World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:1, 2007

An Evaluation of Algorithms for Single-Echo Biosonar Target Classification

Turgay Temel and John Hallam

***Abstract*—** A recent neuro-spiking coding scheme for feature extraction from biosonar echoes of various plants is examined with a variety of stochastic classifiers. Feature vectors derived are employed in well-known stochastic classifiers, including nearest-neighborhood, single Gaussian and a Gaussian mixture with EM optimization. Classifiers' performances are evaluated by using cross-validation and bootstrapping techniques. It is shown that the various classifers perform equivalently and that the modified preprocessing configuration yields considerably improved results.

***Keywords*—** Classification, neuro-spike coding, non-parametric nodel, parametric model, Gaussian mixture, EM algorithm.

I. Introduction

WHEN navigating in their natural habitat, the landmarks available to most bats are trees. The echolocation performed by bats is well established in theory. Iowever, it remains a problematic area how to encode the eceived echoes for landmark recognition. McKerrow showed hat echoes generated from continuous-time frequency nodulated (CTFM) signals contain relevant cues for classifier lesign [1]. Frequency components are extracted by lemodulating the echo with the same transmitted signal and ndicate the distribution of echo scatterers with depth. The utcome is bundled into a database profile, which can be used or classification purposes. Kuc [2] presented a study with iomimetic sonar for classifying various objects where the bject identity information is embedded in the echo envelope. Further, he presented a transformation of plant echoes into pseudo-action potentials where temporal differences are epresented by a spatio-temporal field [3]. In [4], Kuc and Müller presented a neuro-spike representation of random echo nsembles and developed a hypothetical neural architecture.

In this study, we examine a more comprehensive yet easy-to-implement neuro-spike code representation of plant echoes as a random process, given by Müller [5]. The feature representation is based on summary statistics of inter-spike time differences at a number of threshold levels.

Manuscript received November 5, 2004. This work is funded by the EU Framework Programme 5 under the contract IST-2001-35144.

T. Temel is with School of Informatics, IPAB, University of Edinburgh, Kings' Buildings, Edinburgh, EH9 3JZ, UK, (tel: +44 131 6513437; fax: +44 131 6513435; e-mail: ttemel@inf.ed.ac.uk).

J. Hallam is with the Maersk Mc Kinney Moller Institute for Production Technology, Campusvej 55, DK-5230, Odense M, University of Southern Denmark, Denmark (e-mail: john@mip.sdu.dk).

II. Biological Signal Processing for Echoes, Feature Extraction and Proposed Preprocessing Scheme

The echo preprocessing operation consists of two stages: cochlear filtering and coding, [5]. In the first stage, the waveform is passed through a fourth order gammatone filter with center frequency f_c and –3 dB quality factor Q_{-3dB} for modeling cochlear filters. This stage is followed by envelope extraction performed as half-wave rectification and low-pass filtering (LPF) with

$$h_{LPF}(t) = \frac{1}{\tau} e^{\frac{-t}{\tau}} \Rightarrow H_{LPF}(z) = \frac{1}{\tau . f_s} \frac{1}{1 - e^{-1/(\tau . f_s)} z^{-1}} \tag{1}$$

where f_s is the sampling rate.

The LPF output is then normalized and searched for first-crossings of a number of thresholds to model spike generation. Müller [5] showed that successive inter-spike time intervals at threshold levels α_m and α_{m+1} with the condition $\Delta(\alpha_m, \alpha_{m+1}) \geq 1.5/f_c$ constitute sufficient statistics for classifying a given echo source. He defined echo feature vectors as 3-tuples comprising

$$\begin{aligned} n &= \sum_{\forall m} I_{[\Delta(\alpha_m, \alpha_{m+1}) \geq 1.5/f_c]} \\ \bar{\alpha} &= \frac{1}{2n} \sum_{\forall m} (\alpha_m + \alpha_{m+1}) I_{[\Delta(\alpha_m, \alpha_{m+1}) \geq 1.5/f_c]} \\ \bar{\Delta} &= \frac{1}{n} \sum_{\forall m} \Delta(\alpha_m, \alpha_{m+1}) I_{[\Delta(\alpha_m, \alpha_{m+1}) \geq 1.5/f_c]} \end{aligned} \tag{2}$$

where $I_{[.]}$ is an indicator function giving 1 when the condition is met otherwise 0.

In this study, we tested two alternative LPF structures. Since the LPF given by (1) introduces varying nonlinear phase response and hence group delay across the frequency range used, features extracted will deviate from the real quantities. A single-pole architecture is also prone to unstable operation i.e., unbounded impulse response. In order to remedy these shortcomings, we propose a discrete-time LPF given by

$$H_{LPF}(z) = \frac{1}{\tau . f_s} \frac{1+\gamma z^{-1}}{1-\gamma z^{-1}}, \quad \gamma = \frac{2\tau\text{-}1/f_s}{2\tau + 1/f_s} \tag{3}$$

The main characteristic of this filter is that it has an almost constant phase response and hence identical group delay. Impulse response will be more bounded compared to previous design.

III. Density Estimation Models

In classification, the objective is to decide the class label which best represents the data, x, hence a minimum error probability, P_e amongst M different classes, C_k, which results in the Bayesian decision rule [6]

$$k = \underset{i}{\operatorname{argmax}}\, P(C_i \mid x) \tag{4}$$

From Bayes' rule, the above posterior probability can be expressed in terms of likelihood densities, or conditional probability density functions (pdf) $p(x|C_k)$ and *a priori* probabilities $P(C_k)$ as

$$P(C_i \mid x) = \frac{p(x \mid C_i)P(C_i)}{\sum_{k=1}^{M} p(x \mid C_k)P(C_k)} \tag{5}$$

Therefore, it is essential to develop the appropriate representations of the likelihood functions $p(x|C_k)$. This can be done non-parametrically or parametrically.

A. Non-parametric Methods

Data is assigned to the class containing the closest training sample. In this study, the nearest-neighborhood algorithm [7] is adopted, which can be described as

$$k = \underset{i}{\operatorname{argmin}} \left\| x - x_j \right\|_{x_j \in C_i} \tag{6}$$

where the distance metric is the Mahalanobis normalization which, for column vector data, is given by

$$\left\| x - x_j \right\|^2_{x_j \in C_i} = (x - x_j)^T \Sigma_i^{-1} (x - x_j) \tag{7}$$

The model parameter Σ is the within-class covariance matrix and its estimation is similar to parametric models, to be discussed next.

B. Parametric Models (1): Single-Gaussian

A single (multi-variate) Gaussian density profile is given by

$$p(x \mid C_k) = \frac{1}{(2\pi)^{d/2} \mid \Sigma_k \mid^{1/2}} \exp[-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)] \tag{8}$$

where d is the feature vector dimension. Model parameters mean, (μ), and covariance matrix, (Σ), are estimated by using training set as

$$\hat{\mu}_k = \frac{1}{N_k} \sum_{\forall x_j \in C_k} x_j$$
$$\hat{\Sigma}_k = \frac{1}{N_k - 1} \sum_{\forall x_j \in C_k} (x_j - \hat{\mu}_k)(x_j - \hat{\mu}_k)^T \tag{9}$$

C. Parametric Models (2): Gaussian Mixture

Each class conditional pdf is expressed as a linear composition of M_k component Gaussian pdfs as

$$p(x \mid C_k) = \sum_{i=1}^{M_k} P(c_i \mid C_k)\, p(x \mid c_i, C_k) \tag{10}$$

with the constraint $\sum_{i=1} P(c_i \mid C_k) = 1$. A suitable number of subclasses, M_k, can be determined by using Akaike's Information Criterion [8], or Rissanen's Minimum Description Length, [9], etc., --- an ongoing topic of clustering research.

Parameters can be optimized in the maximum likelihood sense by using the expectation-maximization (EM) algorithm [10] iteratively until the likelihood function reaches a local minimum or a predefined number of iterations have been used. EM description of the i-th component conditional model parameters at the (j+1)-th iteration with $\beta_i = P(c_i|C_k)$, as follows:

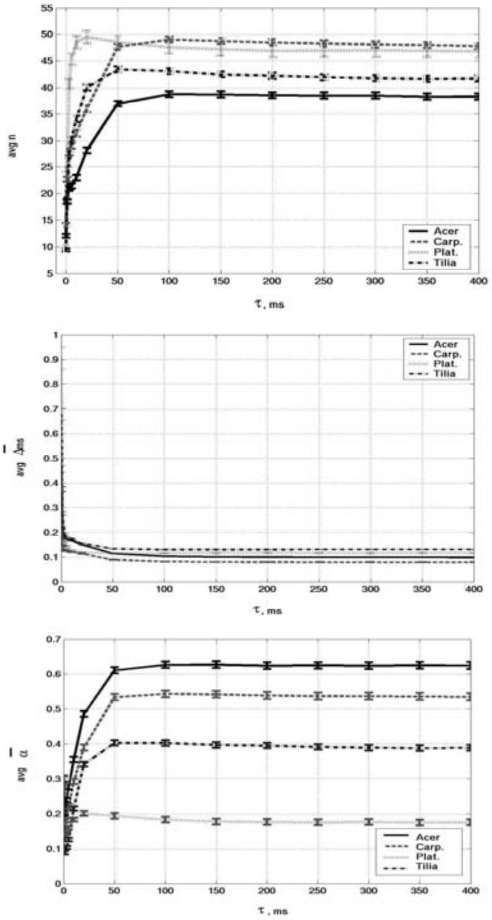
$$\beta_i^{(j+1)} = \frac{1}{N_k} \sum_{\forall x \in C_k} P_j(c_i \mid x, C_k)$$
$$\mu_i^{(j+1)} = \frac{\sum_{\forall x \in C_k} x . P_j(c_i \mid x, C_k)}{N_k . \beta_i^{(j+1)}} \tag{11}$$
$$\Sigma_i^{(j+1)} = \frac{\sum_{\forall x \in C_k} P_j(c_i \mid x, C_k).(x - \mu_i^{(j+1)}).(x - \mu_i^{(j+1)})^T}{N_k . \beta_i^{(j+1)}}$$

IV. Classifier Design and performance measurement

In experiments, we employed 2100 echoes for each of four tree types (acer, carpinus, platanus and tilia) from Müller's database of 85000 echoes. The transmitted signal was a frequency-modulated chirp sweeping linearly from 120 kHz down to 20 kHz in 3 ms. Tree hedges were scanned by two receivers in three dimensions at an almost perpendicular angle and echoes were sampled at f_s=1MHz.

The filterbank consists of a single-channel with a band-pass filter of parameters f_c=50 kHz and Q_{-3dB}=10. Classifier performances are assessed versus the LPF time constant parameter (and the number of chosen components for the mixture models).

Fig. 1 illustrates first order, sample mean, statistics over a randomly chosen 500 echoes for each tree with the proposed discrete LPF characteristic. Reasonable separation of the classes can be seen.



ig.1. Variation of mean feature vector components with 95% confidence ntervals; sample of 500 random echoes for each class

Based on the above feature definitions, nearest ieighborhood, single and mixture of Gaussian models were :onstructed. The mixture Gaussian model is implemented with :M and carried out for (maximum) 1000 iterations. Each :lassifier's performance is evaluated with the leave-one-out :ross-validation technique [11] with 2100 features by using 10 subgroups for each tree. Those 10 distinct sample average results of estimated correct classification probabilities are processed with boot-strapping to obtain a final classification performance within a given 95% confidence interval. The results are shown in Fig. 2-4 where the average proportion of each class correctly classified is plotted against the LPF time-constant. From the results, it is found that optimum component numbers with the EM mixture model for acer, carpinus, platanus and tilia are 2, 4, 4 and 2 respectively.

For the Gaussian mixture classifier with EM algorithm, the components are initialized by using K-Means algorithm [12]. Component model parameters, such as the mean and covariance matrix can be computed by using maximum likelihood method similar to a single-Gaussian model. The initial component probabilities, $P(c_i|C_k)$, are given by relative distribution of subclass members.

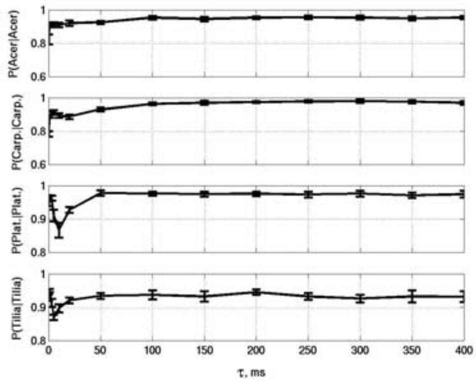


Fig.2. Nearest-neighborhood classifier: average performance on 210 tests with 1890 training instances, and 95% confidence intervals.

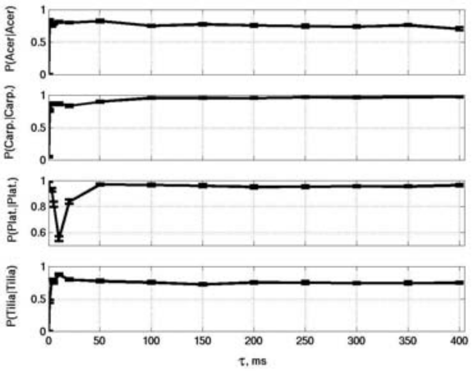
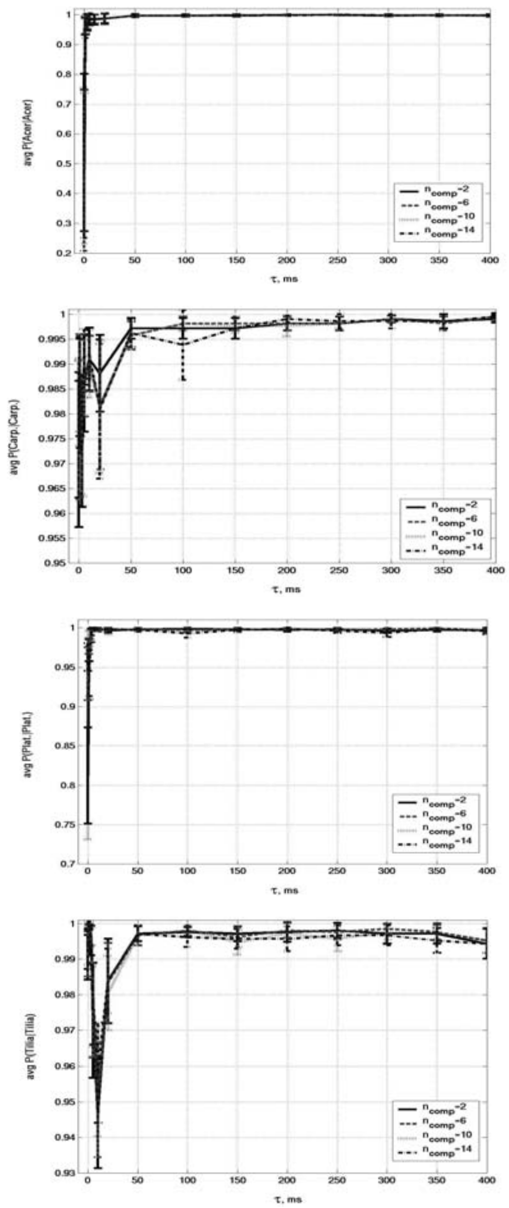


Fig.3. Single-Gaussian classifier: average performance on 210 tests with 1890 training instances, and 95% confidence intervals.

For comparison purposes, a single-Gaussian classifier is designed with the LPF structure used by Müller [5] and its performance is shown in Fig. 5.

All the classifiers tested show similar (good) classification performance except for low LPF time-constants, and all perform equally. The new LPF characteristic results in an average correct decision proportion better than 90% for τ>50 ms with labelled samples. This improves significantly on the classification performance achieved by Müller [5] using the LPF structure of (1) and kernel estimates of the likelihood pdfs computed from the full set of echoes for each tree species in the 85000 echo database.

However, each classifier has pros and cons. For example, nearest neighborhood classifier performance will be subject to the number of training samples, i.e., the larger the number, the longer to train the model and the greater the memory requirements. The single Gaussian model has the problem that the full likelihood pdfs shown in [5] are clearly non-Gaussian. The Gaussian mixture model with EM introduces a convergence problem with small number of training samples. None of these models is straightforward to adapt for on-line class learning.



ig.4. Gaussian mixture model classifier: average performance on 210 tests vith 1890 training instances, plotted for different numbers of mixture omponents against LPF time constant. Confidence intervals are 95%.

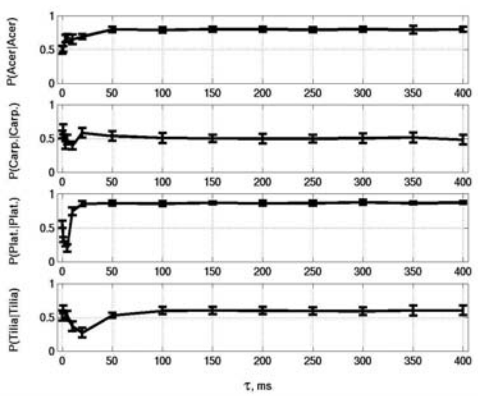


Fig. 5. Single-Gaussian classifier performance using LPF in (1) with 95% confidence interval.

The results demonstrate that the LPF employed with this study brings out better separation of feature statistics, hence, considerably improved classification performance. The effect can be accounted for by the almost constant phase characteristics of the filter, as explained before, while magnitudes are similar for both structures. It should be noted that, since $\tau >> 1/f_s$, both filters operate in the asymptotical region and filter constants should be high precision.

V. Conclusion

Three probabilistic classifier models are examined for classifying various plant echoes with labelled data in a single-shot mode. For improving the feature first-order statistics' separation further, a new LPF structure is proposed. Classifier performances are presented and compared to a previous auditory model [5] that employed kernel-estimated likelihood pdfs and the classical single-pole LPF structure of (1).

A variety of parametric and non-parametric classifiers are tested and found to perform adequately, showing the robustness of the chosen feature set.

Acknowledgement

The authors thank Rolf Müller for his generous sharing of echo data and sample code.

References

[1] P. McKerrow, and N. Harper, "Plant acoustic density profile model of CTFM ultrasonic sensing," *IEEE Sensors Journal*, vol. 1, no. 4, pp. 245-255, Dec. 2001.

[2] R. Kuc, "Bimimetic sonar locates and recognizes objects," *IEEE J. Oceanic Engineering*, vol. 22, no. 4, pp. 616-624, Oct. 1997.

[3] R. Kuc, "Transforming echoes into pseudo-action potentials for classifying plants," *J. Acoust. Soc. Am.*, vol. 110, no. 4, pp. 2198-2206, Oct. 2001.

[4] R. Müller, and R. Kuc, "A parsimonious signal representation of random echoes based on a biomimetic spike code," in *Proc. ICSC Symp. On Intelligent Systems and Applications*, pp. 915-921,Canada, May 2000.

[5] R. Müller, "A computational theory for the classification natural biosonar targets based on a spike code," *Network: Comput. Neural Syst.*, vol. 14, pp. 595-612, May 2003.

[6] C. Therrien, *Decision, Estimation and Classification, An Introduction to Pattern Recognition and Other Related Topics*, John wiley & Sons, 1989.

[7] T. M. Cover, and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Information Theory*, vol. IT-13, pp. 21-27, Jan. 1967.

[8] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automatic Control*, vol. 19, no. 6, pp. 716-723, Dec. 1974.

[9] J. Rissanen, "Stochastic complexity and modeling," The Annals of Statistics, vol. 13, no. 3, pp. 1080-1100, 1986.

[10] A. P. Demspter, N. A. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series B*, vol. 39, pp. 1-38, 1977.

[11] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford Univ. Press, 1995.

[12] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84-95, Dec. 1980.