

Subband Coding of Digital Audio Signals Without Loss of Quality

Raymond N.J. Veldhuis, Marcel Breeuwer, Robbert van der Waal
Philips Research Laboratories, P.O. Box 80.000
5600 JA Eindhoven, The Netherlands

Abstract

A subband coding system for high-quality digital audio signals is described. To achieve low bit rates at a high quality level, it exploits the simultaneous masking effect of the human ear. It is shown how this effect can be used in an adaptive bit-allocation scheme. Results obtained with a low-complexity and a high-complexity system are discussed.

1 Introduction

Transmission and storage of high-quality digital audio is becoming important for the audio industry, for instance in the case of digital radio and new applications for optical disks. The bit rate of a high-quality stereophonic digital audio signal is about 1.4 Mbit/s. For some transmission channels or storage media this is too high and therefore source coding is required. Since digital audio is associated with high quality, a perceptible loss of quality cannot be tolerated.

Source coding of audio signals at low bit rates generally introduces errors. This paper describes a coding system that attempts to keep coding errors inaudible by exploiting the *simultaneous masking effect*. This is the perceptive phenomenon that a weak signal, e.g. quantization noise, is masked (= made inaudible) by a stronger signal, e.g. a pure tone in the audio signal. Simultaneous masking is briefly explained in Section 2.

Simultaneous masking is most effective if both masked and masking signal are in a rather narrow frequency band. This suggests the use of subband coding, where the signal is first split up into frequency bands which are then quantized. The structure of the subband coding system is given in Section 3.

Quantization should be such that the quantization noise is masked by the audio signal. This is achieved by using uniform APCM quantization [1]. The subband signals are split up into blocks. Each block is scaled to a unit level and then quantized by a uniform quantizer. Quantized data and scale factors are transmitted. In this manner the power of the quantization

noise can be controlled by allocating a certain amount of bits to each quantizer.

In a subband coding system we can distinguish *in-band* masking, where both masked and masking signal are in the same subband, and *out-of-band* masking, where masking and masked signal are in different subbands. Both are exploited in the system described in this paper. Section 4 explains how for each subband the maximum power of the quantization noise that is masked, called the *masked power*, can be estimated.

Once the masked powers have been computed for all subbands, bits are allocated to the quantizers. Ideally the amount of bits for each quantizer should be such that the quantization noise is completely masked. However, the masked powers are signal-dependent and therefore the amount of bits needed to ensure complete masking varies in time. Because the coding system described here has a fixed bit rate, the bits must be divided over the subbands in such a way that the audible degradation of the output signal is minimal. This requires an adaptive bit-allocation technique, that is described in Section 5.

There is a trade-off between quality, bit rate, and complexity. Complexity is largely determined by the splitting and merging subband filters. It can be kept low by keeping the number of subbands low and their minimum bandwidth high. At a fixed quality level, the lowest bit rate achievable with a 'low-complexity' system is higher than with more complex systems with more and narrower subbands. This is explained in Section 4. Results obtained with a simple and a complex system are discussed in Section 6.

2 Simultaneous masking

Simultaneous masking is the effect that a weak signal is made inaudible by a simultaneously occurring stronger signal. Masking is discussed in great detail in [2,3]. The use of masking in subband coding is described in [4,5]. First consider the simple case of a pure tone as a masking signal. A signal component with a certain frequency is masked if the ratio of its power and the power of the masking tone is below

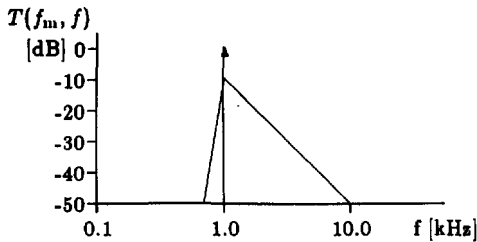


Figure 1: Masking threshold as a function of frequency

the masking threshold. The masking threshold is a function of frequency. Figure 1 shows a stylistic approximation (on a dB scale) of the masking threshold for a pure tone of 1000 Hz at a sound pressure level of 0 dB. In general the masking threshold for a pure tone can be approximated by

$$T(f_m, f) = \begin{cases} T_{\max}(f_m) \left(\frac{f}{f_m}\right)^{28}, & f \leq f_m, \\ T_{\max}(f_m) \left(\frac{f}{f_m}\right)^{-10}, & f > f_m. \end{cases} \quad (1)$$

In this expression f_m is the frequency of the masking signal and $T_{\max}(f_m)$ is the masking threshold at this frequency. To simplify the masking model it is assumed that masking thresholds for tones of all frequencies have the same shape. However, $T_{\max}(f_m)$ depends on the frequency of the masking signal [4]. It is also assumed that the masking threshold is independent of the power of the masking signal. Furthermore, it is assumed that masking is additive: the masking threshold for a signal containing more than one frequency component can be obtained by adding the masking thresholds of the components.

The masking model used here is a simplification of reality. Coding systems based on it may show unexpected and unwanted effects. To avoid this, they must be tested and optimized in extensive listening experiments.

3 Subband coding

It is clear from Figure 1 that masking is strongest for frequencies close to the frequency of the masking signal. This suggests that the masking phenomenon can be well exploited in a subband coding system. In such a system the signal is split up into frequency bands, called subbands, which are then quantized. The splitting of the signal into subbands and the merging of the subbands into a replica of the original signal are done by decimating and interpolating filter banks, such as quadrature-mirror (QMF) or conjugate quadrature-mirror filter banks (CQF) [6,7]. Due to the decimation, the sampling frequency of a subband signal equals twice the subband's bandwidth. Therefore the

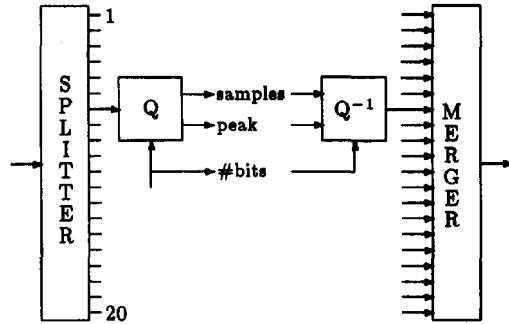


Figure 2: Subband coding system with 20 subbands

total sample rate after splitting is the same as the sample rate at the input. Because the 'bandwidth' of the masking threshold as given by (1) increases with the frequency of the masking signal, the bandwidths of the subbands also have to increase (or be at least non-decreasing) with frequency.

Quantizing signals means adding quantization noise to them. If the filter banks have good separating properties, the additional noise will remain in the subband it was added to. It was assumed in Section 2 that masking occurs if the signal-to-noise ratio is above a certain threshold. This implies that the quantizers must operate at a predetermined signal-to-noise ratio. This can be achieved with uniform APCM [1] quantizers. In this type of quantizer the signal is first divided into blocks. Of these blocks the maximum absolute values, called *peak values*, are computed. By dividing the samples in the blocks by the peak values, they are scaled to a unit level. The scaled blocks are then quantized with a uniform quantizer. After dequantization the signal-to-noise ratio is proportional to the number of bits used in the quantizer. In this way the signal-to-noise ratio of a quantizer can be predetermined by allocating a certain amount of bits to it.

Figure 2 shows a diagram of a coding system with 20 subbands. As can be seen, quantized samples as well as coded peak values and side information to indicate the number of bits used for quantization are transmitted.

In this section the number of subbands and the amount of bits for the quantizers have not been determined. In Section 4 it is shown how for a given division of the signal into subbands the masking model of Section 2 can be used to determine the masked power in the subbands. It is also explained how the final bit rate depends on the division into subbands. In Section 5 it is shown how the number of bits allocated to each quantizer is computed.

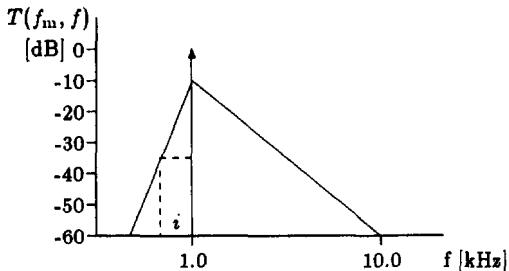


Figure 3: *In-band masking*

4 Masking and subband coding

It can be seen from Figure 1 that signals with a frequency lower than the frequency of the masking signal are hardly masked. Therefore only two kinds of masking are considered: in-band masking, this is masking within a subband, and masking of signals in subbands at higher frequencies. For both cases the masked power is computed as a function of the powers of the subband signals.

Firstly it is assumed that there is only a signal in the subband with index i . This subband ranges from $f_{l,i}$ to $f_{u,i}$. The signal power is $\sigma_{s,i}^2$. The quantization noise is assumed to have a flat spectrum in the subband. The worst-case situation for in-band masking occurs when the masking signal is a pure tone with a frequency $f_{u,i}$. In this case the power of the quantization noise in subband i that is masked by a signal with power $\sigma_{s,i}^2$ in the same subband must be less than $\sigma_{s,i}^2 T(f_{u,i}, f_{l,i})$. This situation is illustrated in Figure 3.

The worst-case situation for the masking of noise in subbands at higher frequencies occurs when the masking signal in subband i is a pure tone with a frequency $f_{l,i}$. For this case the power of the quantization noise in subband j that is masked by a signal with power $\sigma_{s,i}^2$ in subband i must be less than $\sigma_{s,i}^2 T(f_{l,i}, f_{u,j})$. This situation is illustrated in Figure 4.

In this way the contribution of a subband to the masked power in all subbands can be computed. Because masking is assumed to be additive, the masked power in a subband can be obtained by adding all contributions.

The lowest achievable bit rate at a certain quality level depends on the division into subbands. Firstly, the computations of the masked powers are based on worst-case assumptions. The real masked powers can be substantially higher. If the subbands are narrower the results of these computations will, on average, be closer to the real masked powers. Secondly, it follows from Figure 3 that the contribution of in-band masking to the masked power is higher if the subbands are

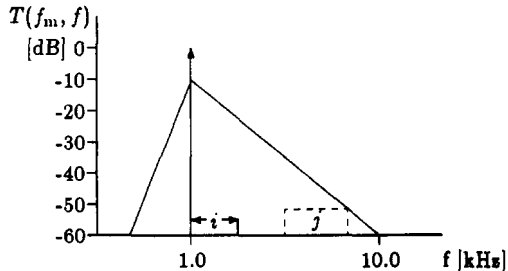


Figure 4: *Masking of subbands at higher frequencies*

narrower. The latter effect, however, is limited because in reality the top of the curve of Figures 1, 3, and 4 is flatter than is depicted [4], so that there is no use in decreasing the bandwidth of the subbands beneath a certain point. Due to the two effects mentioned here, narrower subbands lead to higher masked powers and consequently the amount of bits required for quantization can be lower.

The results of this section are only valid if the distribution of signal power over the subbands is stationary. In reality this is not true. Therefore the computations must be repeated periodically. As a consequence of this instationarity the amount of bits needed to quantize each subband will also vary in time. The allocation of bits to the quantizers on the basis of the masked powers is discussed in Section 5.

5 Bit allocation

The quantizers in the subbands are uniform APCM quantizers. For the block length M a rather arbitrary value of 32 has been chosen. Before quantization, the peak value and the power are computed for each block. The power in a block is obtained as the average of the squares of the samples of the block. The masked powers are now computed for every block, instead of for every subband as was done in the previous section. Before the masked power is computed, the blocks are arranged in an *allocation window*. An allocation window contains all subband samples during a period of time. This period is chosen in such a way that it contains one block of samples from the most decimated subband. This is in general the subband at the lowest frequency. An example of an allocation window for a 20-band system is shown in Figure 5. If the input sample frequency is 44100 Hz, subbands 1-8 have a bandwidth of 689 Hz, subbands 9-20 have a bandwidth of 1378 Hz.

It can now be computed how much each block in an allocation window contributes to its own masked power and to the masked powers in the blocks in

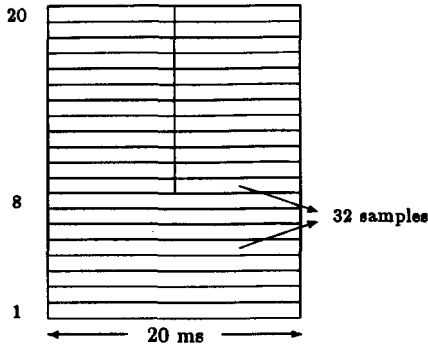


Figure 5: Allocation window for a 20-band system

higher subbands. A block only contributes to the masked power in blocks that lie within the time-interval of the masking block.

Before allocating bits to the blocks, blocks with powers below their masked powers can be excluded. The signals in those blocks will be masked. Only codes indicating that they are empty have to be transmitted. Assume that the amount of blocks in an allocation window is N and that the blocks are numbered from 1 to N . The masked power in the i^{th} block is denoted by $\sigma_{m,i}^2$, and the peak value by p_i . Ideally, one would choose the amount of bits b_i for this block such that the quantization noise is completely masked, which means that

$$\frac{1}{12} \left(\frac{2p_i}{2^{b_i} - 1} \right)^2 \leq \sigma_{m,i}^2.$$

This leads to a varying amount of bits per allocation window that can be higher than what is available. Therefore the bits must be allocated under the constraint

$$\sum_{i=1}^N b_i = B, \quad (2)$$

where B is the number of bits available. This number can be derived from the desired bit rate, taking into account that also a small number of bits is required to code peak values and to code the number of bits allocated to each block.

The allocation procedure is such that the total noise-to-mask ratio, given by

$$\sum_{i=1}^N \frac{1}{12} \left(\frac{2p_i}{2^{b_i} - 1} \right)^2 \frac{1}{\sigma_{m,i}^2}$$

is minimized under the constraint (2). A further constraint is that all b_i must be integers with $2 \leq b_i \leq 16$. The solution to this constrained integer minimization problem is given in [8].

6 Results

The ideas explained in this paper have been applied in two coding systems, a complex system splitting up the signal into 26 subbands, approximately one third of an octave wide, and a simpler 20-band system of which the bandwidths were given in Section 5. In both systems the adaptive bit-allocation method of [8] is used. Both systems have been designed for coding stereophonic 16-bit compact disc signals with a sample frequency of 44.1 kHz. Left and right channel are coded independently. With the 26-band system high-quality results can be obtained at bit rates of 220 kbit/s. With the 20-band system similar results can be obtained at bit rates of 360 kbit/s. The complexity of the systems is largely determined by the memory requirements of the filter banks. These are substantially higher for the 26-band system.

The filtering and coding delay is determined by the maximum decimation factor, the filter banks used, and the quantisation block length. For the 26-band system the decimation factor is 256 and the total delay can be as high as 800 ms. For the 20-band system it is 32 and a typical value for the total delay is 80 ms. Other types of filter banks and shorter quantisation blocks may give lower values.

References

- [1] N.S. Jayant, and P. Noll. *Digital Coding of Waveforms*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
- [2] E. Zwicker, and R. Feldtkeller, *Das Ohr als Nachrichtenempfänger*, S. Hirzel Verlag Stuttgart, 1967.
- [3] B. Scharf, in *Foundations of Modern Auditory Theory*, ed. J.V. Tobias, Academic Press, NY, 1970.
- [4] M.A. Krasner, 'The critical band coder', *Proc. of ICASSP 1980*, pp. 327-331.
- [5] G. Theile, M. Link, and G. Stoll, 'Low bit rate coding of high-quality audio signals', Preprint 2431 (C-1), 82nd convention of the AES, London, 1987.
- [6] P.P. Vaidyanathan, 'Quadrature Mirror Filter Banks, M-Band Extensions and Perfect-Reconstruction Techniques', *IEEE-ASSP Magazine*, vol.4-3, 1987, pp. 4-20.
- [7] M.T.J. Smith, and T.P. Barnwell, 'Exact reconstruction techniques for tree-structured subband coders', *IEEE Trans. on ASSP*, vol.34-3, 1986, pp. 434-441.
- [8] Y. Shoham, and A. Gersho, 'Efficient bit allocation for an arbitrary set of quantizers', *IEEE Trans. on ASSP*, vol.36-9, 1988, pp. 1445-1453.