

CLASSIFICATION OF GEOPHYSICAL ACOUSTIC WAVEFORMS USING TIME-FREQUENCY ATOMS

Naoki Saito, Schlumberger-Doll Research

Schlumberger-Doll Research, Old Quarry Road, Ridgefield, CT 06877 USA

Key Words: Waveform Classification, Geophysics, Local Feature Extraction, Time-Frequency Analysis

1. INTRODUCTION

Acoustic measurements have long been used in geophysical well logging to infer petrophysical properties or the lithology of subsurface formations [12]. In sonic logging, an acoustic pulse is generated at the transmitter of a measurement tool lowered into a borehole. Then, this pulse propagates through the surrounding formations. Finally, the pressure field is recorded at the receiver located in the upper portion of the same tool (about 9 feet above the transmitter). This process is repeated until the tool is raised to a certain depth. A typical recorded waveform (digital signal), as shown in Figure 1, consists of three types of localized wave components: P wave (a refracted compressional wave), S wave (a refracted shear wave), and the Stoneley wave (a guided surface wave). The P and S waves follow paths that minimize the traveltimes between the transmitter and the receiver. The Stoneley wave, which is guided by the fluid-rock interface, travels more slowly than the P and S waves and is the dominant event at later times in the waveform. Traditionally, velocities of these three wave components (with or without their amplitudes) have been used to infer petrophysical/lithologic properties of the surrounding formations such as porosity, mineralogy, grain contacts, fluid saturation, volume percentages of various rocks such as sandstone, shale, and limestone, etc. See [12], [13] and references therein for the detailed physics behind these relationships.

Extracting velocity information of each wave component, however, is not necessarily an easy task. There are two popular approaches for estimating velocities of wave components: one is a semi-automatic tracking of the first zero-crossing of the wave component; the other is based on the semblance and coherency of the wave components which is similar to the localized version of the Radon transform [7]. Both have drawbacks: the former method often requires manual editing since the po-

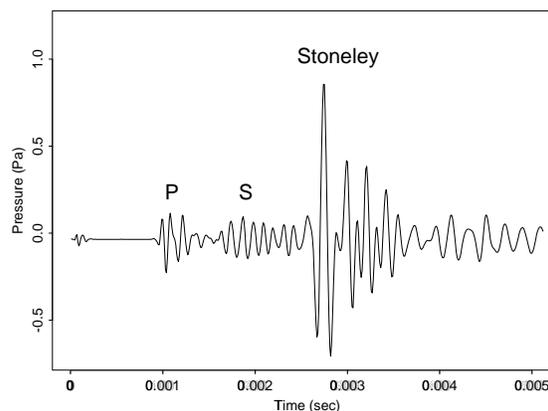


Figure 1: A typical acoustic waveform recorded down-hole. The surrounding subsurface formation consists of shale in this case. The Stoneley wave component normally has a dominant energy.

sitions of these zero-crossings vary (sometimes wildly) from trace to trace; the latter method is computationally expensive.

The velocity and amplitude information of a particular wave component, are just part of the information contained in the entire waveform shape. Thus, it is expected that the entire waveform shape contains more information about the lithology of the formation. In fact, the empirical relationship between the shapes of the waveforms and the lithology has been long recognized. There have been several attempts to infer this information using the entire waveform shape [5], [6]. Most of these attempts have been based on statistical pattern recognition techniques because building an exact mathematical or physical model is complicated and difficult.

In this paper, we view the problem of inferring lithologic information from the entire waveforms as a classification problem, i.e., classification of the rock types using waveform shape information. In particular, we apply recently developed method, the so-called *local dis-*

criminant basis (LDB) [3], [9, Chap. 4], [10] to this inference problem. This is a supervised learning method with automatic feature extraction capability: Given a training dataset (i.e., waveforms and their associated rock type information at specific depth levels), the LDB method *automatically* extracts features (the so-called *time-frequency atoms*) that are well localized in both time and frequency and that are useful for this classification task. This local nature of the time-frequency atoms makes interpretation of the classification results easier and more intuitive compared with the conventional methods (such as linear discriminant analysis or classification trees) directly applied to the original waveforms.

The organization of the paper is as follows. In Section 2, we review the time-frequency atoms and the LDB method. In Section 3, we give some background information of the field dataset on which we test our methods. Then we present the results and their interpretation in Section 4. Finally we present our conclusions in Section 5.

2. REVIEW OF TIME-FREQUENCY ATOMS, DICTIONARY, AND LOCAL DISCRIMINANT BASES

2.1. Time-frequency atoms and dictionaries

Most geophysical signals of interest consist of transients, edges, and/or local oscillations. To use them efficiently and have easily interpretable results for various tasks, one should have tools which can not only analyze the signals but can also *synthesize* the signal components and features useful for those tasks. The traditional Fourier transform is not efficient to handle such local phenomena: it uses global oscillations to analyze local ones. *Time-frequency atoms* are mathematical building blocks (basis vectors) well localized both in time and in frequency; they permit one to decompose the signals into such atoms in a computationally efficient manner, to analyze them, and then to extract and synthesize useful features. Wavelets, wavelet packets, and local trigonometric functions are examples of such atoms and recently have drawn considerable attention from such diverse fields as signal and image processing, numerical analysis, and statistics (see e.g., [14], [9]).

Let $\mathbf{x} \in \mathbb{R}^n$ be an input signal and let $\Omega_{0,0} = \mathbb{R}^n$ represent the input signal space and $B_{0,0} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ be the standard (or canonical) basis of \mathbb{R}^n . Then wavelet packets and local trigonometric transforms split this original space into two mutually orthogonal subspaces

smoothly and recursively, i.e.,

$$\Omega_{j,k} = \Omega_{j+1,2k} \oplus \Omega_{j+1,2k+1},$$

for $j = 0, 1, \dots, J$, $k = 0, \dots, 2^j - 1$, and $J(\leq \log n)$ is the maximum level of recursions specified by the user. (Unless otherwise mentioned, \log is taken as that of base 2 throughout this paper.) Here, we have $n_j \triangleq \dim \Omega_{j,\cdot} = n/2^j$. The wavelet packet transforms recursively split the frequency domain via the so-called conjugate quadrature filters [14, Chap. 5]. The local trigonometric transforms recursively split the time axis and then analyze each time segment by the discrete trigonometric transforms [14, Chap. 4]. These splits naturally generate a set of subspaces with the binary tree structure. Let $\alpha = (j, k)$ be an index to specify a node (i.e., a subspace with its basis set) of this binary tree. The index j specifies the depth of the binary tree; this is an index of the *width* of frequency bands for wavelet packets and that of time windows for local trigonometric bases. The index k specifies the *location* of the frequency bands for wavelet packets and that of the time windows for local trigonometric bases. Let $\Omega = \{\Omega_\alpha\}$ be such a collection of subspaces and let $\mathfrak{D} = \{B_\alpha\}$ be the corresponding set of basis vectors where $B_\alpha = (\mathbf{w}_{\alpha 1}, \dots, \mathbf{w}_{\alpha n_j})$ is a set of basis vectors that spans Ω_α . Each basis vector in \mathfrak{D} is called a time-frequency atom, and the whole set \mathfrak{D} is referred to as a *time-frequency dictionary* or a *dictionary of orthonormal bases* [14], [9]. Expanding a signal of length n into such a dictionary is fast; it costs $O(n \log n)$ for a wavelet packet dictionary and $O(n[\log n]^2)$ for a local trigonometric dictionary. These dictionaries contain many orthonormal bases; if the depth of the tree is J , each dictionary contains more than $2^{2^{(J-1)}}$ different bases [14].

One of the key questions is, then, how to pick a basis which performs “best” for one’s task from a large number of bases in the dictionary. In order to compare the performance or quality of each basis, we need a measure of efficiency or usefulness of a basis for that particular task. If one’s task were to compress a given signal, an information cost such as entropy [4] may be appropriate since entropy measures the flatness of the distribution of the signal’s energy among the coordinates. Therefore, a basis which minimizes entropy is very efficient for compression; most of the signal energy is concentrated in a few coordinates. Classification problems, however, are quite different from the compression problems. Important coordinates for compression, which try to capture the main features of signals, may be completely irrelevant for the classification problems where we need coordinates to see the “differences” among classes.

2.2. Local Discriminant Bases

The LDB method was developed to see these “differences” [3], [9, Chap. 4], [10]. It selects a complete and useful orthonormal basis from a time-frequency dictionary for the classification task. In order to select such a basis from \mathfrak{D} , we need to “prune” the binary tree: evaluate the nodes (i.e., the subspaces and their bases) and get rid of useless ones for discrimination tasks and only retain the useful ones whose union still spans the input signal space. This basis evaluation is the key in the LDB algorithm. The original LDB method uses “distance” measures among the time-frequency energy distributions of signal classes for basis evaluation [3], [9, Chap. 4], [10]. More precisely, let $\Gamma_{\alpha\ell}^{(y)}$ be a normalized total energy of class y signals along the direction $\mathbf{w}_{\alpha\ell}$:

$$\Gamma_{\alpha}^{(y)}(\ell) \triangleq \frac{\sum_{i=1}^{N_y} |\mathbf{w}_{\alpha\ell} \cdot \mathbf{x}_i^{(y)}|^2}{\sum_{i=1}^{N_y} \|\mathbf{x}_i^{(y)}\|^2}, \quad (1)$$

where $\{\mathbf{x}_i^{(y)}\}_{i=1}^{N_y}$ is a set of all class y signals in the training dataset and $\mathbf{w}_{\alpha\ell} \cdot \mathbf{x}_i^{(y)}$ is the standard inner product in \mathbb{R}^n . We refer to the tree-structured set of such normalized energies as the *normalized time-frequency energy map* of class y . The time-frequency energy distribution of class y signals at the node α is defined as $\Gamma_{\alpha}^{(y)} \triangleq (\Gamma_{\alpha}^{(y)}(1), \dots, \Gamma_{\alpha}^{(y)}(n_j))$. The original LDB method measures the “distances” among vectors $\Gamma_{\alpha}^{(1)}$ and $\Gamma_{\alpha}^{(2)}$ at each node α . For the LDB method using the distances among probability density functions of expansion coefficients, see [11]. In this paper, we examine the following distance measures based on the time-frequency energy distributions (there are many more choices for such a measure; see e.g., [1]):

- *Relative entropy* (or *Kullback-Leibler divergence*):

$$D(\Gamma_{\alpha}^{(1)}, \Gamma_{\alpha}^{(2)}) \triangleq \sum_{\ell=1}^{n_j} \Gamma_{\alpha}^{(1)}(\ell) \log \frac{\Gamma_{\alpha}^{(1)}(\ell)}{\Gamma_{\alpha}^{(2)}(\ell)}, \quad (2)$$

with the convention, $\log 0 = -\infty$, $\log(x/0) = +\infty$ for $x \geq 0$, $0 \cdot (\pm\infty) = 0$.

- *Symmetric relative entropy* (or *J-divergence*):

$$J(\Gamma_{\alpha}^{(1)}, \Gamma_{\alpha}^{(2)}) \triangleq D(\Gamma_{\alpha}^{(1)}, \Gamma_{\alpha}^{(2)}) + D(\Gamma_{\alpha}^{(2)}, \Gamma_{\alpha}^{(1)}), \quad (3)$$

- *Hellinger distance*:

$$H(\Gamma_{\alpha}^{(1)}, \Gamma_{\alpha}^{(2)}) \triangleq \sum_{\ell=1}^{n_j} \left(\sqrt{\Gamma_{\alpha}^{(1)}(\ell)} - \sqrt{\Gamma_{\alpha}^{(2)}(\ell)} \right)^2. \quad (4)$$

- ℓ^2 distance:

$$W(\Gamma_{\alpha}^{(1)}, \Gamma_{\alpha}^{(2)}) \triangleq \|\Gamma_{\alpha}^{(1)} - \Gamma_{\alpha}^{(2)}\|^2. \quad (5)$$

Once specified the distance measure, the LDB algorithm compares the discriminant power of each parent node with that of the union of the two corresponding children nodes. If the parent node carries more discriminant information, we retain it and prune the children nodes and vice versa. (Here, the “discriminant power” of a node means a value evaluated by the specified distance measure at that node.) To describe the LDB algorithm more precisely, let A_{α} be the LDB over Ω_{α} which we are after. This “good” basis set A_{α} may be just B_{α} (i.e., the basis set of Ω_{α} in the dictionary \mathfrak{D}) or a concatenation of the basis vectors of the subspaces descended from Ω_{α} . we use the following divide-and-conquer (also known as split-and-merge) algorithm to prune the binary tree \mathfrak{D} :

Step 0: Specify a time-frequency dictionary \mathfrak{D} and the distance measure \mathcal{D} .

Step 1: Expand each training signal in the training dataset and construct two binary trees of the time-frequency energy distributions $\{\Gamma_{j,k}^{(1)}\}$ and $\{\Gamma_{j,k}^{(2)}\}$.

Step 2: At the bottom level J of the tree, set $A_{J,k} = B_{J,k}$ for $k = 0, \dots, 2^J - 1$.

Step 3: For $\Omega_{j,k}$, $j = J - 1, \dots, 0$, $k = 0, \dots, 2^j - 1$, apply the following rule:

If $\mathcal{D}(\Gamma_{j,k}^{(1)}, \Gamma_{j,k}^{(2)}) > \mathcal{D}(\{\Gamma_{j+1,2k}^{(1)}; \Gamma_{j+1,2k+1}^{(1)}\}, \{\Gamma_{j+1,2k}^{(2)}; \Gamma_{j+1,2k+1}^{(2)}\})$,

Then $A_{j,k} = B_{j,k}$.

Else $A_{j,k} = A_{j+1,2k} \oplus A_{j+1,2k+1}$.

Here, the measure \mathcal{D} is chosen from (2)-(5). Once a complete basis (LDB) is selected, we further choose $m (< n)$ atoms from the LDB. We do “feature compression” for classification here. The simplest way to choose m atoms from n atoms is to sort them in the order of decreasing discriminant power and to retain the first m atoms. This is what we followed in our experiments in Section 4. Finally, classifiers based on these features are constructed. We use traditional classifiers here such as linear discriminant analysis (LDA) [8, Chap. 11] and classification tree (CT) [2]. The performance of each classifier is assessed by supplying test waveforms which have not been used for constructing the LDB. This test procedure is necessary to determine whether the feature extractor and classifier have learned a “rule” or “law” of this classification problem. Thanks to the extracted local features, our

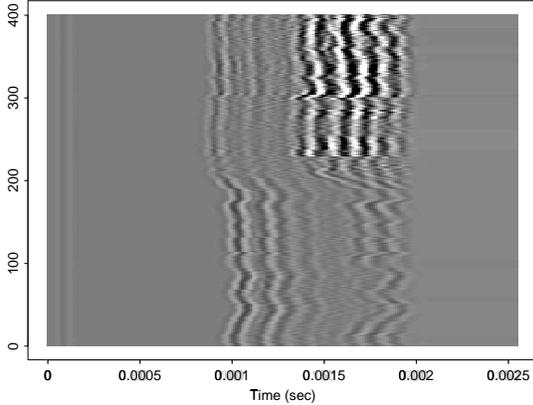


Figure 2: The dataset used in this study. The acoustic waveforms (recorded at different depth levels) are shown as a gray scale image (dark to light shades correspond to negative to positive amplitudes). The bottom 201 waveforms are “shale waveforms” whereas the top 201 waveforms are “sand waveforms.” The Stoneley wave components have been smoothly eliminated.

approach enables one to classify the lithologic information or rock types quickly without picking the velocities or amplitudes of the wave components manually, and to make physically-intuitive interpretations of the prediction results.

3. DATA DESCRIPTION

In this study, we use 402 acoustic waveforms recorded in a certain well at various depth levels. Each waveform consists of 512 time samples with a sampling rate of 10^{-5} second. Because of the sensitivity to the borehole conditions, we smoothly taper the Stoneley wave component from each waveform and consider only the earlier part of the waveforms (i.e., the number of time samples is 256 for each waveform). Along with the waveforms, we also have lithologic information or rock types obtained from other geophysical measurements. The region where the well is located consists mainly of a sequence of sandstone layers and shale layers. Figure 2 shows the dataset under study. Let us call the waveforms propagated through sandstone layers “sand waveforms” and those propagated through shale layers “shale waveforms.” We *observe by visual inspection* the following waveform features from Figure 2:

- The S-wave components in the sand waveforms have much stronger energy and faster speed than those in the shale waveforms.
- Velocities of the P-wave components in the sand waveforms are higher than those in the shale waveforms. On the other hand, their amplitudes are slightly weaker than those in the shale waveforms.

The physics of wave propagation suggests that the velocities of P and S waves are sensitive to the fluid content and the mineralogy [13].

4. RESULTS

Since the velocity information is important in this study, a natural choice of the time-frequency dictionary is the local trigonometric dictionary rather than the wavelet packet dictionary: it is easier to manipulate the time (consequently velocity) information in the local trigonometric transforms than in the wavelet packets. Hence, we use the local sine transform (LST) in this study.

Given these 402 waveforms, we decided to use 10-fold cross-validation. We first split 402 waveforms randomly into 10 groups (nine of them have 40 waveforms and the last one has 42), and then repeated the training and test procedure 10 times by using each group as a test dataset and the remaining waveforms as a training dataset. Finally, the average misclassification rates were computed.

In this study, we only used the linear discriminant analysis (LDA) and the classification tree (CT) as the classifiers. We computed four LDBs with different distance measures (2)–(5) to examine the dependency of the misclassification rates on these distance measures. For the number of features to be fed into the classifiers, we examined the top 5 to 100 LDB coordinates in steps of 5 to see the effect of the feature dimensions on the misclassification rates. We also constructed the classifiers from the original signals represented in the standard coordinate system for comparison.

The average misclassification rates using LDA as a classifier are summarized in Figure 3 and those using CT are summarized in 4.

From these figures, we observe several interesting points:

- Using more than top 35 LDB features, LDA performs very well regardless of the choice of the distance measures (2)–(5). Although the differences among these measures are small, the best result (0.24%) was obtained by the top 35 LDB vectors chosen by the Hellinger distance criterion (4).

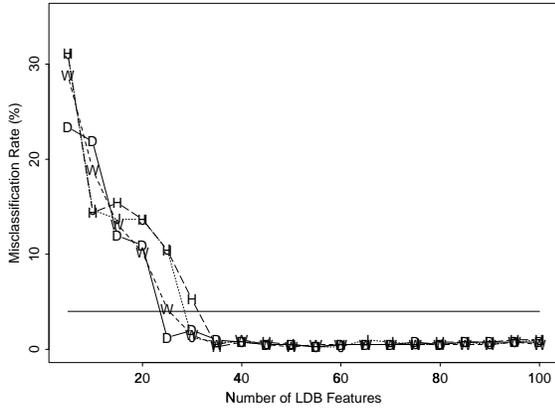


Figure 3: Misclassification rates using LDA as a classifier versus the number of the top LDB features retained. The plots with symbols D, J, H, and W correspond to the results using the distance measures (2), (3), (4), and (5), respectively. The constant level line about 4% indicates the performance of the LDA directly applied to the signals represented in the standard coordinate system (of 256 time samples).

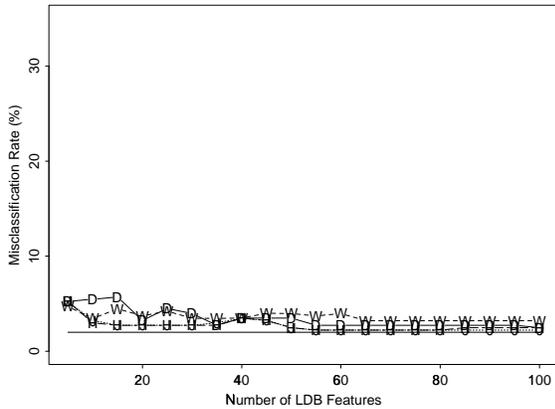


Figure 4: Misclassification rates using CT as a classifier versus the number of the top LDB features retained. The constant level line about 2% indicates the performance of the CT directly applied to the signals represented in the standard coordinate system (with 256 time samples).

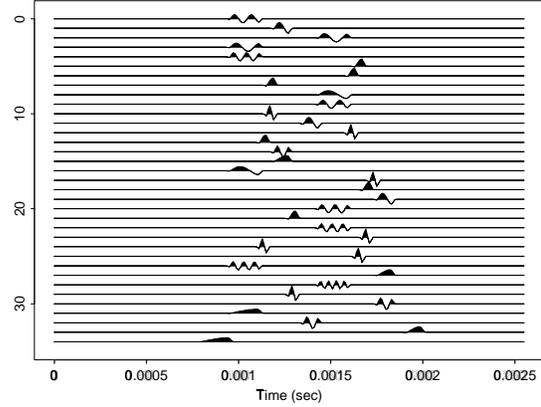


Figure 5: Top 35 LDB vectors which performed best in our experiments (i.e., LDA on the top 35 LDB features with the Hellinger distance criterion). The P and S wave components are analyzed by separate sets of the LDB vectors here.

- The LDA misclassification rates slightly increase as the number of LDB features is increased more than 50. On the other hand, less than 35 LDB features are clearly not enough to capture the truly discriminant information between these two classes, e.g., in particular, the results with less than 20 LDB features are worse than the one with the standard coordinate system.
- In general, the results of classification trees are much worse than those of LDA. This implies that the LDB features must be “oblique” and must be combined linearly to improve the classification performance. In fact, the CT result on the standard coordinate system is better than those on the LDB features.
- The P and S wave components were split naturally by the LDB vectors as shown in Figure 5. Therefore it is clear that their amplitudes and frequencies were analyzed separately in the different time windows. This is the main reason for good classification performance compared with the ones using the standard coordinate system whose time resolution is too fine and whose frequency resolution is too coarse.

Figure 5 shows the top 35 LDB vectors using the Hellinger distance criterion.

5. CONCLUSION

In this paper, we applied the LDB methods developed in [9], [3], [10] to classify the lithology or rock types of subsurface layers from acoustic well-logging waveforms propagated through them. Using the LDB methods, we could successfully extract the features useful for predicting such information in an automatic manner. Our methods also allowed us to make interpretation in an intuitive manner by the use of local time-frequency information.

As we saw in our experiments, the choice of the number of features m and the choice of the classification methods (LDA or CT) made significant difference in performance, but the choice of distance measures we examined were relatively insensitive to their performance.

6. REFERENCES

- [1] M. Basseville, *Distance measures for signal processing and pattern recognition*, Signal Processing **18** (1989), no. 4, 349–369.
- [2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Chapman & Hall, Inc., New York, 1993, previously published by Wadsworth & Brooks/Cole in 1984.
- [3] R. R. Coifman and N. Saito, *Constructions of local orthonormal bases for classification and regression*, Comptes Rendus Acad. Sci. Paris, Série I **319** (1994), no. 2, 191–196.
- [4] R. R. Coifman and M. V. Wickerhauser, *Entropy-based algorithms for best basis selection*, IEEE Trans. Inform. Theory **38** (1992), no. 2, 713–719.
- [5] R. E. Hoard, *Sonic waveform logging: a new way to obtain subsurface geologic information*, Soc. Prof. Well Log Analysts, 1983, Paper XX.
- [6] K. Hsu, *Wave separation and feature extraction of acoustic well-logging waveforms using Karhunen-Loeve transformation*, Geophysics **55** (1990), no. 2, 176–184.
- [7] C. V. Kimball and T. L. Marzetta, *Semblance processing of borehole acoustic array data*, Geophysics **49** (1984), no. 3, 274–281.
- [8] K.V.Mardia, J.T.Kent, and J.M.Bibby, *Multivariate analysis*, Academic Press, San Diego, CA, 1979.
- [9] N. Saito, *Local Feature Extraction and Its Applications Using a Library of Bases*, Ph.D. thesis, Dept. of Mathematics, Yale University, New Haven, CT 06520 USA, Dec. 1994, Available via World Wide Web, <http://www.math.yale.edu/pub/wavelets/papers/lfeulb.tar.gz>.
- [10] N. Saito and R. R. Coifman, *Local discriminant bases and their applications*, J. Mathematical Imaging and Vision **5** (1995), no. 4, 337–358, Invited paper.
- [11] ———, *Improved local discriminant bases using empirical probability density estimation*, ASA Statistical Computing Proceedings, Amer. Statist. Assoc., 1996.
- [12] J. Tittman, *Geophysical Well Logging*, Academic Press, Orlando, FL, 1986.
- [13] J. E. White, *Underground Sound: Applications of Seismic Waves*, Methods in Geochemistry and Geophysics, vol. 18, Elsevier, New York, 1983.
- [14] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A K Peters, Ltd., Wellesley, MA, 1994, with diskette.