International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:1, 2007

Evaluation of Algorithms for Sequential Decision in Biosonar Target Classification

Turgay Temel and John Hallam

***Abstract*—** A sequential decision problem, based on the task of identifying the species of trees given acoustic echo data collected from them, is considered with well-known stochastic classifiers, including single and mixture Gaussian models. Echoes are processed with a preprocessing stage based on a model of mammalian cochlear filtering, using a new discrete low-pass filter characteristic. Stopping time performance of the sequential decision process is evaluated and compared. It is observed that the new low pass filter processing results in faster sequential decisions.

***Keywords*—** Classification, neuro-spike coding, parametric model, Gaussian mixture with EM algorithm, sequential decision.

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. However, it remains a problematic area how to encode the received echoes for landmark recognition [1,2]. Estimation of an accurate statistical model for the classes to be discriminated imposes further difficulties.

In most cases, the statistical models do not provide optimality. In such cases, decision-making or classification employing dynamic programming for the posteriori probabilities of Sequential Ratio Probability Test, SPRT, [3,4] can be adopted. However, in multivariate modeling, error or discrimination surfaces are difficult or impossible to determine. As a practical approach, a safe decision-threshold probability can be chosen and relevant stopping times can be estimated.

In this study, we examine sequential decision processes that classify echo sources based on single and mixture Gaussian models of the conditional likelihood distribution of a neuro-spike code representation of plant echoes as a random process [5]. The feature extraction is based on summary statistics of the inter-spike time differences at a number of threshold levels. In the preprocessing stage, we applied a different low-pass filter structure from [5], such that class feature statistics become more clearly separated with respect to each other.

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).

Sequential classification performances are observed to improve substantially.

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

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

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 the 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

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

The classification process therefore depends critically on the representations of the likelihood functions $p(x|C_k)$. These can be non-parametric or parametrical; in this study we use only parametric representations.

In sequential decision processes, the class label, e.g., k, is determined with the earliest time stamp NTH which gives rise to a posterior probability above a specific decision-threshold value P_{TH} [3,4]. This can be written

$$NTH = \underset{j \ge 1}{\operatorname{argmin}}\ \pi_k^{(j)} \ge P_{TH} \tag{6}$$

where NTH is the class k stopping time for a threshold probability and $\pi_k^{(j)} = P(C_k \mid x_j)$ with $\pi_k^{(0)} = P(C_k)$. The *a posteriori* probability for class C_k at sample x_j can be obtained iteratively using

$$\pi_k^{(j)} = \frac{\pi_k^{(j-1)} p(x_j \mid C_k)}{\sum_{i=1}^{M} \pi_i^{(j-1)} p(x_j \mid C_i)} \tag{7}$$

as each successive data item (echo) is received.

There are two possible sources of error in this process. First, the asymptotic value of the posterior probability may not exceed the threshold for any class, so NTH is infinite. Second, the decision-threshold may be exceeded first for an incorrect class. In both cases, the error is due to imperfections in the representations of the class-conditional likelihoods.

Two parametric models for these class-conditional density functions are studied: single Gaussian and Gaussian Mixture.

A. Single Gaussian.

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

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

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

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

B. Gaussian Mixture

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

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

with the constraint $\sum_{i=1}^{M_k} P(c_i \mid C_k) = 1$. A suitable number of subclasses, M_k, can be determined by using Akaike's Information Criterion [7], or Rissanen's Minimum Description Length [8], for instance.

Parameters can be optimized in the maximum likelihood sense by using the expectation-maximization (EM) algorithm [9] 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. EXPERIMENTS AND RESULTS

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).

First order statistics, sample mean, over a randomly chosen 500 feature vectors for each tree with the proposed new LPF are graphed in an accompanying paper [10].

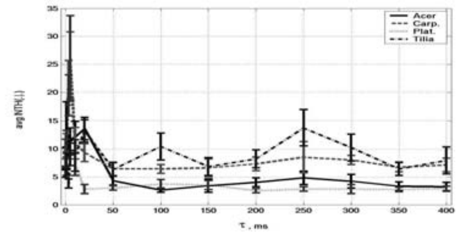
Based on the above feature definitions, single and Gaussian Mixture models were constructed and employed as sequential decision classifiers with P_{TH}=0.99. The Gaussian Mixture model is implemented with EM carried out for (maximum) 1000 iterations. Each classifier's performance is evaluated with the leave-one-out cross-validation technique [11] with 2100 features by using 10 subgroups for each tree. Those 10 distinct sample average results of stopping times 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 model 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.

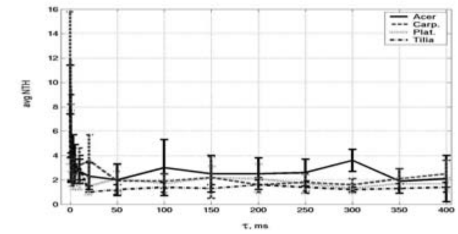
Since stopping time is a random variable, it will be more convenient to express it as a marginal density, e.g., a histogram. For this purpose, a set of experiments carried out with parameters $\psi=\tau$ and $\psi=(\tau, M_k)$ for single- and Gaussian Mixture models, respectively, with 1000 randomly chosen test vectors and 1000 training feature vectors. Each experiment is repeated 100 times, yielding a new random variable

$$\overline{NTH}(\psi) = \frac{1}{100}\sum_{i=1}^{100} NTH_i(\psi) \qquad (12)$$

Fig. 4 illustrates the histograms of $\overline{NTH}$ with two LPFs considered.



(a)



(b)

Fig.2. Single-Gaussian sequential classifier performances designed using (a) LPF in (1), and (b) LPF in (3), with 95% confidence interval.

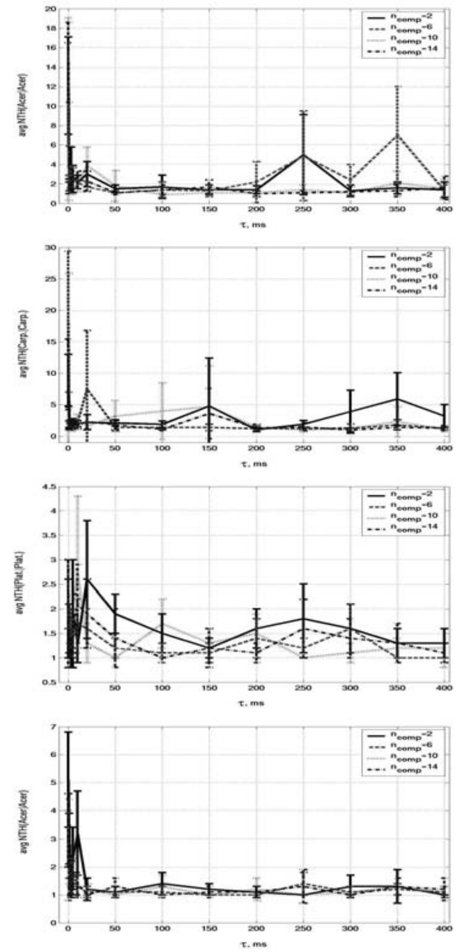
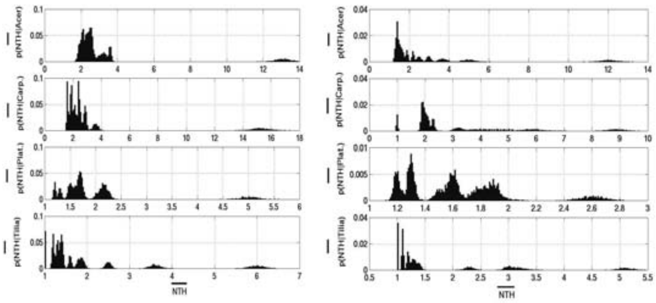
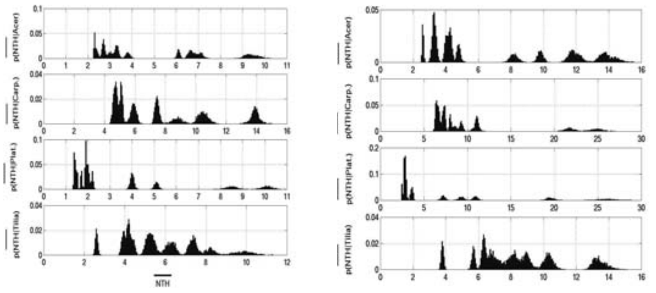


Fig.3. Gaussian mixture model with EM sequential classifier performance designed using LPF in (3) for different component numbers with 95% confidence interval.



(a)



(b)

Fig.4. Single and EM-mixture Gaussian models stopping time distributions using (a) LPF in (3), (b) LPF in (1).

The sequential decision classifiers designed with new LPF operate with smaller mean and variance or confidence interval for the chosen P_{TH}=0.99. Table I summarizes some of the characteristics for each classifier model with LPFs considered where E and δ_{95} denote the estimated ensemble average and 95% confidence interval, respectively.

TABLE I
STATISTICAL QUANTITIES WITH LPF

Tree	Single Gaussian, E/δ_{95}		Mixture Models, E/δ_{95}	
	LPF (3)	LPF (1)	LPF (3)	LPF (1)
Acer	2.7/1.8	4.0/7.6	2.2/2.3	3.9/9.7
Carpinus	2.4/2.5	7.7/7.8	2.3/2.6	7.0/11.8
Platanus	1.7/1.7	2.8/6.9	1.7/1.7	2.6/9.5
Tilia	1.6/2.4	5.6/6.8	1.4/1.9	7.3/7.2

The improvement in sequential classifier performance 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

Probabilistic classifier models are examined for classifying various plant echoes with labeled data in sequential decision. For improving the feature first-order statistics separation further, a new LPF structure is proposed. Classifier performances are presented and compared to the model presented in [5] by Müller who used kernel estimated class-conditional likelihood models based on the full database of echo data.

Our results show that excellent classification performance can be obtained using parametric likelihood models (which are more easily learned than the kernel estimated models) and that the novel LPF structure improves the efficiency of classification substantially.

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. 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.

[3] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Trans. Information Theory*, vol. 40, no. 6, Nov. 1994.

[4] A. G. Tartakovsky, X. R. Li, and G. Yaralov, "Sequential detection of targets in multichannel systems," *IEEE Trans. Information Theory*, vol. 49, no. 2, Feb. 2003.

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

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

[9] 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.

[10] T. Temel, and J. Hallam, "An evaluation of algorithms for single-echo biosonar target classification," submitted to ICCI'04.

[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.