ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

A MULTI-RATE SPEECH AND CHANNEL CODEC:

A GSM AMR HALF-RATE CANDIDATE

S.Villette, M.Stefanovic, A.Kondoz

Centre for Communication Systems Research
University of Surrey, Guildford GU2 5XH, Surrey, United Kingdom.
S.Villette@ee.surrey.ac.uk

10.21437/Eurospeech.1999-453

ABSTRACT

The current GSM standard is a fixed rate system which has been optimised to give good performance in all channel conditions. However since the channel conditions in a terrestrial mobile communication network such as GSM vary significantly between the best and the worst case, the existing GSM speech and channel coding performance can be improved by incorporating the dynamic channel conditions into the codec design. Under the initiative of the European Telecommunications Standards Institute (ETSI) such a system has been launched. This new standard is called AMR for Adaptive Multi-Rate: the source and channel coding rates can be adapted depending on the state of the channel, thus providing optimal balance between them at any time. The University of Surrey has submitted a candidate for this competition through the Mobile VCE. This candidate was the only one amongst eleven to use a vocoder in the half-rate GSM channel instead of a CELP based coder and tests ranked it among the best. This paper presents the system submitted for the half-rate channel as well as the results of the testing.

1. INTRODUCTION

The design requirements for the AMR competition were that the proposed system was able to operate both in the current full-rate and half-rate GSM channel to allow the reuse of the existing infrastructures [3]. All the candidates opted for a CELP based system similar to the EFR system, to be used in the full-rate channel at 22.8 kb/s. Two to four rates were used in the various systems submitted. The characteristics of the full-rate GSM channel in terms of bandwidth and channel errors make such a CELP based solution the best choice.

However, in the half-rate channel, only 11.4 kb/s are available, and the system was expected to maintain good speech quality even at a C/I ratio of 4 dB. In order for the channel coding to keep the number of corrupt frames to a manageable level, we felt desirable to have a source rate below 4 kb/s.

It is difficult to provide acceptable speech quality with CELP at such a low bit rate. Moreover the effect of corrupt frames at low C/I ratios can be difficult to deal with in a CELP system. On the other hand a vocoder based system can give better speech quality at this bit rate, and since it does not use long term prediction, the degradation induced by lost frames would be easier to cope with.

For these reasons we have decided to use a vocoder system in the half-rate, based on our existing Split-Band LPC Vocoder. All the other candidates have selected a CELP based system in the half-rate channel. This half-rate system will be described in this paper.

The vocoder is briefly described in Section 2, together with the improvements made to the model. In Section 3 the complete system is detailed. The results of the qualification phase testing are shown in Section 4. Finally, we summarize the results and present our future work in Section 5.

2. SPLIT BAND LPC VOCODER

2.1 General characteristics

The Split-Band LPC Vocoder has been presented in detail in [1,2,4]. It has been initially a 2.4 kb/s vocoder, whose name derives from the way the voicing is quantised: it assumes all harmonic bands below a certain cut-off frequency to be voiced and the rest unvoiced.

Since the quality of the speech at 2.4 kb/s was not high enough to meet the requirements of the ETSI AMR competition, the update rates of the parameters have been modified to allow higher quality at higher bit rates.

Some improvements have been made, including a varying length windowing, new quantisers, a revised pitch algorithm as well as a new voicing algorithm. The diagrams for the revised encoder and decoder are given in [4].

2.2 Improvements to the model

The basic model has seen several improvements, especially with regard to the background noise performance of the pitch and voicing algorithm, and to the windowing.

In order to improve the pitch and voicing determination under noisy background conditions, the level of background noise is estimated. This is then used to bias the pitch and voicing determination to obtain optimum performance in both clean and noisy background. Results indicate that the voicing algorithm performs very well even at 0dB SNR. At such a signal-to-noise ratio (typically in military applications) only a few pitch errors, amounting to less than 0.1%, were detected. They are due to significant and sudden increases in the background noise coinciding with either speech on-sets or off-sets.

Variable length windowing has also been introduced. The chosen analysis window should cover at least two pitch cycles, which is the minimum required for the correct analysis. A longer analysis window can affect the accuracy in determining the parameters.

2.3 Higher bit rate versions

In order to get the higher bit rates and higher quality needed to meet the ETSI requirements, it has been necessary to change the update rates of the parameters. Instead of updating the various parameters (LP coefficients, pitch, voicing, residual spectral amplitudes) once every 20 ms, the coder updates all of them every 10 ms, apart from the LP coefficients which can be updated either every 10 or 20 ms.

It has been then necessary to design new improved quantisation schemes, capable of providing various bit rates. For the spectral amplitudes, two quantisers have been designed, using either 8 or 21 bits, the latter offering near-transparent quantisation.

The pitch and voicing can also be quantised with various number of bits. For a 20 ms frame, the pitch value corresponding to the less important half of the frame can be differentially quantised with regards to the other pitch, in order to reduce the bit rate while losing very little quality.

Finally a new LSF quantiser has been designed. In order to cope with the tough requirements for channel error resilience, no predictive coding has been used. A simple split-vector quantiser has been designed, using 28 bits. The first 3 LSFs are vector-quantised using 10 bits, the next 3 with 9 bits and the last 4 with 9 bits. This bit allocation reflects well the relative importance of the LSFs.

3. HALF-RATE SYSTEM

3.1 Operating rates

As mentioned earlier, the need for MOS scores close to 4 implies the use of bit rates higher than the basic 2.4 kb/s. Since the system is supposed to operate in the half-rate GSM channel, the overall bit rate including channel coding and control bits has to be 11.4 kb/s. This allows us to use source bit rates of around 4 kb/s and still be able to protect them efficiently in very bad channel conditions, down to 4 dB C/I, and go up to around 7 kb/s for the higher quality version in good channel conditions

Combining the previously described update rates and various quantisers, three coders have been designed, operating respectively at bit rates of 3.9, 5.2 and 6.8 kb/s. The bit allocations and update rates of these three rates are described in Table 1, along with the basic 2.4 kb/s coder.

Bit Rate	2.4 kb/s		3.9 kb/s		5.2 kb/s		6.8 kb/s	
Update rate	20		20		20		20	
(in ms)	10	10	10	10	10	10	10	10
LPC	28		28		28		28	28
Pitch	7		7	5	7	5	7	7
Voicing	3		4	4	4	4	5	5
RMS energy	6		7	7	7	7	7	7
Spectral amplitudes	4		8	8	21	21	21	21
Total for 20 ms:	48		78		104		136	

Table 1: Bit allocation for the different rates of the Split-Band LPC Vocoder

3.3 Channel coding

Preliminary studies have shown that the best suited rates to be used in the GSM Half-Rate channel are 3.9, 5.2 and 6.8 kb/s rates, as there is no gain to be made from the 2.4 kb/s version in this application.

A specific channel coder has been designed for each rate, together with bit prioritisation and CRC checks. Also 6 bits have been added at the beginning for the rate adaptation scheme and a trailing sequence added to flush the bits. This is finally passed through a convolutional encoder with rate ½ and length 7. The resulting bits are either punctured or duplicated depending on the rate to add up to 11.4 kb/s.

The three modes described will from now be referred to as "Sblpc 1,2 or 3" as shown in Table 2. All rates add up to 11.4 kb/s, which is the capacity of the GSM half-rate channel.

	source coding	channel coding + rate adaptation
Sblpc 1	3.9 kb/s	7.5 kb/s
Sblpc 2	5.2 kb/s	6.2 kb/s
Sblpc 3	6.8 kb/s	4.6 kb/s

Table 2: Bit rates selected for the AMR competition

3.4 Error detection

Since the channel is bursty and the bits are protected by a half-rate convolutional code, it is expected that in the event of the code not being able to cope with the errors on the channel, whole chunks of data will be completely lost (as opposed to a random channel where it is much less likely that the code would fail for the same bit error rate).

Hence it is necessary to detect such instances and avoid using the corrupted bits at the decoder. Instead, the corrupted parameters are replaced based on the previously received frames, to ensure the smoothness of the speech output.

Each of the three rates has a different number of CRC checks, of various length, applied to different parameters. This is detailed in Table 3.

	Number of bits protected / length of the CRC					
Parameter	Sblpc 1		Sblpc 2		Sblpc 3	
LPC	28	5	28	6	28*2	
pitch	7+5	4+3	7+5	3	+7*2	
V/UV	4+4	4	4+4	3	+5*2	5*2
RMS energy	7+7	4+4	7+5	3	+7*2	
Spectral amplitudes			+7+5	+3	+5*2	

Table 3: CRC length per parameter for each rate

For the rate 1, which is to be used in bad channel condition, all the parameters are heavily protected by CRC, apart from the spectral shape which is not as important as the other parameters for the speech quality. Rate 2 is similar, with less powerful CRCs which also protects some bits of the spectral amplitudes quantiser joined with the RMS energy bits. This is due to the fact that the amplitude quantiser for rates 2 and 3 is more sensitive to errors than the one used in the first rate. For the last rate, only 2 CRCs are used, each on the 52 most important bits of each half-frame. This is sufficient since this rate is not supposed to be used when many errors are expected.

This all gives a good indication of whether the decoded parameters are valid or not. Only valid parameters are then used in the decoder.

3.5 Error concealment

When a parameter gets corrupted, it is necessary to replace it by a value which is not going to cause problem. For example if an RMS energy parameter is lost, it is much better to simply reuse the previous value than use the corrupted one, which may create a blast in the output signal. It is also important not to decode LSFs sets which are unstable.

In the first two rates, the strategies used for replacing the lost parameters depend on how many parameters have been lost, which parameter has been lost, and on how many times in a row this parameter has been lost.

When the energy bits are lost, instead of simply using the previous energy, the variations of the energy over the previous frames are used to extrapolate the value: if the energy decreases, it is likely to be an offset, so the extrapolated value is chosen to be lower than the last correctly decoded value. The opposite is true for an onset. If too many frames in a row are lost, then the energy is decreased down to silence, to avoid creating long unnatural noises.

Similar techniques are used for the other parameters. If too many of the parameters are lost, i.e. 4 or more out of 6 CRCs detect an error, then it is assumed that the whole frame is lost, and the 2 correct CRCs are just not reliable. In this case all the parameters are replaced.

Using these techniques makes a single frame loss usually not audible, and only under heavy error condition will the speech be degraded.

For the highest rate, it can be assumed that few frames will be lost. Hence only one CRC per half frame is used, and all parameters are replaced if one half-frame is corrupted. This is enough to ensure good speech quality.

By using these parameter recovery techniques, the influence of channel errors on the speech quality is much reduced, and good speech quality can be maintained even at very low C/I ratios.

3.6 Rate adaptation scheme

In order to always use the most appropriate rate at any time, on the up- and on the down-link, a rate adaptation scheme has to be designed.

It has to synchronize the transmitting rate with the reception rate to allow communication between the base station (BS) and the mobile station (MS) and to decide based on an estimation of the channel quality which rate is to be used. As both links are independent and the estimated bit error rate (EBER) of a link is obtained at the receiving end, EBER and the rate switching decisions have to be exchanged between the MS and the BS [4].

The performance of the complete system depends heavily on the speed of the rate adaptation scheme. If it is too slow and a sudden increase in BER occurs, the system may not be able to change rates towards a more channel robust rate, hence losing frames and degrading the quality. In this system, a total delay of 160 ms is needed for a rate adaptation to take place.

The operation of the system is illustrated in Figure 1. In this Figure the up-link is simulated, with the complete system in operation. It can be seen that the rate adaptation scheme follows closely the bit error rate of the channel, and adapts the rate so that the best rate is used at all times.

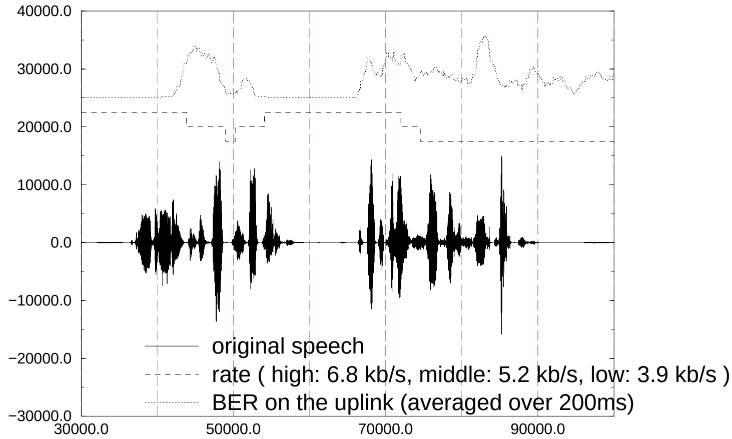

Figure 1: rate adaptation on the uplink

4. SYSTEM PERFORMANCE

4.1 Conditions of testing

The AMR competition consists of several phases. The first phase required each proponent to carry out internal tests on their candidate against various existing standards, following strict guidelines given by ETSI. After this qualification phase, each proponent submitted its solution along with the results. The results of our candidate are presented here.

The testing has been carried out internally following ETSI's Recommendations [3]. It consisted of four parts, each testing a particular aspect of the complete proposed system. The testing conditions are detailed in [4].

4.2 Experiment 1: The effect of errors under clean speech conditions

Coder	Sblpc 1	Sblpc 2	Sblpc 3	EFR	FR
Bit rate	11.4	11.4	11.4	22.8	22.8
clean	3.67	3.96	3.85	4.29	N/A
19 dB	3.63	3.67	4.04	N/A	N/A
16 dB	3.81	3.69	3.77	N/A	N/A
13 dB	3.63	3.71	3.38	N/A	3.35
10 dB	3.81	3.58	2.67	4.08	3.40
7 dB	3.25	2.85	1.65	3.69	2.98
4 dB	2.10	1.31	1.00	2.00	2.00
1 dB	1.10	1.02	1.06	N/A	N/A

Table 4: MOS score for the Experiment 1

In this experiment, the AMR candidate performance in the half-rate channel (11.4 kb/s) is compared to EFR and FR GSM standards performance (both operating at 22.8 kb/s using their own channel codec and error resilience techniques) under various error conditions corresponding to C/I ranging from 1 dB to 19 dB, plus a clean channel condition, using MOS.

4.3 Experiment 2: The effect of background noise for static conditions

The subjects were asked to mark the degradation perceived between the original noisy speech and the processed version of the original, using DMOS. The processing can also include channel errors, in the same conditions as for Experiment 1.

Coder	Sblpc 1	Sblpc 2	Sblpc 3	FR	G729
clean	3.98	3.79	3.86	3.77	3.88
13 dB	3.57	3.86	3.71	3.88	N/A
7 dB	3.36	3.05	2.04	3.57	N/A

Table 5: DMOS scores for Experiment 2, street noise.

Coder	Sblpc 1	Sblpc 2	Sblpc 3	FR	G729
clean	3.84	3.82	3.91	3.88	3.81
13 dB	3.57	3.46	3.49	3.71	N/A
7 dB	3.26	2.49	1.50	3.45	N/A

Table 6: DMOS scores for Experiment 2, car noise.

4.4 Experiment 3: The effect of switching, speech input level and tandeming under clean speech conditions

The AMR candidate is expected to cope with various input level. Part of this experiment consisted of checking the proper operation of the candidate with an input level 10 dB higher or lower than the nominal input level. The candidate was designed to have its nominal level at -26 dB from overload. The scaling of the speech samples was performed using the tools provided by the ETSI for this purpose.

coder	Sblpc1	Sblpc2	Sblpc3	FR	G728	G729
-16 dB	3.69	3.79	3.85	N/A	4.02	3.85
-26 dB	3.23	3.33	3.31	3.23	3.52	3.67
-36 dB	2.50	2.60	2.75	N/A	3.06	2.88
tandem	2.40	2.60	2.92	3.04	3.46	3.10

Table 7: MOS scores for Experiment 3

Switching between rates has also been tested. The test has shown that switching between rates does not produce any artifacts.

4.5 Experiment 4: The effect of dynamic error patterns

The aim of this experiment is to check the performance of the complete system and to validate the concept of Adaptive Multi-Rate. The complete system is simulated using the dynamic error patterns provided on both links and the rate adaptation scheme controls the switching to optimize the performance. Five different scenarios (called Dynamic Error Condition, DEC) are used, all representative of typical mobile channels.

Coder	Sblpc, 11.4 kb/s	GSM FR, 22.8 kb/s
DEC 1	3.63	3.67
DEC 2	3.57	3.67
DEC 3	3.63	3.66
DEC 4	2.98	2.77
DEC 5	2.82	2.81

Table 8: MOS scores for Experiment 4

4.6 Overall performance

There were eleven candidates for the qualification phase, all of whom were well established companies with only one University: the University of Surrey/Mobile VCE Ltd candidate. After checking that all design constraints had been met by the proponents, the results of the listening tests for the qualification phase were tabulated and the proponents ranked to select the promising ones.

According to the initial figure of merit, our candidate was placed third overall, with the best figure of merit for the half-rate GSM channel.[3]

5. CONCLUSION

In this paper we have detailed the main design parameters and results obtained from the candidate submitted by the University of Surrey through Mobile VCE Ltd to the ETSI AMR competition.

Test results of the ETSI/AMR qualification phase have shown that using a vocoder instead of a CELP coder in the half-rate channel was a viable approach, as our candidate was ranked amongst the best.

In addition to the new speech coding scheme, the complementary channel coding schemes and bad frame substitution techniques performed well contributing significantly to the overall performance obtained. The link adaptation algorithm enabled independent link rate variations between the up- and down-links. It tracked the channel variations very well by incorporating up- and down-link channel state estimators.

6. REFERENCES

[1] I.Atkinson, S.Yeldener, A.Kondoz, "High Quality Split-Band LPC Vocoder Operating at Low Bit Rates" *ICCASP 97 Proceedings*, Volume 2, pp 1559

[2] S.Villette, M.Stefanovic, I.Atkinson, A.Kondoz "High Quality Split Band LPC Vocoder and its Fixed Point Real Time Implementation" *Eurospeech 97 Proceedings*, Volume3, pp 1243-1246

[3] ETSI AMR Qualification Phase Documentation, 1998

[4] S.Villette, M.Stefanovic, A.Kondoz ,"Split Band LPC Based Adaptive Multi-Rate GSM Candidate", *ICASSP 99 Proceedings*, Volume 1, paper 1798