and submitted to the qualification phase of the GSM AMR competition. The source coder is based on Code-Excited Linear Prediction (CELP), and the channel coder is based on punctured convolutional codes. The coder also includes a novel method for monitoring channel conditions and communicating the channel measurements and codec mode commands between the base station and the mobile station.

2. SYSTEM OVERVIEW

The coder is designed to operate in both the GSM full-rate channel mode at a total bit-rate of 22.8 kb/s and the GSM half-rate channel mode at 11.4 kb/s. In each channel mode, the coder supports two codec modes. The available bits are allocated differently among source coding, channel coding and signaling in each codec mode. Table 1 illustrates the allocation of bits in the four codec modes.

Codec Mode	Source (kb/s)	Channel+ signaling (kb/s)	Total (kb/s)
Half Rate Mode 0	7.45	3.95	11.4
Half Rate Mode 1	5.15	6.25	11.4
Full Rate Mode 0	11.85	10.95	22.8
Full Rate Mode 1	7.45	15.35	22.8

Table 1: Rate allocation for codec modes

3. SPEECH CODING

Each codec mode uses a CELP coder based on the one we developed for the GSM enhanced full rate standardization activity in 1995 [2], and many features are the same across codec modes. The frame size for all source coders is 20 ms with lookahead of 5 ms for LPC analysis. The LPC parameters are coded once per frame in the Line Spectral Frequency (LSF) domain using a 4-stage, 26-bit multi-stage vector quantizer (MSVQ) which is searched using an M-Best search algorithm. A perceptual weighting function is used to reflect the importance of the Bark scale for

---

† Juan Carlos De Martin is currently with CENS-CNR at the Polytechnic of Turin, Italy. E-mail: demartin@polito.it.

LSF quantization [3]. The remaining parameters are updated once per subframe. All source coders have four 5 ms subframes, with the exception of Half Rate Mode 1 where there are two subframes of 10 ms each. In all codec modes, the pitch lag is coded using a delta-search, adaptive codebook search algorithm where the first pitch lag in each frame is coded using 8 bits and the remaining lags are coded differentially with respect to the previous lag with 5 bits each. The fixed excitation is obtained from a sparse ternary codebook searched using an M-best algorithm, and the pulse locations and signs are encoded and transmitted. The fixed and adaptive excitation gains are jointly vector quantized with a 7-bit codebook, where the fixed excitation gain component is coded differentially with respect to a predicted gain estimated from previous gain values.

Full Rate Mode 1 and Half Rate Mode 0 use an identical CELP coder at a bit rate of 7.45 kb/s. Full Rate Mode 0 is a similar coder, with a higher rate used for fixed excitation coding resulting in a rate of 11.85 kb/s.

Half Rate Mode 1 operates at a bit rate of 5.15 kb/s. It uses a source-controlled multimodal CELP coder where each input speech frame is classified into one of two source coding modes based on a voiced/unvoiced decision. The voiced mode is coded in the same way as in the other codec modes. In the unvoiced mode, no adaptive codebook is used since unvoiced signals do not contain a periodic component. The fixed excitation is encoded with a stochastic codebook, using gain-matched analysis-by-synthesis [4]. The fixed excitation gain is coded using the same codebook as the one used in the voiced mode. The mode information is not transmitted explicitly, but is signaled using a reserved value of the pitch lag of the first subframe in each frame. Since unvoiced frames in Half Rate Mode 1 require fewer source coding bits than voiced frames, the excess bits are reserved for future use. Table 2 illustrates the source bit allocation in all four codec modes.

Parameter	Half Rate 0	Half Rate 1		Full Rate 0	Full Rate 1
		voiced	unvoiced		
LPC	26	26	26	26	26
Pitch Lags	23	13	8	23	23
Fixed Excitation	72	50	24	160	72
Gains	28	14	14	28	28
Total bits/frame	149	103	72	237	149
Rate (bits/s)	7450	5150	3600	11850	7450

Table 2: Bit Allocations

4. CHANNEL CODING AND ERROR CONCEALMENT

The channel coding for each codec mode uses rate-compatible punctured convolutional codes, as well as a CRC protecting the most important bits in each frame. In each codec mode, the source bits are divided into two or three classes, numbered 0, 1, and 2 in order of decreasing perceptual importance. Bits in class 0 include the first two stages of the LSF MSVQ, the most significant bits of the pitch lags and the codebook gains, as well as the in-band signaling of the codec mode command and channel measurement described in Section 5. The class 0 bits are encoded with the highest bit rate (punctured) convolutional code and are also protected by the 7-bit CRC, which acts as a parity check. When the CRC signals a bad frame, all of the previous frame's parameters are repeated and muted, with the exception of the fixed excitation indices which are still decoded from the bit stream. The class 1 and 2 bits are coded with (punctured) convolutional codes with lower bit rates.

5. SIGNALING AND LINK ADAPTATION

An overview of the AMR coding system, including both mobile station and base station, is shown in Figure 1. In general, adaptation depends on the current state of the communication channel. Since channel estimation is done at the decoder, the receiver needs to signal to the encoder through the reverse link some information needed for mode selection. The rate control mechanism varies depending on the direction of transmission, due to a constraint that the codec mode control mechanism must be located in the base station.

5.1. Channel Analysis and Mode Selection

The adaptation algorithm is based on the channel measurement which is an estimate of the carrier to interference ratio (C/I). This estimate is based on the soft-values for the received bits as provided by the demodulator/equalizer. These values are good indicators of the reliability of the bits. We have found that a moving average of the absolute values of the soft bits is a good estimator of the current C/I of the channel. Codec mode decisions are made by comparing this moving average value to a predetermined threshold, and by using additional hysteresis rules designed to ensure smoother codec mode transitions. Because of their different characteristics, the full rate and half rate channels require the various parameters of the adaptation mechanism to be tuned separately.

5.2. In-Band Signaling

The signaling of all information needed for codec mode adaptation is done in-band, using some of the bits normally available for source and channel coding. Adaptation requires the transmission of two different kinds of information: a codec mode command sent from base station to mobile via the downlink channel, and channel measurement information sent from mobile to base station via the uplink channel.

For uplink transmission, the base station monitors the channel condition and decides which mode the mobile station should use. The base station communicates this information in the form of a **codec mode command**, transmitted in the downlink. Upon reception, the mobile station encoder switches to the indicated mode.

The objective of our codec mode transmission scheme is to send the mode information accurately and frequently enough to make the adaptation mechanism work effectively, but using as few bits as possible to minimize overhead. We have chosen to send the codec mode command by means of a variable-length code, using one information bit per frame. This variable length code is shown in Table 3. This table applies to full rate and half rate modes sepa-

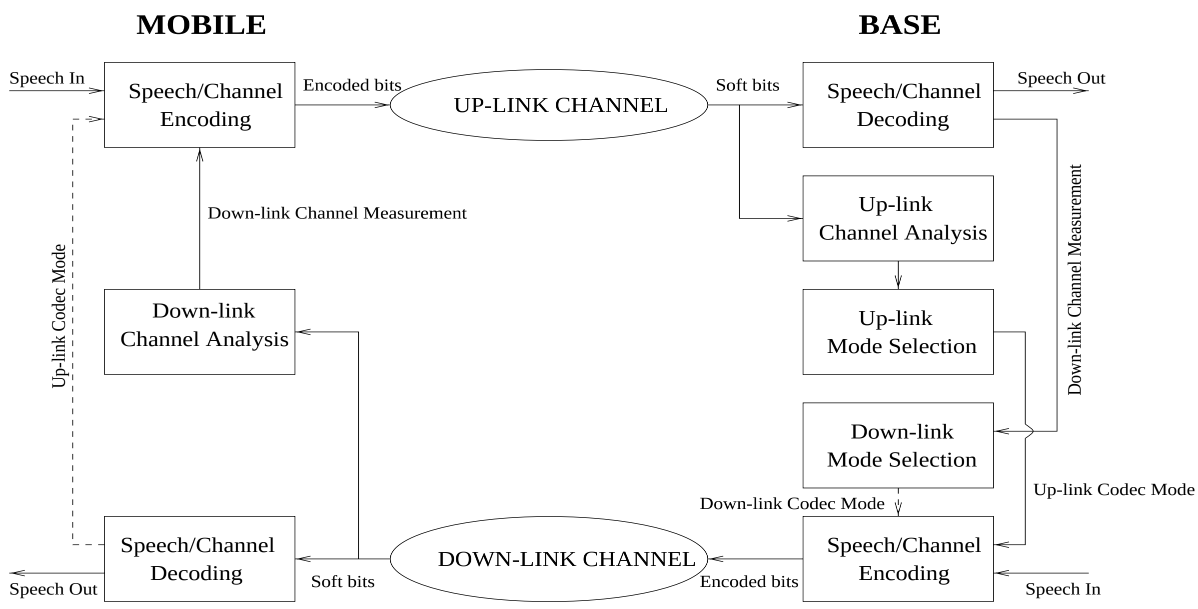

Figure 1. Overview of AMR Coding Scheme

rately. Since this signaling bit is important for reliable operation, it is included in the class 0 bits of the channel coding. Notice that in addition to the two AMR modes, the codec mode command can also signal switching to any number of **extended modes**, which include the existing GSM standards, as well as future options such as wideband coding.

AMR 0	0
AMR 1	10
AMR Wideband 0	110
AMR Wideband 1	1110
GSM FR	11110
GSM EFR	111110
GSM HR	1111110

Table 3: Variable Length Codec Mode Command

For downlink transmission, based on the received bits and possibly other information that may be available, the mobile station computes a downlink **channel measurement** which is representative of the state of the channel. The mobile station cannot autonomously decide which mode to use. Hence, this measurement is quantized and transmitted back on the uplink to the base station. This is done in-band using one-bit delta modulation. The base station then decides which codec mode it will use for the downlink transmission of the next frame.

5.3. Mode Information

One problem in designing an AMR coder is that the channel decoder must know which mode has been used to encode a given frame before it can successfully decode it. We have chosen to solve this problem by explicit transmission of the **codec mode index** as a header to the channel bitstream for each frame. There are only two codec modes in each channel mode, so only one bit is required to transmit this index. Since this bit is not protected by channel coding, a 3-bit repetition code is used to provide robustness to bit errors.

To handle the extended modes shown in Table 3, a “codec mode beacon” is also sent with each frame, both up- and downlink. This beacon uses a variable length code to signal the mode used to code the current frame, including extended modes. The beacon is also sent using one channel bit per frame, and the variable-length code is the same as the one used to code the codec mode command. Since this bit goes into the channel unprotected, the decoder must wait for multiple frames of new beacon mode information before switching to a different extended codec mode.

6. LISTENING TEST RESULTS

The coder was extensively tested in accordance with the GSM AMR qualification test plan by an independent laboratory. Both full rate and half rate coders were tested in four experiments. All tests were done using the Mean Opinion Score (MOS) rating scale, except for the background noise tests which were scored on the degradation MOS (DMOS) rating scale. In these experiments, a distinction was made between static and dynamic error conditions. In static tests, for each condition in the test, the C/I ratio of the channel was held constant. Here, each codec mode of the AMR candidate was tested and the score for the AMR coder in a given condition was taken to be the score of the best codec mode in that condition. In dynamic tests, realistic yet challenging communication scenarios were simulated, resulting in error conditions where the C/I ratio varies drastically during a 50-second time interval.

A subset of the static test results are summarized in Tables

4 and 5. All the scores in the tables are for flat, clean speech, except for the tandem condition where the input material was IRS filtered. In the full rate channel, it can be seen that the AMR candidate is essentially equivalent to the GSM Enhanced Full Rate (EFR) in the clean channel and in tandem, but that it easily outperforms it in degraded channels, thanks to the large amount of channel protection available in Full Rate Mode 1. In fact, the AMR candidate is equivalent to the 16 kb/s ITU G.728 standard for both C/I=10 dB and C/I=7 dB.

Condition	AMR FR	EFR	G.728
No Errors	4.21	4.42	4.06
C/I = 10 dB	4.19	3.79	-
C/I = 7 dB	3.94	3.35	-
C/I = 4 dB	3.48	1.81	-
Tandem	4.00	3.98	-

Table 4: Clean Speech in Full Rate Channel

In the half rate channel, our AMR coder provides high quality for clean speech as demonstrated by the fact that it is statistically equivalent to G.728 for a single encoding and to G.729 in tandem. In all three error conditions listed below, our AMR coder still provides adequate performance as it is at least statistically equivalent to the GSM Full Rate coder.

Condition	AMR HR	GSM FR	G.728	G.729
No Errors	4.08	-	4.27	-
C/I = 10 dB	3.60	3.52	-	-
C/I = 7 dB	2.90	3.10	-	-
C/I = 4 dB	2.13	1.75	-	-
Tandem	3.38	3.35	-	3.56

Table 5: Clean Speech in Half Rate Channel

Static tests were also performed for two different types of acoustic background noise, namely street and car noise, with flat source material. These tests showed similar performance improvement for the AMR candidate as compared to the non-adaptive reference coders.

The dynamic test results are tabulated below for the full rate and half rate channels. Each row corresponds to one of five simulated channel scenarios. It can be seen that in the Full Rate channel the AMR candidate significantly outperforms EFR for the same channel, sometimes by as much as 1.4 on the MOS scale. This clearly shows the advantage of dynamic adaptation in changing channel conditions. The half rate AMR candidate, on the other hand, is essentially equivalent to GSM FR at half the bit rate.

Condition	AMR FR	EFR	ΔMOS
Dynamic EP 1	4.29	3.67	+0.62
Dynamic EP 2	4.21	3.73	+0.48
Dynamic EP 3	3.86	3.01	+0.85
Dynamic EP 4	4.25	3.61	+0.64
Dynamic EP 5	4.16	2.75	+1.41

Table 6: Dynamic Conditions in Full Rate Channel

Condition	AMR HR	GSM FR	ΔMOS
Dynamic EP 1	3.63	3.56	+0.07
Dynamic EP 2	3.55	3.59	-0.04
Dynamic EP 3	2.92	2.68	+0.24
Dynamic EP 4	3.41	3.56	-0.15
Dynamic EP 5	2.94	2.78	+0.16

Table 7: Dynamic Conditions in Half Rate Channel

7. CONCLUSIONS

We have developed a complete AMR solution for both full-rate and half-rate GSM channels. Extensive formal testing has shown that this coder is clearly superior to non-adaptive reference coders for realistic channel conditions and meets the GSM AMR qualification requirements.

8. REFERENCES

[1] A. Gersho and E. Paksoy, “Variable Rate Speech Coding”, in *Proceedings of the Seventh European Signal Processing Conference*, 1994, Edinburgh.

[2] W. LeBlanc, C. Liu, V. Viswanathan, “An Enhanced Full Rate Speech Coder for Digital Cellular Applications”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Volume 1, pp. 569-572, 1996, Atlanta.

[3] A. McCree and J. C. De Martin, “A 1.7 kb/s MELP Coder with Improved Analysis and Quantization”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Volume 2, pp. 593-596, 1998, Seattle.

[4] E. Paksoy, A. McCree and V. Viswanathan, “A Variable-Rate Multimodal Speech Coder With Gain-Matched Analysis-by-Synthesis”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 751-754, 1997, Munich.