

SUBBAND CODING OF HIGH-FIDELITY QUALITY AUDIO SIGNALS AT 128 KBPS

Do-Hui Teh, Ah-Peng Tan

A/V Information Research Centre, Asia Matsushita Electric (S) Pte Ltd, Singapore

Soo-Ngee Koh

School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore

ABSTRACT

This paper considers the design and performance of a 24-band subband coder for the encoding of high-fidelity quality 20 kHz audio signals at 128 kbps. The slightly different performances of the subband coder employing two different adaptive bit allocation algorithms, namely the modified Ramstad algorithm and the algorithm based on noise masking threshold, are discussed and compared in terms of segmental signal to noise ratio, mask to noise ratio and the probability of noise measures.

In this paper, a subband coder, similar to the Layer 2 of the ISO/MPEG audio coding algorithm [4] was used as the basis for comparison between two different dynamic bit allocation schemes; one using perceptual coding while the other does not. The approach of the paper is firstly to review the psychoacoustics applied in the perceptual coder. This is followed by a description of the subband coder, as well as the bit allocation algorithms and lastly, the performance of the two schemes are discussed.

INTRODUCTION

There has been an increasing interest in the area of high-fidelity quality audio coding in recent years ([1], [2], [3]) and this has been fueled by demands from practical applications in digital audio transmission and storage. The coders developed attain a significant reduction in bit rate while preserving the excellent quality of the compact disc digital audio. Standardization effort in ISO/MPEG (Moving Pictures Experts Group) started in 1989, has also resulted in a layered compression algorithm for audio signals.

PSYCHOACOUSTICS

The main concern of a perceptual coding algorithm is to obtain an efficient bit rate reduction by exploiting the properties of the human ear. The perceptual coder works on the masking phenomenon of a signal over another in the frequency domain, as perceived by the ear.

A few algorithms modelling the various psychoacoustical effects are already available. In this paper, the psychoacoustic model 1 of [4] was used for the perceptual coder.

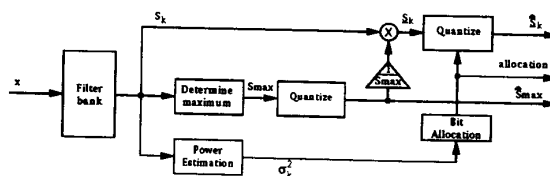


Figure 1a Block diagram of the non-perceptual coder

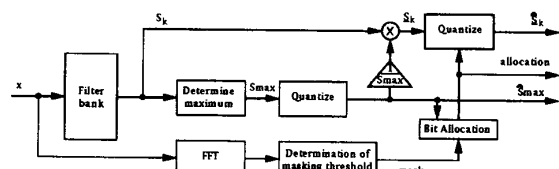


Figure 1b Block diagram of the perceptual coder

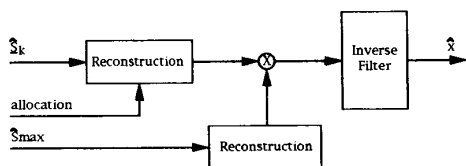


Figure 2 Block diagram of the decoder

THE SUBBAND CODERS

Figure 1a shows the block diagram of the encoder with a simple adaptive bit allocation algorithm while Figure 1b shows the encoder using the perceptual criteria. Both encoders are configured similarly except the block computing the criteria for the bit allocation. The block diagram of the decoder is shown in Figure 2. The same decoder is used to decode the encoded bitstream for the two different encoding schemes.

The filterbank splits the input samples into 24 subbands over a bandwidth of 24 kHz. Quantization of the subband samples is done in blocks of 6 msec segments. This is chosen in view of the pre-masking duration. At 48 kHz sampling, each subband block consists of 12 samples. Three of these blocks of 12 samples are quantized and sent together as a frame to the decoder. S_{max_k} , the absolute peak value is determined by

$$S_{max_k} = \text{maximum } [S_k(i)] \quad \begin{matrix} i = 1, \dots, 12 \\ k = 1, \dots, 24 \end{matrix}$$

Each of the S_{max} value is quantized with a six bit uniform quantizer and sent to the decoder as side information. The subband samples $S_k(i)$ are then normalised with the absolute peak value through

$$S_k(i) = S_k(i)/S_{max}$$

such that overloading of the quantizer does not occur. A uniform quantizer with adaptive step sizes is used to quantize $S_k(i)$. The step size used to quantize a particular block of $S_k(i)$ in a subband is controlled by the bit allocation algorithm. These step size indices are sent to the decoder as side information so that a close replica of the original samples can be reconstructed at the decoder. The complexity of the decoder in this case is kept to the minimum. The reconstructed samples are then passed through the inverse filterbank, where the subbands are synthesised into the output time domain samples.

BIT ALLOCATION ALGORITHMS

Firstly, the bit allocation of the non-perceptual coder is discussed. From three blocks of 12 samples, the power is calculated.

$$\sigma_k^2 = \frac{1}{N} \sum S_k^2(i) \quad i = 1, \dots, N$$

The allocation of bits for the quantization of the subband samples is done adaptively following the procedure outlined below.

1. The subband b , which has the largest power value among the subbands is allocated L number of bits and the power, σ_b^2 is divided by a factor γ .
2. The scaled value σ^2/γ is taken as the new power for that band and step 1 above is repeated until all the available bits for quantization are exhausted.

It was found that when $\gamma = 6$, the best subjective listening quality of the decoded audio was obtained. L is varied according to a table, where it is restricted to a lower range for higher frequency bands and a larger range for lower frequency bands in order to give a better resolution to the lower frequency bands.

In the perceptual coder, better frequency resolution is obtained by performing a Fast Fourier Transform (FFT) on the Hanning-windowed blocks of input samples. The spectral values are then grouped into critical bands where the masking signals are calculated and summed for every masker based on the masking theory of tones and narrow-band noise signals. The result of this masking computation is a set of masking thresholds for every band. These thresholds serve as the criteria for bit allocation.

The basis for this bit allocation procedure is to control the quantization noise such that it is below the masking threshold. As the samples within the a subband is uniformly quantized with L_i number of bits, the power of the quantization noise in subband i can be assumed to be

$$\sigma_{(r,i)}^2 = \frac{1}{12} \left[\frac{2S_{max_i}}{2^{L_i}} \right]^2$$

If $\sigma_{(m, i)}^2$ represents the mask threshold, the mask to noise ratio γ_i is given by

$$\gamma_i = 3\sigma_{(m, i)}^2 \left[\frac{2L_i}{S_{\max_i}} \right]^2$$

The steps for this bit allocation procedure is given below.

1. The subband with the lower γ_i is searched and L number of bits according to the table is allocated to it.
2. The mask to noise ratio is recalculated and taken as the new value for that subband. Step 1 is then repeated until all the available bits are used.

PERFORMANCE OF THE CODERS

The segmental signal to noise ratio is frequently used in the evaluation of a coder's performance. However, this measure need not necessary correlate to the listening results [5].

Two other measures namely the noise to mask ratio and the 'masking flag' indicating the probability of noise occurring were mentioned in [5] and a simple version of both measures were developed to evaluate the outputs of the coders.

A variety of audio sequences were used for evaluation. The quality of the output from the perceptual coder is very good with the impairments almost imperceptible for all the audio sequences tested. The non-perceptual coder while giving very good audio quality for some sequences, has slight impairments in a few of the sequences.

Figure 4 shows the plot of mask to noise ratios versus frequency at a selected frame of the female vocal sequence where the noise in the non-perceptual coder is audible. The plot shows that from 0 to 3 kHz, the MNR is negative, indicating that the quantization noise could not be completely masked, thus resulting in the noise heard. The MNR plot of the perceptual coder on the other hand, shows that the mask margin of the higher frequencies has been reduced to give more margin to the lower frequencies, thereby giving a more balanced response.

Figure 3 indicates that the MNR of the perceptual coder is better than the non-perceptual and this is clearly the case for all the sequences tested. The probability of audible noise measured for the sequences are tabulated in Table 1 as a comparison against the segmental signal to noise ratio (SNR) results. These probability results again affirm the superiority of the perceptual coder. However, it is interesting to note that the figures for the non-perceptual coder is rather good considering its lower complexity.

The segmental SNR results (Table 1) on the other hand indicated that the non-perceptual coder is only slightly better than the perceptual coder, as shown in the SNR plot for the female vocal sequence in Figure 5. In several sequences however, the SNR's recorded are comparable between both coders.

CONCLUSIONS

The quality of audio produced by both coders are very good for most of the audio sequences tested. Very little degradation in the quality from the original audio was heard. However, for some critical sequences, the impairments in the output of the non-perceptual coder is perceptible. The impairments in these critical sequences are hardly noticeable for the output of the perceptual coder. This shows the superiority of the perceptual coder in giving excellent audio outputs. However, the

Table 1 Comparison between the perceptual and the non-perceptual coder

Description	Perceptual Coder		Non- perceptual coder	
	Probability of noise (%)	SNR (dB)	Probability of noise (%)	SNR (dB)
Bass guitar	0.211	36.42	6.32	36.40
Drums & cymbals	0	24.99	5.09	24.89
Glockenspiel	0	28.04	0.46	29.97
Female vocal	0	30.50	7.24	30.6
Trumpet & orchestra	0	43.05	0.09	43.44

Fig 3 MNR comparison (female vocal)

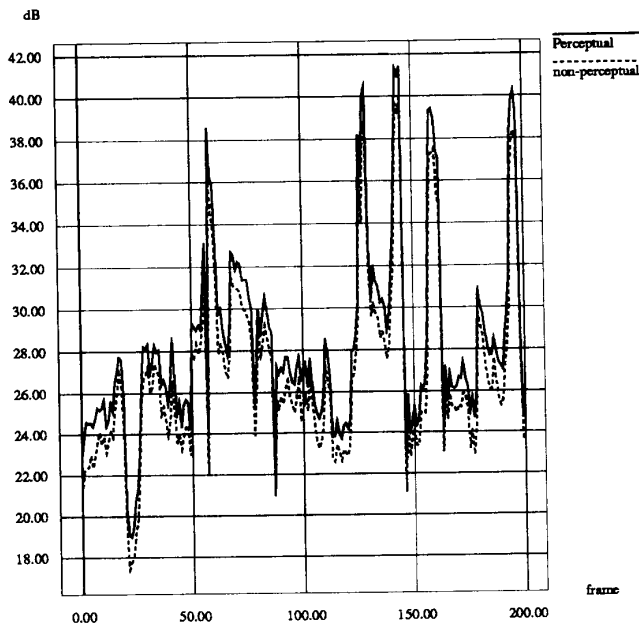


Fig 5 SNR comparison (female vocal)

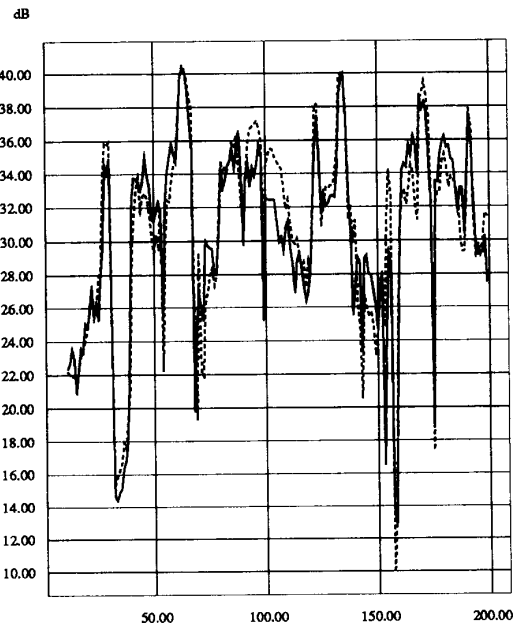
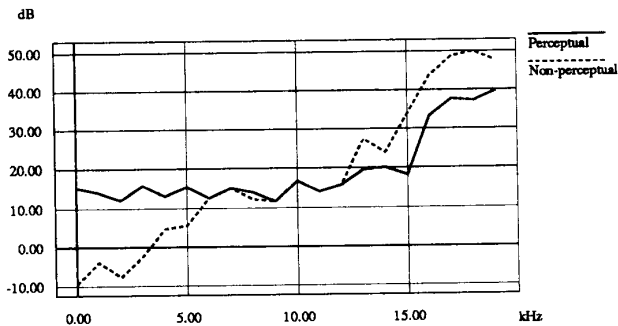


Fig 4 MNR versus frequency (female vocal)



drawback of the perceptual coder lies in its complexity and this is especially crucial in the considerations for hardware implementations in products. The non-perceptual coding algorithm is less complex and is more feasible in applications which are more cost sensitive. While the non-perceptual coder fare marginally better in the segmental signal to noise ratio measure, it is poorer in terms of mask to noise and probability of noise measures. The latter measures correlate better with the subjective quality of the audio signals.

REFERENCES

- [1] JD Johnston, "Transform coding of audio signals using perceptual noise criteria", IEEE J on Sel. Areas in Comm, vol. JSAC-6, Feb 1988.
- [2] G Thiele, M Link, G Stoll, "Low bit rate coding of high quality audio signals", AES Preprint 2432, 1987.
- [3] T A Ramstad, "Subband coder with a simple adaptive bit allocation algorithm - a possible candidate for digital mobile telephony", Proc. of ICASSP '82, Paris, pp. 203-206, May 1982.
- [4] Second draft of proposed standard on information technology coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbits/s, document ISO/IEC JTC1/SC2/WG11 MPEG 90/001, Sept. 1990.
- [5] K Brandenburg, D Seitzer, "Low bit rate coding of high quality digital audio : Algorithms & evaluation of quality", Proceedings of the AES 7th International Conference, May 1989, pp. 201-209.