A Nearly-transparent Low Delay Audio Coder

D. MARTÍNEZ MUÑOZ (*); F. LÓPEZ FERRERAS (**); M. ROSA-ZURERA (**);
N. RUÍZ-REYES (*)
(*) Departamento de Electrónica
Universidad de Jaén
C/ Alfonso X El Sabio, 28; 23700 Linares
SPAIN

(**) Departamento de Teoría de la Señal y Comunicaciones
Universidad de Alcalá
SPAIN

Abstract: - Nowadays, it does not exist an audio coding standard for getting nearly-transparent quality with low delay. The standard ISO-MPEG is profusely used in audio for getting high quality [1]. It uses a perceptual model that requires high frequency precision and introduces high delay. However, in applications where the delay is a critical parameter, e.g. when is important to use a feedback channel, the ISO-MPEG is not adequate and there is not a standard for this case.

This paper describes the different steps followed for the obtainment of an audio coder with nearly-transparent quality and low delay. The complexity was another design criterion because the final objective is the real time implementation of the proposed coder on a commercial DSP.

A comparative analysis of low delay codification schemes is done (ADPCM, Subband, and Subband-ADPCM coding). For each technique we modify its different parameters (e.g. numbers of bits of the quantizer, number of bands, bit allocations criterion, aliasing filters, etc) in order to get the best configuration.

The proposed coder uses a Subband-ADPCM hybrid structure with dynamic bit allocation. In such sense, we propose a very low complexity bit allocation strategy with similar results to the classic strategy [2][3] and no delay.

The coder output is scalable from high quality at higher bit-rate to lower quality at lower bit rates, supporting a wide range of service and resource utilization. Nearly-transparent quality is accomplished with 2.5 bits/sample. The less than 2 ms algorithm delay is due only to the perfect reconstruction filter bank [4].

Key-Words: - Subband, ADPCM, bit allocation, audio coding (low delay), adaptive quantizer Proc.pp..2081-2085

1 Introduction

In last years a lot of work has been dedicated to the difficult issue of data compression either for video or for audio. Consequently, different standards have appeared. The most important are the ISO-MPEG algorithms for video and audio. For audio, the ISO-MPEG coders apply a perceptual model that introduces high delay. The delay in some applications may be a critical requirement. In this case there is not a standard that offers high quality and low delay. In this paper we present a nearly-transparent low delay audio coder.

The plan of the paper is organized as follows:

a) In section 2 we introduce the standards used in the commercial coders.
b) The tests performed with different codification techniques are described in section 3. The results for each evaluated technique will be commented.
c) The structure of the proposed coder is presented in section 4, as well as a bitstream framing suggested to transmit the information to the decoder.
d) Finally, in section 5, we outline the conclusions and future works in order to improve the obtained coder.

2 Commercial Audiocoders

Currently, coders that we can find in the market are multinorm: they permit apply several audio coding standards according to the requirement of our application.

These standards, depending on the required quality are [5]:

a) Telephone quality (frequency range from 300 until 3.400 Hz):

- G.711: used from the 60's. It is a PCM (Pulse

Code Modulation) coder. The output bit rate is 64 Kbps (8 bit/sample). The delay is practically void. The PCM is the simplest coding system.

- G.721: apply an ADPCM (Adaptive Differential PCM) coding. The output bit rate is 32 Kbps. This algorithm has been extended to 24 and 40 Kbps in the standard G.723. As in the previous case, the delay is void.
- G.728: correspond to a low delay CELP coder. The output bit rate is 16 Kbps (2 bit/sample). The delay is only 2.5 ms.

b) Quality AM (frequency range until 7 Khz):

- G.722: it based on a two-band sub-band coder, with ADPCM coding of each subband. The low and high frequency subbands are quantized using 6 and 2 bits per sample, respectively. The filter banks used for analysis and synthesis achieve a communications delay of about 1.5 ms. The output bit rate is 64 Kbps.
- ISO-MPEG2 Layer-2: applies a subband decomposition and dynamic bit allocation. The delay is very high (upper to 150 milliseconds). The bit rate is variable from 8 until 160 Kbps.

c) Quality FM (frequency range until 15 Khz) and compact Disc (CD, until 20 Khz):

- ISO-MPEG1 Layer-2: as in the case of ISO-MPEG2, uses a subband scheme and dynamic bit allocation. The final bit rate is variable from 32 until 384 Kbps. It codifies 1 or 2 audio channels in mono, dual, stereo or joint-stereo mode. The codification delay is also very high.
- ISO-MPEG1 Layer-3: makes use of a transform coding technique and a perceptual model for bit allocation. Also, it exhibits a high codification delay, although inferior to the previous.

Besides of the previously mentioned standards, there exists another commercial coder known as APTX. The company APT is owner of this coder and it does not supply information on this. If we were interested in using it, it is necessary to buy an integrated circuit that provides the company. This coder requires 4 bits for sample, being the delay very small.

As we have seen previously, in applications where the delay is a critical factor, the existing standards do not offer a high quality, remaining the frequency range limited to 7 Khz and requiring 4 bits/sample (G.722 algorithm).

Our purpose here is to find a low delay audiocoder that covers this lack. It must cover a frequency range ≥ 15 Khz and require less than 4 bit/sample.

3 Accomplished Tests

To achieve a low delay coder, a series of codification techniques have to be rejected. Between these, we have:

- Wavelet transform or wavelet packet, because the different decomposition levels suppose an excessive delay for the labeled objective.
- Application of perceptual models since these techniques require a considerable frequency resolution that implies using of high FFT block size.

The techniques here considered are:

- ADPCM coding, according to the G.721 algorithm, using quantizers with different bit number. Also, we use a 6 cells lattice prediction.
- Subband coding and dynamic bit allocation. It has been tested for different subband number, aliasing filters and dynamic bit allocation strategies.
- Subband-ADPCM coding. As in the previous case, it has been evaluated for different band number and bit allocation strategies.

For evaluation purpose, we have used 4 audio files, corresponding to cord instruments (cord), wind instruments (wind), piano note (piano) and vocal passage (vocal). The sampling frequency used is 44.1 Khz.

We have used the segmental SNR to measure the objective quality.

Herein after we present the results obtained with each one from the previously mentioned techniques.

3.1 ADPCM Coding

As already it has been referred, we have used two types of ADPCM coders:

- According to the G.721 algorithm.
- With lattice prediction.

The results for both cases are similar, resulting that it does not exist a far cry between both coding schemes.

It has been used a non uniform scalar quantizer. Prior to quantization, variance estimation is used to scale the error prediction. Their adjustment is backward. The quantization levels are matched to a Gaussian distribution.

The subjective quality obtained with 2 bit/sample is very poor. If the bit number is increased to three, the quality improvement is very important with respect to the two bits case (11 dB on average). Anytime one bit is added to the quantizer, the quality improvement reduces and tends gradually to 6 dB. This is explained considering that the improvement reason is not only the bit increase in the quantizer, but also the better performance of the prediction and variance estimation stages. Once it has been obtained a certain quality, these stages do not suffer any significant improvement. The improvement obtained when increasing one bit for each audio file is shown graphically in figure 1.

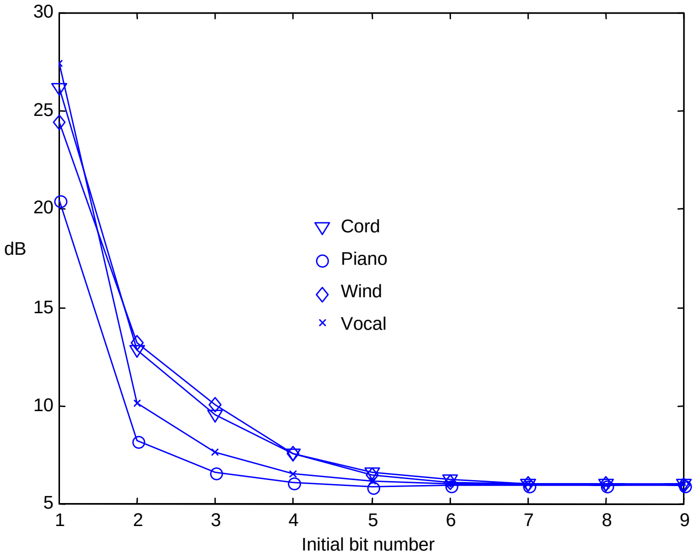

Figure 1. Improvement for increasing one bit.

For one bit quantizer, the segmental SNR obtained is practically zero because predictor adaptation and variance estimation are very bad. As is made clear by figure 1, using an ADPCM scheme, we need around 5-6 bit/sample for maximally profit this structure.

3.2 Subband Coding

We have used perfect reconstruction filters with 2, 4, 6, 8, 12, 16 and 32 bands. Also, for a fixed number of bands, we have proved filter banks with different taps and aliasing levels.

The quantizer used is the same that in ADPCM.

In order to minimize the delay and reduce the bit overhead, our bit allocation strategy has been accomplished each 256 samples, maintaining the obtained bit allocation for the following 256 samples. The achieved results are not far from the obtained for a one-sample bit allocation.

It has been used a simple but effective allocation criterion for reducing the operations number. Compared to the classic assignment criteria [2][3], the results are practically similar. The bit allocation criterion employs the variance estimation used for the subband quantizer and applies the following steps:

- For each subband, it calculates the variance estimation in dB and reset the bit counter.
- It applies the following algorithm:

  1. Assign one bit to the band with maximum variance, add one to the bit counter and rest one to the total number of bits to distribute.
  2. For the band selected in the previous step, modify, subtracting 6, the variance estimation value (in dB).
  3. Verify if there are some bits left to assign. If it is true, returns to step 1. Otherwise, it concludes the allocation process.

The implementation of this algorithm on a fixed point DSP is relatively simple comparing it to the classical bit allocation criteria that require calculations (divisions and geometric average) with high computational cost.

It has been accomplished tests varying the number of bands. On average, the best results have been obtained for 8 bands.

To obtain nearly-transparent quality was necessary to assign an amount of 3-4 bits for sample.

3.3 Subband-ADPCM Coding

This scheme is achieved combining the two previous codification techniques: for each subband output there is an ADPCM coder.

It has been used an ADPCM coder according to the G.721 algorithm.

In this hybrid structure, all previously tested parameters for the individuals schemes have been proved, e.g. total bit number, number of bands (2, 4, 6, 8, 12, 16 and 32 bands), tap number, aliasing filter, bit allocation strategies.

As conclusion of all the simulations done, we propose a coder that is described in next section.

4 The Proposed Coder

In this paper we present a coder that responds to a hybrid structure subband-ADPCM. In order to get nearly transparent quality a rate of 2.5 bit/sample is required. The encoder, shown in figure 2, consists of four main components: analysis filter bank, bit allocation, ADPCM coder, and bitstream formatter.

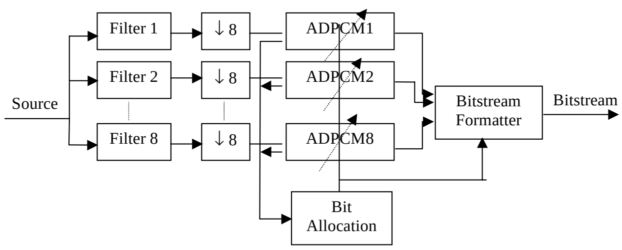

Fig. 2. Block diagram of subband-ADPCM encoder.

4.1 Filter Bank

We use an eight-band 64-tap finite impulse response (FIR) perfect reconstruction filter bank. Figure 3 shows its frequency response.

The complexity of the filtering operation is approximately 2.8M operations/s. The filter bank introduces a delay of less than 1.5 ms. It provides around 50 dB of stopband attenuation. This poor stopband attenuation is a handicap in schemes that apply a perceptual model because it can become a significant source of perceptual distortion due to aliasing of quantization noise. However, in our scheme the stopband attenuation is not a critical parameter.

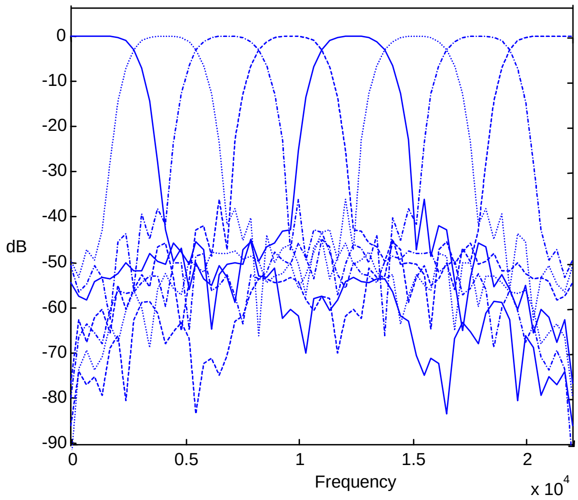

Figure 3. Filter Bank frequency response.

4.2 Bit Allocation

The dynamic bit allocation determines how many bits are allocated to quantize the prediction error in the ADPCM coding of each subband. The total bit number, N, is established according to the available resources. The output bit rate due only to the quantizers is:

$$R_b = N \times \frac{f_s}{8} \qquad (1)$$

Changing the total bit number, the bit rate and quality is modified.

The bit allocation algorithm has been already described in the subband coding section. It is applied every 256 samples and its allocation result is used for the next 256 input samples (32 subband samples). The information relative to this 256 input samples is conveyed using a bitstream frame. This bit allocation strategy does not introduce delay. The results obtained with this procedure, as in the subband coding case, are very similar to the achieved with a bit allocation made sample to sample.

4.3 ADPCM Coder

The ADPCM coder is displayed in figure 4. It corresponds to the G.721 algorithm. The prediction error is quantized in two stages: first is normalized using the root of the variance value, then by quantizing the normalized prediction error. The statistical distribution of the normalized prediction error closely matches a normalized Gaussian distribution. The normalized prediction error is then quantized using a Max-Lloyd [6] quantizer matched to a Gaussian distribution. The bit number of the quantizer that corresponds to the obtained in the bit allocation block, is changed every 32 subband samples.

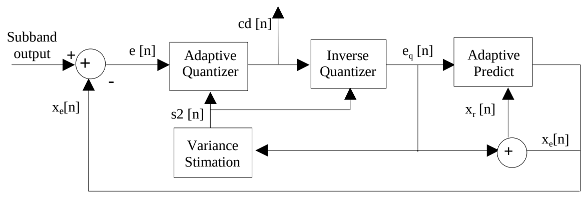

Figure 4. Block diagram of the ADPCM coder.

The variance estimation, $s2[n]$, is calculated every subband sample, in the way:

$$s2[n] = (1 - ALPHA) * s2[n-1] + ALPHA * e_q[n]^2 \qquad (2)$$

The ALPHA value used is 0.03125.

4.4 Bitstream formatter

Figure 5 shows the structure of the proposed bitstream framing. The overhead consists of the alignment word, the subband bit allocation and one parity bit.

The alignment word is a seven-bit sequence and is used for frame synchronization in the reception side.

The bit allocation information consists of 24 bits, three per subband. This is the most important information in each frame.

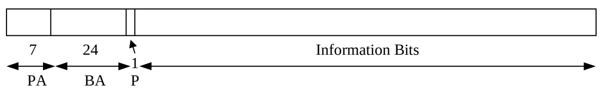

Figure 5. Bitstream frame structure.

The parity bit is calculated over the bit allocation section. It permits to detect an odd error number over these bits. If an error is detected, the last correct frame is repeated. Channel coding techniques more sophisticated, as interleaving, are not used because of the delay increment.

The prediction error of each subband is quantized using the quantization process described in section 4.3 and the bits resulted are conveyed in the frame information section.

The output bit rate is proportional to the total bit number, N, in the way

$$Bit\ rate = N \times \frac{f_s}{8} + 32 \times \frac{f_s}{256} = (N+1) \times \frac{f_s}{8} \qquad (3)$$

4.5 Results Obtained

Table 1 shows the segmental SNR obtained for each one of the audio files. We compare the Greedy bit allocation algorithm [2] with our bit allocation strategy in the case of sample to sample allocation and allocation every 256 samples. The total bit number, *N*, is 20 bits corresponding to a rate of 2.5 bits for sample.

	Cord	Piano	Wind	Vocal
Greedy S/S	41.18	27.75	41.44	36.75
Our S/S	41.20	27.77	41.45	36.78
Greedy 256	41.16	27.79	41.42	36.84
Our 256	41.16	27.71	41.46	36.78

Table 1. Segmental SNR for different allocation strategies.

This table concludes that both bit allocation algorithms are virtually identical in results. Informal listening tests determine that for a rate of 2.5 bits/sample the achieved quality is transparent.

5 Conclusions and future works

A low delay audio coder is presented with the following characteristics:

- Subband-ADPCM coding.
- ADPCM type G.721.
- Filter bank of 8 bands.
- Flexible bit rate according to the service or resources.

A low complexity bit allocation strategy is used. The total delay is less than 2 milliseconds. The complexity can be assumed by a commercial fixed point DSP.

Future works must be oriented in two ways:

a) To perform more simulations at high level using new techniques as vectorial quantization and

b) implementation in real time on a DSP of the proposed coder.

References:

[1] Brandenburg, Stoll: *"ISO-MPEG-1 audio: A generic standard for coding of high-quality digital audio"*. J. Audio Eng.Soc. Oct. 1994.

[2] Gersho, Gray: *"Vector Quantization And Signal Compression"*, Kluwer Academic Publishers, 1992

[3] Jayant, Noll: *"Digital Coding of the Waveform"*. Prentice-Hall, 1984.

[4] Vaidyanathan, P. P.: *"Multirate Systems and Filter Banks"*. New Jersey, Prentice-Hall, 1993.

[5] N. Jayant, *"Signal compression: Technology targets and research directions"*, IEEE J. Select. Topics Commun., vol. 10, pp. 796-818, June 1992.

[6] J. Max, *"Quantizing for minimal distortion"*, IRE Trans. Inform. Theory, vol. IT-6, pp. 7-12, Mar. 1960.