Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

Audio Communication Coder

Aníbal J. S. Ferreira[1,2], and Deepen Sinha[2]

[1] *University of Porto, Portugal*

[2] *ATC Labs, USA*

Correspondence should be addressed to A. Ferreira (ajf@atc-labs.com, a.j.ferreira@ieee.org)

ABSTRACT

3G mobile and wireless communication networks elicit new ways of multimedia human interaction and communication, notably two-way high-quality audio communication. This is inline with both the consumer expectation of new audio experiences and functionalities, and with the motivation of Telecom Operators to offer consumers new services and communication modalities. In this paper we describe the design and optimization of a monophonic audio coder (Audio Communication Coder -ACC) that features low-delay coding (< 50 ms) and intrinsic error robustness, while minimizing complexity and achieving competitive coding gains and audio quality at bit rates around 32 kbit/s and higher. ACC source, perceptual and bandwidth extension tools are described and an emphasis is placed on ACC structural and operational features making it suitable for real-time, two-way audio communication. A few performance results are also presented. Audio demos are available at http://www.atc-labs.com/acc/ .

1. INTRODUCTION

Several proprietary and standardized audio coding schemes are successfully used in many application areas involving audio broadcast, streaming, messaging or download services. In these areas, compression efficiency and audio quality are the most important competitive factors. Usually, high compression efficiency is reached by compromising on such coder characteristics as algorithmic complexity and end-to-end coding delay. Furthermore, interframe coding is frequently used which lowers the intrinsic error robustness of the coding process, and raises significantly the recovery time after a communication failure. However, with the advent of 3G mobile and wireless communication networks, a new communication modality is emerging where these aspects are very critical and may be more relevant than strict compression efficiency: real-time, two-way high-quality audio communication.

This paper presents new advances to the design and optimization of a new audio coder (Audio Communication Coder -ACC) whose conceptual approach has been presented recently [1]. In order of maximize its intrinsic error robustness, ACC focuses on intra-frame coding techniques only (making error concealment more effective and) taking advantage of different opportunities for signal compression, namely:

- source/parametric coding techniques,
- perceptual coding techniques,
- bandwidth extension techniques [2].

The paper is organized as follows. In section 2 the structure of the ACC encoder is described highlighting the most significant optimizations recently introduced. The ACC decoder is briefly described in section 3. Section 4 reviews some of the most important structural features of ACC, and section 5 reviews several ACC operational features making it suitable for real-time audio communication. Section 7 concludes this paper with few ideas on possible application areas.

2. ACC ENCODER STRUCTURE

The ACC encoder structure is represented in Fig. 1 and follows closely that already presented in [1]. It shares the concept of *adaptive transform coding* and combines source coding tools (transform coding, parametric coding, entropy coding, PDF-optimized quantization [3]), perceptual coding tools (psychoacoustic modeling and noise shaping), and bandwidth extension tools [2]. The audio signal is first analyzed in the time domain for strong non-stationary events. In this case, a window switching mechanism is activated (section 4.2) in order to improve the time resolution of the coding process. The signal is then transformed to the Odd-Discrete Fourier domain (by means of an Odd-DFT -ODFT) where it is further analyzed in order to detect sinusoids or harmonic structures of sinusoids, and in order to estimate the Threshold of Masking using a psychoacoustic model. If relevant sinusoids are detected or if a relevant structure of sinusoids harmonically related is identified, they are modeled, parametrically encoded (frequency, magnitude, and phase), and are subtracted from the (complex) ODFT domain using the same sinusoidal synthesis model existing at the decoder.

The spectral residual is then normalized by a model of the Threshold of Masking, and is transformed to the MDCT domain. This model is obtained as a short-pass liftered version of the real cepstrum of the Threshold of Masking [4]. Non-linear quantization follows that implements PDF-optimized quantization concepts [5]. Rate and distortion are simultaneously controlled by means of a single parameter (α -ALPHA) that sets the degree of overcoding or undercoding of the coding process.

Huffman coding is used with spectral and cepstral coefficients, and with the magnitudes of sinusoids, in order to improve the compression efficiency.

A major optimization has involved the operations of spectral subtraction and normalization by a model of the Threshold of Masking. In fact, spectral subtraction is performed before the normalization in order to take full advantage of the accuracy of the sinusoidal synthesis [6] and to deliver, as a consequence, a smaller residual. Two other important optimizations concern rate/distortion control and Huffman coding. In fact, as will be described in sections 4 and 5, the operation of the rate/distortion loop has been improved in order to minimize the number of iterations needed in approaching constant bit rate coding (CBR), and new Huffman tables have been designed to better match statistical distributions of quantized coefficients, cepstral coefficients, and magnitudes of harmonic sinusoids.

3. ACC DECODER STRUCTURE

The ACC decoder structure is represented in Fig. 2, corresponds closely to that already presented in [1], and reflects the major structural improvements introduced at the encoder. The major processing steps at the decoder consist of Huffman decoding, inverse quantization, signal reconstruction and bandwidth extension. The signal reconstruction involves denormalization of the (quantized) spectral residual by a model of the Threshold of Masking, combination with in-band sinusoids (after being synthesized to the MDCT domain), and transformation to the ODFT domain[1]. Bandwidth extension is separately implemented for stationary noise and sinusoids using the bandwidth extension tools of FSSM

[1]This transformation is influenced by the Time Domain Aliasing Cancellation mechanism of the MDCT [7].

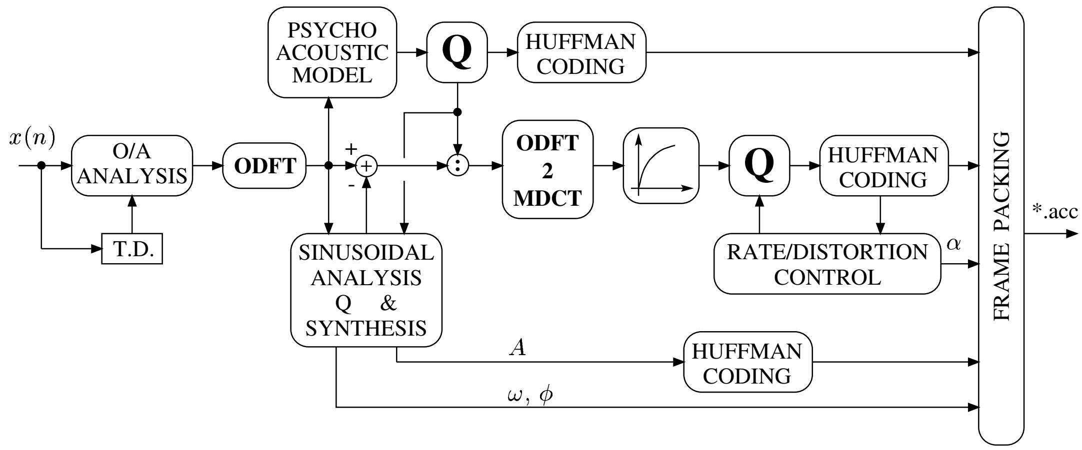

Fig. 1: Block diagram of the ACC encoder. In this diagram T.D. denotes transient detector.

and ASR [8, 9, 2]. Sinusoids are bandwidth extended in the ODFT domain and therefore are not subject to the TDAC mechanism of the MDCT. On the other hand, bandwidth extension of sinusoids is only implemented when REGULAR, START and STOP long windows are used.

4. STRUCTURAL FEATURES

4.1. Parametric coding of sinusoids

Each sinusoid is parametrically represented using a quantized frequency, magnitude and phase, that are transmitted to the decoder. If relevant sinusoids exist in a given audio frame such that their coding and subtraction from the MDCT spectrum is likely to provide a coding gain, they may be coded as individual sinusoids or as harmonic structures. The first case implies that all individual frequencies [10] are transmitted, which is reasonable only when the number of non-harmonic sinusoids is small. The second case is more bit-efficient since only the frequency of the fundamental is transmitted to the decoder. Interestingly, this is the most frequent situation as typical audio material has a rich content of harmonic structures of sinusoids [11]. The synthesis of each sinusoid in the ODFT/MDCT domain takes advantage of accurate spectral reconstruction models that encompass several frequency bins of the ODFT/MDCT [12, 13, 6].

4.2. Window switching

Typical audio material contains both quasi-stationary and non-stationary regions. These elicit different signal quality criteria since it is known that when judging quality, the human auditory system focuses on spectral detail of the audio signal in the first case, and focuses on temporal detail of the audio signal in the second case. For this reason, it is important to code quasi-stationary regions using high frequency resolution, and to code non-stationary regions using high-time resolution. As several other coding schemes, ACC implements this by means of a window switching mechanism as illustrated in Fig. 3. This mechanism allows the switch between LONG windows and SHORT windows, improving by a factor of eight the normal segmentation of the audio signal. Signal processing constraints require the use of two special windows, START and STOP, as illustrated in Fig. 3, in order to implement the transition.

From a purely signal processing perspective, high frequency resolution coding is more interesting since transform coding using a higher number of bands provides a higher coding gain [14]. Fortunately, the need to code non-stationary regions with high time resolution (and therefore, poor frequency resolution, which means a lower coding gain) does not imply, on average, a significant penalty to the compression efficiency since highly non-stationary regions repre-

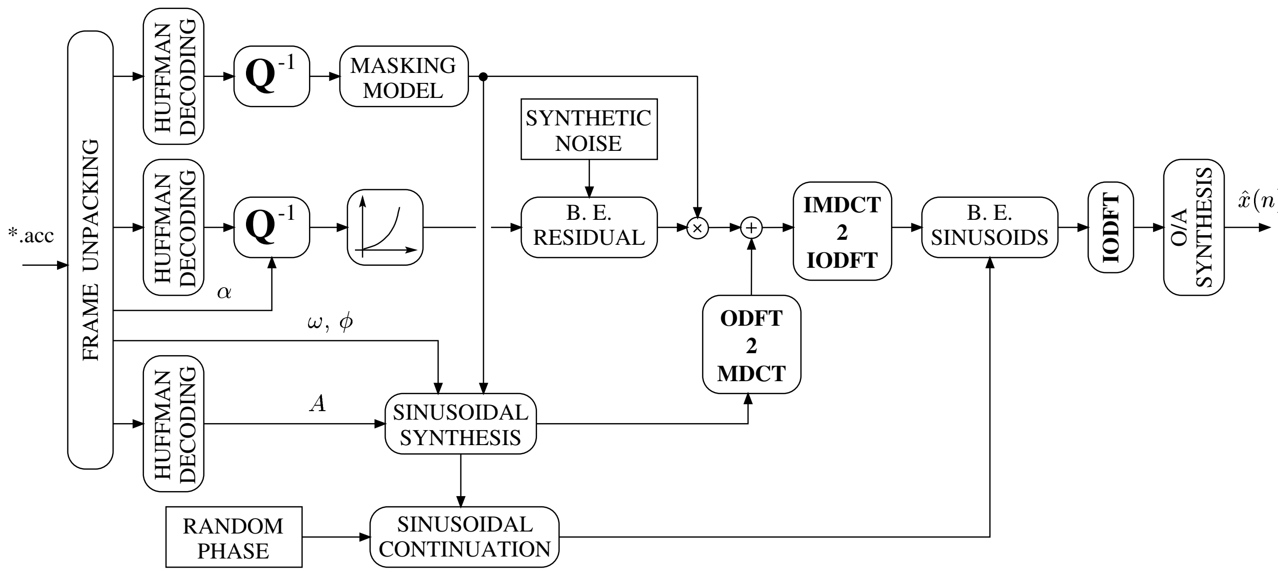

Fig. 2: Block diagram of the ACC decoder. In this diagram B.E. denotes bandwidth extension.

sent about 5% of the duration of typical audio material[2] [11]. However, the local impact on the bit rate usage may be significant as high time resolution coding may require as much as two times the number of bits used in high frequency resolution coding. Several coding schemes such as MPEG-Audio Layer 3 (MP3) deal with this problem by means of a *bit reservoir* which is a bit stream buffer allowing adjacent frames to share a common bit pool. This way, a surplus of bits may be assigned to frames coded with SHORT windows, at the cost of less bit demanding LONG windows. The negative impact of this is coding delay: the larger the bit stream buffer, the higher the coding delay.

In order to minimize the coding delay, ACC uses a residual or even no bit stream buffer at all. Under these circumstances, and so as to avoid poor coding when window switching takes place, ACC minimizes the cost of side information due to short blocks. This is achieved by transmitting to the decoder, at most two Threshold of Masking models (instead of eight) for each sequence of eight short windows. The opportunity for this lies in the fact that typically, the Threshold of Masking pertaining to short blocks in non-stationary regions of the audio signal, varies in different ways before and (over and) after the beginning of the non-stationarity (or 'attack'). In fact, most often, as illustrated in Fig. 4, it is possible to model the Threshold of Masking of short blocks lying before the signal attack using a single (average) shape. On the other hand, it is possible to model the Threshold of Masking of short blocks lying over and after the signal attack, using another single (average) shape. The specific vertical position of this shape for each block may be controlled by means of a displacement factor (D) which may be positive or negative.

The scheme illustrated in Fig. 4 requires that the start of the non-stationarity be accurately identified. ACC also uses this information to provide additional control to the insertion of short blocks. In fact, the window transition states, which are depicted in Fig. 5 are primarily controlled by a time-domain transient detector, as illustrated in Fig. 1. However, in low bit rate coding, difficulties in the time-domain aliasing cancellation between the last SHORT window and the STOP window, may give rise to noticeable artifacts. Using the Threshold of Masking information of SHORT blocks, these artifacts may conveniently be prevented by acting on the window

[2]Therefore, quasi-stationary regions represent about 95% of the duration of typical audio material.

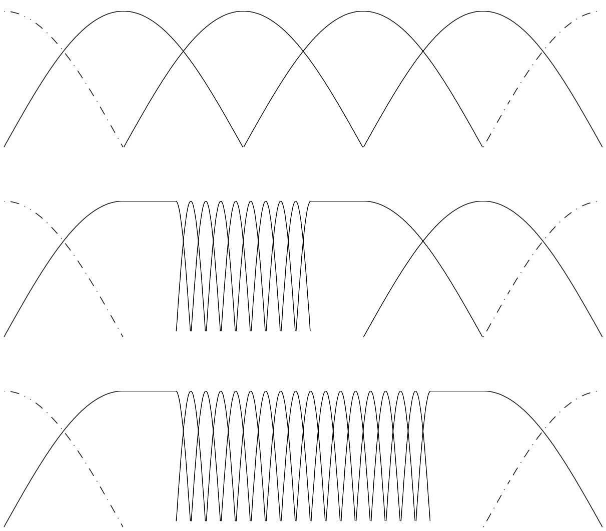

Fig. 3: Audio segmentation alternatives: normal (top) using LONG windows, and switching to SHORT windows using one sequence (middle) or two sequences (bottom) of eight high time-resolution windows. Transition constraints require the use of two special START and STOP windows.

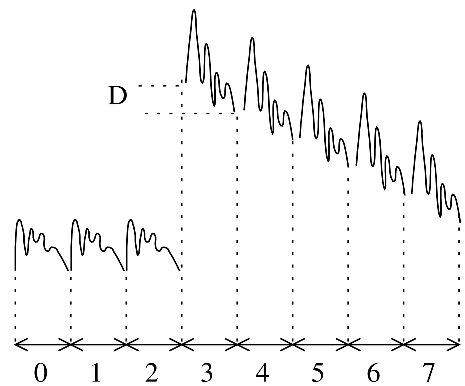

Fig. 4: Modeling of the Threshold of Masking of SHORT blocks. The illustration presumes that the signal 'attack' takes place on the fourth block.

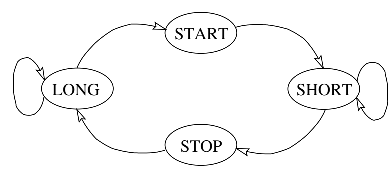

Fig. 5: Allowed transitions among different time windows.

transition mechanism so as to insert a second set of SHORT windows, as illustrated in Fig. 3.

4.3. PDF-optimized quantization

The spectral residual that is obtained after spectral subtraction and normalization, is perceptually 'white' or equally loud in frequency. In order to minimize the average power of this noise, PDF-optimized quantization is implemented through companding and uniform quantization [14]. The flattened MDCT coefficients exhibit a well defined PDF function that determines the optimum companding function [3].

4.4. Bandwidth Extension

Bandwidth extension tools allow for the replacement of a spectral region of the original signal, that has not been explicitly coded or that has been lost, by a synthetic but perceptually similar sound signal, at the cost of a very parsimonious description (and therefore bit-rate efficient) of perceptually important features such as spectral envelope and harmonicity. ACC implements bandwidth extension separately for noise and sinusoids, as described in section 3. The specific ACC bandwidth extension techniques (FSSM and ASR) have been described recently [8, 9] and are also described on another paper [2].

4.5. Huffman coding

Statistical redundancies in the quantized MDCT spectral coefficients, in the quantized cepstral coefficients of the model of the Threshold of Masking, and in the quantized differential magnitudes of sinusoids; are reduced by Huffman coding. ACC uses in total nine different tables, seven of which are specialized to code spectral coefficients (COF1-COF7), and the remaining two are specialized to code cepstral coefficients (CEPS) and sinusoidal magnitudes (SINM). Table 1 identifies all Huffman tables and characterizes their size. In complexity sensitive applications the size of those tables may be further reduced as indicated, with a minor performance penalty, taking as a reference the results presented in section 6.

As one example, Fig. 6 shows the PDF of the symbols of Huffman table COF7. This particular table

Table 1: ACC Huffman Tables. The regular size of each Table may be further reduced in complexity sensitive applications.

table	*size*	*MINsize*
COF1	81	16
COF2	625	81
COF3	81	25
COF4	169	49
COF5	361	100
COF6	625	169
COF7	289	289
CEPS	33	33
SINM	45	45

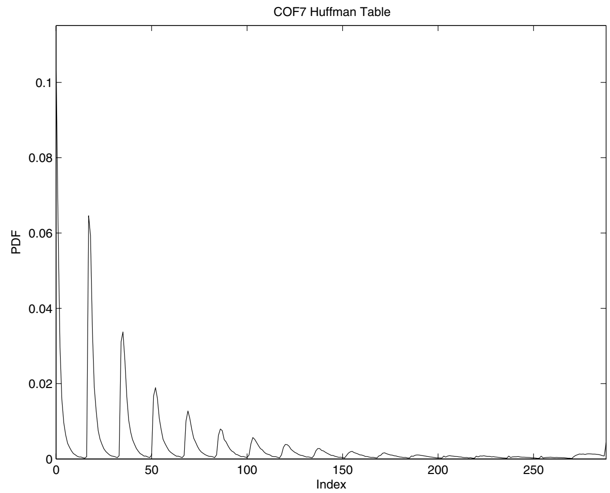

Fig. 6: PDF of Huffman codes allowing to represent sections of quantized spectral coefficients with large Maximum Absolute Value.

admits ESCape codes such that large quantized values may always be represented in the bit stream.

4.6. Bit stream structure

The ACC bit stream structure is represented in Fig. 7 and consists of four main parts. The first part has a fixed bit size and includes the following information:

- a synchronization pattern,
- the number of audio channels,
- the length of LONG windows,
- the window switching option (*i.e.*, if window switching is allowed and, in this case, the ratio of LONG/SHORT window sizes),
- the average coding bit rate,
- the sampling frequency.

The second part has a variable bit size and includes the following information:

- the window type (in case window switching is allowed);
- the assignment scheme of Huffman code books to sections of spectral coefficients.

The third part has also a variable bit size and may have two possible structures. One structure is used when the frame regards a LONG, a START or a STOP window. In this case, if existing, sinusoidal information includes the following fields:

- the number of individual sinusoids or harmonic partials,
- the frequencies of the individual sinusoids or the fundamental frequency,
- information about harmonic discontinuities if existing,
- the phases of all sinusoids,
- the magnitudes of all sinusoids,
- the cepstral coefficients.

Another structure is used when the frame regards the coding of a sequence of SHORT windows. In this case information is included regarding the short window organization, as illustrated if Fig. 4, as well as the associated cepstral data.

Finally, the fourth part of the ACC bit stream comprises the undercoding/overcoding factor (ALPHA), and Huffman codes of the quantized spectral coefficients.

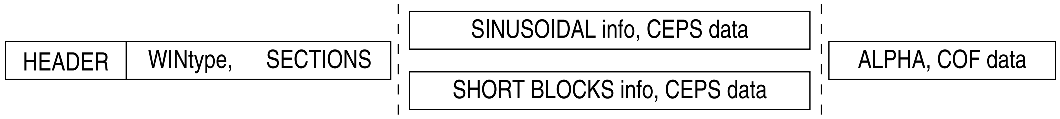

Fig. 7: ACC bit stream structure.

5. OPERATIONAL FEATURES

5.1. No interframe coding

Although interframe coding is frequently used to improve the coding gain, it has also a negative impact in several regards:

- it accentuates error propagation across several frames making more difficult the resynchronization and the communication recovery after a failure,
- it makes difficult the application of error concealment techniques in the sequence of a communication failure,
- it lowers the time resolution when accessing bit streams,
- it may inhibit fast-rewind and fast-forward play modes.

Since ACC targets real-time, two-way audio communication such as in wireless communication[3], intrinsic error robustness is required and therefore ACC does not use any scheme of interframe coding. This however has implied a careful optimization of ACC intraframe coding techniques in order to reach competitive compression ratios.

5.2. Bit stream buffer

In order to minimize the coding delay, ACC uses a residual bit stream buffer, or not buffer at all. As discussed in section 4.2, a strong argument requiring a bit stream buffer is the asymmetry in bit demand among LONG frames and SHORT frames. In order to address this problem, ACC has been designed to moderate the typical high bit demand of SHORT frames. Another strong argument requiring a bit stream buffer is related to difficulties of the rate/distortion control loop in converging to a desired constant bit rate coding[4]. As will be shown in the next section, this problem has been addressed in ACC through a careful design of the rate/distortion loop making the iterative convergence to a desirable bit rate not only smoother, but also predictable, which means faster.

5.3. Rate-distortion control

By their very nature, perceptual coders are variable bit rate coders since the Perceptual Entropy [15] of an audio signal may vary significantly with time. In fact, because perceptual coders decide the bit assignment using an estimate of the Threshold of Masking, the coding delivers constant subjective quality while requiring a variable number of bits in time. In order to force perceptual coders to operate in constant bit rate coding, a bit rate loop is implemented that iteratively adjusts the coding quality such that the bit requirement is uniform in time.

Normally, this adjustment procedure is complicated by the fact that techniques implementing overcoding (*i.e.*, using more bits to code a signal than those strictly necessary for transparent coding) may differ significantly from techniques implementing undercoding (*i.e.*, using less bits to code a signal than those required for transparent coding). In fact, the concern in the first case is to insure an appropriate safety margin, while in the second case the goal is to minimize the audibility of higher levels of quantization noise than those corresponding to transparent coding[5].)

Another complicating factor arises as a result of Huffman coding since a non-linear correspondence

[3] Which is typically influenced by a hostile communication channel.

[4] Because perceptual coders are by nature variable bit rate coders.

[5] This is also referred in the literature as a graceful degradation.

may result between the number of bits effectively used to code a given frame, and the degree or overcoding/undercoding.

Several solutions devised to implement rate/distortion control in perceptual coders are known, some of which include two iterative procedures, one for rate control, and another one for distortion control. However, to our knowledge, a complete survey of their performance is not available.

While off-line coding of audio can cope with many iterations of the rate/distortion control loop, in real-time coding a fast convergence to the target rate is absolutely required, not only to make real-time coding viable, but also to insure that the coding quality is the best possible using the available number of bits.

ACC addresses these concerns by controlling rate and distortion simultaneously, in an iterative procedure using a single quality parameter (ALPHA), that is transmitted to the decoder as side information [16], and that decides the degree of overcoding or undercoding. The encoder operates at constant quality coding (and variable rate coding) if that parameter is constant in time, or operates at constant bit rate coding (and variable coding quality) if that parameter is made to vary in time in a suitable way.

The relation between the value of ALPHA and the resulting bit rate and coding quality, is discussed in the next section. It will be shown for example that the relation is quite smooth which helps the convergence to the desired bit rate. Furthermore, it has been concluded that an appropriate start of the iterative process is quite helpful in terms of speed of convergence and efficient use of the available bits. Also in this regard, the smoothness of that relation is very convenient.

The ACC iterative procedure of rate/distortion control has also been optimized to avoid 'dead locks' and to overcome non-linear convergence trends. A study on the convergence speed has been carried out using a large audio file with different types of audio material. Table 2 presents the results obtained for different bit rates. These results depend strongly on the departing value of α (ALPHA) and also include the quantization effect on the value of α.

Table 2: Average (AVGiter) and maximum (MAXiter) number of iterations of the rate/distortion loop when coding at constant bit rate. The initial α is also indicated.

bit rate	*AVGiter*	*MAXiter*	*initial* α
32 kbit/s	2.30	8	-15
48 kbit/s	2.70	8	-13
64 kbit/s	2.23	7	-10
80 kbit/s	1.89	9	-7

6. PERFORMANCE EVALUATION

Evaluating the performance of an audio compression scheme implies taking into consideration not only the relation of coding quality *versus* bit rate, but also other structural and operational aspects such as algorithmic complexity, coding delay, and intrinsic error robustness (*e.g.*, if interframe coding is allowed or not). Therefore, evaluating different coding schemes using only coding quality as a function of bit rate, is not an honest exercise and can give rise to misleading conclusions[6].

Regarding coding delay, ACC has an algorithmic delay of about 21 ms at 48 kHz sampling frequency. Considering look-ahead (for transient detection) and framing, the ACC practical coding delay is less than 50 ms, either for 48 kHz, 44.1 kHz, or 32 kHz sampling frequency.

Regarding coding quality, our evaluation has concentrated on subjective quality, but also on objective quality in order for example to study and express objectively the impact of ALPHA, or the difference between CBR and VBR coding. The objective quality assessment has been performed using the procedure illustrated in Fig. 8. This procedure allows to conclude on the Signal to Coding Noise Ratio (SNR). Using this procedure, we have studied the coding quality as a function of ALPHA, under VRB coding. For each test, the value of ALPHA was kept fixed, the allowed coded bandwidth was 15 kHz, and no bandwidth extension was used. As audio material, we have used a large audio file sampled at 32 kHz and including excerpts of *harpsichord,*

[6]For this reason it does make sense for example to compare MP3 and AC-3 using only coding quality as a function of bit rate.

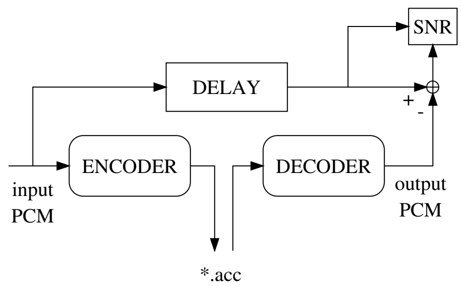

Fig. 8: Procedure allowing to evaluate the Signal to Coding Noise Ratio (SNR).

castanets, Suzanne Vega, Sting, male speech, female speech, pitch pipe and acoustic guitar solo. Results are depicted in Fig. 9. This figure reveals that

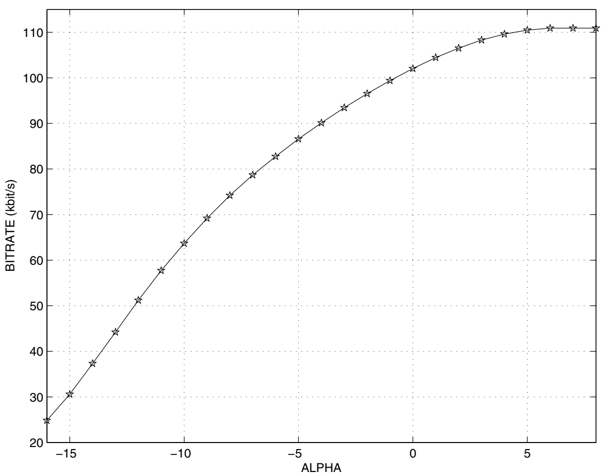

Fig. 9: Bit Rate as a function of ALPHA, in VBR coding.

there is a very smooth proportionality in the relation between bit rate and the rate/distortion parameter ALPHA. The subjective quality also increases with bit rate. Transparent quality is achieved for the majority of audio excerpts at bit rates between 60 kbit/s and 80 kbit/s.

While performing the tests just described, the SNR has also been computed for each value of ALPHA. Results are represented in Fig. 10. Interestingly, this figure shows a trend that is similar to that of Fig. 9, with the difference that the saturation effect starts a little bit earlier. This is a consequence of

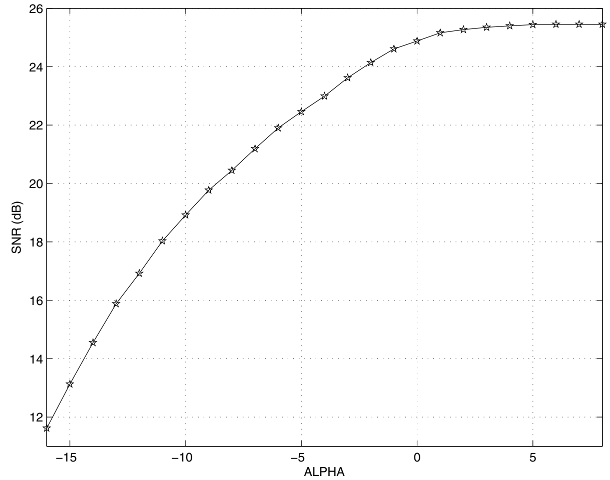

Fig. 10: SNR of coded audio material, as a function of ALPHA, in VBR coding.

the fact that the degree fo overcoding, which affects mainly the high frequency content, leads to an increase of the required bit rate, but does not improve significantly the SNR given that the power spectral density of the high frequency region is rather low. The knee of the saturation effect is determined by the allowed degree of overcoding after companding and before the quantization. This aspect has been calibrated in ACC given the expected range of operation. It may however be displaced further to the right in case higher bit rates are tolerated.

By combining Fig. 9 and Fig. 10, a new figure is obtained that relates SNR with bit rate under constant quality coding (or VBR coding). This figure is represented in Fig. 11. A new set of tests using the same large audio file was carried out with ACC configured as CBR coding (and therefore variable quality coding). Each frame was allowed to use unused bits from the previous frame. This is synonymous with allowing a residual bit stream buffer that however does not impact significantly on the coding delay.

Only those bit rates that are allowed in the bit stream were used. In order to obtain results that can be compared to those obtained previously (in VBR coding), the coded bandwidth was set as 15 kHz for all bit rates, and no bandwidth extension was used. The resulting SNR versus bit rate curve

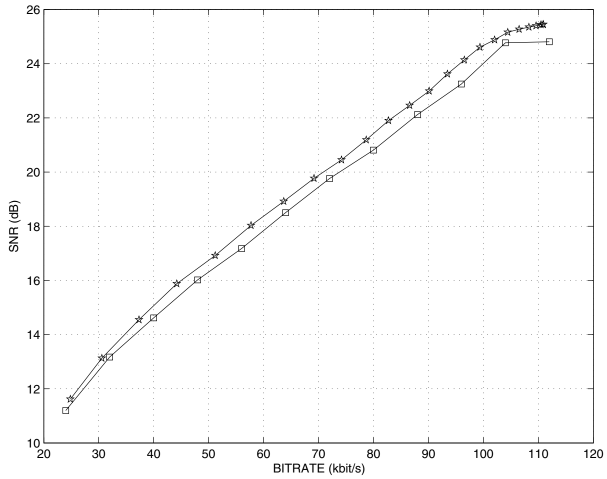

Fig. 11: SNR of coded audio material as a function of coding bit rate, in VBR coding (solid line with pentagram symbols) and in CBR coding (solid line with square symbols).

is also represented in Fig. 11. The value of ALPHA used at the beginning in the rate/distortion iteration loop, was obtained from the results in Fig. 9.

Objectively, the results in Fig. 11 indicate that the quality obtained under CBR or VBR and for comparable average bit rates, is similar, although there is a small advantage for the VBR results. This also indicates that the ACC design and efficient intraframe coding techniques make that the audio quality in CBR coding is not strongly dependent on a large bit stream buffer.

Subjectively, for similar bit rates the audio quality under CBR or VBR coding does not differ significantly, even for an expert listener.

Audio demos that are illustrative of the ACC coding quality under CBR or VBR coding, are available at http://www.atc-labs.com/acc .

7. CONCLUSION

We have presented the structure, described the operation and characterized the performance of a new source/perceptual audio communication coder that has been recently optimized. An emphasis has been placed on those structural and operational features that make it suitable for real-time, two-way high quality audio communication. Other possible application scenarios include Music over IP (MoIP), audio conferencing, mobile audio communication, and B-channel audio communication. Audio demos are available at http://www.atc-labs.com/acc . Due to the use of powerful parametric coding techniques and no interframe coding, ACC may also easily pave the way for the realization of special effects such as gender transformation in the case of voice or singing, and such as fast-rewind and fast-forward play modes.

8. REFERENCES

[1] Aníbal J. S. Ferreira and Deepen Sinha, “A New Low-Delay Codec for Two-Way High-Quality Audio Communication,” *119th Convention of the Audio Engineering Society*, October 2005, Paper 6572.

[2] Deepen Sinha and Aníbal J. S. Ferreira, “A Novel Integrated Audio Bandwidth Extension Toolkit (ABET),” *120th Convention of the Audio Engineering Society*, May 2006, Paris, France.

[3] Aníbal J. S. Ferreira, “Optimum Quantization of Flattned MDCT Coefficients,” *115th Convention of the Audio Engineering Society*, October 2003, Paper 5988.

[4] Aníbal J. S. Ferreira, “Combined Spectral Envelope Normalization and Subtraction of Sinusoidal Components in the ODFT and MDCT Frequency Domains,” in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 21-24 2001, pp. 51–54.

[5] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, 1992.

[6] Aníbal J. S. Ferreira, “Perceptual Coding Using Sinusoidal Modeling in the MDCT Domain,” *112th Convention of the Audio Engineering Society*, May 2002, Paper 5569.

[7] J. P. Princen, A. W. Johnson, and A. B. Bradley, “Subband/Transform Coding Using Filter Bank Designs Based on Time Domain

Alias Cancellation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1987, pp. 2161–2164.

[8] Deepen Sinha, Aníbal J. S. Ferreira, and Deep Sen, "A Fractal Self-Similarity Model for the Spectral Representation of Audio Signals," *118th Convention of the Audio Engineering Society*, May 2005, Paper 6467.

[9] Aníbal J. S. Ferreira and Deepen Sinha, "Accurate Spectral Replacement," *118th Convention of the Audio Engineering Society*, May 2005, Paper 6383.

[10] Aníbal Ferreira and Deepen Sinha, "Accurate and Robust Frequency Estimation in the ODFT Domain," in *2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 16-19 2005, submission accepted.

[11] Aníbal J. S. Ferreira, *Spectral Coding and Post-Processing of High Quality Audio*, Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto-Portugal, 1998, http://telecom.inescn.pt/doc/phd_en.html.

[12] Aníbal J. S. Ferreira, "Accurate Estimation in the ODFT Domain of the Frequency, Phase and Magnitude of Stationary Sinusoids," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 47–50.

[13] Aníbal J. S. Ferreira, "Combined Spectral Envelope Normalization and Subtraction of Sinusoidal Components in the ODFT and MDCT Frequency Domains," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 51–54.

[14] N. S. Jayant and Peter Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.

[15] J. D. Johnston, "Method of Estimating the Perceptual Entropy of an Audio Signal," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1988.

[16] Aníbal J. S. Ferreira, "Efficient Intraframe Coding of Monophonic Audio," *116th Convention of the Audio Engineering Society*, May 2004, Paper 6166.