Classification of Underwater Mammals Using Feature Extraction Based on Time–Frequency Analysis and BCM Theory

Quyen Q. Huynh, Leon N Cooper, Nathan Intrator, and Harel Shouval

***Abstract*—Underwater mammal sound classification is demonstrated using a novel application of wavelet time–frequency decomposition and feature extraction using a Bienenstock, Cooper, and Munro (BCM) unsupervised network. Different feature extraction methods and different wavelet representations are studied. The system achieves outstanding classification performance even when tested with mammal sounds recorded at very different locations (from those used for training). The improved results suggest that nonlinear feature extraction from wavelet representations outperforms different linear choices of basis functions.**

***Index Terms*—Classification, nonlinear feature extraction, time–frequency analysis, wavelets.**

I. INTRODUCTION

DETECTION, classification, and localization are among the most important and challenging goals of underwater signal analysis. A cocktail of sounds, which includes biological sounds (dolphins, sperm whales, shrimp, etc.), is mixed with environmental sounds (estuaries, cracking of ice, rain, etc.) and manmade sounds (torpedoes, submarines, surface ships, etc.) dramatically reduces recognition performance.

It is well known that the features presented to a classifier play a crucial role on its performance. Indeed, the feature set selected may be more important than the classifier architecture itself. Recently, with advances in time–frequency analysis (wavelet packet, local trigonometric basis, Gabor expansions, etc.), different feature extraction methodologies [1]–[3] have been proposed based on the localization properties of the time–frequency basis functions. It has been shown that using a wavelet representation of the acoustic signals, we can achieve improved classification [3]. This has led to the increased interest in methods for feature extraction from this data representation.

Wavelet representation is merely a different full representation of the same signal. Although it suggests natural ways to reduce representation dimensionality by keeping only the highest energy coefficients (which is similar to keeping only the first few principal components or Fourier coefficients of the signal), there is no rigorous result showing that these will be a useful representation for the purpose of signal classification and detection. The need for dimensionality reduction is clear; it follows from the *curse of dimensionality* [4], namely, the fact that the number of data points needed for a robust parameter estimation of the data density grows exponentially with the dimensionality. The problem of feature extraction is fundamental in information science. We look for an efficient and compact representation of data that leads to new insight into the problem to be solved. Under some conditions, features extracted with an unsupervised learning procedure may be more robust and general than those extracted by a supervised learning procedure. This is because the unsupervised algorithm must focus on the underlying structure of the data and not on preassigned labels that may not reveal the full structure of the data (especially with a small training set). The Bienenstock, Cooper, and Munro (BCM) theory was developed to understand and model the plasticity of the mammalian visual cortex. This model has recently been extended to a lateral inhibition network [5], and a statistically motivated variant of it has been used in various high-dimensional feature extraction tasks [6], [7].

In this paper, we use a network of BCM neurons for optimal feature extraction from a wavelet representation, leading to improved classification of underwater acoustic signals. We emphasize here that the BCM network is not playing the role of a classifier; rather, its role is feature extraction.

A. Feature Extraction from Wavelet Representations

Previous approaches to feature extraction from wavelet representation were based on signal energy [1]–[3]. Although this is not necessarily the best statistic of the signal for the purpose of classification, it was a must in the methods that have been used for feature extraction. In [1] and [2], the training set was analyzed using the time–frequency energy map of the wavelet packet decomposition tree. Coifman and Saito [1] used statistical considerations to determine the optimal wavelet packet basis for classification, which they termed the "local discriminating basis" (LDB). Unknown signals were then projected onto this LDB, and classification of the unknown signals was based on the time–frequency coefficients of only those

Manuscript received February 15, 1997; revised November 30, 1997. This work was supported by the Office of Naval Research. The associate editor coordinating the review of this paper and approving it for publication was Prof. Yu-Hen Hu.

Q. Q. Huynh is with the Department of Physics and the Institute for Brain and Neural Systems, Brown University, Providence, RI 02912 USA, on leave from the Coastal Systems Station, Naval Surface Warfare Center, Panama City, FL 32407-7001 USA.

L. N Cooper, N. Intrator, and H. Shouval are with the Department of Physics and the Institute for Brain and Neural Systems, Brown University, Providence, RI 02912 USA, on leave from the Computer Science Department, Tel-Aviv University, Tel-Aviv, Israel.

Publisher Item Identifier S 1053-587X(98)03269-3.

1053–587X/98$10.00 © 1998 IEEE

Authorized licensed use limited to: BROWN UNIVERSITY. Downloaded on April 13, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

basis functions in the LDB with the largest "discriminating power." Willsky *et al.* [2] determined relevant features from a *time-averaged* energy map that did not necessarily correspond to a single wavelet packet basis. For each signal class in the training set, an energy matrix was constructed, and the singular vectors of this matrix were used to identify the dominant energy pattern of each class. The features were then selected from the energy bins of the wavelet packet basis that corresponded to the peak values of the "primary singular" vectors. Huynh *et al.* [3] approached the binary classification problem by searching the wavelet packet library for another "discriminating basis" (LDB-2) using the "best basis" paradigm of Coifman and Wickerhauser [8] to find the basis that best approximated the *difference* of the two classes of signals. LDB-2 was thus the basis that maximized the *separation* of the two classes. Unknown signals were then projected onto the LDB-2 and classified by feeding a fixed number of the largest time–frequency coefficients of the LDB-2 (along with their corresponding time and frequency indices) into a standard classifier such as the back propagation artificial neural network [9].

II. Projection Index for Classification: The Unsupervised BCM Neuron

Exploratory projection pursuit theory [10], [11] tells us that search for structure in input space can be approached by a search for deviation from normal distribution of the projected space.[1] Furthermore, when input space is clustered, a search for deviation from normality can take the form of search for multimodality since when clustered data is projected in a direction that separates at least two clusters, it generates multimodal projected distributions.

It has been recently shown that a variant of the BCM neuron [12] performs exploratory projection pursuit using a projection index that measures multimodality [5]. This neuron allows modeling and theoretical analysis of various visual deprivation experiments [5] and is in agreement with the vast experimental results on visual cortical plasticity [13]. A network implementation that can find several projections in parallel while retaining its computational efficiency was found to be applicable for extracting features from very high-dimensional vector spaces [14], [15].

In the single neuron case, the neuronal activity (in the linear region) is given by $c = m \cdot d$, where d is the input vector, and m is the synaptic weight vector (including a bias). The essential properties of the BCM neuron are determined by a modification threshold Θ_m (which is a nonlinear function of the history of activity of the neuron) and a ϕ function that determines the sign and amount of modification Θ_m. The synaptic modification equations are given by

$$\frac{dm_i}{dt} = \mu\phi(c, \Theta_m)\, d_i$$

where, in a simple form, $\Theta_m = E[(m \cdot d)^2]$ and $\phi(c, \Theta_m) = c(c - \Theta_m)$.

[1] In a neural net architecture, this is the space generated by the hidden unit activity of the feedforward network.

In the lateral inhibition network of nonlinear neurons, the activity of neuron k is given by $c_k = m_k \cdot d$, where m_k is the synaptic weight vector of neuron k. The *inhibited* activity and threshold of the kth neuron is given by

$$\tilde{c}_k = \sigma\left(c_k - \eta \sum_{j \neq k} c_j\right)$$
$$\tilde{\Theta}_m^k = E[\tilde{c}_k^2]$$

for a monotone saturating function σ.

The projection index for a single neuron is given by

$$R(w_k) = -\{\tfrac{1}{3}E[\tilde{c}_k^3] - \tfrac{1}{4}E^2[\tilde{c}_k^2]\}.$$

The total index is the sum over all neurons in the network. The resulting stochastic modification equations for a synaptic vector m_k (the negative gradient of the index) in the network are given by

$$\dot{m}_k = \mu\left[\phi(\tilde{c}_k, \tilde{\Theta}_m^k)\sigma'(\tilde{c}_k) - \eta \sum_{j \neq k} \phi(\tilde{c}_j, \tilde{\Theta}_m^j)\sigma'(\tilde{c}_j)\right] d.$$

This network is a first-order approximation to a lateral inhibition network (using a single-step relaxation). Its properties and connection to a lateral inhibition network, as well as some related statistical and computational issues, are discussed in [5].

Under reasonable assumptions, the BCM algorithm (with k BCM neurons) produces k weight vectors that converge iteratively to fixed points corresponding to states of "maximum selectivity." In other words, for a single BCM neuron, the converged weight vector becomes orthogonal to all cluster centers except one. The feature set of the BCM algorithm is formed by the convolutions of the k weight vectors with the unknown data.

Lateral inhibition in the network allows the construction of an array of feature-selective cells in which the same feature is not selected more than once, and all features of the data set are represented in an orderly fashion.

III. Feature Extraction Based on Time–Frequency Analysis and BCM Theory

Our previous work [3] on using wavelet transforms for feature extraction have shown good results in the classification of marine mammals (dolphins, sperm whales, and porpoises). Modern time–frequency techniques (wavelet packet, local trigonometric basis, Gabor expansions) are considered as tools for providing an efficient data representation to transform the original data set to a preliminary feature set. However, classification may be improved if a dimensionality reduction takes place before the classification stage (*curse of dimensionality* [4]). In this case, applying the BCM algorithm to the preliminary feature set (time–frequency transformed data) reveals important clues about the underlying structure of the data. The use of wavelet representation is supported by the fact that classification results obtained by feature extraction from the raw signal are worse than those obtained from the wavelet representation (Table II).

Authorized licensed use limited to: BROWN UNIVERSITY. Downloaded on April 13, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

We approach the problem of building a robust classifier that combines the virtues of modern adaptive time–frequency techniques and BCM optimal selectivity as follows.

1) Choose an efficient coordinate system (library of orthogonal and nonorthogonal bases) to transform the original data set to a preliminary feature space.
2) Construct a network of connected k BCM neurons with lateral inhibition.
3) Train the k BCM neurons on the transformed data to produce k stable weight vectors.
4) Extract k crucial features that are the convolution outputs of the k weight vectors with the transformed unknown data.
5) Present the k features as inputs to a classifier, e.g., the back propagation classifier [9].

A. Signal Description

The types of signals explored in this study are the marine mammal sounds, namely, porpoise and sperm whale, which were recorded at a sampling rate of 25 kHz at various locations such as the Gulf of Maine, the Mediterranean Sea, and the Caribbean sea. We consider large original data files where sounds consist intermittently of mammal sounds and background noise. Note that each of these large original files contain whale or porpoise sounds and not both. Six data sets of length 32 768 samples, each corresponding approximately to 1.3 s, were extracted from these large files for each class. Three data sets from each class were used for training, and another three data sets (of sounds reordered at different geographical locations) were used for testing (12 data sets altogether). These data sets contained mammal sounds (of one or more animals) mixed with background noise.

B. Projections on Wavelet Space

As a first step in our approach, we choose to project each of the sound vectors on an orthonormal wavelet basis. Since the sound files are sequences of discrete numbers, we adopt the compactly supported wavelets Daubechies 4 [16], which are based on discrete-time filter banks. Let $f = \{f_k\}_{k=0}^{K-1}$ be the discrete version of the input signal $f(t)$ of length $K = 2^n$. In the fast discrete wavelet transform, the signal f is first decomposed into low- and high-frequency bands by the convolution-decimation (subsampling by two) operations of f with a lowpass filter $G = \{g_k\}_{k=0}^{L-1}$ and a highpass filter $H = \{h_k\}_{k=0}^{L-1}$. The filters G and H satisfy the orthogonality conditions

$$GH^* = HG^* = 0, \quad \text{and} \quad G^*G + H^*H = I.$$

G and H are called quadrature mirror filters (QMF's). The QMF's allow perfect reconstruction. The decomposition process continues iteratively on the resulting low-frequency bands, and each time, the high-frequency bands are left intact. The iteration stops with one low-frequency coefficient and one high-frequency coefficient. As a result, the frequency axis is partitioned *smoothly* and *dyadically*. On the time–frequency (phase), the signal is decomposed in an octave-band fashion (Fig. 1). The entire phase plane is covered by disjoint cells of equal area, which we call the Heisenberg cells. The uncertainty principle can be interpreted as a rectangular cell located around (t, f) that represents an uncertainty region associated with (t, f). The total number of cells is equal to the dimension of the input vector.

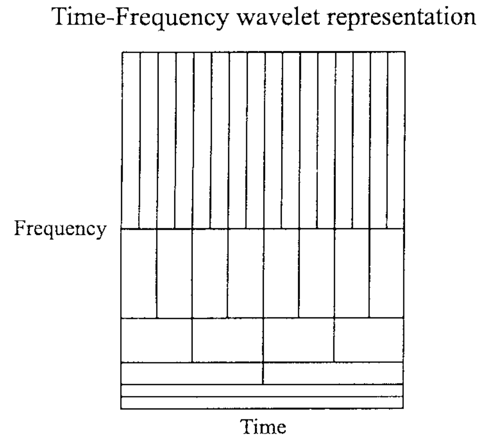

Fig. 1. Dyadic time–frequency tiling of the phase plane. The frequency axis is partitioned in an octave-band fashion. Low-frequency bands with low temporal resolution are at the bottom, whereas higher frequency bands with high temporal resolution are toward the top of the figure. The entire phase plane is covered by disjoint rectangles of equal area. On the time–frequency plane, the highest frequency bin was 6.25–12.5 kHz, and there were 16 384 wavelet coefficients spanning over the bandwidth of the signals in the time domain. The next frequency bin was 3.125–6.25 kHz, and there were 8192 wavelet coefficients. The third frequency bin of 1.562–3.125 kHz contained 4096 wavelet coefficients. Toward the lower frequency bands, each successive frequency bandwidth is reduced by half.

This representation is different than a Fourier representation and a windowed Fourier transform (WFT). Fourier basis has optimal frequency localization but no time localization. It is thus not practical for mammal sounds representation, which are of transient nature. The WFT of these signals derived from signals with 512 samples was generated by 32 overlapped windows of equal length. This corresponds to a cover of the time–frequency plane with congruent Heisenberg cells whose width Δt is the window width. In contrast with a wavelet representation, WFT does not maintain the same uncertainty for all frequencies.

C. Construction of Training Examples

We applied the wavelet transform to 12 porpoise and whale signals, each of length 32 768 samples, sampled at 25 kHz (see Section III-A). Two different approaches to construct the training data were used. The more conventional one is described on the right side of Fig. 2; here, we randomly choose small chunks of acoustic signal (512 consecutive samples) and apply wavelet analysis to get a new representation of this 512-dimensional data. Then, we extract ten features from the wavelet representation. The less conventional method is described on the left side of Fig. 2. Here, we first transform the full 32 768 samples of the raw signal into a wavelet representation (details of the representation are in Fig. 1). The two-dimensional representation is then converted into a single 32 768-dimensional vector. From this vector, we randomly choose a chunk of 512 samples starting at a random location and use this 512-dimensional vector for feature extraction.

Authorized licensed use limited to: BROWN UNIVERSITY. Downloaded on April 13, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

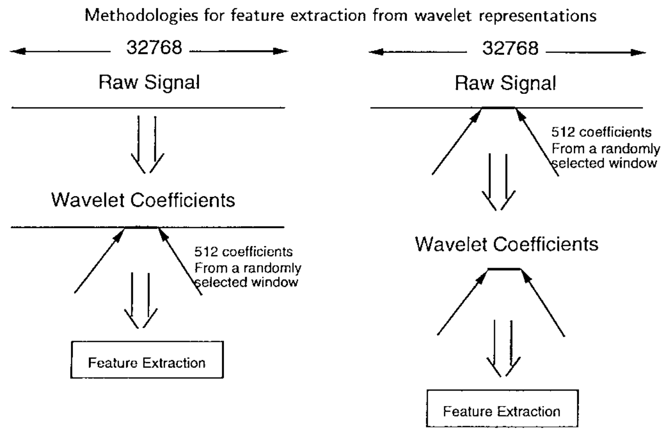

Fig. 2. Application of BCM and PCA feature extraction to the wavelet representation; On the left, the raw signal with 32768 components was run through the Daubechies 4 discrete wavelet transform. From right to left are the different levels of hierarchy (16385–32768, 8193–16384, 4907–8192, etc.) correspond respectively to frequency bandwidths 6.25–12.5, 3.12–6.25, 1.56–3.12 kHz.

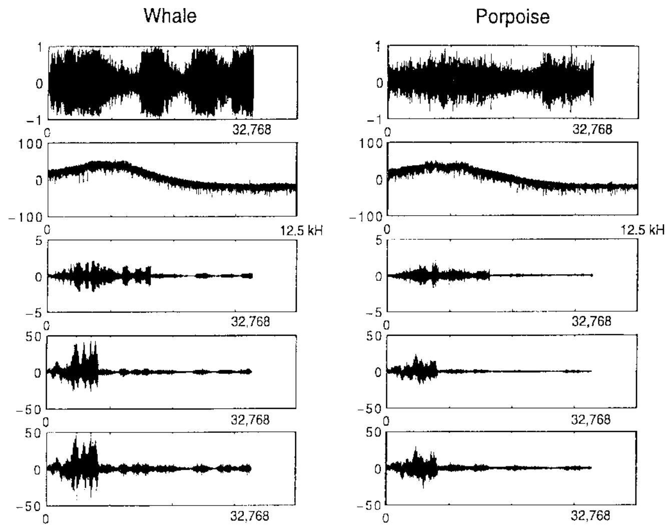

Fig. 3. Various representations for the acoustic signal based on different preprocessing methods: (First row) Raw signal: 32 768 consecutive samples representing approximately 1.3 s of signal sampled at a rate of 25 kHz (horizontal axis is sample number). (Second row) Fourier representation of the signal (horizontal axis represents frequency). Note that although the raw signal does show some differences between a porpoise signal and a whale signal, the Fourier representations are very similar, indicating the difficulty of the classification problem. (Third row) Wavelet representation (transformed to a vector) of the signal (horizontal axis represents time and frequency). This one-dimensional signal is a concatenation of time and frequency information (see Fig. 1) so that the low-frequency coefficients with low temporal resolution appear at the left, followed by high frequency with higher temporal resolution. It can be seen that the high-frequency part carries less information (see also Fourier representation) compared with the lower frequency part. This fact is emphasized in the fourth and fifth rows, where the convolutions of two BCM neurons with the wavelet signals are depicted (horizontal axis is the same as in the wavelet representation). It is clear that BCM found discriminating information in the low-frequency range at a frequency band of 1.562–3.125 kHz. We can then view the BCM neuron as a matched (nonlinear) filter designed to increase discrimination between the signals.

The next step of our approach was to train the k BCM neurons on the wavelet transformed data to produce k stable weight vectors. We used ten BCM neurons that were fully connected and formed a network with lateral inhibition. Each neuron was represented by one weight vector of dimension 512. The neurons were trained simultaneously on wavelet transformed signals of porpoises and whales. It took several hundred thousand iterations to converge to ten fixed points.

Fig. 3 presents various processings of the acoustic signals. There are 32 768 consecutive measurements of the raw data (top panel), a Fourier representation (which looks very similar for both signals), a wavelet representation of the same signal, and a convolution with two BCM neurons (bottom two panels). It can be seen that the convolution between the BCM and wavelet representation of the whale signals indicates that the BCM neurons (all ten of the network) respond only within the frequency bandwidth of 1.562–3.125 kHz at different time locations. There are no responses in the porpoise cases.

IV. Classification Results

We have used 300 examples of whale signals and 300 examples of porpoise signals for the training of the classifier. Each example was in a vector form with ten components representing ten features extracted by the BCM feature extraction network. For a baseline comparison, we also extracted the first ten principal components (PCA) from the same data. PCA is much used in signal processing as it is very simple to apply and extracts second-order statistics, which is sufficient for many applications [17].

A feedforward neural network with ten input nodes was used as a classifier. The architecture of the network consisted of one hidden layer with eight nodes and one output node. The network was trained to over 90% correct classification on the training data.

When using the large wavelet representation for feature extraction, we have noticed that classification performance could be improved if we do not train the classifier from signals that were taken from the same frequency band (for both species). Although this may sound odd, it is actually very reasonable and demonstrates a unique property of the BCM feature extraction. The selective response of BCM neurons to a specific frequency band was mainly seen for the whale signals due to the feature vectors becoming orthogonal to the class of porpoise sounds. The orthogonality to the other class of signals makes it difficult for the classifier to converge, as there is no *error* signal due to the inputs being close to zero. We have therefore used the frequency bin 1.562–03.125 kHz, which contains 4096 wavelet coefficients for the porpoise signal. During *testing* of the classifier, only the *same* frequency band was used for both species (since we does not know *a priori* to what animal the signal belongs). Thus, the “Different freq. bins” referred to in Table I corresponds to the training methodology only.

The results presented in Tables I and II are for test data that was recorded from different oceans, thus representing a different acoustic environment and possibly different species types. These results are therefore not comparable to results shown in [2], where training and testing was done from the same geographical location and possibly the same animal. We have performed such analyses as well and got results in the range of 95–100% correct classification.

Authorized licensed use limited to: BROWN UNIVERSITY. Downloaded on April 13, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

TABLE I
PERCENT CORRECT CLASSIFICATION USING PCA AND BCM FEATURE EXTRACTION FROM DAUBECHIES FOUR BASIS REPRESENTATION. RESULTS ARE PRESENTED BASED ON FEATURES EXTRACTION DIRECTLY FROM THE COEFFICIENTS OR FROM THE SQUARE OF THE COEFFICIENTS (THE ENERGY). RESULTS ARE ALSO PRESENTED FOR TRAINING THE CLASSIFIER BASED ON FEATURES EXTRACTED BY BCM FROM THE WHOLE WAVELET REPRESENTATION, NAMELY, FROM ALL FREQUENCY BANDS, OR BASED ON FEATURES EXTRACTED ONLY FROM LOCATIONS BCM WAS SELECTIVE TO (SEE TEXT FOR DETAILS). TEN FEATURES WERE EXTRACTED IN EACH OF THESE METHODS

	Porpoise	Sperm Whale
PCA, squared wavelet	76.7	32
BCM, same freq. bins (orig. wavelet)	92	74
BCM, different freq. bins (orig. wavelet)	96	88
BCM, same freq. bins (wavelet energ.)	100	81
BCM, different freq. bins (wavelet energ.)	100	91

TABLE II
PERCENT CORRECT CLASSIFICATION BASED ON VARIOUS SIGNAL REPRESENTATIONS (SEE TEXT FOR DETAILS). BCM APPLIED TO THE RAW DATA IS PERFORMED BY EXTRACTING TEN FEATURES WHILE TRAINING ON RANDOMLY CHOSEN SEQUENTIAL CHUNKS OF 512 SAMPLES FROM THE 32 768-SAMPLE RAW DATA. LDB ON WAVELET PACKET EXTRACTS TEN BEST DISCRIMINANT BASIS FUNCTIONS BASED ON COIFMAN'S ALGORITHM [1]. HIGHEST ENERGY CORRESPONDS TO EXTRACTING TEN HIGHEST ENERGY COEFFICIENTS WITH THEIR LOCATION (20 FEATURES TOTAL) FROM DAUBECHIES FOUR BASIS. THE LAST ROW REPRESENTS CLASSIFICATION PERFORMANCE ON TEN BCM FEATURES EXTRACTED FROM DAUBECHIES FOUR BASIS REPRESENTATION

	Porpoise	Sperm Whale
BCM applied on raw signals	32	95
LDB on wavelet packet	98	51
Highest energ. from Daub. 4	72	47
BCM extraction from Daub. 4	99	76

A. Importance of BCM Feature Extraction

In this case, we have studied feature extraction from the compactly supported wavelet Daubechies 4 representation. We have compared the BCM feature extraction with PCA feature extraction from this representation. As is seen in Table I, the performance of PCA here is worse, suggesting that additional structure beyond second-order statistics is required for the classification task. We have further tested whether the squared coefficients were more informative than the coefficients themselves, as is often assumed. It turned out that the squared coefficients that correspond to the energy in a particular time–frequency location are more informative, as is seen in Table I.

B. Importance of the Wavelet Representation

The Fourier representation of the data was not useful for discrimination as it was very similar for both species (see the second panel from top of Fig. 3). The usefulness of wavelet representation for classification of underwater sounds has been extensively studied and briefly reviewed in Section I-A. We have thus not attempted to compare classification performance based on a wavelet to performance based on other representations. However, since we have been using a novel feature extraction method for these signals, we evaluated the performance of the BCM feature extraction based on the wavelet representation and compared it with performance on feature extraction via BCM from the raw signal.

Table II presents classification results from the more conventional way of extracting features from this data, which is a method that allows comparison with (LDB) [18]. The first row represents results of feature extraction taken directly from the raw signal, namely, choosing randomly 512 consecutive measurements from the raw signal and using them as input to the BCM feature extraction. The high sensitivity to the whale signal is in contrast to the high sensitivity of the other methods to the porpoise signal. This suggests a possible combination between these two signal representations in the future. We have also compared two different wavelet representations: the compactly supported wavelet Daubechies[2] 4 [16] and the wavelet packet representation with the LDB feature extraction of Coifman and Saito [1]. LDB yields the closest results to classification from BCM features.

Support for the usefulness of the wavelet representation is provided by comparison with the performance of a WFT representation (see Section III-B). In this representation, the ten largest Fourier coefficients with their corresponding ten locations (20 features total) were fed into the classifier. Classification error on the test data were 60% for porpoises and 32% for whales. Although WFT representation is much more appropriate for the signal than a Fourier representation, it appears that the nonuniform covering of the time–frequency plane with Heisenberg cells whose width Δt changes inversely with the frequency, maintaining a fixed uncertainty at all frequencies, provides a better result.

V. CONCLUSIONS

We have shown that feature extraction from a wavelet representation has a profound effect on the classification results. While wavelet representations are appropriate for these acoustic signals, the detail in the resulting representation is not directly appropriate for classification, due to its size. We have shown the useful properties of an efficient nonlinear feature extraction method for classification from wavelet representations.

The BCM feature extraction that performs nonlinear unsupervised dimensionality reduction was found to be more practical than PCA on one hand and supervised discriminant pursuit on the other. Rather than looking for the projections that minimize the ratio of the within-class distance versus the between-class distance (as is done in discriminant analysis) [19], BCM looks for a direction that is mostly orthogonal to one group of signals (without knowing if they belong to the

[2] The third row represents classification from the ten highest energy coefficients of the wavelet representation.

Authorized licensed use limited to: BROWN UNIVERSITY. Downloaded on April 13, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

same class or not) while retaining selectivity to another set of signals.

We have also demonstrated the ability of this method to extract features from the huge full-signal wavelet representation. This is a unique feature that cannot be performed by linear discrimination [18]. Classification based on this feature extraction achieved outstanding results on test data that was recorded at the same location as well as data based on a mixture of locations.

ACKNOWLEDGMENT

The best discriminant basis software was kindly supplied by R. Coifman.

REFERENCES

[1] R. R. Coifman and N. Saito, "Constructions of local orthonormal bases for classification and regression," *Comptes Rendus Acad. Sci. Paris, Serie I*, vol. 319, no. 2, pp. 191–196, 1994.
[2] R. E. Learned and A. S. Willsky, "A wavelet packet approach to transient signal classification," *Appl. Comput. Harmonic Anal.*, vol. 2, no. 3, pp. 265–278, 1995.
[3] Q. Huynh, W. Greene, and J. Impagliazzo, "Feature extraction and classification of underwater acoustic signals," in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt, Eds. Dordrecht, The Netherlands: Kluwer, 1995, pp. 183–188.
[4] R. E. Bellman, *Adaptive Control Processes*. Princeton, NJ: Princeton Univ. Press, 1961.
[5] N. Intrator and L. N Cooper, "Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions," *Neural Networks*, vol. 5, pp. 3–17, 1992.
[6] N. Intrator and J. I. Gold, "Three-dimensional object recognition of gray level images: The usefulness of distinguishing features," *Neural Comput.*, vol. 5, pp. 61–74, 1993.
[7] N. Intrator, D. Reisfeld, and Y. Yeshurun, "Face recognition using a hybrid supervised/unsupervised neural network," *Pattern Recognit. Lett.*, vol. 17, pp. 67–76, 1996.
[8] R. R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–719, 1992.
[9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, 1986, pp. 318–362, vol. 1.
[10] P. J. Huber, "Projection pursuit. (with discussion)," *Ann. Stat.*, vol. 13, pp. 435–475, 1985.
[11] J. H. Friedman, "Exploratory projection pursuit," *J. Amer. Stat. Assoc.*, vol. 82, pp. 249–266, 1987.
[12] E. L. Bienenstock, L. N Cooper, and P. W. Munro, "Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex," *J. Neurosci.*, vol. 2, pp. 32–48, 1982.
[13] E. E. Clothiaux, L. N Cooper, and M. F. Bear, "Synaptic plasticity in visual cortex: Comparison of theory with experiment," *J. Neurophys.*, vol. 66, pp. 1785–1804, 1991.
[14] N. Intrator, J. I. Gold, H. H. Bülthoff, and S. Edelman, "Three-dimensional object recognition using an unsupervised neural network: Understanding the distinguishing features," in *Proc. 8th Israeli Conf. AICV*, Y. Feldman and A. Bruckstein, Eds., 1991, pp. 113–123.
[15] N. Interator, "Feature extraction using an unsupervised neural network," *Neural Comput.*, vol. 4, pp. 98–108, 1992.
[16] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. XLI, pp. 909–996, 1988.
[17] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[18] J. Buckheit and D. L. Donoho, "Improved linear discrimination using time-frequency dictionaries," Tech. Rep., Stanford Univ., Stanford, CA, 1995.
[19] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.

Quyen Q. Huynh received the Ph.D. degree in numerical analysis from Uppsala University, Uppsala, Sweden. His doctoral research, under the guidance of Prof. B. Engquist, was on developing fast iterative method for shock computation.

He has held visiting positions in the mathematics departments of the University of California, Los Angeles, the Massachusetts Institute of Technology, Cambridge, and Yale University, New Haven, CT. He is a senior research scientist at the Coastal Systems Station, Naval Surface Warfare Center, Panama City, FL. He has also joint appointments as adjunct professor at the Institute for Brain and Neural Systems and the Division of Applied Mathematics at Brown University, Providence, RI. His research interests are computational harmonic analysis, statistical computing, and hyperbolic conservation laws.

Leon N Cooper received the B.S., M.A., and Ph.D. degrees from Columbia University, New York, NY.

He spent a year at the Institute for Advanced Study and taught at the University of Illinois, Urbana-Champaign, and The Ohio State University, Columbus, before coming to Brown University, Providence, RI, in 1958. He has carried out research at various institutions, including the Princeton Institute for Advanced Study and the European Organization for Nuclear Research (CERN), Geneva, Switzerland. He is the Thomas J. Watson Senior Professor of Science at Brown University and Director of the Institute for Brain and Neural Systems. He has been involved in the study of learning and memory for over 20 years. Among his main contributions to the field is a detailed theory of learning and memory (now called the BCM theory), which has recently partially validated by experimental data.

Prof. Cooper is a Nobel Laureate, having been awarded the Nobel Prize in 1972 jointly with Bardeen and Schrieffer for their work on superconductivity. He is a fellow of the American Physical Society and the American Academy of Arts and Sciences, a Member of the National Academy of Sciences and the American Philosophical Society, and an Associate of the Neurosciences Research Program. He was an NSF Postdoctoral Research Fellow, an Alfred P. Sloan Research Fellow from 1959 to 1966, and a Guggenheim Fellow from 1965 to 1966. He has been a member of the Defense Science Board and is currently on the board of journals *Neural Computation* and *Neural Networks*.

Nathan Intrator, for photograph and biography, see this issue, p. 1201.

Harel Shouval received the Ph.D. degree in physics from Brown University, Providence, RI.

He now holds a research position at the Institute for Brain and Neural Systems at Brown University. He has both simulated and analyzed receptive field formation for the PCA and BCM learning rules in an environment composed of natural images and compared the findings to experimental results. He has analyzed the effects of lateral connectivity of networks of interacting neurons and how such models relate to simpler phenomenological models. Currently, he is combining his theoretical work with experimental work on related issues.

Authorized licensed use limited to: BROWN UNIVERSITY. Downloaded on April 13, 2009 at 23:15 from IEEE Xplore. Restrictions apply.