An Adaptive Multirate Speech Codec Proposed for the GSM

Wen Xu[1], Stefan Heinen[2], Marc Adrat[2], Peter Vary[2], Thomas Hindelang[3], Maximilian Schmautz[3], Joachim Hagenauer[3]

[1]Dept. of Mobile Phone Development, Siemens AG, Hofmannstr. 51, 81359 Munich

[2]Inst. of Comm. Systems and Data Processing (IND), RWTH Aachen, 52056 Aachen

[3]Inst. for Communications Eng. (LNT), TU München, Arcisstr. 21, 80290 Munich

***Abstract* - An adaptive multirate (AMR) speech codec proposed for the GSM is described. By dynamically splitting the gross bit rates between source (speech) and channel coding according to the channel quality, almost wireline speech quality even for poor channel conditions, and higher speech quality for relatively good conditions can be achieved. The CELP based speech codec and the RSC based channel codec were designed as homogenous/hierarchical as possible for a robust transmission. Novel algorithms for mode bit coding and error concealment were used. One of our design philosophy employed was that the source and channel codecs are jointly instead of individually optimized.**

1. INTRODUCTION

Presently, the so-called adaptive multirate (AMR) (narrowband) speech codec is standardized for GSM (Global System for Mobile communications) within ETSI (European Telecommunications Standards Institute). With the AMR concept, the ratio between source bit rate and error protecting redundancy is adaptively chosen according to the channel conditions. When the channel is bad, the source encoder operates at low bit rates with lower speech quality, allowing more bits to be used for a powerful forward error correction (FEC). The higher rates of speech encoder are used for good channels, since in this case weak error protection is sufficient.

In this paper, we present a GSM AMR codec including algorithms for speech/channel coding, channel quality estimation and rate adaptation, as well as the subjective test results. In section 2, we give a brief description. The variable rate code excited linear prediction (VR-CELP) speech coding algorithm is described in section 3 and the corresponding error concealment algorithm in section 4. The FEC using recursive systematic code (RSC) as well as inband transmission of the mode bits is dealt with in section 6. Channel estimation for the rate adaptation is considered in section 7. In section 8, we present the subjective results, and in section 9, we give a short conclusion.

2. OVERVIEW OF THE PROPOSED GSM AMR

Contrary to a fixed single rate codec like the existing GSM FR (fullrate), HR (halfrate), EFR (enhanced fullrate) codecs, an AMR codec consists of a family of modes (rates). Each mode actually corresponds to a fixed rate speech and channel codec. To achieve the best possible performance (in terms of speech quality and channel capacity), not only the fixed rate codec corresponding to each single mode itself should be optimized, but the mode switching according to the actual channel quality must work sufficiently well. Here the issues like inband transmission of the mode bits, rate adaptation and channel quality estimation play a crucial role.

In order to track the varying channel conditions four speech codec modes, 13.3, 9.5, 8.1 6.1 (6.3) kbit/s were designed, where the modes 8.1 and 6.1 (6.3) kbit/s are used for both FR and HR channel, and 13.3 and 9.5 kbit/s only for FR channel.

As required by the ETSI AMR standardization group, the switching between FR or HR channel will be realized by a GSM intra-cell handover and within FR or HR channel, the signaling of the AMR codec modes must be inband, i.e. the rate indicating mode bits must be transmitted together with speech data within a traffic channel (TCH). In this way, no extra control channel is necessary for the rate and channel quality indication. The signaling shall cover the AMR (narrowband) codec modes, the existing GSM speech codecs (FR, EFR, and HR) and at least one future AMR wideband mode. For this reason we reserved as the mode bits 3 bits in FR channel and 2 bits in HR channel. This will allow to signal 2^3=8 modes in FR channel (e.g. 4 modes for AMR, 2 for the existing GSM FR and EFR, and 2 for the future wideband AMR), and 2^2=4 modes in HR channel (2 for AMR and 1 for the GSM HR and 1 for the future wideband AMR).

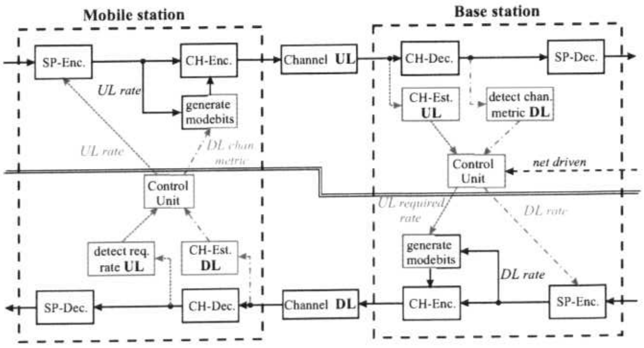

Fig. 1 Block diagram of the AMR codec for the GSM.

The basic concept of the GSM AMR speech transmission is depicted in Fig. 1, where the dashed-dotted lines indicate the downlink signal flow, and the dashed lines the uplink signal flow. Both the base station (BS) and the mobile station (MS) contain speech encoder/decoder (SP-Enc/SP-Dec), channel encoder/decoder (CH-Enc/CH-Dec), channel estimation entity (CH-Est) and control unit for the mode adaptation.

In the studied system, the codec control entity is located in the network, which means an asymmetrical link. The BS is the master and decides about the rates in up- and downlink. The MS will decode the rates which are used in the up- and downlink, and send the estimated channel metric to the BS. A channel quality parameter derived from the soft output of the equalizer is used to control the mode adaptation in up- and down-link.

- *Uplink (UL):* After initialization, the mobile starts transmission with the lowest speech bit rate, ensuring a secure transmission. The mode bits and the DL channel metric (information about the DL quality) are sent to the channel encoder and transmitted inband within the traffic channel. At the receiver (i.e. BS) appropriate channel decoding is first done and then speech will be decoded according to the detected mode bits. In parallel a measurement of the UL channel quality is carried out by the BS channel estimator.

The measured UL channel quality and the detected DL channel quality metric are fed to the BS control unit, which determines the current DL rate (based on DL channel metric analysis) and the requested UL rate.

- *Downlink (DL):* The DL mode bits as well as the requested UL mode bits are transmitted inband within the TCH to the MS. The MS does channel decoding and speech decoding according to the detected DL mode. Similar to the UL, a DL channel quality measurement is done by the estimator of the MS, and the requested UL rate is decoded from the received bit stream.

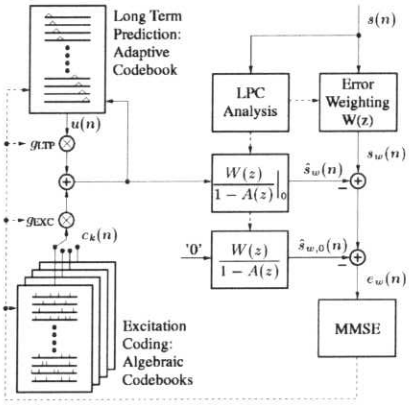

Fig. 2 VR-CELP speech encoder.

3. VARIABLE RATE CELP SPEECH CODING

The proposed variable rate CELP (VR-CELP) coding is based on CELP concept [7] and the encoder operates as follows [4]. The sampled input speech signal $s(n)$ is partitioned into segments of 20 ms (160 speech samples) duration and a linear predictor is computed for each speech segment. The coefficients of this predictor are used to build an LPC (linear predictive coding) synthesis filter $1/(1-A(z))$ describing the spectral envelope of the signal (see Fig. 2).

An analysis-by-synthesis procedure is employed to find the excitation that minimizes the weighted Minimum Mean Square Error (MMSE) between the synthesized and the original signal. The applied weighting filter $W(z)$ is derived from the LPC synthesis filter and takes into account the psychoacoustic effect that quantization noise in the spectral neighborhood of the formants is less perceptible.

For complexity reasons adaptive and fixed codebook are sequentially searched for the best entry, first the adaptive contribution, and then the fixed one. The adaptive codebook consists of time-shifted versions of past excitation sequences, to describe the long-term characteristic (periodicity) of the speech signal.

- *Short-Term Prediction:* Like the existing GSM EFR, a 10th order LPC analysis is performed on the actual speech frame using a split-Levinson approach. This algorithm represents an efficient means to compute the direct prediction coefficients a_i (i=1...10), the reflection coefficients k_i as well as pairs of line spectral frequencies (LSF). In order to avoid sharp transitions between successive coefficient sets we use a 5 ms lookahead for the LPC analysis in combination with an interpolation of the LSF coefficient set for subframes of 40 samples.

- *LSF quantization:* A memory based two-stage split (predictive) vector quantization (VQ) was used to quantize the LSF information. The prediction (with a order of 4) is carried out independently on each component of the LSF vector. To increase the flexibility of the prediction we trained two different coefficient sets, one optimized for stationary speech segments, the other for transient ones. For each LSF vector the optimum of these coefficient sets is selected in the MMSE sense. One extra bit is transmitted to specify the selected predictor.

As a result, only 22 bits are required to quantize the LSF vector. This leads to an average spectral distortions of 1.2 to 1.4 dB compared to the unquantized spectral envelope, a distortion usually imperceptible.

- *Long Term Prediction (LTP):* Here we utilize a 1 tap adaptive codebook. For short pitch periods we introduced a fractional pitch in the range of pitch periods from 20 up to 84 samples by steps of 1/3. The encoding of the complete pitch range requires 8 bits per 5 ms subframe. The LTP gain factor shows a limited dynamic range, therefore a 3-bit Lloy-Max quantizer was employed.

Rate	Length	Tracks	Pulses	Pulse type
13.3	20	7	7	ternary
9.5	40	3	6	ternary
8.1	40	5	5	binary
6.1	40	2	2	binary

Tab. 1 Excitation codebook properties for different modes.

Rate	LPC	LTP	g_{LTP}	g_{EXC}	EXC	Total
13.3	9+2x6+1	4x8	4x3	8x4	8x21	266
9.5				4x4	4x27	190
8.1					4x20	162
6.1					4x10	122

Tab. 2 Overall bit allocation for different modes.

- *Excitation Coding:* Different rates were realized by using exclusively different fixed excitation (sparse algebraic) codebooks, i.e. different algebraic codebook indices (EXC) and gains (g_{EXC}), while leaving the coding schemes for all other speech parameters invariant (see Tab. 1 and 2). This is due to the following reasons:

- The different codec modes share most of the software and tables. The overall codec thus has a program memory and ROM size comparable to a single mode codec.

- Seamless mode switching can be realized by simply changing the excitation codebooks.

- The VR-CELP exhibits a very robust behavior against mode misdetection since a misinterpretation of fixed excitation usually leads only to minor distortions in the reconstructed speech.

- *Post-Processing:* The noise shaping postfilter consists of a long term part, short term part, tilt compensation and an automatic gain control. Two different postfilter adjustments were employed, one for the rates 13.3 and 9.5 kbit/s and another for 8.1 and 6.1 kbit/s. This is a tradeoff between speech naturalness and quantization noise.

4. ERROR CONCEALMENT

The error concealment technique used here is based on estimation theory. It makes use of reliability information about the channel decoded bits (provided by the SOVA channel decoder) as well as *a priori* information about the sent speech codec parameters [1]. The method works in a principle different from those based on the CRC (like the GSM EFR) and can be implemented in a GSM mobile station and even base station [5]. It aims at the individual speech parameters instead of the whole frame. With such an error concealment algorithm, no frame error detection such as one based on the CRC is necessary. The bits usually used for the CRC can be used for error protection as well, which leads to a further channel coding gain.

The parameters considered include LSF, LTP pitch and gain, as well as the fixed codebook gain. Our analysis showed that, especially for these important parameters, certain residual redundancy (like non-uniform distribution) still exist after speech encoding. Here no attempt was made to further remove such redundancy by using a sophisticated compression algorithm. Instead we keep it in the bit stream and utilize it as *a priori* information in the error concealment.

For the various parameters specific statistical models (e.g. zeroth or first order Markov process) and proper estimators are designed. The actually sent parameters are estimated by using the minimum mean square or the maximum *a posteriori* criterion. An automatic muting mechanism is inherently contained in the estimator when subsequent frames are corrupted.

The performance of this error concealment concept can be further enhanced by adding explicit redundancy to the data to be transmitted. This redundancy may be generated e.g. by a linear block code [4]. Notice that such explicit redundancy is not directly used for error correction but as *a priori* information in the error concealment.

We employed this method to the lowest rate 6.1 kbit/s, in which the first LPC index was protected by a shortened (13, 9) Hamming code. Because of the additional 4 bits, the codec rate is increased to be 6.3 kbit/s. However, as shown in our subjective tests, the reconstruction quality under bad channel conditions is dramatically increased although the corresponding FEC was weaker.

5. HIERARCHICAL UEP WITH RSC

After speech coding the generated bits are not equally important for audibility. Therefore, an unequal error protection (UEP) scheme must be used. The most important bits such as the mode bits and the bits of LPC coefficients should be highly protected.

- *Recursive Systematic Code (RSC):* A convolutional code of memory m=4 and rate 1/3 was used for error protection. The generator polynomials are defined as $[1, G_1/G_0, G_2/G_0]$, where $G_0 = 1 + D^3 + D^4$, $G_1 = 1 + D + D^3 + D^4$ and $G_2 = 1 + D^2 + D^4$ are the generator polynomials. Note that we used recursive systematic code (RSC). Compared with the non-systematic code (NSC), the use of RSC leads to a lower bit error rate (BER) in typical mobile radio channels (where BER $> 10^{-4}$). Since the systematic bits in RSC coding are directly transmitted, all information bits are available in the received bit stream before channel decoding. So, the *a priori* information of the information bits can be estimated (e.g. by using the decoded bits) and added outside the Viterbi decoder. That means, a normal Viterbi algorithm (VA) can work as an Apri-VA [3] to exploit *a priori* knowledge to improve the decoding performance.

In general, the RSC decoding can be realized by using the NSC decoder for the equivalent non-systematic polynomials. For the encoder polynomials $[1, G_1/G_0, G_2/G_0] = 1/G_0\ [G_0, G_1, G_2]$, the RSC decoder can then be realized by an NSC decoder for the polynomials $[G_0, G_1, G_2]$, followed by a rate 1 encoder with the feedback polynomial G_0. The advantage of such an implementation is that the old hardware (e.g. the standard Viterbi decoder) for NSC can further be used to decode the RSC. The additional computational complexity for the rate 1 encoder remains small.

- *Hierarchical UEP:* Due to the secure initial and ending state (when termination) of the convolutional decoder, the first and last bits in the data bit block are better protected than those in the middle. Therefore, we place the very important mode bits at the very beginning of the data block and perform a joint encoding/decoding of the mode and speech bits. This ensures a very secure inband transmission of mode bits. In contrast, employing a separate block code to protect the mode bits is less powerful since a block code does not make use of a secure initial state of the decoder.

We used a hierarchical scheme to jointly code the mode and speech bits, where at least the first part, typically $5m$ (in our case 23) bits, are coded using the same coding scheme [6]. As such, the complete trellis of the first part can be built and the mode bits can be decoded without knowing the actual blocksize. With the mode bits determined, the blocksize and the used rates are now known so that the rest of the received bit stream can be decoded properly. In addition, a plausibility check of the decoded mode bits can also be carried out by comparing the decoding metrics in conjunction with the estimated channel quality.

Tab. 3 and 4 show the code rates used for different bit groups, where within a group the bits have about the same sensitivity against bit errors. The code rates 1/5 and 1/4 are realized by repetition of information bits. The rates higher than 1/3 are realized by a puncturing in the rate-compatible way (the rate 1/2 is realized by puncturing all G_2/G_0 bits) [2]. As a result, 456 code bits (i.e. 22.8 kbit/s) suitable for the GSM FR channel are generated. An important advantage of the use of the rate-compatible punctured convolutional (RCPC) codes is that the same encoder and decoder can be used for all RCPC codes of the same memory by controlling the metric memory access. This greatly simplifies its implementation.

Rate (kbit/s)	UEP for Bit# (channel code rate used)
A (13.3)	**0..2 (1/5), 3..6 (1/4), 7..22 (1/3),** 23..44 (1/2), 45..96 (8/14), 97..268 (8/11), *269..272 (1/1)*
B (9.5)	**0..2 (1/5), 3..6 (1/4), 7..22 (1/3),** 23..36 (2/5), 37..78 (1/2), 79..123 (9/16), 124..164 (1/2), 165..180 (2/5), 181..196 (1/3), *197..200 (1/2)*
C (8.1)	**0..2 (1/5), 3..6 (1/4), 7..22 (1/3),** 23..24 (1/3), 25..48 (2/5), 49..109 (1/2), 110..133 (2/5), 134..154 (1/3), 155..170 (3/7), *171..174 (2/5)*
D (6.3)	**0..2 (1/5), 3..6 (1/4), 7..22 (1/3),** 23..38 (1/3), 39..62 (2/5), 63..93 (1/3), 94..134 (1/4), *135..138 (1/3)*

Tab. 3 The AMR UEP in FR channel.

Rate (kbit/s)	UEP for Bit# (channel code rate used)
A (8.1)	**0..1 (1/3), 2..5 (2/5), 6..22 (1/2),** 23..42 (10/16), 43..97 (11/16), 98..163 (1/1)
B (6.3)	**0..1 (1/3), 2..5 (2/5), 6..22 (1/2),** 23..52 (10/16), 43..107 (1/2), 108..127 (1/1)

Tab. 4 The AMR UEP in HR channel.

In the FR channel, the mode bits are additionally protected, namely in rate B (9.5 kbit/s) by a 4-bit CRC, in rate C (8.1 kbit/s) by two 3-bit CRC's and in rate D (6.3 kbit/s) by two 3-bit CRC's, respectively. This leads to data of 266 bits (rate A), 194 bits (rate B), 168 bits (rate C) and 132 bits (rate D), respectively. The 3 mode bits, together with these data bits and 4 tail bits are then coded by the RSC defined above.

In the HR channel, the same RSC (but without termination) is used to encode the block of the mode and speech bits to generate 228 bits (i.e. 11.4 kbit/s) necessary for GSM HR channel.

At the receiver side, the transmitted speech coding rate is first detected and the rest data are then decoded. The main decoding routine is a soft output Viterbi algorithm (SOVA), where the soft output is necessary for the error concealment algorithm outlined above.

6. CHANNEL QUALITY ESTIMATION

The channel quality in a typical mobile radio transmission is determined by two different physical effects, namely the short term fading due to scattering, reflection and interference, and the long term fading due to the slow change of geographical surroundings. The influence of the short term fading can effectively be reduced by using an interleaver, hence only the effect of the long term fading has to be estimated in the AMR system.

The proposed channel estimator makes use of the soft output from the equalizer. It first calculates for every speech frame of 20 ms a channel state information out of 2 (halfrate) or 4 (fullrate) bursts, describing the momentary short term quality of the channel. From each burst the arithmetical mean of several (10 in our simulation) most unreliable equalized soft bits is calculated and the average of 2 or 4 bursts is taken, respectively. The result is fed to a FIR lowpass filter to obtain an output proportional to the long term fading. The order of the filter (28 in our simulation) must be chosen as short as possible for fast rate adaptation. From the filtered output a two bit metric is derived by threshold comparison and can indicate four different channel quality values.

The current metric and four previous metrics are summed up in the BS for both UL and DL adaptation and, depending on the resulting value the new speech mode can be determined.

7. COMPLEXITY AND DELAY

- *Complexity:* The complexity of the proposed AMR codec is estimated on the basis of a floating point ANSI C source code. Tab. 5 and 6 show the complexity estimations in terms of weighted million operations per seconds (WMOPS), RAM and ROM. The GSM HR and EFR are used as references.

	WMOPS	RAM	ROM	FOM
HR	2.2 wMOPS (≈ 0.8×HR)	2.7 kwords (≈ 0.9×HR)	1.6 kwords	3.4 (≈ 0.85×HR)
FR	5.2 wMOPS (≈ 1.9×HR)	4.4 kwords (≈ 1.4×HR)	3.1 kwords	7.3 (≈ 1.83×HR)

Tab. 5 Complexity of the channel codec.

WMOPS	RAM	ROM	FOM
21 wMOPS (≈ 1.4×EFR)	8.6 kwords (≈ 1.8×EFR)	14.9 kwords (≈ 2.8×EFR)	23.5 (≈ 1.4×EFR)

Tab. 6 Complexity of the speech codec.

Furthermore, the program size of the proposed channel codec is about 2 times that of the HR channel codec program ROM, and the program size of the proposed source codec is about 1.5 times that of the EFR codec program ROM.

- *Transmission delay:* In GSM, the MS-to-MS algorithmic round-trip delay for a speech codec is defined as

$$D_{round\text{-}trip} = 2(T_{sample} + T_{rftx}) + T_{Abisu} + T_{Abisd}$$

Where T_{sample} is the analysis frame length delay (duration of the segment of PCM speech operated on by the speech transcoder), T_{rftx} the interleaving and de-interleaving delay (of a speech frame over the air interface), T_{Abisu} the uplink Abis delay (time needed to transmit the minimum amount of bits over the Abis interface that are required at the speech decoder to synthesise the first output sample), and T_{Abisd} the downlink Abis delay (time required to transmit all the speech frame data bits over the Abis interface in the downlink direction that are required to encode one speech frame).

The round trip delay of the proposed AMR codec is shown in Tab. 7. Since in FR channel (assuming 16 kbit/s sub-multiplexing) the delay for 9.5, 8.1 and 6.3 kbit/s code modes is less than that of 13.3 kbit/s, only the result for 13.3 kbit/s is given. For the similar reason, only the results of 8.1 kbit/s mode (16 kbit/s Abis sub-multiplexing) and of 6.3 kbit/s mode (8 kbit/s sub-multiplexing) are given in HR channel.

AMR Delay	T_{sample}	T_{rftx}	T_{abisu}	T_{abisd}	Total [ms]
FR 13.3	25.0	37.5	5.8125	17.875	148.6875
HR 8.1 (16 kbit/s Abis)	25.0	32.9	4.1875	11.375	131.3625
HR 6.3 (8 kbit/s Abis)	25.0	32.9	7.625	18.25	141.675

Tab. 7 Round trip delay of the proposed AMR codec.

For 13.3, 8.1 and 6.3 kbit/s, T_{abisu} includes the 83, 57 and 51 bits necessary to reconstruct the first transmitted sample, respectively, plus 10 control and synchronisation bits in FR and HR channel. T_{abisd} includes 20 additional control bits. With the complexity and delay stated above, all constraints set by ETSI are fulfilled.

8. TEST RESULTS

Subjective tests in German were carried out to evaluate the proposed AMR codec under the conditions defined by the ETSI AMR standardization group. The tests include the following four experiments. The opinion scale used for the experiments 1, 3 and 4 was the Absolute Category Rating (i.e. 5=Excellent, 4=Good, 3=Fair, 2=Poor and 1=Bad). For the experiment 2, the modified Degradation Category Rating was used, with 5=Degradation not perceived or even some improvement, 4=Degradation perceived but not annoying, 3=Degradation slightly annoying, 2=Degradation annoying and 1=Degradation very annoying.

For the experiments 1 and 2, the ETSI qualification requirement is that, unless otherwise stated, the performance of at least one of the tested rate (A, B, C or D) is "as good as" the reference which is shown as a thick solid line ("Spec.") in Figs. 3 - 5.

- *Exp. 1:* Influence of errors and speech input level in clean speech conditions. In Exp. 1a (for FR channel), the reference for (carrier-to-interference ratio) C/I ≥ 13 dB is *EFR no error*, for C/I = 10/7 dB is *ITU G.728 no error*, and for C/I = 4 dB is *EFR at 10 dB*. In Exp. 1b (for HR channel), the reference for C/I ≥ 16 dB is *ITU G.728 no error*, for C/I = 13/10/7/4 dB is *FR at 13/10/7/4 dB*, respectively.

- *Exp. 2:* Influence of background noise in static C/I conditions (Exp. 2a for street noise and Exp. 2b for vehicle noise). In FR channel, the reference for C/I ≥ 13 dB is *EFR no error*, for C/I = 10/7 dB is *G. 729 and FR no error*, for C/I = 4 dB is *FR at 10 dB*. In HR channel, the reference for C/I ≥ 16 dB is *G. 729 and FR no error* (where the requirement is "better than" instead of "as good as"), for C/I = 13/10/7/4 dB is *FR at 13/10/7/4 dB*, respectively.

- *Exp. 3:* Influence of mode switching, speech input level and tandeming under clean speech conditions. The requirements are: 1) *no annoying artifacts when switching between the bit-rates within a channel.* 2) *the highest bit rate A should perform as well as EFR in FR channel and as well as G. 729 and FR in HR channel, all at the same tandeming and input level condition.*

- *Exp. 4:* Performances in dynamic conditions (Exp. 4a for FR channel and Exp. 4b for HR channel). For Exp. 4a, the requirement is that *the AMR performance is same or better than the EFR under the same conditions, and also the same or better than all the AMR FR tested modes under the same conditions.* For Exp. 4b, the requirement is that *the AMR performance is same or better than the FR under the same conditions, and also the same or better than all the AMR FR tested modes under the same conditions.*

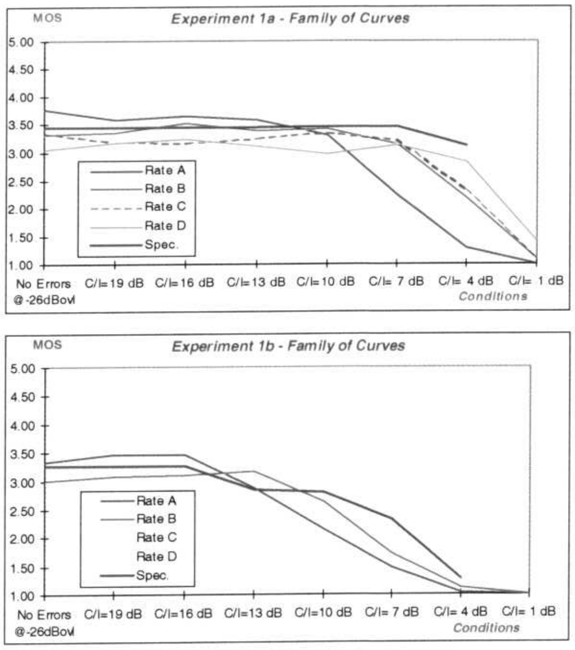

Fig. 3 Subjective test results for clean speeches with experiment 1a for FR channel and experiment 1b for HR channel.

The test results showed that the AMR codec passed all test conditions (requirements) except the following failures: 1 failure in Exp. 1b (at C/I= 7 dB); 3 failures in Exp. 2a (in HR channel at no error, C/I=19, 7 dB); 4 failures in Exp. 2b (in FR channel at C/I=4 dB, in HR channel at no error, C/I=19, 7 dB). According to the ETSI AMR qualification rules the codec should be qualified for a further selection phase. It can be noted that the proposed AMR codec failed to meet the requirements mainly in HR channel when channel error *and* background noise are present. In fact, these are the conditions known to be very challenging for a GSM AMR codec and among the 11 AMR candidate codecs submitted to ETSI for standardization competition, none of them fulfilled all these conditions.

The subjective performances, i.e. mean opinion scores (MOS) for different test conditions, of the proposed codec and the corresponding references are illustrated in Fig. 3 - 6, where ECx means the Error Condition for static C/I conditions with C/I = x dB, and DECi means the Dynamic Error Condition #i for Dynamic C/I conditions. Notice that in comparison to a fixed bit rate codec like the GSM EFR, significant gains are obtained especially in dynamic conditions, where the AMR rate adaptation is activated ("Y_{test}" in Fig. 6).

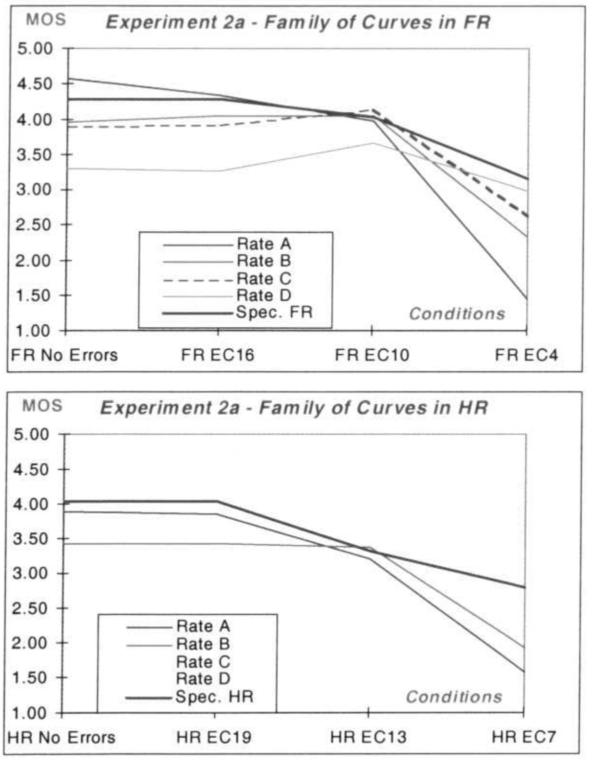

Fig. 4 Subjective test results for noisy speeches (street noise).

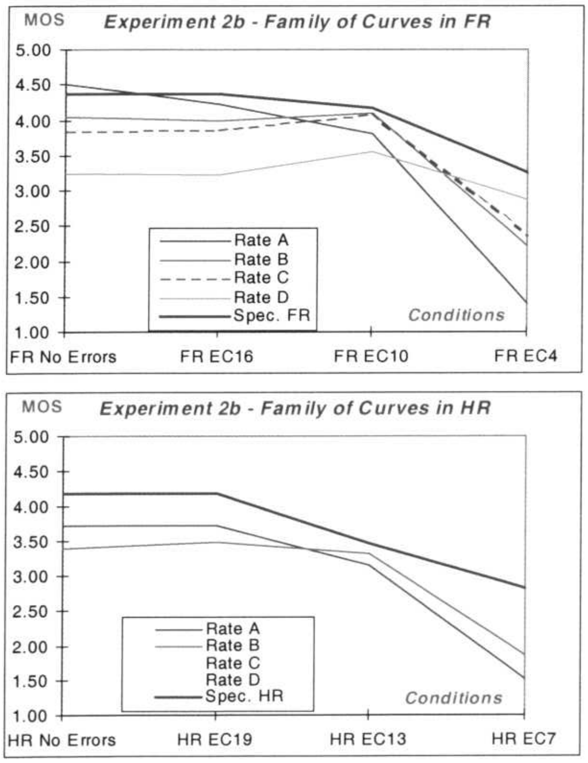

Fig. 5 Subjective test results for noisy speeches (vehicle noise).

9. CONCLUSION

We described an AMR speech codec including variable rate CELP speech coding, unequal error protection, error concealment and adaptation algorithms. The codec was, together with other 10 candidates, submitted to ETSI for standardization. Although the codec proposal described here did not win the competition, some novel techniques like the use of RSC for AMR channel coding has been accepted by the ETSI standardization group and integrated in the current AMR standard.

ACKNOWLEDGEMENTS

We want to thank Mr. O. Steil, with RWTH Aachen, Dr. S. Oestreich, Dr. J. Paulus, Dr. I. Varga and Mr. Ch. Erben with Siemens AG, for many fruitful discussions, and T. Ansahl for his engagement in carrying out the subjective test.

REFERENCES

[1] T. Fingscheidt and P. Vary, "Robust speech decoding: A universal approach to bit error concealment," in *ICASSP'97*, vol. 3, pp. 1667 - 1670, 1997.

[2] J. Hagenauer, "Rate-compatible punctured convolutional (RCPC) codes and their applications," *IEEE Trans. on Commun.*, vol. 36, pp. 389-400, 1988.

[3] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, pp. 2449-2457, 1995.

[4] S. Heinen, M. Adrat, O. Steil, P. Vary and W.Xu, "A 6.1 to 13.3 kb/s variable rate CELP codec (VR-VCELP) for AMR speech coding," In *ICASSP'99*, 1999.

[5] S. Heinen, M. Adrat, P. Vary and W.Xu, "Softbit speech decoding with re-quantization," In *Proc. of 3rd ITG Conf. Source and Channel Coding*, Jan. 17-19, 2000.

[6] T. Hindelang, M. Schmautz, J. Hagenauer and W. Xu, "Improved channel coding and estimation for adaptive multirate (AMR) speech transmission," submitted to *VTC'2000*.

[7] M.R. Schroeder and B.S. Atal, "Code-excited linear Prediction (CELP): high-quality speech at very low bit rates," in *ICASSP'85*, pp. 937-940, 1985.

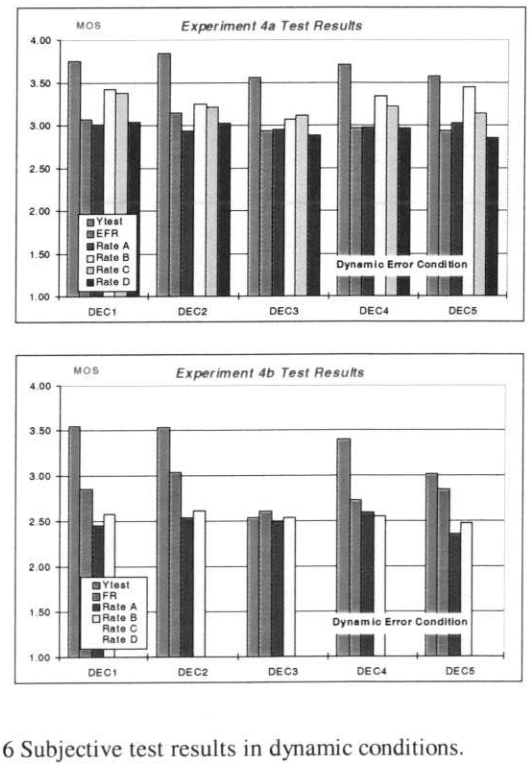

Fig. 6 Subjective test results in dynamic conditions.