Integration of a CELP Coder in the ARDOR Universal Sound Codec

Balázs Kövesi, Dominique Massaloux, David Virette and Julien Bensa

France Telecom/Division R&D/ TECH/SSTP
2 Av. Pierre Marzin, 22300 Lannion – France
balazs.kovesi;david.virette;dominique.massaloux@francetelecom.com

10.21437/Interspeech.2006-64

Abstract

This paper describes the CELP coding module within the Adaptive Rate-Distortion Optimized sound codeR (ARDOR). The ARDOR codec combines coding techniques of different nature using a rate-distortion control mechanism, and is able to adapt to a large range of signal characteristics and system constraints. The implemented CELP codec is derived from the 3GPP AMR-WB codec. Adaptations were necessary to match the ARDOR structure constraints and several new features have been added to improve the codec performance in this context. Listening test results are given to illustrate the behavior of the final codec compared to state-of-the-art coders.

Index Terms: sound coding, CELP, IST project ARDOR

1. Introduction

The ARDOR codec is a universal sound coding prototype. It has been developed within the IST ARDOR project. The objective of this project was the development of a universal codec as an answer to the need created by the emergence of time-varying heterogeneous networks [1]. In function of the input signals nature and the given constraints the control unit, based on a rate-distortion (R-D) optimization mechanism, configures the sound codec and allocates the available bit budget among the selected coding techniques in an optimal way. An advanced perceptual distortion measure provides the perceptual criterion for the R-D mechanism.

One of the coding techniques available in the ARDOR sound codec is a Code-Excited Linear Predictive (CELP) module that is derived from the 3GPP AMR-WB codec [2]. This paper focuses on the main difficulties met by introducing a CELP coder into the ARDOR structure and on the solutions proposed. The challenge was to combine several coding strategies with different features such as time segmentation, sampling frequency or delay. Adaptations to the AMR-WB codec were necessary. Several new features were also added to improve the perceptual quality of the CELP module in this context.

A general description of the ARDOR codec is given in section 2. Section 3 highlights the problems related to the introduction of a CELP module into the ARDOR structure. Section 4 describes the ARDOR CELP module developed to solve those issues. Section 5 shows the results of formal listening test involving the CELP module, section 6 addresses complexity issues in the present context and section 7 concludes this paper.

2. The ARDOR codec

The principle of the ARDOR codec, shown in Figure 1, is to decompose the input signal into an additive set of signal components, each encoded with a specific coding technique. The signal decomposition is based on a R-D control mechanism that segments the input signal, selects the involved coding techniques and their operation order for each segment and distributes the available bits among them [3]. The R-D optimization framework gives the coder its versatility: it allows the coder to adapt to the input signal and to constraints such as maximum bit rate, delay, or maximum distortion [1].

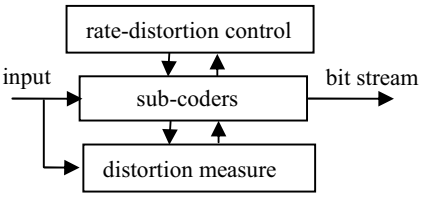

Figure 1 *R-D optimized coder scheme.*

Due to complexity reasons in the following only fixed segmentation is considered. However, within a segment, each sub-coder can determine its own flexible sub-segmentation and the distribution of the allocated bit budget among the sub-segments. A sub-segment consists of one or several ARDOR frames where the frame size was chosen equal to 128 samples at 48 kHz sampling frequency (2.67 ms).

The sub-coders work in a sequential way. The input signal is encoded by the first specified sub-coder and the residual signal is used as an input for the following coder. After testing few properly selected configurations with given bit allocations and evaluating their quality using the distortion measure, the control determines the optimal configuration for the segment.

2.1. Distortion measure and sub-coders

The decisions of the control unit are based on the perceptual distortion measure, which is a fundamental component of the R-D optimization mechanism. A new family of distortion measures has been designed to produce accurate and meaningful perceptual quality estimation [4].

Four different coding techniques were integrated in the ARDOR universal sound codec: CELP coding, Transform coding [5], Sinusoidal coding [6] and Noise coding [7]. The input signal is first encoded by a combination of CELP, Transform and Sinusoidal coders. Then, the Noise coder was used to complete the global synthesis signal with synthetic noise signal to account for missing parts in the spectrum. This noise synthesis is based on the excitation pattern [5, 7] that is transmitted to the decoder. The excitation pattern is a perceptually meaningful representation of the spectral envelops of the original signal; derived from the spectral energy distribution across auditory filters modeling the human auditory system.

3. Introduction of the CELP module

The role of the CELP module in the ARDOR codec was to offer good quality speech coding at low bitrates (< 20 kbit/s). Yet the introduction of a time domain coder into a complex structure involving parametric and transform coding schemes generated a lot of problems, such as:

- The sampling frequency of the ARDOR codec is 48 kHz while CELP coding works generally at maximum 16 kHz sampling rate. Special care was needed to handle the delay caused by the down- and upsampling filtering. This is detailed in section 4.2
- The configuration of the ARDOR codec can change during the encoding while switching between different coding schemes without annoying effect is not trivial. CELP coding is highly recursive. After a configuration change its memories have to be correctly rebuilt to assure smooth transition(see section 4.3)
- At lower bitrates, Transform and Sinusoidal coding encode first the most important parts of the spectrum, the remaining less important parts can be efficiently replaced by the synthetic noise produced by the noise coder. This is not true for the CELP codec where only the lower band is encoded and the missing higher band can still contain important information. Replacing this missing part by synthetic noise tends to cause annoying effects like pre-echo. That is why a bandwidth extension module was designed for the CELP codec (section 4.4)
- The R-D optimization control needs a finely graduated convex R-D curve as input, so a muti-bitrate CELP codec was necessary. A scheme deriving from the ACELP part of the AMR-WB codec has been developed. This choice was motivated by the high level performances of this coder and large number (eight) of available bitrates (see section 4). A combined encoding procedure was designed (section 4.5) to avoid running eight times the full process. The problem of the non convexity of the resulting R-D curve has also been addressed (section 4.6).

4. The ARDOR CELP module

First a very brief description of the AMR-WB codec is given focused on the main relevant features. For more details see [2].
The AMR-WB codec splits the frequency band in two parts: a 50-6400 Hz lower band (ACELP part) at 12.8 kHz sampling frequency and a 6400-7000 Hz high band. The ARDOR CELP sub-coder module was derived from only the ACELP part..
The frame length of this part is 256 samples (20 ms). The linear prediction (LP) analysis is performed once per frame using a 30 ms asymmetric window with 5 ms of look-ahead.
The adaptations and new features added to this coding scheme are described in the next sections.

4.1. Sampling frequency, frame length and bitrates

The sampling frequency of ARDOR codec is 48 kHz while that of the CELP module was chosen 12 kHz to make down- and upsampling easier. Though the CELP part of the AMR-WB was designed for 12.8 kHz sampling it was decided to use it as it is: informal listening tests have shown that there is no significant perceptual degradation due to this slight mismatch. In this way, at 48 kHz sampling frequency the CELP frame length is 1024 samples (21.33 ms). The ARDOR CELP module uses this fixed sub-segmentation of 1024 samples (8 ARDOR frames).

Table 1 summarizes the bit allocation of the eight CELP modes and the corresponding bitrates.

Table 1. *ARDOR CELP sub-coder operating modes*.

Mode	LPC filter	Lowp. filter	Pitch delays	Fixed CB	Gains	Total / frame	Bitrate (kbit/s)
8	46	4	30	352	28	460	21.56
7	46	4	30	288	28	396	18.56
6	46	4	30	256	28	364	17.06
5	46	4	30	208	28	316	14.81
4	46	4	30	176	28	284	13.31
3	46	4	30	144	28	252	11.81
2	46	0	26	80	24	176	8.25
1	36	0	23	48	24	131	6.14

4.2. Look-ahead

The downsampling from 48 to 12 kHz of the CELP input signal and the upsampling from 12 to 48 kHz of the CELP output signal require the use of anti-aliasing filters. To preserve phase information, a 129 coefficient FIR filter has been used. This filter introduces a delay, both for the input and the output signals. To ensure alignment of the upsampled CELP output and input signals, 128 samples of look-ahead are necessary. This creates a problem in the ARDOR framework: if the CELP sub-coder is working at a second stage, these look-ahead samples are not available. To overcome this problem, the missing signal is extrapolated by prediction, performing a pitch estimation and a LPC analysis on the previous samples.
Contrary to the ACELP part of AMR-WB codec, no look-ahead was used for the LP analysis to avoid increasing the number of unavailable samples. Again, no perceptually significant degradations were noticed due to this change.

4.3. Memory update

As mentioned in section 3, after a change in configuration, , CELP memories have to be updated to ensure smooth transitions. For that, the knowledge of the past input and output signals is necessary. However these signals are not always available. This is the case, for instance, when the CELP module has worked on the first stage for a frame but on the second stage for the next frame, encoding the residual of a sinusoidal coder. The nature of the CELP input signal is significantly different for these two frames. The past input and output signals of the CELP codec, corresponding to this residual encoding position, are not available either. Below is a summary of the different situations which may occur and the corresponding solutions:

- CELP is on the first stage: the input and output of the ARDOR codec for the previous frame can be considered as the past input and output of the CELP module, but these signals are only available at 48 kHz sampling frequency. The delay of the down-sampling filtering creates holes (corresponding to the half length of the filter) in the past output signal at 12 kHz. This has been solved by extrapolating the past output signal using a prediction process similar to that described in section 4.2.
- When CELP is not the first stage two cases may occur:
  - Past signals that correspond to the current CELP configuration are available from the previous segment (e.g. previous segment: Sinusoidal/Transform sub-coders were working, current segment: Sinusoidal/CELP sub-coders are working, i.e. the first stage remains the same). The past

signals are available (i.e. past input and output of the Transform sub-coder) and memories are recomputed in the same way as in the first stage case.

- Past signals that correspond to the current CELP configuration are not available from the previous segment (e.g. switch from CELP/Transform to Sinusoidal/CELP configuration). CELP memories are set to zero in this case, at least continuity is ensured in this way. The drawback is that the CELP coder will need some time (generally less than half a CELP frame) to rebuild its memories and fully reproduce the current residual. To correct this artifact, the synthesis signal is post-processed: a "backward" prediction is applied. The first, low energy part is replaced by backward predicted samples as shown in Figure 2, obtained after performing a pitch estimation and a LPC analysis on the second part of the synthesized frame. A fading function is applied on the first samples to prevent discontinuities.

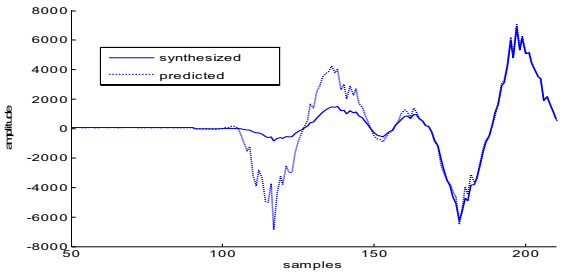

Figure 2 *Example of "backward" prediction.*

4.4. Bandwidth extension

As mentioned in section 3, a bandwidth extension module is activated when the CELP codec is on the first stage. This feature operates in a similar way as that in the AMR-WB+ codec [8].
At the decoder, once the CELP output at 48 kHz is computed, a linear prediction (LP) analysis is performed on the full band spectral envelope extracted from the Noise coder parameters. To avoid over-voicing effects, this spectral envelope amplitude is slightly decreased above 6 kHz. A lower band (6 kHz bandwidth) residual signal is obtained by filtering the CELP output by the inverse LP filter. This residual is replicated four times in the frequency domain, producing a 24 kHz bandwidth excitation signal. The full band extended CELP synthesis signal is obtained by filtering this excitation signal by the LP filter. Finally, the Noise coder is used to fill and smooth the spectrum.

4.5. Combined encoding procedure

The eight modes of the ARDOR CELP sub-coder rely on the same scheme. As shown in table 1, the differences between the eight modes are coming from:

- the two different quantization settings of the LPC parameters.
- the possibility to introduce an optional fixed low-pass filter (to avoid a possible "over-voicing" effect) for higher bit rates modes.
- the three different precisions in the fractional delay of the adaptive codebooks.
- the use of different algebraic codebooks for each mode.
- the two different gain quantizers.

To reduce the computational complexity of the CELP parameters for the different operating modes a process was designed to retain in memory common parameters which are useful for another mode computation. The complexity reduction obtained in this way is approximately 25%.

4.6. Choice of the CELP operating mode

A nine point initial Rate-Distortion (R-D) curve is calculated for each frame where the first point corresponds to the 0 bit mode (CELP module not active) and the others correspond to the distortion of the eight modes of the CELP coder calculated using the perceptual distortion measure coming from the perceptual module (see section 2.2). An analysis of the R-D curve is performed and some of the eight R-D estimates are discarded, to ensure that the curve is monotonously decreasing and convex. However, discarding the first mode (6.33 kbits/s) has been forbidden to ensure that at least one low bit rate is available. Then the slope of the rate-distortion curve is compared to the constraint value given by the R-D control module to determine the best R-D point.

5. Listening Test Results

Several listening tests were conducted to evaluate the quality of the ARDOR codec in different configurations [1]. A comparison test between variable and constant CELP bit rate is presented as well as the results of a MUSHRA test, evaluating different configurations of the ARDOR codec at low bit rate.

5.1. Variable versus constant bit rate

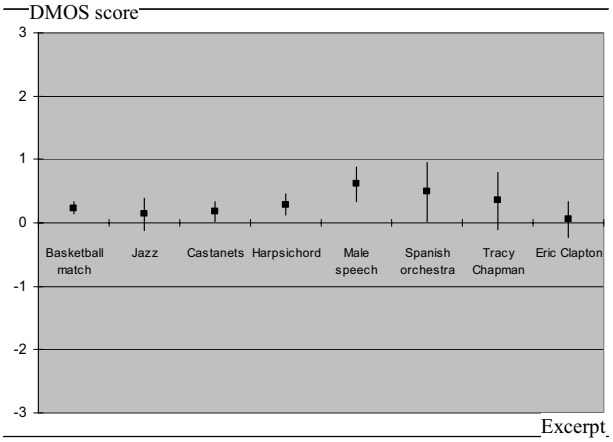

Figure 3 *CELP encoder at 12 kbit/s variable versus constant bit rate.*

Using the CELP sub-coder alone in the ARDOR codec leads to a source controlled R-D optimized variable bit rate CELP coder. The variable bit rate encoding, operating at an average of 12 kbit/s was compared to the constant 12 kbit/s encoding. Excerpts of different nature were listened (speech, music, speech with noise). Pairs of excerpts were presented in random order, listeners rated the second sample with respect to the first one on a -3…3 DMOS scaling rate. Figure 3 gives the obtained mean DMOS scores and the confidence interval for each excerpt. Positive value shows preference for the R-D optimized encoding. The variable CELP coder in the ARDOR structure

hence clearly outperforms the fixed rate CELP, especially for speech signals, which is the targeted mode for this module.

5.2. MUSHRA test results

A MUSHRA test (ITU-R BS.1534 [9]) was conducted to evaluate the ARDOR codec quality at 20 kbit/s. Three ARDOR configurations were tested: CELP alone, Transform alone and CELP/Transform (hybrid). Two state-of-the-art codecs were also chosen for their proven qualities: AMR-WB+ and the MPEG HE-AAC (coming from the Nero software package 6.6.0.14). Five audio excerpts were selected: two speech excerpts, two music excerpts and one speech with noise.

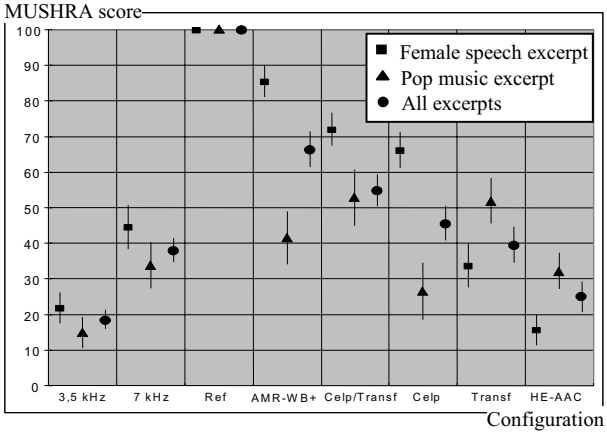

Figure 4 *MUSHRA test results, 20 kbit/s.*

Figure 4 extracts the results for a speech excerpt, a music excerpt and the mean of all excerpts, for each tested configuration. The hybrid ARDOR codec performs better than the individual sub-coders. This shows that the R-D control and the distortion measure work well. Yet for speech excerpts, the hybrid ARDOR codec does not reach the quality level of the AMR-WB+. This is mainly due to the CELP sub-coder module which does not perform as good as the AMR-WB+ codec for the higher frequencies: Informal tests have shown that both codecs are equivalent in the lower band (50-6000 Hz) and that therefore, the bandwidth extension algorithm used in the ARDOR CELP sub-coder is still too crude to achieve the high level of quality provided by the AMR-WB+ at this bit rate: the noise coder which is used to smooth the spectrum and avoid over-voicing effects still introduces pre-echo. Another point which may explain this difference is that the bandwidth extension requires approximately 4 kbit/s (noise coder bit rate) for the CELP ARDOR module whereas in the AMRWB+ case, only 0.8 kbit/s is used to encode the high frequency bands.

6. Complexity

The computation complexity of ARDOR encoding grows rapidly with the degrees of freedom of the ARDOR codec. The main challenge of the ARDOR project was to prove that a universal and flexible sound codec can perform as well or even better than a codec specially designed for a given condition. This concept has been proved, but the overall complexity of the structure remains high. Yet several ideas were identified to decrease the processing time like the use of property vectors to estimate R-D curves. For the CELP coder, combined processing as described in section 4.5 also allows to reduce complexity. This shows that solutions exist to overcome this problem.

7. Conclusions

This paper presented the solutions to the problems met during a multi bitrate CELP codec in a very flexible universal sound codec. Tests show that the performance of the hybrid configurations is superior or equal to that of the individual components. This proves that the control module combined with a novel objective measure works well and also that the switching problems due to configuration changes were successfully solved..

8. Acknowledgements

The ARDOR project was supported by the E.U. grant IST-2001-34095 (www.hitech-projects.com/euprojects/ardor). We thank all the members of the ARDOR project, especially Nicolle H. van Schijndel (Philips), project leader, Catherine Colomes (France Télécom) and Steven van de Par (Philips) for performing the listening tests and Stéphane Ragot (France Télécom) for his advises.

9. References

[1] ARDOR final report, E.U. grant no IST-2001-34095, 2001 (www.hitech-projects.com/euprojects/ardor).

[2] Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; General Description, 3GPP TS 26.190, 2004.

[3] N. H. van Schijndel and S. van de Par, "Rate-Distortion Optimized Hybrid Sound Coding," in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, 2005.

[4] S. van de Par, A. Kohlrausch, G. Charestan, and R. Heusdens, "A new psychoacoustical masking model for audio coding applications," in Proc. IEEE ICASSP, vol.2, pp. 1805-1808, Orlando, 2002.

[5] O. Niemeyer and B. Edler, "Efficient Coding of Ecitation Patterns Combined with a Transform Audio Coder," in Proc. 118th AES Convention, Barcelona, paper 6466, 2005.

[6] R. Heusdens and S. van de Par, "Rate-Distortion optimal sinusoidal modeling of audio and speech using psychoacoustical matching pursuits," in Proc. IEEE ICASSP, vol.2, pp. 1809-1812, Orlando, 2002.

[7] S. van de Par, V. Kot, and N. H. van Schindel, "Scalable noise coder for parametric sound coding," in Proc. 118th AES Convention, Barcelona, paper 6465, 2005.

[8] J. Makinen, B. Bessette, S. Bruhn, P. Ojala, R. Salami, A. Taleb, "AMR-WB+: a new audio coding standard for 3rd generation mobile audio services," in Proc. IEEE ICASSP, vol.2, pp.1109-1112, 2005.

[9] ITU-R Recommendation BS.1534, "Method for the subjective assessment of intermediate quality level of coding systems"; approved in 2001-06