NEAR-BEST BASIS SELECTION ALGORITHMS WITH NON-ADDITIVE INFORMATION COST FUNCTIONS

*Carl Taswell**

Scientific Computing and Computational Mathematics
Bldg 460 Room 314, Stanford University, Stanford, CA 94305-2140

ABSTRACT

Search algorithms for finding signal decompositions called near-best bases using decision criteria called non-additive information costs are proposed for selecting bases in wavelet packet transforms. These new methods are compared with the best bases and additive information costs of Coifman and Wickerhauser [1]. All near-best and best bases were also compared with the matching pursuit decomposition of Mallat and Zhang [2]. Preliminary experiments suggest that for the application of time-frequency analysis, a wide variety of results can be obtained with the different methods, and that for the application of data compression, the near-best basis selected with non-additive costs may outperform the best basis selected with additive costs.

1. INTRODUCTION

Coifman and Wickerhauser [1] presented an algorithm for the selection of the best basis in a library of bases generated by the wavelet packet transform or other similar transforms (such as local trigonometric transforms) which can be represented and searched as balanced binary trees. They defined the best basis to be that which minimized an information cost function $\mathcal{M}$ and chose the Shannon entropy as their archetype for $\mathcal{M}$. Wickerhauser [3] discussed several additional information cost functions including the ℓ^p norm and the logarithm of energy. However, experimental results have been reported only for the Shannon entropy. Furthermore, the Coifman-Wickerhauser best basis search algorithm requires the key restriction of additivity for the information cost function. In this report, I propose a new search algorithm (that does not require this restriction) for use with several new non-additive information cost functions and provide preliminary experimental results comparing the various methods on artificial transient and real speech test signals.

2. ADDITIVE COSTS

Recalling definitions from Coifman and Wickerhauser [1]:

Definition: A map $\mathcal{M}^{\text{add}}$ from sequences $\{x_i\}$ to R is called an *additive information cost function* if $\mathcal{M}^{\text{add}}(0) = 0$ and $\mathcal{M}^{\text{add}}(\{x_i\}) = \sum_i \mathcal{M}^{\text{add}}(x_i)$.

Definition: The *best basis* relative to $\mathcal{M}^{\text{add}}$ for a vector $\mathbf{x}$ in a library $\mathcal{B}$ of bases is that $\mathbf{B}$ for which $\mathcal{M}^{\text{add}}(\mathbf{Bx})$ is minimal.

Here, for a vector $\mathbf{x} \in R^N$ and orthonormal matrix $\mathbf{B} \in R^{N\times N}$, then $\mathbf{y} = \mathbf{Bx}$ and $\mathcal{M}^{\text{add}}(\mathbf{y})$ are respectively the coefficient vector and additive information cost scalar for $\mathbf{x}$ in the coordinate system represented by the basis $\mathbf{B}$. Now scale $\mathbf{y}$ to the unit vector $\mathbf{u}$ in the ℓ^2 norm with $u_i = y_i/\|\mathbf{y}\|_2$ so that the transform vector has energy $\|\mathbf{u}\|_2^2 = 1$. Then define the additive information cost functions

$$
\begin{aligned}
\mathcal{M}_1^{\text{add}}(\mathbf{u}) &= \mathcal{H}(\mathbf{u}) = \sum_{i:u_i\neq 0} u_i^2 \ln u_i^2 \\
\mathcal{M}_2^{\text{add}}(\mathbf{u}) &= \mathcal{E}(\mathbf{u}) = \sum_{i:u_i\neq 0} \ln u_i^2 \\
\mathcal{M}_3^{\text{add}}(\mathbf{u}) &= \ell^p(\mathbf{u}) = \left(\sum_i |u_i|^p\right)^{(1/p)}
\end{aligned}
$$

which are respectively the Shannon entropy, the log energy, and the ℓ^p norm (*cf.* [3]).

3. NON-ADDITIVE COSTS

Additive costs and best bases can be extended to non-additive costs and near-best bases.

Definition: A map $\mathcal{M}^{\text{non}}$ from sequences $\{x_i\}$ to R is called a *non-additive information cost function* if it serves as a selection criterion for a basis search algorithm and it is not an additive information cost function $\mathcal{M}^{\text{add}}$.

Definition: The *near-best basis* relative to $\mathcal{M}^{\text{non}}$ for a vector $\mathbf{x}$ in a library $\mathcal{B}$ of bases is that $\mathbf{B} \in \mathcal{B}^* \subset \mathcal{B}$ for which $\mathcal{M}^{\text{non}}(\mathbf{Bx})$ is minimal.

Here $\mathcal{B}^*$ is the proper subset of library bases which are searched by the Coifman-Wickerhauser algorithm. Searching $\mathcal{B}^*$ yields the optimal or best basis for an *additive* information cost function $\mathcal{M}^{\text{add}}$ (*cf.* proof [1, page 717]). However, since $\mathcal{B}^* \neq \mathcal{B}$, this search is not exhaustive and cannot guarantee the selection of a *best* basis for a *non-additive* information cost function $\mathcal{M}^{\text{non}}$. The selected basis is therefore called a *near-best* basis. Prototypes for $\mathcal{M}^{\text{non}}$ can be constructed with the sorted vector $[v_k(\mathbf{u})]$ where

$$v_1(\mathbf{u}) = |u_{i_1}| \geq \cdots \geq v_N(\mathbf{u}) = |u_{i_N}|$$

so that $v_k(\mathbf{u}) = |u_{i_k}|$ is the k^{th} largest absolute value element of the unit vector $[u_i]$. The decreasing-absolute-value sorted vector $[v_k]$ suffices to define the weak-ℓ^p norm (*cf.* [4]). However, constructing the decreasingly sorted, powered, cumulatively summed, and renormalized vector $[w_k(\mathbf{u}, p)]$ where

$$w_k(\mathbf{u}, p) = \left(\sum_{j=1}^{k} v_j^p(\mathbf{u})\right) / \left(\sum_{j=1}^{N} v_j^p(\mathbf{u})\right)$$

*Internet email: taswell@sccm.stanford.edu; Telephone: 415-723-4101; Facsimile: 415-723-2411.

makes it convenient to define several other $\mathcal{M}^{\text{non}}$. (Note that $0 \le w_k(\mathbf{u}, p) \le 1$ because of the normalization.) Thus, with $[v_k(\mathbf{u})]$ and $[w_k(\mathbf{u}, p)]$ obtained from $[u_i]$, define the non-additive information cost functions

$$\begin{aligned}
\mathcal{M}_1^{\text{non}}(\mathbf{u}) &= \mathcal{W}\ell^p(\mathbf{u}) = \max_k \; k^{(1/p)} v_k(\mathbf{u}) \\
\mathcal{M}_2^{\text{non}}(\mathbf{u}) &= \mathcal{N}_f^p(\mathbf{u}) = \arg\min_k |w_k(\mathbf{u}, p) - f| \\
\mathcal{M}_3^{\text{non}}(\mathbf{u}) &= \mathcal{A}^p(\mathbf{u}) = N - \sum_k w_k(\mathbf{u}, p)
\end{aligned}$$

which are respectively the weak-ℓ^p norm, data compression number, and data compression area. Here the power p and fraction f can be taken from the intervals $0 < p \le 2$ and $0 < f < 1$. The functions $\mathcal{N}_f^p$ and $\mathcal{A}^p$ were designed to yield scalar values that could be meaningfully minimized in a basis search algorithm and were named according to their natural or geometric interpretation. For example, choosing $p = 2$ and $f = 0.99$ and then using $\mathcal{N}_{0.99}^2$ yields the minimum number of vector coefficients containing 99% of the energy of the entire vector. The data compression number $\mathcal{N}_f^p$ and area $\mathcal{A}^p$ can be contrasted by observing that the number $\mathcal{N}_f^p$ is a local measure with varying "sensitivity" to different intervals of the w_k versus k curve whereas the area $\mathcal{A}^p$ is a global measure of the entire curve. The minimum values attainable represent maximum compression. They are readily computed for a Kroniker delta vector δ with unit energy: $\mathcal{N}_f^p(\delta) = 1$ and $\mathcal{A}^p(\delta) = 0$.

4. BEST BASIS SEARCH

Wickerhauser [3] provided notes for an implementation of the best basis search. This search algorithm is presented here with some changes in terminology and notation and with an emphasis on data structure implementation using matrices and vectors. A discrete packet transform is considered to be any multiresolution transform (such as a wavelet packet transform or local trigonometric transform) that yields a table of transform coefficients which can be organized as a balanced binary tree. The table is called a discrete packet table $\mathbf{P}$ with levels l and blocks b of the table corresponding to levels l and branches b of the tree. For both tables and trees, the finest and coarsest resolution scales are indexed levels 0 and L respectively. There are 2^l blocks on each level and thus $K = 2^{(L+1)} - 1$ blocks in the entire table. Within each block b on level l, there are $2^{-l}N$ cells c where N is the length of the original signal $\mathbf{x}$. Thus each coefficient in the packet table $\mathbf{P}$ can be specified as the 4-vector $[a, l, b, c]$ where a is the packet's amplitude and l, b, and c are its level, block, and cell indices.

To exploit modularity, it is necessary to build two trees for each packet table $\mathbf{P}$: the additive information cost tree $\mathbf{C}^{\text{add}}$ and the basis selection tree $\mathbf{S}$. In WavBox 4.1 © 1994 Carl Taswell, the functions *dpt2ict* and *ict2bst* perform these mappings from discrete packet table to information cost tree and from information cost tree to basis selection tree, respectively. This modularity permits 1) the output of various $\mathbf{C}^{\text{add}}$ for the same $\mathbf{P}$ input to *dpt2ict* with various choices of $\mathcal{M}^{\text{add}}$ as second argument, and 2) the output of various $\mathbf{S}$ for the same $\mathbf{C}^{\text{add}}$ input to *ict2bst* with various choices of basis selection method[1] as second argument.

[1] The only basis selection method presented here for additivie costs is the best basis search; others include the best level search and the restricted best basis search [5].

Finally, to compare various decompositions, it is convenient to convert discrete packet tables $\mathbf{P}^{\text{table}}$ to discrete packet lists $\mathbf{P}^{\text{list}}$ representing the selected bases. Each list contains M packets specified as row 4-vectors $[a_i, l_i, b_i, c_i]$ with rows $i = 1, ..., M$ ordered so that $|a_1| \ge \cdots \ge |a_M|$. In WavBox 4.1, the function *dpt2dpl* performs this restructuring of the data via the mapping $\mathbf{P}^{\text{list}} = \text{dpt2dpl}(\mathbf{P}^{\text{table}}, \mathbf{S})$. To study the complete basis decomposition, we must examine the entire list where $M = N$; however, we may also study subsets of the list where $M < N$, for example, where we choose $M = \mathcal{N}_{0.99}^2 < N$.

Thus, there are four data structures: $\mathbf{P}^{\text{table}} \in R^{N \times (L+1)}$, $\mathbf{C}^{\text{add}} \in R^K$, $\mathbf{S} \in \chi^K$ where $\chi = \{0, 1\}$, and $\mathbf{P}^{\text{list}} \in R^{M \times 4}$. Since tables and trees are implemented respectively as matrices and vectors, table blocks and corresponding tree branches indexed by (l, b) are respectively vectors and scalars; they are denoted $P_{lb}^{\text{table}} \equiv P_{i_{lb}, j_{lb}}^{\text{table}}$, $C_{lb}^{\text{add}} \equiv C_{k_{lb}}^{\text{add}}$, and $S_{lb} \equiv S_{k_{lb}}$ where for $l \in \{0, 1, \ldots, L\}$ and $b \in \{0, 1, \ldots, 2^l - 1\}$, the row and column vector indices i_{lb}, j_{lb} are for level l block b in a table matrix, and the scalar index k_{lb} is for level l branch b in a tree vector. Since the i^{th} packet in $\mathbf{P}^{\text{list}}$ will be denoted $P_i^{\text{list}} \equiv [a_i, l_i, b_i, c_i]$, it should be clear from context that P_i is from the list $\mathbf{P}^{\text{list}}$ while P_{lb} is from the table $\mathbf{P}^{\text{table}}$.

Now with $C_{lb}^{\text{add}} = \mathcal{M}^{\text{add}}(P_{lb})$ already computed for all l and b, and S_{lb} initialized to 1 for all b on level L and to 0 elsewhere, then the selection step of the best basis search can be expressed as

if $C_{lb}^{\text{add}} \le C_{l+1,2b}^{\text{add}} + C_{l+1,2b+1}^{\text{add}}$
then $S_{lb} = 1$
else $C_{lb}^{\text{add}} = C_{l+1,2b}^{\text{add}} + C_{l+1,2b+1}^{\text{add}}$

and the search is performed breadth-first and bottom-up through the tree. Retaining only the top-most selected branches of $\mathbf{S}$ by resetting any lower selected branches to 0 (*ie.,*pruning descendant lines) yields the best basis selection tree $\mathbf{S}$ with $S_{lb} = 1$ indicating a selected branch.

5. NEAR-BEST BASIS SEARCH

The same sequence of comparisons of basis blocks' information costs are performed for the near-best basis search as for the best basis search. However, $\mathcal{M}^{\text{add}}$ is replaced by $\mathcal{M}^{\text{non}}$. This substitution invalidates the modular independence separating computation of costs from selection of bases described in Section 4. It is therefore necessary to combine the basis selection with the cost computation. So with $C_{lb}^{\text{non}} = \mathcal{M}^{\text{non}}(P_{lb})$ already computed for all b on level L, and S_{lb} initialized to 1 for all b on level L and to 0 elsewhere, then the selection step of the near-best basis search can be expressed as

$C_{lb}^{\text{non}} = \mathcal{M}^{\text{non}}(P_{lb})$
if $C_{lb}^{\text{non}} \le \mathcal{M}^{\text{non}}(P_{l+1,2b} \oplus P_{l+1,2b+1})$
then $S_{lb} = 1$
else $P_{lb} = P_{l+1,2b} \oplus P_{l+1,2b+1}$

and the search is performed breadth-first and bottom-up through the tree with pruning of descendant lines as described in Section 4. Although not detailed here, it is possible to implement this algorithm *without* repeating for the same coefficients the sorts and powers required by $\mathcal{M}^{\text{non}}$. In WavBox 4.1, the function *dpt2bst* performs this mapping from discrete packet table to basis selection tree. The additional computational cost of *dpt2bst* with $\mathcal{M}^{\text{non}}$ relative to *dpt2ict* and *ict2bst* with $\mathcal{M}^{\text{add}}$ is essentially the cost of the sorting required by the examples $\mathcal{W}\ell^p$, $\mathcal{N}_f^p$, and $\mathcal{A}^p$ of $\mathcal{M}^{\text{non}}$ described in Section 3.

6. EXPERIMENTS

For both time-frequency and data-compression analyses, wavelet packet decompositions by best basis search with additive costs (WPDB($\mathcal{M}^{\text{add}}$)) or near-best basis search with nonadditive costs (WPDB($\mathcal{M}^{\text{non}}$)) were compared with wavelet packet decompositions by matching pursuit (WPDP). Although not reviewed in detail here, the nonorthogonal matching pursuit method of Mallat and Zhang [2] decomposes an N-coefficient signal $\mathbf{x}$ into an M-packet list $\mathbf{P}^{\text{list}}$ usually for which $M << N$. Therefore, to compare WPDB($\mathcal{M}^{\text{add}}$), WPDB($\mathcal{M}^{\text{non}}$), and WPDP decompositions, M was chosen to be variable and representing equal energy for time-frequency distribution comparisons, while fixed and representing equal bit-rate for data-compression distortion comparisons.

Experiments were performed on the test signals (artificial "transients" with $N = 512$ and the spoken word "greasy" with $N = 5632$) studied in [2]; these signals were kindly provided by S. Mallat. They were analyzed by WPDB($\mathcal{M}^{\text{add}}$), WPDB($\mathcal{M}^{\text{non}}$), and WPDP using wavelet packet libraries constructed from boundary-adjusted wavelets [6] of order 2–5 (with interior wavelet filters of length 4–10) and from circular-periodized wavelets [7] of order 8–10 (with length 16–20) where all of these wavelets were derived from Daubechies' orthogonal least asymmetric family [8]. The test signals were analyzed to maximum level $L = 5$ for "transients" and $L = 8$ for "greasy". All tables of results are shown for the test signal "transients" analyzed with circular-periodized wavelets of order 8. In each table's left-most column listing all the different decompositions, the names WPDB($\mathcal{M}^{\text{add}}$) and WPDB($\mathcal{M}^{\text{non}}$) were abbreviated to just the particular choice of $\mathcal{M}^{\text{add}}$ or $\mathcal{M}^{\text{non}}$.

6.1. Time-Frequency Analysis

For the comparison of time-frequency distributions, all $\mathbf{P}^{\text{list}}$ were truncated to M packets with $M = \mathcal{N}^2_{0.99}$ different for each $\mathbf{P}^{\text{list}}$. Histograms of these variable-M equal-energy packet lists were then computed for total energy, cells, and blocks per individual level l (*cf.* Table 1). These histograms demonstrate wide variations in the resulting distributions of energies, cell-numbers, and block-numbers across levels for the different decompositions. Because the distribution of selected blocks across levels of a packet decomposition corresponds to the distribution of sizes and shapes of tiles in a time-frequency tiling plot of cell energies a_i^2 (i^{th} packet's squared amplitude), the observed numerical differences in histograms were also visually apparent in these time-frequency plots. In an attempt to quantitate these visual differences, all variable-M equal-energy packet lists were displayed as rectangular tilings of the time-frequency plane represented as an image matrix of size 512 × 512. Then relative mean-square errors (MSE) and 2-dimensional cross-correlations (CC2) were computed for each of the WPDB($\mathcal{M}^{\text{add}}$) and WPDB($\mathcal{M}^{\text{non}}$) relative to the WPDP (*cf.* Table 2). Results for these statistics again demonstrate wide variation in time-frequency distributions for the different decompositions. Furthermore, for the particular example shown here, WPDB($\mathcal{N}^{1.0}_{0.999}$) yielded a higher quantitative correlation with WPDP than did WPDB($\mathcal{H}$).

6.2. Data Compression

For the comparison of data-compression distortions, all $\mathbf{P}^{\text{list}}$ were truncated to M packets with M fixed the same for all $\mathbf{P}^{\text{list}}$. This fixed M was chosen to be that $M = \mathcal{N}^2_{0.99}$ obtained from the WPDP. The packet amplitudes were then uniformly quantized at each of the bit rates $r = \{0, 1, 2, 3, 23\}$ for each

Table 1. Time-frequency distribution histograms.

Decomp.	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
	Energy of selected cells per level l.					
$\mathcal{H}$	0	0	.305	.392	.099	.196
$\mathcal{E}$	0	0	.142	.501	.268	.082
$\ell^{0.5}$	0	0	0	0	0	.993
$\ell^{1.0}$	0	0	0	.696	.138	.158
$\ell^{1.5}$	.993	0	0	0	0	0
$\mathcal{W}\ell^{0.5}$	0	0	0	.501	.268	.224
$\mathcal{W}\ell^{1.0}$	0	0	0	.481	.213	.299
$\mathcal{W}\ell^{1.5}$	0	0	.364	.410	.157	.062
$\mathcal{A}^{0.5}$	0	0	0	.413	.498	.082
$\mathcal{A}^{1.0}$	0	0	0	.413	.467	.113
$\mathcal{A}^{2.0}$	0	0	0	.697	.099	.196
$\mathcal{N}^{1.0}_{0.900}$	0	0	0	.218	.663	.113
$\mathcal{N}^{1.0}_{0.990}$	0	0	.142	.479	.334	.038
$\mathcal{N}^{1.0}_{0.999}$	0	0	.993	0	0	0
$\mathcal{N}^{2.0}_{0.900}$	0	0	0	.481	.285	.226
$\mathcal{N}^{2.0}_{0.990}$	0	0	0	.413	.420	.159
$\mathcal{N}^{2.0}_{0.999}$	0	0	0	.413	.498	.082
WPDP	.100	.104	.278	.123	.247	.148
	Number of selected cells per level l.					
$\mathcal{H}$	0	0	67	107	40	49
$\mathcal{E}$	0	0	66	93	67	32
$\ell^{0.5}$	0	0	0	0	0	282
$\ell^{1.0}$	0	0	0	161	53	40
$\ell^{1.5}$	343	0	0	0	0	0
$\mathcal{W}\ell^{0.5}$	0	0	0	94	76	80
$\mathcal{W}\ell^{1.0}$	0	0	0	116	63	84
$\mathcal{W}\ell^{1.5}$	0	0	92	91	65	28
$\mathcal{A}^{0.5}$	0	0	0	86	129	32
$\mathcal{A}^{1.0}$	0	0	0	87	111	48
$\mathcal{A}^{2.0}$	0	0	0	164	40	50
$\mathcal{N}^{1.0}_{0.900}$	0	0	0	54	145	48
$\mathcal{N}^{1.0}_{0.990}$	0	0	66	109	74	14
$\mathcal{N}^{1.0}_{0.999}$	0	0	298	0	0	0
$\mathcal{N}^{2.0}_{0.900}$	0	0	0	120	76	66
$\mathcal{N}^{2.0}_{0.990}$	0	0	0	87	111	48
$\mathcal{N}^{2.0}_{0.999}$	0	0	0	86	129	32
WPDP	17	22	26	36	41	39
	Number of selected blocks per level l.					
$\mathcal{H}$	0	0	1	3	3	6
$\mathcal{E}$	0	0	1	3	4	4
$\ell^{0.5}$	0	0	0	0	0	32
$\ell^{1.0}$	0	0	0	5	4	4
$\ell^{1.5}$	1	0	0	0	0	0
$\mathcal{W}\ell^{0.5}$	0	0	0	3	5	10
$\mathcal{W}\ell^{1.0}$	0	0	0	4	4	8
$\mathcal{W}\ell^{1.5}$	0	0	1	3	4	4
$\mathcal{A}^{0.5}$	0	0	0	3	8	4
$\mathcal{A}^{1.0}$	0	0	0	3	7	6
$\mathcal{A}^{2.0}$	0	0	0	5	3	6
$\mathcal{N}^{1.0}_{0.900}$	0	0	0	2	9	6
$\mathcal{N}^{1.0}_{0.990}$	0	0	1	3	5	2
$\mathcal{N}^{1.0}_{0.999}$	0	0	4	0	0	0
$\mathcal{N}^{2.0}_{0.900}$	0	0	0	4	4	8
$\mathcal{N}^{2.0}_{0.990}$	0	0	0	3	7	6
$\mathcal{N}^{2.0}_{0.999}$	0	0	0	3	8	4
WPDP	1	2	4	8	15	22

Table 2. Time-frequency distribution summary statistics.

Decomp.	$\#l$	$\#b$	M	$\lvert a_M \rvert$	MSE	CC2
$\mathcal{H}$	4	13	263	0.0135	0.7050	0.7155
$\mathcal{E}$	4	12	258	0.0136	0.8581	0.5296
$\ell^{0.5}$	1	32	282	0.0114	0.9582	0.3579
$\ell^{1.0}$	3	13	254	0.0138	0.8589	0.5349
$\ell^{1.5}$	1	1	343	0.0139	0.9382	0.4651
$\mathcal{W}\ell^{0.5}$	3	18	250	0.0135	0.8807	0.4999
$\mathcal{W}\ell^{1.0}$	3	16	263	0.0134	0.9043	0.4699
$\mathcal{W}\ell^{1.5}$	4	12	276	0.0124	0.7563	0.6601
$\mathcal{A}^{0.5}$	3	15	247	0.0146	0.8808	0.4953
$\mathcal{A}^{1.0}$	3	16	246	0.0141	0.8847	0.4897
$\mathcal{A}^{2.0}$	3	14	254	0.0135	0.8617	0.5313
$\mathcal{N}^{1.0}_{0.900}$	3	17	247	0.0138	0.9030	0.4518
$\mathcal{N}^{1.0}_{0.990}$	4	11	263	0.0128	0.8198	0.5771
$\mathcal{N}^{1.0}_{0.999}$	1	4	298	0.0124	0.5945	0.8057
$\mathcal{N}^{2.0}_{0.900}$	3	16	262	0.0119	0.8948	0.4782
$\mathcal{N}^{2.0}_{0.990}$	3	16	246	0.0141	0.8884	0.4839
$\mathcal{N}^{2.0}_{0.999}$	3	15	247	0.0146	0.8808	0.4953
WPDP	6	52	181	0.0131	0	1.0000

coefficient's mantissa. Mean square errors (MSE) of reconstruction were computed as a distortion measure in this fixed compression rate experiment comparing the different decompositions (*cf.* Table 3). The results demonstrate different distortion rankings of the decompositions at different bit rates. In particular, WPDP produced significantly less distortion at all but the lowest bit rate $r = 0$; and WPDB($\mathcal{A}^{1.0}$) performed better than WPDB($\mathcal{H}$) by as much as approximately 15% reduction in MSE.

Table 3. Reconstruction MSE at bit rate r.

Decomp.	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 23$
$\mathcal{H}$	0.6251	0.2559	0.2162	0.1999	0.1913
$\mathcal{E}$	0.5274	0.2739	0.2095	0.1932	0.1865
$\ell^{0.5}$	0.5910	0.2820	0.2352	0.2203	0.2126
$\ell^{1.0}$	0.5161	0.2659	0.1994	0.1810	0.1731
$\ell^{1.5}$	0.5531	0.3618	0.3208	0.3092	0.3057
$\mathcal{W}\ell^{0.5}$	0.5044	0.2616	0.1955	0.1778	0.1697
$\mathcal{W}\ell^{1.0}$	0.5217	0.2695	0.2074	0.1877	0.1804
$\mathcal{W}\ell^{1.5}$	0.5475	0.2906	0.2295	0.2136	0.2053
$\mathcal{A}^{0.5}$	0.5270	0.2547	0.1996	0.1786	0.1717
$\mathcal{A}^{1.0}$	0.5273	0.2543	0.1990	0.1772	0.1703
$\mathcal{A}^{2.0}$	0.5198	0.2648	0.1977	0.1791	0.1709
$\mathcal{N}^{1.0}_{0.900}$	0.5740	0.2487	0.1965	0.1769	0.1695
$\mathcal{N}^{1.0}_{0.990}$	0.5195	0.2677	0.2119	0.1940	0.1865
$\mathcal{N}^{1.0}_{0.999}$	0.6696	0.2987	0.2651	0.2515	0.2437
$\mathcal{N}^{2.0}_{0.900}$	0.5263	0.2560	0.2033	0.1816	0.1735
$\mathcal{N}^{2.0}_{0.990}$	0.5279	0.2557	0.1975	0.1788	0.1714
$\mathcal{N}^{2.0}_{0.999}$	0.5270	0.2547	0.1996	0.1786	0.1717
WPDP	0.6265	0.1912	0.1430	0.1152	0.0985

7. DISCUSSION

Although tables of results have been shown consistently throughout this report for the same example (the "transients" signal analyzed to level 5 with circularly-periodized wavelets of order 8), analogous results were obtained for all examples that were investigated (both the "transients" and "greasy" signals with wavelets of various orders). These results can be summarized as follows: the decompositions WPDB($\mathcal{M}^{\text{add}}$), WPDB($\mathcal{M}^{\text{non}}$), and WPDP produced time-frequency distributions that varied dramatically and data-compression distortions that varied more or less significantly depending on the bit rate r.

Further research should extend these preliminary experiments to investigate the statistical performance of these methods for various classes of signals. Other non-additive costs could be studied: *eg.,* an idea suggested by S. Mallat[2] would be to consider functionals constructed not from the transform coefficients but rather from their density. Moreover, these methods should be examined more carefully in the context of other schemes: *eg.,* in the application of these methods to the study of lossy compression, they could be combined with more sophisticated data-compression and bit-allocation methods such as that of Ramchandran and Vetterli [9] rather than the simple uniform quantization studied here. Finally, these methods can be extended from 1-D signals to 2-D images and other higher-dimensional signals.

In contrast to the approach of always using one selection criterion (such as entropy) assumed "best" for all applications, exploring a versatile suite of criteria enables the investigator to determine the method most suited for his particular application. In fact, it may prove true that the various information cost functions described here for selecting different bases in discrete packet transforms may have a significant impact on the subsequent data processing for that application whether it be time-frequency analysis, signal extraction from noise, signal reconstruction from compressed data, or pattern recognition.

REFERENCES

[1] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, pp. 713–718, Mar. 1992.

[2] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.

[3] M. V. Wickerhauser, "Inria lectures on wavelet packet algorithms," tech. rep., INRIA, Roquencourt, France, 1991. minicourse lecture notes.

[4] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *Applied and Computational Harmonic Analysis*, vol. 1, pp. 100–115, Dec. 1993.

[5] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wave packets," tech. rep., Yale University, 1990.

[6] A. Cohen, I. Daubechies, and P. Vial, "Wavelets on the interval and fast wavelet transforms," *Applied and Computational Harmonic Analysis*, vol. 1, pp. 54–81, Dec. 1993.

[7] C. Taswell and K. C. McGill, "Wavelet transform algorithms for finite-duration discrete-time signals," *ACM Transactions on Mathematical Software*, vol. 20, pp. ?–?, Sept. 1994.

[8] I. Daubechies, "Orthonormal bases of compactly supported wavelets. ii. variations on a theme," *SIAM Journal on Mathematical Analysis*, vol. 24, pp. 499–519, 1993.

[9] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Transactions on Image Processing*, vol. 2, pp. 160–175, Apr. 1993.

[2] personal communication