Bit Allocation Algorithms for Frequency and Time Spread Perceptual Coding

Ricky Der[1] *Peter Kabal*[1] *Wai-Yip Chan*[2]

[1] Electrical & Computer Engineering
McGill University
Montreal, Quebec H3A 2A7

[2] Electrical & Computer Engineering
Queen's University
Kingston, Ontario K7L 3N6

Abstract

We examine the problem of bit allocation when time-spread and frequency-spread perceptual distortion criteria are used. For such measures, standard incremental techniques can fail. Two algorithms are introduced for bit allocation; the first a multi-band version of the greedy algorithm, and the second an inverse greedy algorithm initialized by the bit allocation of a forward algorithm driven by a non-spread metric. Experimental results show the second algorithm outperforms the first.

1 Introduction

A transform coder consists of a set of quantizers $\mathbf{q} = \{q_i\}$ acting on transform coefficients $\mathbf{x} = \{x_i\}$ to give quantized coefficients $\tilde{\mathbf{x}} = \{q_i(x_i)\}$. Fidelity is measured with a criterion $D(\mathbf{x},\tilde{\mathbf{x}})$ which is minimized, typically with a bit allocation algorithm, subject to a rate constraint. A classical example for D is the noise-to-mask ratio (NMR) distortion, which in one manifestation reads

$$D(\mathbf{x},\tilde{\mathbf{x}}) = \left(\sum_i \frac{(x_i - \tilde{x}_i)^2}{M_i} \right)^{1/2}, \qquad (1)$$

where $\{M_i\}$ is the masking threshold.

A few observations are in place. First, the criterion is of the form $D(\mathbf{x},\tilde{\mathbf{x}}) = \|\mathbf{x}-\tilde{\mathbf{x}}\|$ where $\|\cdot\|$ is a norm. These types of distortion functions define metric spaces, and have been well-studied in information theory. Second, D is a function of the component distortions: $D = F(D_1,\ldots,D_n)$, where $D_i = (x_i - \tilde{x}_i)^2/M_i$. In the case where F is a sum, efficient algorithms exist for finding the optimal bit allocation minimizing D via dynamic programming or Lagrangian search [1].

Recently, there has been interest in defining measures directly in perceptual spaces. These types of distortion functions are of the form

$$D(\mathbf{x},\tilde{\mathbf{x}}) = \|T(\mathbf{x}) - T(\tilde{\mathbf{x}})\|, \qquad (2)$$

where T is generally a non-linear, non-invertible mapping from the transform domain to the perceptual domain. The function (2) defines a distance between the "internal representations" for the coded and uncoded signals, bypassing the construction of a noise signal $\mathbf{e} = \mathbf{x} - \tilde{\mathbf{x}}$, and, compared with (1), offers a more general approach to modelling psychoacoustical effects [3]. Typical examples for T include the mapping from spectrum to excitation pattern, or loudness pattern. In general, such distortion measures are not weighted Euclidean measures, cannot be decomposed as a function of component distortions, and do not even define metric spaces on the set of transform coefficients $\mathbf{x}$.

This paper studies the use of incremental bit allocation algorithms for a distortion measure of the form (2). We highlight the peculiarities and differences for the rate allocation of (2) as opposed to (1), showing that a standard incremental algorithm can fail to halt when operating in a constant-distortion coding mode. Moreover, the spreading of frequency components in the mapping T can lead to very suboptimal solutions.

To combat the halting problem, a multi-band version of the greedy algorithm is introduced. We also show that under certain model assumptions for T, an algorithm involving "reverse" incremental allocation may be defined. In essence, a forward incremental algorithm on the transform coefficients $\mathbf{x}$ is used as an initialization for the allocation, and then a reverse incremental algorithm on the spread-domain coefficients $T(\mathbf{x})$ removes redundant bits. Many perceptual models involve a mapping T that introduces *time-spreading*. In this case, a dependent allocation problem arises across time, in addition to frequency. The reverse allocation algorithm can be extended, under certain model assumptions, to account for this case, by concatenating another inverse incremental algorithm driven to remove redundancy in the time-spread perceptual coefficients. Experimental results show that the reverse allocation algorithm significantly outperforms the multi-band greedy algorithm under a distortion target imposed on time-spread excitation patterns.

2 Framework

2.1 Constant Distortion Coding

A frame-based audio coder operating under a constant distortion target solves the following optimization problem for every block of incoming data: minimize the rate R subject to the distortion constraint $D(\mathbf{x},\tilde{\mathbf{x}}) < K$. This mode of operation is useful for a variety of reasons. First, transparent coding can be expressed as a constant-distortion criterion: $D(\mathbf{x},\tilde{\mathbf{x}}) = K_0$ for some threshold K_0. Second, a qualitative understanding of supra-threshold distortion incurred from using a particular psychoacoustic model can only be evaluated subjectively when every frame is coded to the same distortion. Third, a distortion-constrained scheme can form the kernel for a rate-constrained scheme. In particular, a constant-distortion engine can find a *relative* allocation of bits b_i, while an outer algorithm adjusts the *absolute* sizes of b_i to meet the rate constraint. This process occurs in the MPEG coding standard, for example, where the inner loop adjusts the relative step-sizes of the quantizers q_i to meet a distortion constraint (masking threshold), while an outer loop varies a global gain factor to meet the rate target. We assume the constant-distortion framework for the remainder of the paper.

2.2 Perceptual variables

One of the most successful concepts in psychoacoustics, excitation patterns provide a unified way of accounting for a dis-

parate range of auditory phenomena, including loudness, masking, just-noticeable changes in amplitude and frequency, and absolute threshold of hearing, to name a few. We shall take them as fundamental perceptual variables. The mapping T from power spectrum to excitation can be broadly summarised in the steps (1) pointwise transformation on the power spectrum (2) frequency spreading, and (3) time spreading.

In what follows, we assume the following generic model: (1) the transformation between the transform coefficient x_i and the pattern B_i just prior to frequency spreading is given by a pointwise mapping $B_i = g_i(x_i)$, and (2) frequency spreading is performed in a p-power law domain, so that the excitation variables E_i at frequency i are produced by

$$E_i = \left(\alpha_{i1} B_1^p + \cdots + \alpha_{in} B_n^p\right)^{1/p}. \tag{3}$$

The subject of additional time-spreading is taken up in Section 3.3. Our procedure is consistent with a number of excitation models, including the ones of [2] and [5].

2.3 Distortion Function

Having selected the perceptual variables, a distortion function must be chosen. For simplicity of discourse, we will work with the function

$$\mathcal{D}_i = 10\log_{10}\left(\frac{\tilde{E}_i}{E_i}\right)$$
$$D = \max_i |\mathcal{D}_i|. \tag{4}$$

with E_i and $\tilde{E}_i$ the excitation patterns corresponding to the unquantized and quantized transform coefficients $\mathbf{x}$ and $\tilde{\mathbf{x}}$ respectively. We call $\mathcal{D}_i$ the distortion pattern. This definition falls into the class (2), and is a very natural measure, since the theory of just-noticeable differences can be formulated with it. In particular, Zwicker's criterion [2] states that two excitation patterns are perceptually indistinguishable if $D < 1$ (dB). Note, however, many of the subsequent results and observations, in particular the theorems of Section 3.2 will still hold if the logarithmic function is replaced with an increasing compressive nonlinearity g.

3 Incremental Allocation Algorithms

Consider the problem of achieving the distortion constraint $D(\mathbf{x},\tilde{\mathbf{x}}) < K$ with minimum rate. One method is to use the so-called greedy algorithm, which, beginning with an initial distribution of zero, allocates one additional bit to the component resulting in the largest decrease in D at each iteration. In general, it is possible that the distortion will not change or actually increase with every possible test allocation. In such a case, the test quantum is incremented to two bits, three bits, and so forth, until one such allocation achieves a decrease in distortion. The procedure continues until the distortion target K is met.

Now, suppose that the fidelity criterion is of the form $D = F(D_1,\ldots,D_n)$ for a function F increasing in each dimension, and where the individual distortions D_i are metrics on $(x_i,\tilde{x}_i)$. If the rate-distortion pairs (b_i,D_i) are such that $D_i \to 0$ as $b_i \to \infty$, then it is easy to see that, at each iteration, there exists a band i for which the allocation increase $b_i = b_i + n$ decreases D, for some n. This is because D_i decreases for large enough b_i and hence D decreases for that choice as well. The greedy algorithm, for this case, produces a monotonic decrease in distortion at every iteration.

The above discussion applies, in particular, to the NMR distortion function of (1). The function $F = \sum_i D_i^2/M_i$ is an increasing function in each of the individual error differences $D_i \equiv |x_i - \tilde{x}_i|$. A well-designed sequence of quantizers $q_{i,b}$ is one for which $\lim_{b\to\infty} q_{i,b}(x) = x$, which ensures that $D_i \to 0$ as $b_i \to \infty$.

Consider now the application of the greedy algorithm to the distortion function of (4). The presence of the spreading function (3) makes it impossible for the distortion D to be written as any function of component distortions $D_i(x_i,\tilde{x}_i)$. Another way of stating this is that the excitation distortion pattern $\mathcal{D}$ in band i is not only a function of the quantizer q_i, but the quantizers in the vicinity of band i. In general, the allocation of an extra bit to a single band will have effects on the distortion pattern $\mathcal{D}$ of multiple bands.

As an example, consider a bit allocation which reaches the excitation distortion pattern $\mathcal{D}$ of Fig. 1. In this scenario, the distortion is negligible in all bands except for two adjacent bands n and $n+1$. From the graph, it is clear that the transform coefficient corresponding to band n is quantized to a value with larger excitation strength than the original, while the coefficient of band $n+1$ is quantized to a value with smaller power than the original. Now, if the reconstructed coefficient in band n is improved (in any metrical sense), then the excitation distortion in band n reduces — but because of spreading, the power leaking into band $n+1$ also reduces, *increasing* the distortion in band $n+1$. Alternatively, if the reconstructed coefficient in band $n+1$ is improved, then the power of band $n+1$ increases, leading to a decrease in distortion in that band, but an *increase* in distortion for band n. Thus the incremental algorithm experiences a deadlock: no improvement in overall distortion results from an allocation to either band n or $n+1$. The consequences can be two-fold: (1) allocation eventually occurs in a region of non-interest, such as a band with negligible distortion, leading to extremely sub-optimal solutions, or (2) the algorithm enters an infinite loop, in the case that no allocation leads to a decrease in overall distortion.

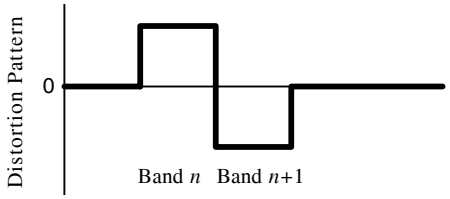

Fig. 1 Deadlock

The above situation, though seemingly factitious, arises in approximate form sufficiently often in practice to give qualitatively similar outcomes. It is not dependent upon the particular form of the distortion function (4), but rather a general outcome of spread-domain operation. For example, though exacerbated by the presence of the $\max_i|\cdot|$ function of (4), deadlock can still occur even when the operator is exchanged with the more general integrative function $(\sum_i |\cdot|^p)^{1/p}$.

3.1 Multi-band Greedy Algorithm

We have seen that spread domain distortion functions pose serious problems for the standard greedy algorithm. The crux of the issue involves the frequency spreading which no longer confines local quantization error within the band. An incremental algorithm improving single bands at each iteration can fail to halt. This immedi-

ately suggests, however, that a multi-band version of the foregoing algorithm, allowing for the simultaneous update of multiple quantizers, might avoid these pitfalls.

The introduction of multi-band allocation introduces a large complexity increase. There are in general $\binom{n}{k}$ ways to select k quantizers out of a set of n. The complexity increase becomes tolerable, however, by observing that the spreading function decreases monotonically at an exponential rate from its maximum. Changes in the quantization of a single band only affect a *contiguous* group of bands. By allowing only updates to contiguous groups, the number of test allocations is constrained considerably, to $n-k+1$. An algorithm description follows.

Algorithm 1: Multi-band Greedy
The bit distribution is initialized to zero. The following parameters are introduced: a mandatory improvement factor at each iteration $\beta < 1$, and an upper integer bound $\eta > 0$ on the size of the allocation quantum before increasing the band-range. At each iteration, a bit is allocated to the quantizer producing the largest decrease in overall distortion D. If the distortion does not decrease by at least the factor β with any such allocation, the search is repeated with the test quantum increased to two bits, three bits, and so forth, up to a maximum of η bits. Should no such allocation still produce the desired distortion, the entire procedure is repeated with the test quantum reset to one, except contiguous groups of two bands are tested simultaneously. If the target decrease is still not reached, groups of three, four etc. are tested, up to, if required, a simultaneous refinement in all quantizers. Once the mandatory distortion reduction β is achieved, the algorithm proceeds with the next iteration, continuing until the final target K is attained.

As stated, this multi-band algorithm will terminate not only with the distortion function of (4), but more generally with the criterion $D = \|T(\mathbf{x}) - T(\tilde{\mathbf{x}})\|$. The allocation procedure is guaranteed to decrease the distortion by the factor β every iteration, since, in the worst case, all n bands can be simultaneously refined so that $\sum_i |x_i - \tilde{x}_i| < \delta$ for large enough b_i. Both the excitation transformation T and the norm are continuous in the metric $\sum_i |x_i - y_i|$ for finite-dimensional spaces, implying then $D < \varepsilon$ for any $\varepsilon > 0$ with sufficiently large b_i. Finally, we note that the single-band greedy algorithm is a special case of the multi-band algorithm in the limits $\beta \to 1$ and $\eta \to \infty$.

3.2 Reverse Allocation Algorithm

An interesting bit allocation algorithm can be derived by restricting attention to (4). To develop the algorithm, first consider the following ostensibly unrelated lemma.

Lemma 1 *Let $\alpha_1, \alpha_2, \beta_1, \beta_2$ be positive real numbers. Then the following holds:*

$$\min\left\{\frac{\alpha_1}{\alpha_2}, \frac{\beta_1}{\beta_2}\right\} \le \frac{\alpha_1+\beta_1}{\alpha_2+\beta_2} \le \max\left\{\frac{\alpha_1}{\alpha_2}, \frac{\beta_1}{\beta_2}\right\} \tag{5}$$

Proof: We will only show the case where $\frac{\alpha_1}{\alpha_2} \le \frac{\beta_1}{\beta_2}$; the reverse case has an identical proof because of symmetry. In this scenario, the right inequality holds if and only if $\beta_2(\alpha_1+\beta_1) \le \beta_1(\alpha_2+\beta_2)$, which holds if and only if $\alpha_1\beta_2 \le \beta_1\alpha_2$, which is true. The left inequality holds if and only if $\alpha_1(\alpha_2+\beta_2) \le \alpha_2(\alpha_1+\beta_1)$, holding if and only if $\beta_2\alpha_1 \le \beta_1\alpha_2$; true of course, hence we are done. ■

The lemma's usefulness comes in the proof for Theorems 1 and 2.

Theorem 1 *Let α_i, B_i, $\tilde{B}_i$, $i = 1 \ldots n$ be three sequences of positive real numbers. Let K and p be positive real numbers, and suppose that*

$$10^{-K/10} < \frac{B_i}{\tilde{B}_i} < 10^{K/10} \tag{6}$$

for every i. Then the following holds:

$$10^{-K/10} < \left(\frac{\alpha_1 B_1^p + \cdots + \alpha_n B_n^p}{\alpha_1 \tilde{B}_1^p + \cdots \alpha_n \tilde{B}_n^p}\right)^{1/p} < 10^{K/10} \tag{7}$$

Proof: The case $n = 1$ is trivial. For $n = 2$, assume that $\frac{B_1}{\tilde{B}_1} < \frac{B_2}{\tilde{B}_2}$ with no loss of generality. By Lemma 1, we have:

$$\left(\frac{B_1}{\tilde{B}_1}\right)^p < \left(\frac{\alpha_1 B_1^p + \alpha_2 B_2^p}{\alpha_1 \tilde{B}_1^p + \alpha_2 \tilde{B}_2^p}\right) < \left(\frac{B_2}{\tilde{B}_2}\right)^p \tag{8}$$

$$10^{-Kp/10} < \left(\frac{\alpha_1 B_1^p + \alpha_2 B_2^p}{\alpha_1 \tilde{B}_1^p + \alpha_2 \tilde{B}_2^p}\right) < 10^{Kp/10} \tag{9}$$

as desired. An induction argument generalises the result to all n. ■

The above theorem can be interpreted as follows: the B_i's are the values of the unspread perceptual pattern used as an argument in (3), and the numerator and denominator of (7) represent spread excitation patterns in the power domain p of the original and coded signals, respectively. The bounds of (6) are equivalent to the requirement that

$$D' = \max_i |10\log_{10}(B_i/\tilde{B}_i)| < K \tag{10}$$

or, expressed in words, that the unspread reference and reproduced patterns are within K dB of another. The theorem then states that reference and reproduced excitation patterns are also within K dB of another.

Thus matching the *unspread* patterns to within K dB, *suffices* to match the respective (spread) excitations to within K dB. The key point is that the B_i's are obtained by a simple pointwise mapping $g_i(\cdot)$ on the transform coefficients, and hence the distortion function of (10) is not in a spread domain. The standard single-band greedy algorithm can then be expected to find a reasonable solution. Once the target of (10) is achieved, the excitation distortion of (4) is automatically bounded by K. In general, the patterns will tend to overshoot the bound: they will be over-coded. They will not be over-coded a great deal, however, since the bounds of (9) will be achieved by some frames. An *inverse* incremental algorithm can then proceed to perform a correction: redundant bits are removed until the excitation distortion just meets the constraint K. The foregoing can be summarised as follows:

Algorithm 2: Reverse Allocation
Initialize the bit distribution to zero. A standard greedy algorithm allocates bits driven by the distortion function and the target of (10). When the algorithm terminates, the final bit allocation is used as an initialization for the reverse algorithm. At each iteration, the bit is removed from the quantizer which results in the smallest updated excitation distortion, as computed by (4). The process continues until the constraint $D < K$ is first breached; the last allocation for which the target is achieved is retained.

It is important to observe that though a single-band incremental algorithm driven by a spread distortion function is used

for the removal of bits, it does not suffer from the same issues as a forward algorithm. For instance, the halting problem does not occur here since a decrease in distortion in the reverse allocation—though generally unexpected—is a positive result, whereas a generally unexpected increase in distortion in the forward algorithm is a negative outcome. Indeed, the reversal of priorities in the inverse algorithm can transform the weaknesses of the forward greedy algorithm into strengths in the reverse case.

3.3 Extension to Time Spreading

The inverse incremental algorithm has the interpretation of removing the bits from components which are partially masked by adjacent frequency components, as a result of the spreading. It is natural to ask whether an analogous result to Theorem 1 can be obtained for temporal masking, as expressed by time spreading. It turns out that this is possible.

To begin, we will assume that the time-spread excitation pattern $F_n(i)$ at frequency i and time n is given by the time-varying autoregressive system

$$F_n(i) = a_n(i)F_{n-1}(i) + b_n(i)E_n(i), \tag{11}$$

where $E_n(i)$ is the non-time-spread excitation, and parameters $a_n(i), b_n(i) \geq 0$. The equation generalises a number of models for time-smearing, among them [4]. Now we have the following theorem.

Theorem 2 *Suppose that*

$$10^{-K/10} < \frac{F_{n-1}(i)}{\tilde{F}_{n-1}(i)} < 10^{K/10}, \tag{12}$$

$$10^{-K/10} < \frac{E_n(i)}{\tilde{E}_n(i)} < 10^{K/10} \tag{13}$$

for all i. Then the following holds:

$$10^{-K/10} < \frac{F_n(i)}{\tilde{F}_n(i)} < 10^{K/10}, \quad \forall i \tag{14}$$

Proof: The model (11) implies

$$\frac{F_n(i)}{\tilde{F}_n(i)} = \frac{a_n(i)F_{n-1}(i) + b_n(i)E_n(i)}{a_n(i)\tilde{F}_{n-1}(i) + b_n(i)\tilde{E}_n(i)}. \tag{15}$$

The statement is then immediate by Lemma 1. ■

This proposition is a time-domain version of Theorem 1. In particular, with the initial conditions $F_{-1}(i) = \tilde{F}_{-1}(i) = 0$, we have by induction the corollary that matching the non-time-smeared patterns E_i, $\tilde{E}_i$ to within K dB suffices to automatically bring the time-smeared patterns $F(i)$, $\tilde{F}(i)$ to within K dB. This suggests that an algorithm for finding an allocation satisfying distortion constraints on patterns spread in both frequency and time can be formulated as follows:

Algorithm 3: Reverse Allocation for Frequency and Time-Spread Patterns

For each frame n, use Algorithm 2 to determine a bit allocation to be used as an initialization. Since the non-time-spread excitations are now within the distortion constraint, so must the time-spread excitations, given that the previous frame's time-spread excitations satisfy (12). A second inverse greedy algorithm removes redundant bits, driven by the distortion function of (4) with variables $E_i, \tilde{E}_i$ replaced by the time-spread counterparts $F(i), \tilde{F}(i)$. The procedure continues until the target is just achieved. Now the bounds of (12) are satisfied for frame n, which prepares the algorithm for the next time frame $n+1$.

4 Simulation Results

The relative performance of the algorithms presented above can be evaluated by applying each to a constrained-distortion audio coder and computing the respective rates. We use a transform coder with scalar quantizers in the discrete Fourier domain, structurally similar to the one presented in [6]. A single bit quantum is associated with an increase or decrease of quantizer step-size by the factor 0.9. The procedure of [4], with small variations, is used to compute the time-spread excitation patterns on a resolution of 1 Bark. Bit allocation algorithms 1 and 3 are applied to each frame of data with the aim of meeting the distortion constraint of (14); or equivalently, to match the time-spread excitation patterns to within K dB using a minimum number of bits. For the multi-band greedy algorithm, some tuning of the parameters β and η are required; we found a good operating point at $\beta = 0.999$ and $\eta = 3$.

The output is constant-distortion file (in the sense of Sec. 2.1), with a different rate for each frame. Overall performance is measured by computing the empirical entropy of each quantizer output across time, and then averaging over all frequency bands. The resulting rate-distortion curves, with the distortion target K swept from 1 dB to 10 dB, are plotted in Fig. 2 for a test sample of 4 speech utterances.

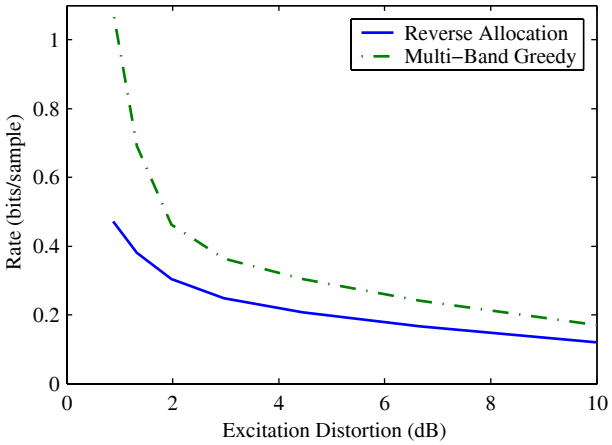

Fig. 2 Empirical Entropy vs. Target Distortion

These curves demonstrate that, for all target distortion values, the reverse allocation algorithm finds a bit distribution meeting the distortion constraint at approximately half the rate of the multi-band incremental algorithm.

References

[1] Y. Shoham, A. Gersho. "Efficient Bit Allocation for an Arbitrary Set of Quantizers," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 36, pp. 1445–1453.

[2] E. Zwicker, H. Fastl. *Psychoacoutics: Facts and Models*, Springer-Verlag, second edition, 1999.

[3] R. Der, P. Kabal, W. Chan, "Towards a New Perceptual Coding Paradigm for Audio Signals", *Proc. ICASSP*, 2003.

[4] ITU-R, Geneva. *Recommendation BS.1387-1, Methods for Objective Measurements of Perceived Audio Quality*, Nov. 2001.

[5] B. Glasberg, B. Moore. "Derivation of auditory filter shapes from notched-noise data", *Hearing Research*, 47, pp. 103–138, 1990.

[6] J. D. Johnston. "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Selected Areas of Commun.*, vol. 6, pp. 314–323, Feb. 1988.