Quantization Noise Power Injection In Subband Audio Coding Using Low Selectivity Filter Banks

D. MARTÍNEZ-MUÑOZ, N. RUIZ-REYES, P. VERA-CANDEAS,
P.J. RECHE-LÓPEZ, J. CURPIÁN-ALONSO
Departamento de Electrónica
Universidad de Jaén
C/ Alfonso X El Sabio, 28; 23700 Linares
SPAIN

Abstract: - The ISO-MPEG standards are widely used for high quality audio coding. They use perceptual models that require high frequency resolution. This implies two characteristics: they use high selectivity filter bank to decompose the audio signal, and the FFT frame length for the frequency masking threshold calculations is long. Both characteristics carried out a very high delay coding. 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 how to overcome the first handicap: to compatibilize the use of short impulse filter bank response and perceptual model application. These filter banks present low selectivity and so the non-ideal filters frequency responses should be taken into account.

Key-Words: Subband audio coding, Noise injection, Filter bank, Psycho-acoustic model

1 Introduction

In last years, a lot of work has been developed in the difficult issue of data compression either for video or audio [1][2]. The main objective that has guided the mentioned projects has been the maximization of the coder efficiency. Other aspects, such as coding delay, have been considered as minor goals. The most important are the ISO-MPEG algorithms for video and audio.

MPEG is a working group in a subcommittee of ISO/IEC (International Organization for Standardization/ International Electrotechnical Commission) in charge of developing international standards for compression, decompression, processing, and coded representation of moving pictures, audio, and their combination. In particular, MPEG defines the syntax of low bitrate video and audio bitstreams of synthetic and natural sources, descriptions of their structure and content, and the operation of conformant decoders of these bitstreams.

The encoder algorithms are not defined by MPEG. This allows for continuous improvement of encoders and their adaptation to specific applications, within the bitstream syntax definition. Along with the video and audio coding, MPEG also defines means to multiplex several video, audio and information streams synchronously in one single bitstream, describes methods to test conformance of bitstreams and decoders to the standard, and publishes technical reports containing software describing the decoder operation and software describing examples of encoder operation. MPEG works in phases. These phases are normally denoted by Arabic numbers (MPEG-1, MPEG-2, MPEG-4, MPEG-7). These phases do not describe different versions of one standardisation effort, but are rather completely different coexisting standards that all handle different aspects of Multimedia communication.

For audio, the coders included in the ISO-MPEG standards apply perceptual models that require high frequency resolution. This implies two characteristics:

1) Coders use high selectivity filter bank (with long impulse response length) to decompose the audio signal.
2) The frame length is high in order to compute the frequency masking threshold by FFT.

Both characteristics carried out a very high delay coding.

Recently, a new high quality audio MPEG standard has been proposed: MPEG2-AAC (Advanced Audio Coding). As in the previous MPEG standards, it achieves high quality but the delay is also high. However, this coder presents a named low-delay mode that introduces around 40 ms of coding delay. In full-duplex communications, the overall delay will be at least 80 ms. The propagation and the implementation delay should also be considered.

The delay in some applications may be a critical requirement. For example, a sport commentator whose speaking is coded and transmitted to the base, where it is mixed with music before returning to his working place in order to be heard by himself. A long delay (greater than 60 ms.) may force the commentator to stutter.

Consequently, nowadays there is not a standard for audio coding that offers high quality and low delay. Actually, in applications where a feedback channel is required, the ITU-T standard for 7 kHz audio (G.722, 64 kbps) is applied. It is 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. This algorithm was developed primarily for ISDN teleconferencing and loudspeaker telephony [3]. Because of the 64 kbps capability, a single "voice-grade" channel on a digital or analog, public-switched telephone network (PSTN) can transport an AM-quality sound program over any distance and yield a broadcast-grade voice program at the receiving end. Its subband-ADPCM coding scheme attains acceptable quality for speech signals but the features for musical signals are not so good.

In this work, we present how to compatibilize the application of a psycho-acoustic model in a subband coder which uses a low selectivity filter bank, in order to reduce the delay coding. In this case, the control of quantization noise introduced in one subband is a difficult task, because it is spread over different frequency bands. Quantization noise introduced in any subband may introduce artifacts into any different frequency band.

The other important problem is the incorporation of psycho-acoustic information: psycho-acoustic models need the estimation of the power spectral density of the audio frame. Therefore, high delay is added due to the necessity of large number of samples in the frame to achieve an accurate spectral estimation. However, this problem is not considered in this paper.

The overall problem may be formulated as follow:

1. Filters in the filter bank are not selective.
2. Quantization noise into any subband spreads over other frequency bands of the reconstructed signal.

In next section we describe the proposed strategy. Finally, we show an example of the previously presented strategy.

2 Proposed strategy

In perceptual subband audio coder, the psycho-acoustic model output is used to compute the masking to noise ratio (SMR) for each band over any frame. Bit allocation is realized according to the obtained SMR value. We need to calculate both the energy signal level and the masking threshold level for every subband in order to find the SMR. In the calculus of both values we must have into account the low selectivity of the filter bank. From here on, we present the calculus we do for every subband.

2.1 Masking Threshold

Over each frame we calculate the masking threshold in the frequency domain applying any psycho-acoustic model. The obtained masking threshold must be modified to consider the low selectivity of the filter bank.

Once it is known the masking threshold, we must concentrate the study in each band. We need to insure that the overall distortion in any subband due to the spreading of the quantization noise introduced in each subband is below the masking threshold. Next, we explain how to determine the maximum quantization noise level that can be injected in each band. We denote the band number of the filter bank as M, the frequency responses in the DFT domain of the analysis and synthesis filter bank as $H[k]$ and $F[k]$, respectively, and the masking threshold in the DFT domain as $U[k]$. Both, $H[k]$ as $F[k]$, are matrixes and $U[k]$ is a vector.

The steps that follow are:

1. For each band, i, we search for the minimum of the masking threshold modified by the frequency response of the synthesis filter bank branch, as follows:

$$Umin_i = \text{minimum}\left[\frac{U[k]}{\left|F_i[k]\right|^2}\right] \quad (1)$$

   In this way, we take into account how much noise can be injected in each subband without perceptual quality loss, considered in isolation. The frequency index where $U\text{min}_i$ was found is also obtained ($K\text{min}_i$).

2. Now we consider the contribution of all the subbands. The overall distortion must be below the masking threshold, $U[k]$. To calculate the maximum noise power that can be injected in

each band, we consider the following suppositions:

- Quantization noise is white: we suppose that the noise produced in each band go into the other according to frequency response of the corresponding filter of synthesis filter bank. As the number of bits assigned to the band is higher, this supposition become true. However, at this point we still do not know the final bit allocation. By applying this supposition, we must guarantee that

$$\sum_{i=1}^{M} U_i \cdot \left|F_i[k]\right|^2 \leq U[k] \text{ for all } k \qquad (2)$$

where U_i is the noise power injected in the i-th subband, which must be calculated.

- The frequency index of the minimum of the overall distortion inside a band is close to that previously found when the band was considered in isolation ($K\min_i$). By applying this second supposition, we obtain the following system of equations

$$\begin{aligned} Umin_1 &= U[K\min_1] = \sum_{k=1}^{M} U_k \cdot \left|F_k[K\min_1]\right|^2 \\ Umin_2 &= U[K\min_2] = \sum_{k=1}^{M} U_k \cdot \left|F_k[K\min_2]\right|^2 \\ &\vdots \qquad\qquad \vdots \\ Umin_M &= U[K\min_M] = \sum_{k=1}^{M} U_k \cdot \left|F_k[K\min_M]\right|^2 \end{aligned} \qquad (3)$$

3. Once U_i is calculated, these values will be used to compute the signal to mask ratio.

2.2 Signal level

The subband signal level for each band, L_i, is calculated according to expression (4).

$$L_i = \text{maximum}\left\{E[k] \cdot \left|H_i[k]\right|^2\right\} \qquad (4)$$

Where $E[k]$ represents the power spectral density of the input samples frame and $H_i[k]$ is the frequency response (in the DFT domain) of the *i-th* filter of the analysis filter bank.

2.3 Signal to mask ratio

Once the masking and the signal levels have been calculated for each subband, the signal to mask ratio, SMR(dB), is obtained as:

$$SMR(i) = 10 * \log 10\left(\frac{L_i}{U_i}\right) \qquad (5)$$

3 Example of application

The coder considered responds to the structure shows in figure 1. It presents a sub-band structure with dynamic bit allocation according to a psycho-acoustic model and adaptive quantization.

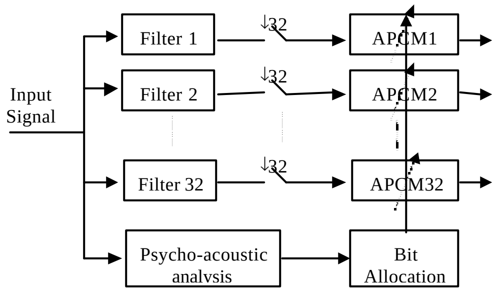

Fig. 1. Block diagram of the considered coder.

The input signal is decomposed in 32 subbands. Each subband is decimated and quantized using adaptive quantization. The number of quantization steps corresponding to each subband is defined according to a psycho-acoustic model. Psycho-acoustic information is estimated using frames of 512 samples each one.

3.1 Pseudo-QMF Filter Bank

We have used pseudo-QMF conventional cosine-modulated filter bank [4,5] for dividing the incoming signal into 32 separate subband signals. This filter banks based on cosine modulation, demand only to design a unique prototype filter $p[n]$. Several methods have been proposed to facilitate the design of the prototype filter and to improve the characteristics of the resultant system [6,7]. We have designed a 256-length prototype filter by using a method that controls the position of the 3 dB cutoff frequency and sets it at $\pi/2M$ [8]. The magnitude response plot for the analysis filter is showed in figure 2. The stopband attenuation is around 50 dB. This poor stopband attenuation must be taken into account because it can become a significant source of perceptual distortion due to aliasing of quantization noise.

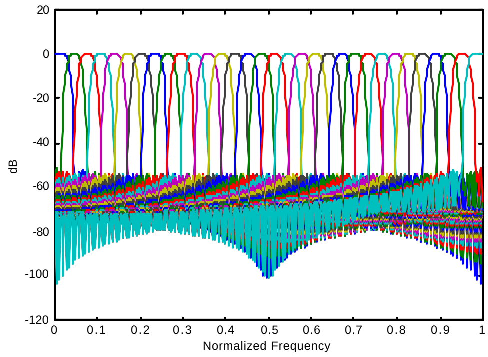

Fig. 2. Filter bank magnitude frequency response.

3.2 Psycho-acoustic Analysis

Once a frame is completed, we apply a psycho-acoustic model to the last input samples. The masking to noise ratio (SMR) is calculated for each band over this frame. This value is used by the bit allocation block. We calculate the SMR values according to the strategy presented in the previous section.

3.2.1 Masking Threshold

Over the frame of 512 samples we obtain the masking threshold in the frequency domain applying the psycho-acoustic model similar to the psycho-acoustic model number 2 of the ISO-MPEG1 standard. This threshold has been modified to consider the low selectivity of the filter bank.

In figure 3 is showed the masking threshold (rectangular line), the U_i value obtained (star point) and the overall distortion (flexible line) when the noise power injected in each band is U_i. Figure 4 is a detail of figure 3. The U_i value is lower than minimum masking threshold for the band *i*. In this way, the overall distortion obtained when the noise power injected in the band *i* is U_i, is lower than the masking threshold.

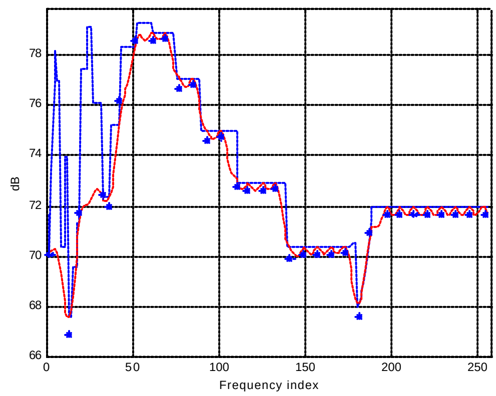

Fig. 3. Masking threshold, U_i and overall distortion.

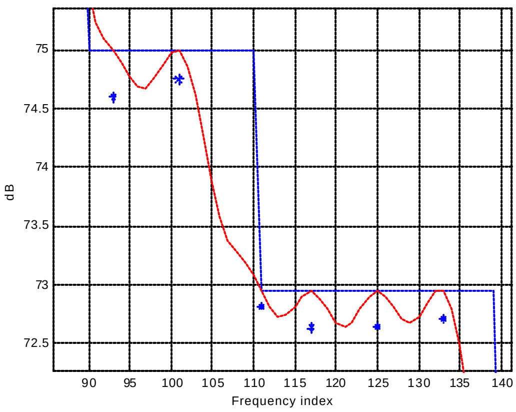

Fig. 4. Figure 3 detail.

4 Results and conclusions

In applications where the delay is a critical parameter the standards MPEG for audio coding are not suitable. They use long impulse filter bank response and the frame length for apply the FFT transform is also long. Both characteristics introduces high delay. In this paper we have presented a strategy to compatibilize the short impulse filter bank response use and the psycho-acoustic applications. Due to the low selectivity of the filter bank, we must have into account the overlapping distortion. The spread of quantization noise introduced in each subband over the others is considered in order to determine the maximum quantization noise power that can be introduced in every subband signal.

We applied this strategy to a subband coder that use a cosine-modulated filter bank with length of the filters impulse response of 256 and adaptive quantization. This low number of taps produces delay as low as 5.8 ms but the filter bank has low selectivity. Applying the presented strategy we get a better value for the noise power that we can inject in each band without overcoming the masking threshold.

References:

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

[2] D. Martínez, F. López, M. Rosa, N. Ruíz: "*Recent Advances in Signal Processing and Communications*". World Scientific Engineering Society, pp. 234-238, 1999.

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

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

[5] J. H. Rothweiler: "*Polyphase Quadrature Filters – A New Subband Coding Technique*", in Proc. of Int. Conf. ASSP 83, Boston (USA), pp. 1280-1283, April 1983.

[6] C. D. Creusere and S. K. Mitra: "*A Simple Method for Designing High-Quality Prototype Filters for M-Band Pseudo QMF Banks*", IEEE Transactions on Signal Processing, vol. 43, No. 4, pp. 1005-1007, April 1995.

[7] Y.-P- Lin and P. P. Vaidyanathan: *"A Kaiser Window Approach for the Design of Prototype Filters of Cosine Modulated Filters Banks", IEEE Signal Proc. Letters, vol. 5, No. 6, pp. 132-134, June 1998.*

[8] F. Cruz, P. Martín, F. López, P. L. López; A. Martín: "*Nearly Perfect Transmultiplexers Using Cosine-Modulated Filters Banks"*. Proc. of 2001 IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP-01), Baltimore-Maryland (USA), Junio 2001.

[9] D. Martínez, M. Rosa, F. López, N. Ruíz: "*Low delay audio coder using adaptive quantization*", in Proc. of X European Signal Processing Conference (EUSIPCO 2000), Tampere (Findland), September 2000.

[10] M. Rosa, D. Martínez, F. Cruz, F. López: "*Low delay audio coding incorporating Psycho-acoustic Information*", in Proc. of 110th Audio Engineering Society Convention (110th AES Convention), Amsterdam (The Netherlands), May 2001.