# PYRAMID VECTOR CODING FOR HIGH QUALITY AUDIO COMPRESSION

*Daniele Cadel*[1] *Giorgio Parladori*[2]

[1]Cefriel, Via Emanueli 15, I-20126 Milano, Italy
[2]Alcatel Telecom, Via Trento 30, I-20059 Vimercate (MI), Italy

## ABSTRACT

Target of this work is the high quality audio coding at low bit rate. It will be shown how the **Pyramid Vector Coding** (PVC) can conveniently replace the classical Huffman Coding technique in audio compression systems, giving also an advantage in the bit allocation procedure. The compression performances can be further improved by fixing an upper limit value of the vector components.

## 1. INTRODUCTION

Recent developments in signal compression algorithms show a general orientation in using the so called *perceptual coding* technique. A conceptual scheme is shown in Figure 1. Using this approach, there are two relevant consequence to take into consideration:

- the compression scheme should represent the signal features according to the considered perceptual model;
- the compression and the entropy coding algorithm should be able to easily control the bit-allocation procedure responsible of noise allocation according to perceptual model.

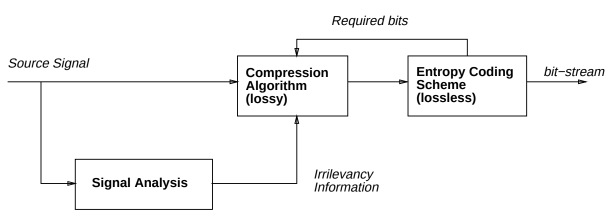


**Figure 1.** Perceptual based coding structure

The use of an entropy coding procedure significantly improve the final compressio-ratio, while introduce an undesirable drawback. In fact, most of the common bit-allocation procedures work on iterative basis. Usually, the entropy coding algorithms cannot forecast the number of used bits, unless actually implementing the coding procedure. As a consequence, in real-time implementation, only few iterations are possible and in many cases the system cannot converge to the optimal solution. The proposed method allows to compute the exact number of required bits with a low complexity procedure, performing the actual coding only at the end of bit-allocation phase. It is worth to highlight that while the Huffman coding requires pre-computed tables, based on the input signal statistics, the presented system results to be quite insensitive to input statistics, while the input data arrangement impact on the effectiveness of the method. In the following paragraphs the basic theory, the use of PVC in a real audio codec and simulation results are respectively presented.

## 2. PYRAMID QUANTIZATION AND CODING

The Pyramid Vector Quantization was proposed by Fischer [1] and it is optimized for vector quantization of Laplacian sources.

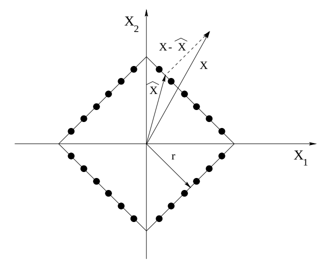


**Figure 2.** In PVQ, quantization error depends on the vector projection on the pyramid and on the lattice density.

A new system has been derived from Fischer's work by substituting the PVQ with the cascade of a Scalar Quantizer followed by an entropy coding block.

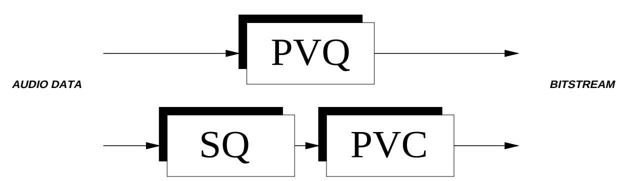


**Figure 3.** The combination of a scalar quantizer and a pyramid coder is proposed as alternative solution of PVQ.

The PVC algorithm [3] enumerates the element of a vector $\mathbf{X}$ with length $L$ and norm $K$ (i.e the sum of absolute value of the vector components), according to the number of solutions of equation Eq.(1).

$$N(L,K): \quad \sum_{i=0}^{L} x_i = K \qquad x_i \geq 0. \tag{1}$$

The transmitted information consist of a code-word and the norm value $K$ for each vector. Without loose generality, it is possible to consider the vector components as positive integers. However, in the present implementation, the sign information is coded with a separate procedure.

Assuming that

$$\sum_{i=1}^{L} x_i \;=\; K; \quad x_i \geq 0; \quad K \geq 0; \quad x_i \text{ and } K \text{ integer.}$$

the number of solution $N(L,K)$ can be easily computed using one of the following expressions:

$$P(L,K) = \begin{pmatrix} L+K-1 \\ K \end{pmatrix}. \tag{2}$$

or

$$P(L,K) = P(L-1,K) + P(L,K-1) \tag{3}$$

In order to keep the codeword length within a reasonable size, two additional features have been introduced:

1. an upper limit value is has been set for each vector components.
2. a *recovery procedure* to manage L and K values which lead to a codeword lengths bigger than a given upper limit;

If, an upper limit is set for each vector component, the enumeration equation is given in (4).

$$\widehat{N}(L,K,max): \qquad \sum_{i=0}^{L} x_i = K \qquad 0 \leq x_i \leq max \tag{4}$$

The number of solutions of this new problem $\widehat{N}(L,K,max)$ are less with respect to the previous case $N(L,K)$. As a consequence we obtain a significant compression gain, taking into account also the transmission of $max$ value. A recursive procedure to solve the Eq.(4) is given by the following expression:

- $max \geq K$:

$$\widehat{P}(L,K) = \widehat{P}(L-1,K) + \widehat{P}(L,K-1) \tag{5}$$

- $max < K$:

$$\widehat{P}(L,K) = \widehat{P}(L-1,K) + \widehat{P}(L,K-1) - \widehat{P}(L-1,K-1-max) \tag{6}$$

The *recovery procedure* is implemented transmitting the LSB of each vector element, which is then divided by 2. The procedure is iterated until the norm assume a value less than the set upper limit. As shown in simulation results paragraph, this procedure substantially do not alter the compression performances of PVC, while reducing the implementation complexity.

The encoding and decoding procedure is illustrated by the following pseudo-code. Let be $P(l,k)$ the matrix which elements $(l,k)$ are

$$\begin{pmatrix} l+k-1 \\ k \end{pmatrix}$$

$l \;=\; 0,\cdots,L$ e $k \;=\; 0,\cdots,K_m$.

- *Coding*
  1. Let be $\overline{x} \;=\; [x(0),\cdots,x(L-1)]$ a vector with positive integer components. Let be K $l_1$ norm of the vector $\overline{x}$ (K $<$ $K_m$ ).
  2. Set the following start conditions:
  $$k \;=\; \text{K}; \quad l \;=\; \text{L}; \quad b \;=\; 0;$$
  3. Set
  $$b = b + P(l,k) - P(l,k-x(L-l)); \quad k \;=\; k - x(L-l); \quad l \;=\; l-1;$$
  4. If $l > 0$,
     then goto (3);
     else b represent the index of vector $\overline{x}$.
- *Decoding*
  1. Let be $K$ the $l_1$ norm of vector $\overline{x} = [x(0),\cdots,x(L-1)]$. and b the transmitted index of vector $\overline{x}$, $b < P(L,K)$.
  2. Set the following start conditions:
  $$l \;=\; L; \quad k \;=\; K;$$
  3. Let be:
  $$x(L-l) \;=\; 0;$$
  4. if $P(l,k) - P(l,k-1-x(L-l)) \leq b$;
     then $x(L-l) \;=\; x(L-l)+1$;
     go to (4);
  5. Set:
  $$b \;=\; b - [P(l,k) - P(l,k-x(L-l))]; \quad k \;=\; k - x(L-l); \quad l \;=\; l-1;$$
  6. if l>1
     then go to (3);
     else $x(L-1) \;=\; k$;
  7. The vector $\overline{x}$ is now decoded.

## 3. PVC IN AUDIO CODING

Referring to a generic MPEG-like subband coding scheme, the PVC replace the Huffman Coding block giving a better compression performances. In addition, PVC coding allows to calculate the needed number of bits before that the entropy coding procedure is actually applied. This feature can efficiently exploited in the *bit allocation procedure* with low computational cost. Finally, this coding scheme do not require any side information to be transmitted at the decoder.

The performances of this technique has been evaluated by integrating PVC module in a MPEG-4 audio coder [2] which uses the classical Huffman procedure.

The data organization of the codec suggests to use the following data structure. Starting from a frame of $96ms$

($48khz$ sampling frequency), we form 64 vectors of 72 samples, by means of a subband filter. As shown in fig.(4) the 72 temporal samples are split into three vectors, obtaining this way three sub-matrices which will be coded by PVC. To better understand fig.(4) the following definitions are needed:

- *Data Matrix*
  $D(i,j), \qquad 0 \leq i \leq 71, \qquad 0 \leq j \leq 63$
- *Norm Matrix*
  $N(n,j) = \sum_{s=0}^{23} \mid D(s+24n,j) \mid,$
  $0 \leq n \leq 2, \qquad 0 \leq j \leq 63$
- *Global Norm Vector*
  $GN(j) = \sum_{n=0}^{2} N(n,j), \qquad 0 \leq j \leq 63$
- *Norm of GN() Vector*
  $NGN = \sum_{j=0}^{63} GN(j)$
- *Max Matrix*
  $M(n,j) = Max\{D(s+24n,j)\},$
  $0 \leq n \leq 2, \qquad 0 \leq j \leq 63$
- *Global Max Vector*
  $GM(j) = \sum_{n=0}^{2} N(n,j), \qquad 0 \leq j \leq 63$
- *Norm of GM() Vector*
  $NGN = \sum_{j=0}^{63} GN(j)$

According with the presented procedure, code-words representing $D(i,j), N(n,j), M(n,j), GN(j), GM(j)$ are transmitted. In addition $NGN$ and $NGM$ are linearly quantized and transmitted.

Additional work has been carried out in order to allows an efficient implementation both in term of memory occupancy and computational complexity, by means of the recovery procedure, which mainly impact on the the $N()$ and $M()$ vector coding.

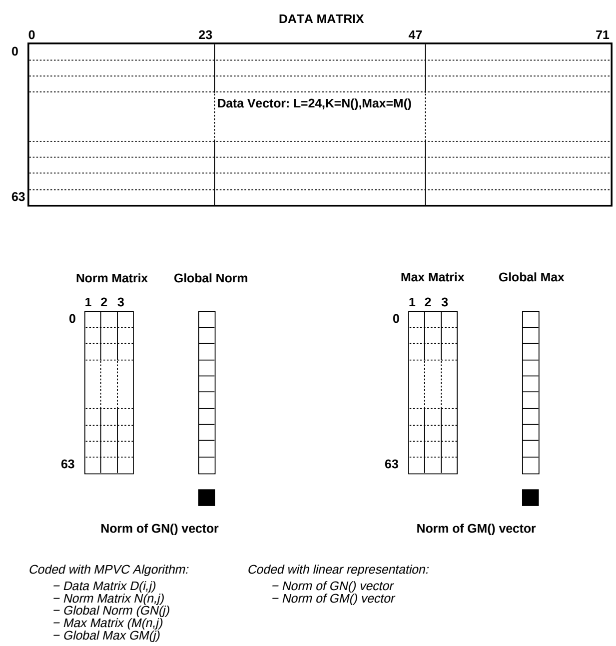


**Figure 4.** Audio data organization for PVC coding.

## 4. SIMULATION RESULTS

Simulations have shown promising results in term of absolute performance and flexibility of the system. As an example, the compression gain of PVC coding for an orchestral sequence of audio data is shown in Fig.(5). Further improvements of the system performances are expected exploiting the correlation present in the $N()$ and $M()$ matrices both in time and in frequency domain.

## 5. CONCLUSION

This paper presents a new coding procedure derived from Fischer's Pyramid Vector Quantizer. It has been proved that PVC coding performs a better compression than the classical Huffman coding giving, at the same time, more flexibility for bit allocation procedure.

## REFERENCES

[1] Thomas R. Fischer. A pyramid vector quantizer. *IEEE Transaction on Information Theory*, 32(4):568–583, July 1986.

[2] Werner Oomen (Philips) Giorgio Dimino (Rai), Giorgio Parladori (Alcatel). Description of the alcatel/philips/rai proposal for mpeg4 audio coding. ISO/IEC JTC1/SC29/WG11 MPEG95/0351 Coding of Moving Pictures and Associated Audio, November 1995.

[3] Roberto Sannino. Transform coding of images using a modified pyramid transform coder. 11th European Conference on Circuit Theory and Design - ECCDT '93, Aug.30 - Sept.3 1993. Davos, Switzerland.

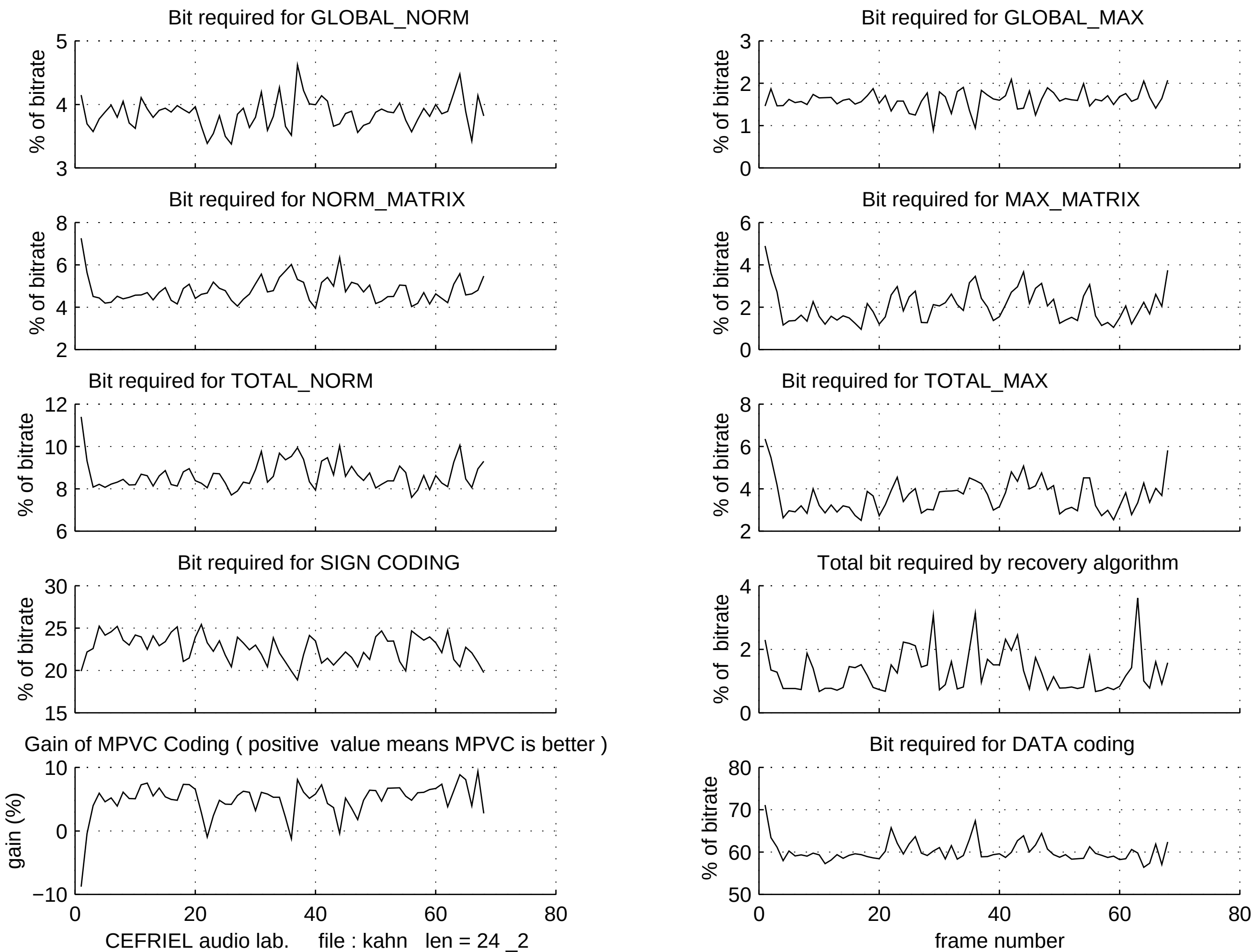


**Figure 5.** Bits usage for the coded information. TOTAL_NORM=GLOBAL_NORM+NORM_MATRIX and TOTAL_MAX=GLOBAL_MAX+NORM_MAX. The bottom left plot shows the gain of PVC scheme with respect to Huffman coding.