THE SpEnt METHOD FOR LOSSY SOURCE CODING†

Jerry D. Gibson, Mark G. Kokes, and Justin Ridge

Department of Electrical Engineering
Southern Methodist University
Dallas, TX 75275-0338
{gibson, kokes, jridge}@seas.smu.edu

ABSTRACT

At present, the most successful methods for lossy source compression are sample-function adaptive coders. Prominent examples of these techniques are the still image compression methods utilizing wavelet expansions and tree structures, such as the zero-tree method or the SPIHT algorithm, and variable rate speech coders that allocate bits to parameters within a frame based upon the classification of the current frame. All of these techniques can be classified as non-linear approximation methods.

In this work, we use Campbell's coefficient rate, and the spectral entropy of the source random process, as a guide to formulate a new non-linear approximation method to lossy source compression. We call this new approach, the spectral entropy (SpEnt) method, and we develop and report on the promise of SpEnt based coders for the lossy compression of still images and wideband speech (50 Hz to 7 kHz).

1. INTRODUCTION

The most successful methods for lossy source compression today are sample-function adaptive coders (also called input-by-input adaptive or realization-adaptive coders). Prominent examples of these techniques are the still image compression methods utilizing wavelet expansions and tree structures, such as the zero tree method [1] or the SPIHT algorithm [2], and variable rate speech coders that allocate bits to parameters within a frame based upon the classification of the current speech frame [3]. In sample function adaptive coders, not only might the *number* of parameters transmitted in each block (frame) vary from block-to-block (frame-to-frame), but for a given number of transmitted parameters, *which* parameters are transmitted in each block may vary. For such coders with a fixed set of basis functions, it is usually said that the coefficients corresponding to the best n basis functions are sent, rather than the first n, and this is called nonlinear approximation in harmonic analysis.

2. CAMPBELL'S COEFFICIENT RATE

Campbell derived the quantity that he called the coefficient rate of a random process in 1960 [4]. To do so, he considered the product of N sample functions of a random process, and showed using an (AEP)[1]–like argument, that a Karhunen–Loeve expansion of this product could be separated into two sets −− a set with average power very close to the average power in the product and the other set with very low average power. Campbell showed that asymptotically in the number of sample functions forming the product, and in the support interval of the process, the average number of terms in the high power set approached a quantity that he interpreted as a coefficient rate given by

$$Q = \exp\left[-\int_{-\infty}^{\infty} S(f) \log S(f) df \right], \tag{1}$$

where we denote the quantity in the exponent as the spectral entropy. No coding theorems were proved and no possible implications of coefficient rate for source compression were stated.

In [5]-[9], two new derivations of Campbell's coefficient rate where developed. One derivation allows coefficient rate to be interpreted with respect to a quantity defined as the *equivalent bandwidth* of a process. The other derivation reveals a new approach to source compression based upon coefficient rate that adapts to each realization of the source. More specifically, by studying the dominant terms in the series expansion of the product of sample functions, it has been shown that in a sequence of N samples of a particular coefficient, the number of coefficient samples that should be coded is proportional to the coefficient variance [9]. Thus, whether a particular coefficient is being coded or not is changing from block-to-block, and therefore, this new technique clearly falls into the class of nonlinear approximation methods.

3. EQUIVALENT BANDWIDTH

Consider a stationary random process $X(t)$ with a normalized spectrum $S(f)$. Since $S(f)$ is normalized, it can be treated as a probability density function and probability can be defined as $P(F) = \int_F S(f)df$, where F is a set defined along the frequency axis and $P(F)$ is the power of the process in the frequency band defined by F. According to the AEP for continuous random variables [10], the

†This research was supported, in part, by NSF Grant NCR-9796255.

[1] Asymptotic Equipartition Property

volume of a N-dimensional set $F^{(N)} \in \mathrm{R}^N$ is defined as

$$\mathrm{Vol}(F^{(N)}) = \int_{F^{(N)}} df_1 df_2 \ldots df_N. \qquad (2)$$

Given the physical meaning of $S(f)$, the volume of set $F^{(N)}$ has a meaning related to signal bandwidth. For example, in one dimension, if F is a continuous set, $\mathrm{Vol}(F)$ is just the bandwidth of the signal; if F consists of several separate subsets along the frequency axis, then $\mathrm{Vol}(F)$ is the sum of the bandwidths of these subsets. Now, let the support of $S(f)$ be the edges of $F^{(N)}$ in all dimensions, then $\mathrm{Vol}(F^{(N)}) = 1$.

The typical set $F_\epsilon^{(N)}$ is then defined such that

$$P(F_\epsilon^{(N)}) > 1 - \epsilon, \qquad (3)$$

that is, $F_\epsilon^{(N)}$ contains most of the power in set $F^{(N)}$. Using the AEP, the volume of the typical set $F_\epsilon^{(N)}$ that contains most of the power satisfies

$$(1-\epsilon)\mathrm{e}^{N(h(S)-\epsilon)} \le \mathrm{Vol}(F_\epsilon^{(N)}) \le \mathrm{e}^{N(h(S)+\epsilon)}, \qquad (4)$$

where $h(S)$ is the differential entropy of the spectrum, or *spectral entropy.* As $N \to \infty$, $(1-\epsilon)^{\frac{1}{N}} \to 1$, and so on a per dimension basis, the equivalent bandwidth of the random process,

$$W_e = \frac{1}{2}[\mathrm{Vol}(F_\epsilon^{(N)})]^{1/N}, \qquad (5)$$

satisfies

$$\mathrm{e}^{h(S)-\epsilon} \le 2W_e \le \mathrm{e}^{h(S)+\epsilon}. \qquad (6)$$

If we call $2W_e$ the equivalent rate of the random process, i.e., $R_e = 2W_e$, then the equivalent rate of the random process is within a small range around $\mathrm{e}^{h(S)}$, which is the coefficient rate derived by Campbell.

4. DOMINANT TERMS IN THE EXPANSION

We now explore Campbell's result further to estimate asymptotically the variance of the surviving terms after truncation of

$$y_\mu(t_1, t_2, ..., t_N) = \sum_{k=1}^{\mu} c^{(k)}\phi^{(k)}(t_1, t_2, ..., t_N) \qquad (7)$$

as $N \to \infty$ and to calculate the number of these terms. Since all of the sample functions are independent, for any random variable $C^{(k)} = C_{i_1}C_{i_2}\cdots C_{i_N}$, in which each C_{i_j} is one of the random variables in the K-L expansion,

$$\begin{aligned}\mathrm{E}[(C^{(k)})^2] &= \mathrm{E}[C_{i_1}^2]\mathrm{E}[C_{i_2}^2]\cdots\mathrm{E}[C_{i_N}^2] \\ &= \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_N}.\end{aligned} \qquad (8)$$

The total average energy of the product is

$$\begin{aligned}\mathrm{E}\big[y(t_1,t_2,\ldots,t_N)^2\big] &= \sum_{k=1}^{M^N}\mathrm{E}\big[(C^{(k)})^2\big] \\ &= \prod_{j=1}^{N}\sum_{i=1}^{M}\lambda_i = T^N.\end{aligned} \qquad (9)$$

There are a total of M^N terms in $\mathrm{E}[y(t_1,t_2,\ldots,t_N)^2]$, but some of them are very small compared to others. As $N \to \infty$, the large energy terms will dominate the sum and small energy terms can be thrown away without significantly affecting the total energy. On the other hand, each of these M^N terms is a product of N energy terms chosen from the M K-L expansion components. Since $M < N$, the product can be rewritten as $\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}$ and we can see that there is some repetition among those M^N terms; therefore, (9) can be rewritten as

$$\mathrm{E}\big[y(t_1,t_2,\ldots,t_N)^2\big] = \prod_{j=1}^{N}\sum_{i=1}^{M}\lambda_i = \Big(\sum_{i=1}^{M}\lambda_i\Big)^N = \sum_{\{n_i:\sum_{i=1}^{M}n_i=N\}}\frac{N!}{n_1!n_2!\cdots n_M!}\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}. \qquad (10)$$

The coefficient before the energy term is the count of repetitions of that energy term. We want to find the largest term of $\frac{N!}{n_1!n_2!\cdots n_M!}\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}$ subject to the constraint that $\sum_{i=1}^{M} n_i = N$. As $N \to \infty$, this term grows much faster than any other term, and it dominates the total energy of $y(t_1,t_2,\ldots,t_N)$. So the problem becomes finding the maximum of the functional

$$I = \log\Big\{\frac{N!}{n_1!n_2!\cdots n_M!}\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}\Big\} + \alpha\sum_{i=1}^{M}n_i. \qquad (11)$$

Using the approximation $\log N! = N\log N - N$ for large N, taking partial derivatives of I with respect to n_i and setting them to zero, and by using the constraint $\sum_i n_i = N$, we obtain[2]

$$n_i = \frac{\lambda_i}{T}N, \qquad i = 1,2,\ldots,M. \qquad (12)$$

That is, the number of λ_i in $\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}$ is proportional to λ_i. This is a key result. Let $p_i = \lambda_i/T$, since $\sum_i \lambda_i = T$, $\sum_i p_i = 1$, and we have $n_i = p_i N$.

We can find the value of these energy terms and the number of them. We have

$$\begin{aligned}\sigma^2 &= \lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M} = \exp\big[\sum_i n_i\log\lambda_i\big] \\ &= T^N\mathrm{e}^{-NH(S)},\end{aligned} \qquad (13)$$

and the number of such terms is

$$\begin{aligned}\mu &= \frac{N!}{n_1!n_2!\cdots n_M!} = \exp\Big[-N\sum_i\frac{n_i}{N}\log\frac{n_i}{N}\Big] \\ &= \exp\Big[-N\sum_i\frac{\lambda_i}{T}\log\frac{\lambda_i}{T}\Big] = \mathrm{e}^{NH(S)},\end{aligned} \qquad (14)$$

where $H(S) = -\sum_i(\lambda_i/T)\log(\lambda_i/T)$, the spectral entropy in discrete form. This agrees with Campbell's result. We also have the asymptotic relation $T^N = \mu\sigma^2$, so we can see that when $N \to \infty$, we can use these μ equal energy terms to approximate $y(t_1,t_2,\ldots,t_N)$, and the energy loss introduced by truncation is negligible.

[2] We have only demonstrated a stationary point here. The divergence inequality can be used to show that Eq. (12) yields a maximum. We omit details due to space limitations.

5. ADAPTIVE SOURCE COMPRESSION

We have presented two alternative derivations of Campbell's coefficient rate result, and each provides insight into the physical situation of interest. To elaborate on these ideas, it is useful to recall the approach used in two-dimensional discrete cosine transform coding of images. The basic approach in classical transform coding is to take samples of an image, apply a two-dimensional discrete transform to a particular image block, assign the number of bits to be used to encode each coefficient, and then quantize and code each coefficient using the allocated number of bits [11, 12]. Coefficients allocated zero bits are not coded at all. Most bit allocation rules assign bits to a coefficient in direct proportion to the coefficient variance −− a higher variance receives a greater share of the bits to be allocated. If the bit allocations are determined once and then held fixed for all encodings, the decoder can be sent this information once, and the rate required for this information is asymptotically negligible.

The bits allocated to each coefficient for a block are called the *side information*, and a binary indicator showing which coefficients are encoded and which are not transmitted at all is called the *significance map*. Most transform (or wavelet-based) coders used today adaptively allocate bits on a block-to-block basis so the side information and significance map are continuously changing. Thus, this information needs to be provided to the receiver relatively often (or surmised from the data structure).

The equivalent bandwidth result in Eq. (6) gives a connection between coefficient rate and the signal bandwidth, but this does not necessarily imply that the coefficient rate is the number of samples one should use for the random process. Instead, the sampling rate may still be the Nyquist rate, but the importance of the different samples may be different; thus, the significance map should be adjusted according to the importance of the samples.

The results on dominant terms in the expansion have a more direct connection to coder design, and in fact, imply a novel coder structure. First, observe that these results have exactly the kind of interpretation that is expected from an AEP approach. Namely, the number of terms in the high power set is related to the entropy of the (spectral) density as given by Eq. (14), and each coefficient in the high power set is about the same and can be found from Eq. (13). The new implication for coding comes from Eq. (12). Equation (12) says that in a sequence of N samples of a particular coefficient, the *number* of coefficient samples that should be coded is proportional to the variance of the coefficient!

There are two new ideas in this implication. First, the coder implied here "looks ahead" at N coefficient samples. This implies a delay of N blocks. Second, in this sequence of coefficients, we are not allocating *bits* according to the coefficient variance, but we are determining how many *coefficient samples* in the sequence to send according to the coefficient variance. Why would this make sense at all? First, from the AEP, each of the terms retained in the expansion for the high power set all have about the same power. Second, the coefficient rate is not a coding paradigm. There is no coding theorem as yet. These ideas and the implied source coder structure are made clearer in the following.

6. SpEnt-BASED CODER DESIGNS

The basic approach to source compression implied by the spectral entropy result is illustrated in Fig. 1. In Fig. 1a, we show M transform coefficients or coefficients of basis functions for N frames of source data, denoted c_{ij}, $i = 1, ..., M, j = 1, ...N$. The coefficients in the first frame or block are c_{i1}, $i = 1, 2, ..., M$, while the coefficients for the second block are c_{i2}, $i = 1, 2, ..., M$, and so on. Therefore, the block index is indicated by the second subscript (j) and the coefficient index in indicated by the first subscript (i). In classical transform based coding, coefficient bit allocation is accomplished on a frame-by-frame or block-by-block basis as illustrated in Fig. 1b. That is, given a particular frame or block (fixed j), a fixed number of bits are allocated across the M coefficients according to their relative energies. The spectral entropy approach implies that each transform coefficient should be considered as a separate sequence, C_{ij}, $j = 1, ..., N$, as shown in Fig. 1c, and the significant values of that coefficient in the sequence should be determined by comparing to a threshold derived from the coefficient energy. The spectral entropy does not specify a coding or quantization method.

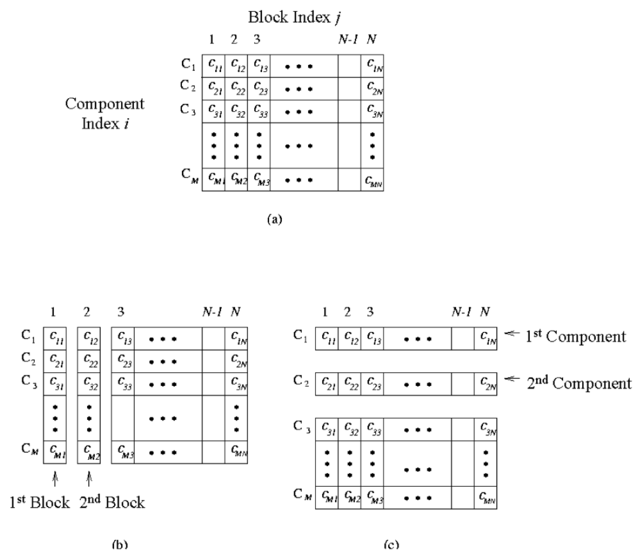

Figure 1: Encoding Transform Coefficients. (a) Coefficients of N Sample Functions, (b) Encoding Smple Function by Sample Function, (c) Encoding Component by Component.

In the remaining sections, we present example coders for still images and wideband speech which are based upon the SpEnt method.

A SpEnt-Based Image Coder

A block diagram of the SpEnt-based image coder can be seen in Fig. 2. In this example, the coder performs an 8×8 DCT for all the data blocks in the entire image (so for a 512×512 image, there are N=4096 data blocks). These data blocks are treated as independent sample functions, although in practice, they are often highly correlated.

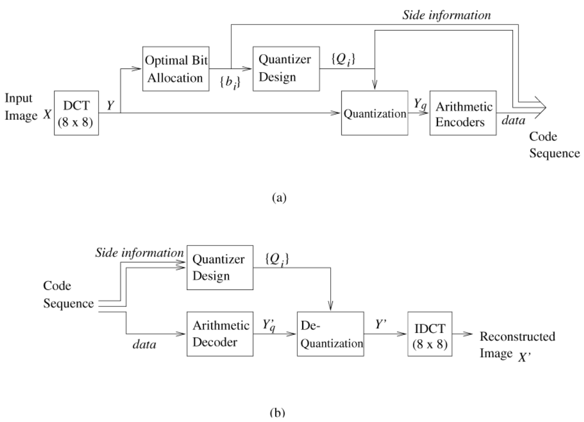

Figure 2: The New Image Coder, (a) Encoder, (b) Decoder.

The spectrum of the transform coefficients is estimated by calculating the variances of the 64 coefficients averaging over all the data blocks. After the spectrum is estimated, the coder determines how many coefficients need to be encoded for each component. The goal is to keep the number of coded coefficients of each component proportional to the average energy of the component, and thus the distribution of the significant coefficients is represented by the normalized spectrum of the source.

After the number of coefficients to be coded is determined, we build the significance map for each data block by selecting the n_i coefficients with the highest energy of each component. For each coefficient to be coded, we use a uniform quantizer. Following quantization, the quantized coefficients are entropy coded using individual adaptive arithmetic coders.

On the decoder side, the decoder first reads in the side information to reconstruct the significance map for each data block, and then the coded data is decoded using the same arithmetic coder, then dequantized. Since the significance map is known from the side information, the decoder knows where to put the decoded data, and the reconstructed image is obtained after the inverse DCT.

When this image coder is applied to the lena image, we obtain up to a 0.75 dB improvement over JPEG (we have not yet compared this coder to JPEG-2000) [9]. This is promising since the significance map required approximately 60% of the transmitted bits! Clearly, more efficient encoding of the significance map will be necessary, or more likely, the spectral entropy method should be applied to a more appropriate basis function expansion.

A SpEnt-Based Wideband Speech Coder

Wideband speech coding is concerned with speech content over the 50 Hz to 7 kHz band and has long been important for videoconferencing applications. Standards activities in wideband speech coding have recently gained momentum with a new international standard at 24 and 32 kbits/s, G.722.1, established in September, 1999 [13], and the stated goal of setting a standard at 16 kbits/s in the near future. We have performed initial investigations into the application of spectral entropy-based techniques for wideband speech coding at 24 kbits/s and 16 kbits/s. The 24 kbits/s coder, seen in Fig. 3, is bit-stream compliant with G.722.1. The 16 kbits/s coder uses more efficient quantization and lossless coding techniques, but is not bit-stream compliant because of these modifications.

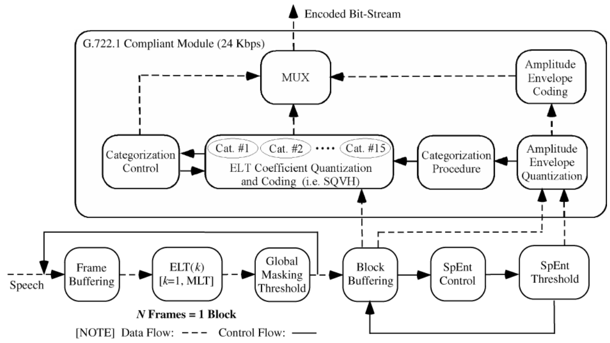

Figure 3: Block Diagram of Spectral Entropy-Based Wideband Speech Codec.

Due to length restrictions, we describe only our G.722.1 compatible wideband speech codec. For the purpose of G.722.1 bit-stream compatibility, we (a) employ a framesize of 320 samples per frame, (b) do not code the upper 40 transform coefficients representing the 1 KHz band from 7-8 KHz, and (c) use the reverse modulated lapped transform (RMLT) [14]-[16] with one framesize of overlap.

The perceptual masking scheme used is a hybrid algorithm drawn from multiple sources [17]-[19]. The transform coefficients, power spectrum, and masking thresholds are calculated on a frame-by-frame basis and stored.

Using the average energy of a subband component as a measure of its overall significance, the SpEnt control module iteratively determines the number of significant coefficients to be coded per subband. Given the following: (a) a user-defined average number of bits per significant coefficient, (b) an average output bitrate constraint, and (c) the assumption that the transform coefficients in a subband are distributed as a zero-mean process, we determine the total number of coefficients allowed for transmission in order to achieve the desired bitrate.

Taking the logarithm of the subband energies, we initiate an iterative procedure to adjust a multiplicative "quality" factor. The quality factor scales the energy representations in order to determine the appropriate number of nonzero coefficients per subband. The process iterates upon the quality factor until the cumulative number of relevant coefficients over all subbands equals the number of coefficients calculated using the bitrate constraint.

The SpEnt thresholding module operates across subbands to achieve the desired number of coefficients allocated by the SpEnt control module. Using the perceptual masks, the SpEnt thresholding module adjusts the stored thresholds so that there are approximately the suggested number of significant coefficients. The SpEnt-thresholded coefficients are then fed to a G.722.1 compliant codec for compression. The resulting output bit-stream of this encoder is decodable by any standard G.722.1 decoder which supports 24 Kbps output.

Subjective listening tests and quantitative analyses (see Table I) reveal that the spectral entropy based coder and G.722.1 achieve comparable performance at 24 kbits/s. At 16 kbits/s listening tests reveal minor artifacts in the spectral entropy coder reconstructed speech, apparently due to band switching between adjacent frames of speech. It is expected that these artifacts can be removed by smoothing of transitions or by more tightly coupling the perceptual masking with the spectral entropy-based thresholding of coefficients [20].

Table I: Quantitative Wideband Speech Coding Results (Average Segmental SNR [AS], Peak Segmental SNR [PS])

Sequence (256 kbits/s)	**SpEnt** (16 kbits/s) AS/PS [dB]	**G.722.1** (24 kbits/s) AS/PS [dB]	**SpEnt** (24 kbits/s) AS/PS [dB]
Male #1	21.527 / 35.834	21.196 / 36.182	21.397 / 36.163
Male #2	21.117 / 26.255	20.747 / 26.436	20.974 / 26.178
Male #3	21.473 / 32.826	21.161 / 33.044	21.250 / 32.992
Female #1	19.328 / 31.289	19.325 / 31.443	19.403 / 31.452
Female #2	20.570 / 26.709	20.361 / 27.220	20.388 / 27.000
Female #3	21.571 / 33.679	21.460 / 33.332	21.491 / 33.880

REFERENCES

[1] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Trans. Signal Processing*, Vol. 41, pp. 3445–3462, Dec. 1993.

[2] A. Said and W. A. Pearlman, "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 6, pp. 243–250, June 1996.

[3] S. A. McClellan and J. D. Gibson, "Variable Rate Tree Coding of Speech," *IEEE Wichita Conf. on Communications, Networking, and Signal Processing*, Wichita, KS, Apr. 1994.

[4] L. Campbell, "Minimum Coefficient Rate for Stationary Random Processes," *Information and Control*, Vol. 3, pp. 350–371, Dec. 1960.

[5] W. Yang and J. D. Gibson, "Spectral Entropy, Equivalent Bandwidth and Minimum Coefficient Rate," *Proc. IEEE Int. Symposium Inform. Theory*, pp. 181, 1997.

[6] W. Yang and J. D. Gibson, "The AEP, Spectral Entropy, and Theoretical Dimension," *Int. Conf. Combinatorics, Information Theory and Statistics*, Portland, Maine, July 18–20, 1997.

[7] W. Yang and J. D. Gibson, "The Coefficient Rate and Equivalent Bandwidth in Source Coding," *Proc. IEEE Inform. Theory Workshop*, Killarney, Ireland, June 22–26, 1998, pp. 24–25.

[8] W. Yang and J. D. Gibson, "Distortion Analysis of the Coding Scheme Based on Coefficient Rate," *Proc. IEEE Int. Symposium on Inform. Theory*, MIT, Cambridge, MA, pp. 224, Aug. 16–21, 1998.

[9] W. Yang, "Coefficient Rate and Adaptive Coding of Side Information," Ph.D. Dissertation, Texas A&M Univerity, May, 1998.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: John Wiley, 1991.

[11] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Boston, MA: Kluwer, 1992.

[12] David L. Donoho, Martin Vetterli, R. A. DeVore, and Ingrid Daubechies, "Data Compression and Harmonic Analysis," *IEEE Trans. Inform. Theory*, Vol. 44, no. 6, pp. 2435–2476, Oct. 1998.

[13] ITU, "ITU-T G.722.1: 7 KHz Audio Coding at 24 and 32 Kbps for Hands Free Operation in Systems With Low Frame Loss," Sept. 1999.

[14] Malvar, H. S., "Fast Algorithm for Modulated Lapped Transforms," *Electronics Letters*, Vol. 27, No. 9, pp. 775-776, Apr., 1991.

[15] Malvar, H. S., "Extended Lapped Transforms: Fast Algorithms and Applications," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. 1797-1800, 1991.

[16] Shlien, S., "The Modulated Lapped Transform, Its Time-Varying Forms, and Its Applications to Audio Coding Standards," *IEEE Transactions on Speech and Audio Processing*, Vol. 5, No. 4, pp. 359-366, July, 1997.

[17] Sen, D., Irving, D. H., and Holmes, W. H., "Use of an Auditory Model to Improve Speech Coders," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, No. 4, pp. 411-414, 1993.

[18] ISO/IEC, "MPEG-2 Standard", Nov., 1993.

[19] Sen, D., Allen, J. B., "A New Auditory Masking Model for Speech and Audio Coders," *IEEE Workshop on Speech Coding For Telecommunications*, pp. 53-54, 1997.

[20] M. Kokes and J. D. Gibson, "Spectral Entropy Based Wideband Speech Coding," submitted to the *Thirty-Fourth Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 29-Nov. 1, 2000.