LOW DELAY AUDIO CODER USING ADAPTIVE VECTOR QUANTIZATION

D. Martínez Muñoz¹, M. Rosa Zurera², F. López Ferreras² and N. Ruiz Reyes¹
1 Departamento de Electrónica, Universidad de Jaén
C/ Alfonso X El Sabio, 28; 23700 Linares, SPAIN
e-mail address: damian@ujaen.es
2 Departamento de Teoría de la Señal y Comunicaciones, Universidad de Alcalá, SPAIN

ABSTRACT

Nowadays, there is not a low delay audio coding standard that achieves nearly-transparent quality. The ISO-MPEG standards are profusely used in audio for high quality coding. They use perceptual models that require high frequency resolution. In applications where the delay is a critical parameter, i.e. when the use of a feedback channel is important, the ISO-MPEG algorithms are not suitable. In this paper we present a new coder that responds to a subband ADPCM structure and incorporates adaptive vector quantization, requiring only 2 bits/sample. Each vector component (prediction error) is weighted according to its estimated standard deviation. The introduced coding delay is less than 2 milliseconds, and is only due to the perfect reconstruction filter bank. In order to obtain the best configuration, results have been obtained varying the number of bands, the number of bits in the quantizer and the analysis filters.

1 INTRODUCTION

The difficult issue of data compression for both video and audio has justified an important research effort. In consequence of this, different standards have appeared. The most important are the ISO-MPEG algorithms [1] for video and audio. For audio, the coders included in the ISO-MPEG standard apply perceptual models that introduce high-delay. The delay in some applications may be a critical requirement. Right now there is not a standard that offers high-quality and low-delay.

In a previous work [2], we have presented a low delay audio coder using a Subband-ADPCM hybrid structure which only requires 2.5 bits/sample for nearly-transparent coding. In this paper we present an improved version that only requires around 2 bits/sample. The paper is organized as follows:
1. In section 2, a discussion about the possibility of implementing a low-delay audio coder using different schemes is included.
2. The structure of the proposed coder is presented in section 3.
3. Finally, in section 4, we outline the conclusions and future works in order to improve the coder here presented.

2 FIRST APPROACH

To achieve a low delay audio coder, different coding techniques have been revised. Wavelet-transform based coders have been rejected, due to the exponential increase of the delay with the analysis filter-bank depth. In the same way, perceptual coders can not be used. Psycho-acoustic models need high-frequency resolution and so, the frame length must be high. This gives rise to high delays.

The techniques here considered are:
• Subband coding [3]. It has been tested for different number of subbands and filters.
• Subband-ADPCM coding. As in the previous case, it has been evaluated for different number of bands and filters.

We use adaptive vector quantization anyway. The adaptation strategy will be described in section 3.3. For evaluation purpose, we have used 4 audio files, corresponding to string instruments (string), wind instruments (wind), piano note (piano) and vocal passage (vocal). The sampling frequencies considered in this work are 32 and 44.1 kHz. We have used segmental SNR to measure the objective quality.

2.1 Subband Coding

Focusing on a scheme that only uses subband coding, we have look for the best number of bands in order to obtain the best subjective quality. We have used perfect reconstruction filter-banks and the number of bands has varied from 2 to 32 bands. Also, for a fixed number of bands, different kings of filters have been tested, with different number of taps and so, different aliasing levels.

The best subjective quality has been obtained with only 8 bands, because when the number of bands is increased, the inter-band aliasing is also increased, giving rise to lower quality if no psycho-acoustic information is considered. Another important result is that for a higher number of bands, the segmental SNR increases but the subjective quality decreases.

2.2 Subband-ADPCM Coding

In order to improve the results obtained with subband
coding, an ADPCM coder has been included to code the resulting subbands. A predictor according to the I.T.U. (International Telecommunication Union) recommendation G.721 with 2 poles and 6 zeros has been used. For this hybrid structure, all previously tested parameters in the Subband scheme have been checked again (number of bands and number of taps in the filter). The results that have been obtained allow extracting the following conclusions:

- As in the Subband case, when the number of subbands is increased, the segmental SNR also increases but the subjective quality decreases when the number of bands is higher than 8.
- Subband/ADPCM improves the obtained results for the Subband case. This improvement is more significant when the considered audio file incorporates vocal component.
- If the number of bands is increased, the improvement in segmental SNR is higher for the hybrid scheme of Subband/ADPCM than for the scheme where only Subband decomposition is used.

### 3 PROPOSED CODER

According to the preliminary results discussed in section 2, the hybrid approach subband-ADPCM has been selected. In order to achieve nearly transparent quality a binary rate of 2 bit/sample is required. The encoder, shown in figure 1, consists of three main components: analysis filter bank, ADPCM coders, and adaptive vector quantizer.

![Figure 1. Block diagram of the hybrid Subband-ADPCM encoder.](image)

**3.1 Filter Bank**

We use a perfect reconstruction filter bank with only eight bands. Finite impulse-response (FIR) filters are used, each one with 64 taps. The complexity of the filtering operation is approximately 2.8 MOPS and the introduced delay is less than 1.5 milliseconds when the sampling frequency is 44.1 kHz. With this filter-bank around 50 dB of stopband attenuation is provided. This poor stopband attenuation is a drawback because it can become a significant source of perceptual distortion due to aliasing of quantization noise. Further improvement is necessary to achieve higher stopband attenuation that will provide the same performance at a lower bit rate.

![Figure 2. Filter Bank magnitude frequency response.](image)

**3.2 ADPCM Coder**

For each subband there is an ADPCM coder. Its block diagram is displayed in figure 3. The predictor corresponds to the G.721 algorithm. The prediction error ($e_q[n]$) in each subband is considered a component of the vector formed for vectorial quantization (VQ).

![Figure 3. Block diagram of the ADPCM coder.](image)

An estimation of the variance ($s_2[n]$) of the quantized prediction error ($e_q[n]$) is also included for each subband, as follows:

$$s_2[n] = (1 - \text{ALPHA}) * s_2[n-1] + \text{ALPHA} * e_q[n]^2$$

The used value for ALPHA is 0.03125. The value of $s_2[n]$ is used as a weighting factor in the VQ process.

**3.3 Vector Quantization**

Vector quantization represents an extension of conventional scalar quantization. Each input vector can be viewed as a point in an $M$-dimensional space. The quantizer is defined by a partition of this space into a set of non-overlapping volumes. The task of the vector quantizer is to determine the volume in which an input
A vector quantizer is defined with two tasks [4]:

a) Code design for performing the multidimensional volume partition.

b) Searching for the particular volume which corresponds to the best description of the source according to some fidelity criterion.

In our case, the code design have been made using two strategy:

- The LBG (Linde, Buzo and Gray) algorithm [5][6].
- Codebook design according to a predefined probability density function.

A study of the statistical distribution of the vector components (subband output in the Subband coder and error prediction for Subband/ADPCM coder) has been made. Previous to vector quantization we normalize the vector components using the estimation of the standard deviation.

In figure 4 we present the histograms corresponding to each one of the normalized subband output for a subband coder with 4 band considering an audio file input with prevalence of vocal component.

![Figure 4. Normalized subband output histograms for a Sub-band coder.](image)

As we can appreciate, the histograms approximate quite well to a normalized Gaussian distribution (continue line). Figure 5 represents the histograms corresponding to the normalized prediction error for a sub-band/ADPCM coder using the same filter-bank and audio file mentioned above.

![Figure 5. Histogram for the normalized prediction error for a Sub-band/ADPCM coder.](image)

As in the previous case, the histograms approximate a Gaussian distribution. The results don’t vary very much with the input audio file or with the number of bands. These results justify the use of the Gaussian distribution for designing the codebook.

Tests performed using a codebook obtained applying the LBG algorithm conclude that the improvement in quality using a LBG codebook instead of a random codebook designed according to a predefined Gaussian probability density function is around 2-4 dB. However, the last codebook is not specific and can be used for any audio file. That is why we use the last method of codebook design.

The measure or the distance between a code vector, \( c_k \), and the normalized prediction error vector is obtained as follows:

\[
\text{Dist}_k = \sum_{i=1}^{K} s^2_i \cdot |e_i(n) - c_{k,i}|^2
\]

In this way, the distance measure is calculated giving a greater weight to the higher power subband.

In our proposed coder, for each subband output, \( i \), the prediction error, \( e_i[n] \), is normalized using the square root of the variance estimation, \( s^2_i[n] \). The overall normalized prediction error forms a vector that is quantized using the codebook defined above.

In reception, it is not necessary to introduce additional information if we use a statistical recovering framing strategy. Therefore, the output bit rate (\( R_b \)) is only due to the number of bits (\( N \)) used to represent each centroid:

\[
R_b = N \times \frac{f_s}{8}
\]

3.4 Results

Tables 1 and 2 show the segmental SNR obtained for each one of the audio files. Sampling frequencies of 32 and 44.1 kHz have been used. In these table we compare the Subband structure with the Subband-ADPCM one corresponding to the proposed coder. We also include results using scalar quantization and dynamic bit allocation applying an optimum criterion [4][6]. The output bit rate is always 2 bits/sample, so 16 bits are used to represent each subband vector.
<table>
<thead>
<tr>
<th></th>
<th>String</th>
<th>Piano</th>
<th>Wind</th>
<th>Vocal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subband BA</td>
<td>31.12</td>
<td>24.87</td>
<td>31.86</td>
<td>30.5</td>
</tr>
<tr>
<td>Subband VQ</td>
<td>36.72</td>
<td>25.85</td>
<td>35.05</td>
<td>34.46</td>
</tr>
<tr>
<td>SB/ADPCM BA</td>
<td>37.17</td>
<td>25.72</td>
<td>32.23</td>
<td>35.67</td>
</tr>
<tr>
<td>SB/ADPCM VQ</td>
<td>38.69</td>
<td>27.25</td>
<td>37.44</td>
<td>37.13</td>
</tr>
</tbody>
</table>

Table 1. Segmental SNR for different techniques (fs=44.1 Khz).

<table>
<thead>
<tr>
<th></th>
<th>String</th>
<th>Piano</th>
<th>Wind</th>
<th>Vocal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subband BA</td>
<td>31.29</td>
<td>22.87</td>
<td>29.62</td>
<td>24.68</td>
</tr>
<tr>
<td>Subband VQ</td>
<td>31.92</td>
<td>24.30</td>
<td>29.52</td>
<td>31.35</td>
</tr>
<tr>
<td>SB/ADPCM BA</td>
<td>31.94</td>
<td>23.37</td>
<td>30.65</td>
<td>24.74</td>
</tr>
<tr>
<td>SB/ADPCM VQ</td>
<td>33.34</td>
<td>25.36</td>
<td>31.62</td>
<td>33.44</td>
</tr>
</tbody>
</table>

Table 2. Segmental SNR for different techniques (fs=32 Khz).

As we can see, when the sampling frequency decreases the objective quality also decreases. Anyway, the vector quantization always improves the scalar quantization with dynamic bit allocation. Informal listening tests determine that using the proposed coder with a binary rate of 2 bits/sample the achieved quality is nearly-transparent.

4. CONCLUSIONS AND FUTURE WORKS

A low delay audio coder has been presented with the following characteristics:

- Subband-ADPCM coding with adaptive vector quantization.
- ADPCM according to the G.721 standard.
- Filter bank with 8 bands.
- Total delay less than 2 milliseconds.
- Nearly-transparent coding with 2 bits/sample.

We have showed that including vector quantization the objective quality is improved around 3 dB, compared with the Subband-ADPCM structure using scalar quantization with dynamic bit allocation.

Future works must be oriented to reduce complexity in order to allow the real time implementation on a commercial DSP of the proposed coder.

References