CRITICAL BAND QUANTISATION ANALYSIS FOR MASKED DISTORTION SPEECH CODING

Paul M. McCourt

Department of Electrical&Electronic Engineering
Queen's University of Belfast
Belfast BT9 5AH, UK
e-mail pm.mccourt@ee.qub.ac.uk

ABSTRACT

This paper presents new results on critical band masked distortion controlled quantisation of a linear transform representation of speech. In particular, fixed rate split vector quantisation of a critical band gain vector is investigated. While shown to be objectively significant in meeting masked distortion criteria, near-transparent quantisation of the critical band gain spectrum is nonetheless achieved at 1.75 kbits/sec. The relevance of this result is explained by a comparative interpretation of the parametric spectral synthesis performed by current analysis-by-synthesis, multi-band excitation and sinusoidal transform coders.

1 INTRODUCTION

Improved understanding and exploitation of psychoacoustic criteria has become recognised as increasingly important in realising future goals for speech and audio compression [1]. Noise shaping with respect to the LPC spectrum is taken into account in code-excited linear predictive coders by 'perceptual filtering' in the analysis-by-synthesis excitation search. Adaptive postfilters [2] increase distortion masking, and have proved significant in the acceptance of low and medium rate CELP coders by standards organisations. Recent techniques for coding of audio signals *directly* shape the noise spectrum based on linear transform quantisation according to a critical band *just noticeable distortion* (JND) objective [3]. 'Perceptual transform coding' has been applied to audio [3,4] and wideband speech [5]. A method for calculating a masked distortion threshold for speech is reported in [6] and a spectral envelope representation based on a critical band decomposition of the Short Term Fourier Transform has been proposed [7]. Quantisation with respect to the critical band distortion threshold is however not explored. In this paper, new results are presented on masked distortion controlled quantisation of a linear transform analysis of speech. In particular, the performance of fixed rate vector quantisation (VQ) of a critical band power spectrum with regard to a masked distortion objective is investigated. The relevance of the results to the envelope quantisation rate requirements of analysis-by-synthesis, multi-band excitation and sinusoidal transform coders is subsequently discussed.

2 CRITICAL BAND DECOMPOSITION

Fig (1) illustrates the critical band power spectrum and threshold of a 16 msec speech segment. The threshold indicates the limiting level of noise power that can be present in each critical band and be effectively masked by the signal.The threshold is calculated according to the method described in [4]. The MDCT (Modified Discrete Cosine Transform), based on a Time Domain Aliasing Cancellation (TDAC) design [8], was used to perform the signal analysis. Quantisation of the MDCT with respect to the critical band threshold was analysed using a formulation of gain-shape vector quantisation. In this framework a 16 dimensional spectral power vector formed by the rms values in each critical band is used to normalise the transform analysis frame. Table(1) defines critical band [9] decomposition of the 128 coefficients of a 16 msec analysis frame. Quantisation of the critical band 'gain' spectrum and the normalised coefficient vectors is performed. The reason for this formulation was motivated by considering the contributions that both quantised components make to the total distortion within each band. Based on this transform quantisation formulation, the distortion *D(b)* in critical band *b* is described by

$$D(b) = \sum_{k=l(b)}^{k=u(b)} (g_n(b)s(k) - \hat{g}_n(b)\hat{s}(k))^2 \qquad (1)$$

where $g_n(b)$ identifies the gain and $s(k)$ the normalised coefficients, with $\hat{g}_n(b)$ and $\hat{s}(k)$ the quantised versions respectively, and $l(b)$ and $u(b)$ the lower and upper coefficient limits of band b. The objective in perceptual transform coding can be defined as performing quantisation in each critical band such that

$$D(b) \le g_{th}(b) \qquad (2)$$

Critical Band *b*	Lower Edge (Hz)	Upper Edge (Hz)	Range *l(b)-u(b)*	Coeff. Width
1	0	100	0 - 3	4
2	100	200	4 - 6	3
3	200	300	7 - 9	3
4	300	400	10 - 12	3
5	400	510	13 - 16	4
6	510	630	17 - 20	4
7	630	770	21 - 24	4
8	770	920	25 - 29	5
9	920	1080	30 - 34	5
10	1080	1270	35 - 40	6
11	1270	1480	41 - 46	6
12	1480	1720	47 - 54	8
13	1720	2000	55 - 63	9
14	2000	2320	64 - 73	10
15	2320	2700	74 - 85	12
16	2700	3150	86 - 100	15

Table(1) Critical Band Decomposition of 16msec MDCT frame

where $g_{th}(b)$ defines the masked distortion threshold. The noise contributions due separately to the quantised gain and quantised normalised coefficient shape vector, with the other unquantised, can alternatively be described by the pair of modified objectives shown by equations (3) and (4).

$$(g_n(b)-\hat{g}_n(b))^2 \le \frac{g_{th}(b)}{\sum_{k=l(b)}^{k=u(b)} s(k)^2} \qquad (3)$$

$$\sum_{k=l(b)}^{k=u(b)} (s(k)-\hat{s}(k))^2 \le \frac{g_{th}(b)}{g_n(b)^2} \qquad (4)$$

These modified objectives suggest that the gain quantisation performance is most directly related to the threshold, with the shape distortion criteria scaled to a much smaller objective. The critical band gain vector and normalised shape vectors can be modelled as providing different information about the transform analysis spectrum. The critical band normalising power vector could be related to providing spectral *envelope* information, with the shape vectors describing the spectral *detail* information.Detailed analyses were performed to assess the separate rate requirements for each element necessary to satisfy the masked distortion objectives. An analysis of normalised coefficient VQ assessed the per-frame total bit rate required to meet the distortion objective by adaptive rate VQ of each intra critical band vector. This analysis is discussed in detail [12]. Despite its smaller contribution to distortion power, a minimum of 11 kbits/sec is shown to be required for shape VQ. This high rate is due to frequency information in a discrete transform analysis being linked to coefficient place, with the basic rate for *full* spectral synthesis therefore necessarily high. Parametric coders of course represent and synthesise the spectral *detail* information more efficiently. The remainder of this paper concentrates on analysing VQ of the critical band gain information.

3 CRITICAL BAND GAIN VQ ANALYSIS

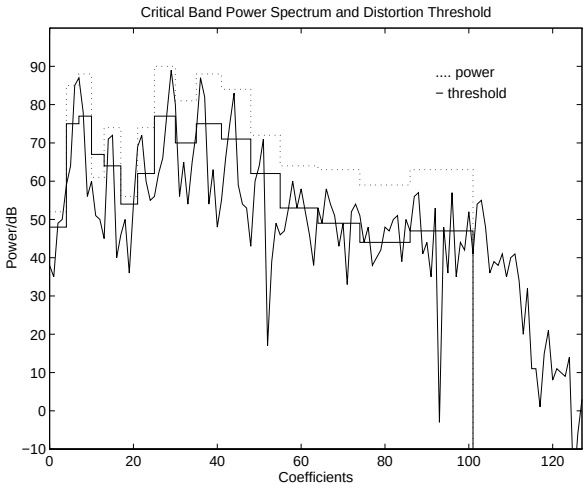
3.1 Analysis Procedure

A detailed analysis was carried out to determine the 'perceptual' performance of fixed rate VQ of the critical band gain vector by assessing the 'occurrence' or probability of masking over a range of quantisation rates for split-vector quantisation of the 16 element vector. The split vector options are specified in Table(2).

gain option	**split vector dimensions**		
	split vector 1	*split vector 2*	*split vector 3*
1	8	8	-
2	4	4	8
3	5	5	6
4	4	5	7

Table(2) Gain Vector Split Options

Fig(2) illustrates, for each gain split-vector option, profiles of the occurrence of successful distortion masking in each critical band over a range of gain codebook sizes. The plots are for a female speech sample of 20 seconds duration. These are representative of results achieved with other speakers. The codebook bit allocation axis indicates the codebook size allocated to *each* split-vector. Slight 'crevices' occur in the transition from the two-stage 12 bit codebook (6+6 bits) to a sin-

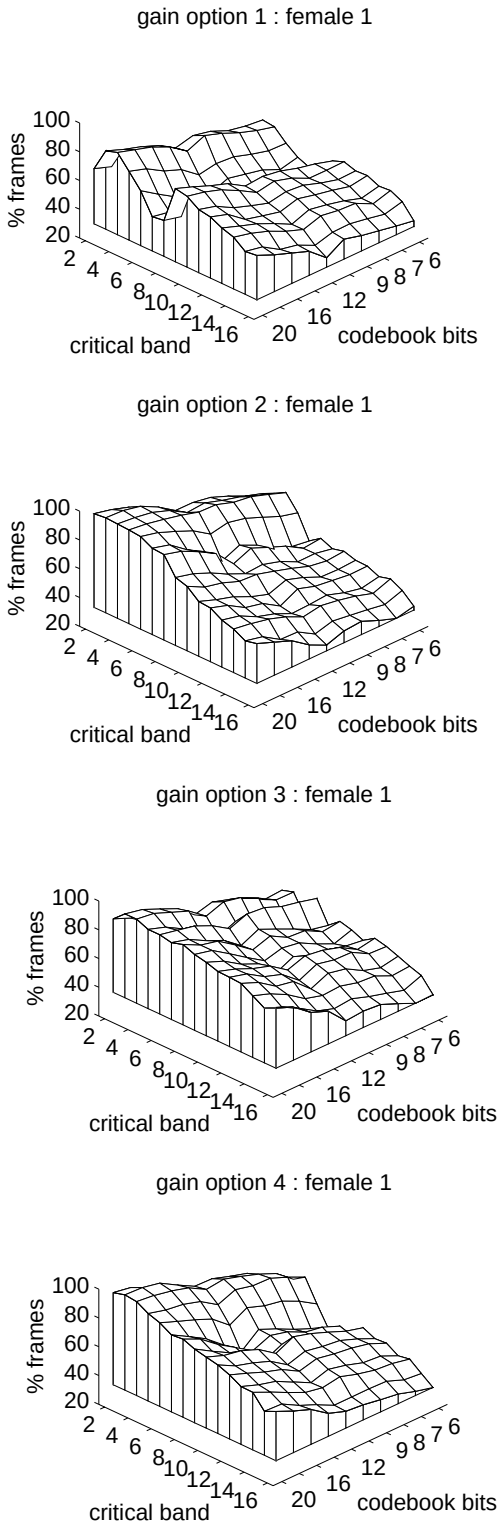


Fig(1) Critical Band Power Spectrum and Distorttion Threshold, 16msec frame

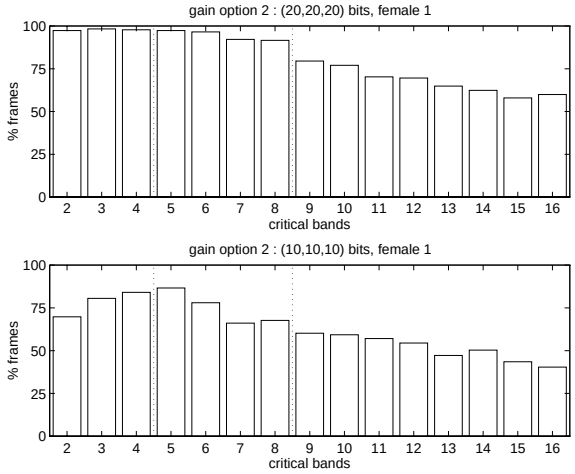
gle stage 10 bit codebook. The reduction in performance is a consequence of the non-optimality of separately designed codebook stages.

3.2 Discussion of Gain VQ Analysis Results

Gain option 1 with only two split vectors has poorest performance while gain option 3 gives the best equitable masking performance over all the critical bands. Gain options 2 & 4 however, achieving higher probability of consistent masking in the lower critical bands, perform better subjectively. The lower probability (0.5-0.6) of meeting the masked distortion objective in the higher bands does not impact significantly on subjective performance. This can be explained by considering the 'meaning' of the critical band threshold. While the threshold indicates the level of narrow-band noise power that can be masked, the speech signal in higher frequency bands is typically modelled by a noise-like unvoiced spectrum. The objective noise criteria therefore appears to have limited subjective validity in the highest critical bands. Fig(3) illustrating cross-sectional plots from Fig(2) for gain option 2, highlights only marginal reduction in masking probability by halving the quantisation rate from 20 to 10 bit codebook allocations. Although the higher quantisation rate ensures consistent probability of meeting the distortion objective in the lower critical bands, the probability of unmasked distortion at the lower rate is reflected in only very slight subjective difference, the coded speech being virtually indistinguishable.This result is significant in that it suggests that for critical band gain information a fixed quantisation rate of 1.75 kbits/sec is sufficient to meet the distortion objective at a probability consistent



Fig(2) Gain VQ Analysis Results



Fig(3) Detail of Gain VQ Results

with near-transparent subjective quality.This value is based on a 10-10-8 bit allocation to split vector 2.

6 FINAL DISCUSSION

The significance of this result can be explained by a comparative interpretation of the spectral *envelope* synthesis performed by current analysis-by-synthesis, multi-band excitation (MBE) and sinusoidal transform coders (STC). In both the MBE and STC coders, envelope analysis is harmonic specific. If the pitch range typical for speech is matched to the critical band scale then, except for the very lowest pitched male speakers, the critical bands contain mostly one or two harmonic components. The results presented here suggest that quantisation of the harmonic amplitudes related to critical band masked distortion objective may allow lower quantisation rates than currently used. These analysis models would in particular provide a ready format for direct implementation of a masked distortion quantisation objective. In comparison, the long term postfilter section of the highly successful adaptive postfilter [2] also effectively forces harmonically resolved noise shaping consistent with a critical band resolution, particularly in the low frequency region. With LPC envelope quantisation at 1.1 kbits/sec, this remains an effective combined route to `indirectly' realising a critical band masked distortion objective than encoder driven shaping.

7 CONCLUSIONS

Quantisation of the Modified Discrete Cosine Transform was analysed with respect to a critical band masked distortion objective in terms of transform spectra decomposed into a critical band gain spectrum and normalised coefficient vectors. Gain quantisation performance is shown to have greater objective significance in matching the masked distortion objective than normalised coefficient VQ. While critical band gain quantisation performance is shown to be objectively significant to ensure distortion masking, a quantisation rate of only 1.75 kbits/sec is nonetheless demonstrated to be sufficient for near transparent coding. The significance of this result was explored with respect to the spectral synthesis models of current low and medium rate speech coders.

Acknowledgement Funding of this project by the Industrial Research and Technology Unit (IRTU) is gratefully acknowledged.

References

[1] N. S. Jayant, “Signal Compression: Technology Targets and Research Directions”, *IEEE Jor. Selected Areas in Communications*, Vol. 10, No. 5, pp. 796-818, June 1992

[2] J.H. Chen and A. Gersho, “Adaptive Postfiltering for Quality Enhancement of Coded Speech”, *IEEE Trans. on Speech and Audio Processing*, Vol. 3, No. 1, January 1995

[3] N. Jayant, J. Johnston and R. Safranek, “Signal Compression Based on Models of Human Perception”, *Proceedings of the IEEE*, Vol. 81, No. 10, pp.1385-1422, October 1993

[4] J. D. Johnston, “Transform Coding of Audio Signals Using Perceptual Noise Criteria”, *IEEE Jor. Selected Areas in Communications*, Vol. 6, No. 2, pp. 314-323, February 1988

[5] S. Quackenbush, “A 7 kHz Bandwidth, 32 kbps Speech Coder for ISDN”, *Proc. IEEE ICASSP*-91, pp. 1-4

[6] D. Sen, D. H. Irving and W.H. Holmes, “Use of an Auditory Model to Improve Speech Coders”, *Proc. IEEE ICASSP*-93,, Vol. II, pp. 411-414

[7] V. R. Algazi et al., “Transform Representation of the Spectra of Acoustic Speech Segments with Applications - II: Speech Analysis, Synthesis and Coding”, *IEEE Transactions on Speech and Audio Processing*, Vol. 1, No. 3, pp. 277-286, July 1993

[8] J. P. Princen, A. W. Johnson and A. B. Bradley, “Subband/Transform Coding Using Filter Bank Designs Basedon Time Domain Aliasing Cancellation ”, *Proc. IEEE ICASSP*-87, pp. 2161-2164

[9] B. Scharf, “Critical Bands”, pp. 157-202, in *Foundations of Modern Auditory Theory, Volume I*, edited by J. V. Tobias, Academic Press, 1970

[10] P. M. McCourt, “Critical Band Masked Disortion Coding of Speech”, *submitted to IEE Proceedings Vision, Image and Signal Processing*