

# A NEW SUBBAND PERCEPTUAL AUDIO CODER USING CELP

*Olivier van der Vrecken<sup>\*</sup>, Laurent Hubaut<sup>\*\*</sup>, Florence Coulon<sup>\*\*</sup>*

Laboratoire de Théorie des Circuits et de Traitement du Signal (TCTS)

<sup>\*\*</sup> Faculté Polytechnique de Mons - Bd Dolez, 31, B-7000 Mons (BELGIUM)

<sup>\*</sup> BaBel Technologies SA – Bd Dolez, 33, B-7000 Mons (BELGIUM)

<sup>\*</sup> Tel: +32 65 37 42 75 – E-mail: [secre@babeltech.be](mailto:secre@babeltech.be)

<sup>\*\*</sup> Tel : +32 65 37 41 78 - E-mail : [hubaut@tcts.fpms.ac.be](mailto:hubaut@tcts.fpms.ac.be)

## ABSTRACT

This paper presents an audio coding system which uses filter banks to decompose, in the frequency domain, the audio signal into constant width subbands. A specific compression is applied in each subband. This compression is achieved by means of CELP coders. In order to obtain a high audio quality, psychoacoustic models allocate dynamically the number of bits needed in each subband.

A particular care has been taken for the elaboration of the filter banks in order to limit the delay and the computational cost of the system. We have implemented several filter banks and tested their influence on the perceptual quality of the output audio signal.

Finally, we show that our proposed coder is capable of delivering excellent audio signal quality at bit rates of 50-60 kbit/s.

## 1. INTRODUCTION

The reduction of the bit rate associated to the audio signal representation has recently become an important objective in the telecommunications arena. The derived applications principally cover the following domains :

- Digital audio broadcasting (DAB),
- Audio compression associated with digital television,
- Audio storage on magneto optical support (DCC, Minidisk,...).

Following this great interest, the International Standards Organization (ISO) and the International Telecommunication Union (ITU) have defined audio compression norms such as NICAM, SB-ADPCM and more recently the MPEG Layer I, II, and III. In order to increase more efficiently the compression rates, the last audio compression schemes integrate the perceptual properties of the human ear by means

of psychoacoustic models leading to transparent quality for bit rates up to 96 kbit/s (MPEG Layer III). On the other hand, perceptual degradations appear for audio signal processed with lower bit rates such as 64 kbit/s.

This paper attempts to bring solutions for the realization of audio compression systems providing a good listening quality at 64 kbit/s and approaching transparent quality at this compression ratio.

This paper is organized as follows: part two describes the psychoacoustic model principles and the different types of filter banks we used. Part three explains the coding schemes associated with filter banks. Finally, we report some tests and conclusions.

Note that the processed audio signal are quantized on 16 bits and sampled at 44.1 kHz (for an original bit rate of 705 kbit/s).

## 2. PSYCHOACOUSTIC MODELS AND FILTER BANK DESIGN

The principles of the psychoacoustic models are based on the next two facts [6][9]:

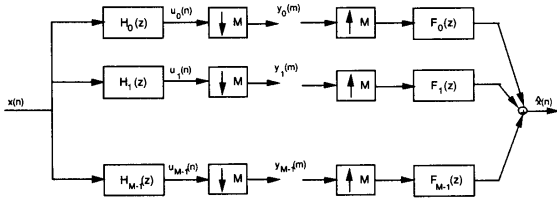
- The human ear can not perceive sounds with an energy lower than an absolute auditory threshold.
- High energy sounds can hide lower energy sounds if they have neighboring frequencies. This is known as the masking phenomena.

Following these two auditory rules, a shape adaptation of the noise spectrum added during signal quantization can be realized. We have implemented for that purpose the psychoacoustic models described in the MPEG norm.

As these models work in the frequency domain, filter banks are used to split the audio signal into several frequency channels [1]. This enables each channel to be

quantized independently so that the quantization noise can be confined to this channel and be easily perceptually masked.

We have first implemented the 32-band uniform filter bank of the MPEG norm, obtained by modulation of a 512-tap prototype lowpass filter. One reason for choosing such a filter bank is that it has a reasonable computational complexity since it can be implemented with a polyphase filter followed by a fast transform [4][11].



**Figure 1.** Decomposition of the audio signal in subbands

The ideal situation is to apply a subband decomposition similar to the critical bands achieved by the human ear [10]. A filter bank that would exactly implement critical band is computationally not feasible. Instead, people can use a tree-structured filter bank. This type of frequency decomposition is achieved by successive filtering and decimation/interpolation operations.

A 32-band tree-structured filter bank has been realized and based on a 32-order Pseudo-QMF filter [4].

### 3. SUBBAND COMPRESSION SCHEMES

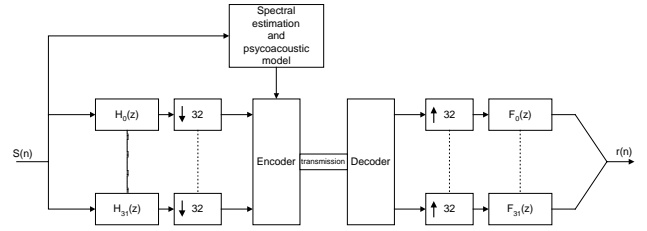
#### 3.1 Requirements attached to the implementation of the coding algorithm

The main goal of coding applications is the storage of audio signal on magneto-optic support. Consequently, the encoding operation can be performed off-line and there is no constraint regarding the computational complexity of the encoder. On the other hand, the complexity of the decoder must be kept as small as possible.

Also, users must have the possibility to fine-tuning the quality of the coded-decoded audio frames. This adjustment can be easily realized by the modification of an auditory quality factor (AQF). This factor can vary from 0.1 (low bit rates) to 2 (high bit rates) according to the desired quality. It is chosen once for all frames.

#### 3.2 Computation of the masking curves

In parallel to the computation of the outputs of the filter banks, a fast Fourier transform is used for spectral estimation. Based on the power spectrum, a masking curve is calculated. Quantization noise is then allocated in the various subbands according to the masking function [7]. This allocation is done on a small block of subband samples. The psychoacoustic model gives, for each subband, a minimum SNR that the subband coding algorithm must reach in order to limit the quantization noise introduced during this compression operation.



**Figure 2.** The encoder and decoder schemes

The quality of the coded-decoded audio frames is adjusted by the *quality ratio coefficients* ( $Qr$ ), defined as,

$$Qr[i] = SNRb[i] * AQF, \quad i = 1..32$$

where  $SNRb[i]$  is the minimum SNR required by the  $i$ th subband. The encoder algorithm associated to the  $i$ th subband uses the  $Qr[i]$  coefficient to determine the compression precision.

#### 3.3 Description of the encoding algorithm

##### 3.3.1 Principle

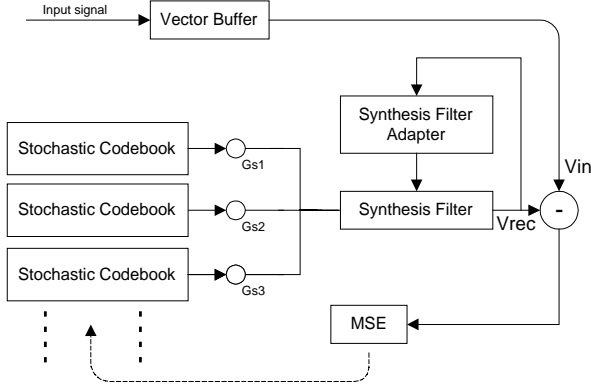
An analysis-by-synthesis coder with backward LP analysis has been chosen to encode input audio frames in each subband [3][8]. These frames are reconstructed by filtering an excitation vector, which is a combination of vectors derived from a variable number of stochastic codebooks, through a short-term predictor synthesis filter,

$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}}$$

where  $a_k$  are the short-term prediction coefficients and  $p$  is the order of the filter.

The synthesis filter coefficients are computed every frame. They result from a 2nd order LPC analysis on the original signal. A particularly small order has been chosen to limit the complexity of the decoding section (see also [2][5]).

The excitation signal is found by minimizing the mean-squared error over several samples, where the error signal is obtained by filtering the difference between the original and the reconstructed audio frames (see Fig 3.).



**Figure 3.** The encoder.

In the case of a 32-band decomposition, the input frames contain 384 samples. Consequently, the filter bank gives 32 subband blocks of 12 samples.

For each subband block, the best complementary excitation vectors are searched in the stochastic codebooks. A parameter, denoted  $Nbr\_Max$ , fixes the maximum number of stochastic codebooks used to model each frame.

Since the analysis involves synthesis, the description of the analysis procedure completely describes the decoder.

### 3.3.2 Construction of the compressed audio signal

The compressed audio signal is composed of 32 sequences including the number of stochastic codebook(s) used ( $Nbr\_Cod$ : 2 bits) and the corresponding shape and gain indexes. In the current implementation of the coder, the same codebook is used for all subbands. This codebook contains 256 vectors of 12 components. The associated gain is scalar quantized on 64 levels (6 bits). The stochastic codebook is trained in a recursive way so as to optimize this codebook to the compression of the audio signal.

The maximum number of stochastic codebooks used to model each frame has a typical value of 4.

The following table summarizes the number of bits required to code each frame of the signal :

Number of codebooks	$Nbr\_band * 2$ bits
Stochastic shapes	$Nbr\_cod * Nbr\_band * 8$ bits
Stochastic gains	$Nbr\_cod * Nbr\_band * 6$ bits

where  $Nbr\_band$  is the number of subband blocks (32 in the case of MPEG norm) and  $Nbr\_Cod$ , the number of codebook(s) used to model one subband block.

## 4. RESULTS

In this section, we take the quality of the audio signal produced by the filter bank without compression as the reference.

The quality of a coded-decoded audio signal can be adjusted by fine-tuning the minimum SNR by modifying the auditory quality factor involving bit rate modifications.

Subjective tests have been realized with a group of 10 persons familiarized with the listening conditions to evaluate the distortions due to the integration of the coding algorithm in the filter banks. The following table gives the compression rates with their listening appreciation scores over several music records in the case of MPEG filter bank.

	a	b	c	d
Ratio	23.3	19.2	14.3	11.5
Bit rate (kbit/s)	30.2	36.8	49.5	61.6
Auditory Quality Factor	0.1	0.4	0.6	0.8
Score	Weak	Good	Good	Near transparency

For a compression rate equal to 14.3, it can be seen that the music coding algorithm presents a good listening quality. The last trial (d) offers a very good quality for a compression ratio of 11.5. It was found to be near transparency by listeners.

The same subjective test has been realized with the tree-structured filter bank. It appears that the listeners attribute the same auditory scores.

A direct application of this audio coding system is the *BaBel Audio project*[12]. The goal of this project is to store, on an internet server, samples of compressed music records. When implemented as a plug-in, the decoder will enable internet users with good connectivity to listen to CD excerpts with CD quality.

## 5. CONCLUSIONS

Several tests, made on different audio sequences, have shown the good performance of our music coding algorithm with rates below 64 kbit/s. The integration of our coder in the tree-structured filter bank and the MPEG filter bank gives similar listening quality for the coded-decoded audio records. The low computational cost and a high compression rate of our coder make it possible to distribute music records via the internet.

More information and examples of compressed music files are available at the URL address :

<http://tcts.fpms.ac.be/coding/coding.html>

## 6. REFERENCES

- [1] P.P. Vaidyanathan, "Multirate Systems and Filter banks", Prentice Hall Signal Processing Series, 1993.
- [2] Olivier van der Vrecken, Nicolas Pierret, Thierry Dutoit, Vincent Pagel, Fabrice Malfre, "A Simple and Efficient Algorithm for the Compression of MBROLA Segment Databases", Eurospeech, Rhodes, September 1997.
- [3] M.R. Schroeder and B. Atal, "Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates", ICASSP 1985, pp. 937-940.
- [4] M. Vetterli, J. Kovacevic, "Wavelets and Subband Coding", Prentice Hall PTR, 1995.
- [5] S. White, "Codeur CELP à débit variable: application au codage de diphones", Proc. 19ème Journée d'Etudes sur la Parole, Montreal, 1990.
- [6] P. Pierrick, "Codage audionumérique et transformation en ondelettes", Ph. D. dissertation, Université de Paris XI Orsay, November 1995.
- [7] X. Wei, Martyn J. Shaw and Martin R. Varley, "Optimum Bit Allocation and Decomposition for High Quality Audio Coding", Proc. ICASSP 97, Munich.
- [8] D. Rowe, W. Cowley, A. Perkis, "A Multi-Band Excitation Linear Predictive Speech Coder", Proc. Eurospeech 91, Genova, pp. 239-243.
- [9] E. Zwicker, R. Feldtkeller, "Psychoacoustique", Ed. Masson, 1981
- [10] J.-P. Renard, "Système de codage en sous-bandes du signal sonore", Ph. Dissertation, Faculté Polytechnique de Mons, 1992.
- [11] N. Moreau, "Techniques de compression des signaux", Ed. Masson, 1994.
- [12] <http://www.babeltech.be/>, BaBel Technologies sa