ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 33-38

Analyses on the temporal patterns of spikes of auditory neurons by a neural network and tree-based models

Tetsuya Takahashi

NMR Center, Massachusetts General Hospital, Bldg. 149, 13th Street,
Charlestown Navyyard, Charlestown, MA 02129, USA
Justsystem Scientific Institute
Aoyama Bldg., Kita-Aoyama 1-2-3, Minato-ku, Tokyo 107, Japan

We estimated the time scale over which information in the primary auditory cortex is processed. An artificial neural network was used to learn the temporal patterns of spikes. After learning, test patterns were input to the network. Comparison of the accuracy of the network with that of the maximum likelihood function computed from the spike count reveals that the temporal patterns of spikes are closely related to stimulus discrimination. Next, we constructed a tree-based model from a subset of the spike trains with a fixed time resolution and validated the model with another. By repeating this for different bin widths, we found that there are no simple models for the time bin width larger than 50 ms. This indicates that the time scale in the auditory cortex is not larger than 50 ms.

1. Introduction

Although the spike count has been widely used in studies based on single unit recordings, there are no clear rationales for a particular choice of the time bin width over which the number of spikes is summed. Since conclusions based on the results of statistical analyses can be affected by the choice made, an accurate estimation of the time scale of neural information processing is necessary. Moreover, the results of studies indicate that neuronal information in some areas of the brain is carried in terms of temporal patterns rather than in terms of the spike count. It is important to estimate the time range of the temporal patterns of spikes in which meaningful information is present.

2. Analyses by a multi-layer perceptron

The experimental procedures are described elsewhere [5,6,7]. Stimuli were pure tone bursts with 11 different frequencies. Twenty spike trains were obtained for a stimulus with a resolution of 1 ms without averaging. Since it is difficult to assume the form of the distribution which the temporal patters of spikes follow, one needs to construct models based on data sets and not assume the form of the models apriori.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 33-38

Artificial neural networks satisfy this requirement. First, we directly confirm that temporal patterns of neuronal spikes in the primary auditory cortex carry more information than the spike count by using spike sequences to train a three-layer perceptron (input elements 35, hidden elements 1 or 2, output element 1, patterns 6) with back-propagation as the learning rule. Although one spike train contains some information over 350 ms, the number of spikes is averaged over 10 ms so that only 35 input elements are necessary. The learning data set consists of 6 spike trains (2 spike trains for 1 kHz, 70 dB SPL, 2 spike trains for 6 kHz, 70 dB SPL, 2 spike trains for 12 kHz, 70 dB SPL). Convergence is considered to be attained when three stimuli are discriminated with a small margin of error. At the same time, the cumulative density function is computed from the spike count of the same learning set. After learning is achieved, 6 spike trains (2 spike trains for 1 kHz, 70 dB SPL, 2 spike trains for 6 kHz, 70 dB SPL, 2 spike trains for 12 kHz, 70 dB SPL) that were not used in the learning phase are input to the network while the stimulus is predicted according to the cumulative density function computed from the spike count. The temporal patterns of spikes were consistently superior to the spike count in terms of stimulus discrimination ability in all but one of the 16 analyses. The percents correct for the spike frequency and the temporal patterns were 33.3 and 43.8, respectively. This indicates that temporal patterns can yield a more reliable prediction in terms of stimulus discrimination than spike counts can.

3. Analyses by Tree-based models

If one decreases the time resolution of measurements or takes a larger time bin width over which spikes are summed, the temporal patterns of responses will carry less information. One can therefore estimate the marginal time bin width for which the temporal patterns of responses are lost. For this purpose, one needs to generate models from given sets of data. Taking the index of the time bin as the prediction variable and the number of spikes occurring in the time bin as the response variable should satisfy this requirement.

3.2. Tree-based model

Although tree-based models can be used to deal with factor response variables as well as numeric response variables, theses should be considered an alternative for standard linear models when it is used for numeric predictors and a numeric response variable [1,2,3,4]. We construct tree-based models in the following manner. Spikes in a spike train are classified into different time bins according to their time stamps. Since spike occurrence is rather sparse, we performed averaging over 20 trials.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 33-38

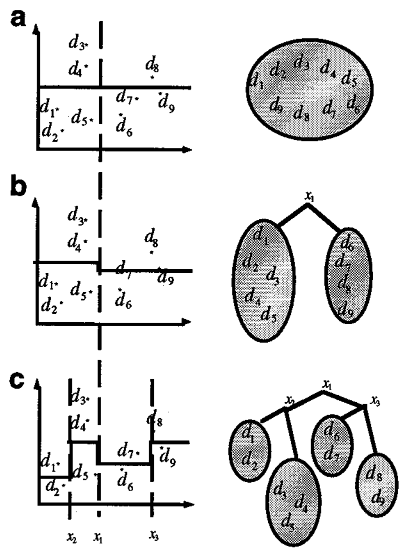


Fig. 1
A tree-based model for a data set which has a numerical response variable (the number of spikes) and a single predictor variable (the index of the time bin) is a step function with a certain property which predicts the number of spikes from a given index of the time bin. To construct the step function, we use the deviance defined by

$$dev(D) = \sum_{i=1}^{n} \left\| y_i - \lambda_i(x_i) \right\|^2 \qquad (1)$$

where D, (x_i, y_i) and λ_i stand for a set of data, a two-dimensional data point, and the value of the step function for the interval in which x_i is contained, respectively. A split on the x-axis is so created on the data that the difference in deviance between a node and its two splits, i.e.,

$$\begin{aligned} \Delta(dev) &= dev(D_{before}) - dev(D_{after}) \\ &= dev(D_{before}) - \left(dev(D_R) + dev(D_L)\right) \end{aligned} \qquad (2)$$

is maximized where D_L (D_R) is a subset of D whose elements lie in the left (right) half of the split (from a to b, Note that the figures are schematically drawn). Then, new splits are created on the terminal nodes which the first split has created such that $\Delta(dev)$ in equation (2) is maximized (from b to c).

Let $D = \bigcup_{i=1}^{N} (x_i, y_i)$ and (x_i, y_i) denote the data set and the i-th data where x_i, y_i are the index of the time bin and the number of spikes (averaged over 20 trials) occurring in the time bin. If the width of the time bin is set to 5 ms, the number of time bins is equal to 70 (5 ms x 70 = 350 ms) in the current study. Since we have

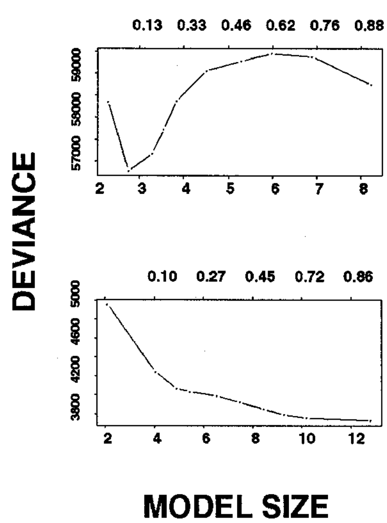
ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 33-38

responses from 11 different stimuli, the total number of data points is equal to 770 (=11x70) when the time bin width is set to 5 ms. Then a tree-based model is constructed so that the data points are approximated by a step which satisfies certain conditions (see legend of Fig. 1).

3.2. Cross-Validation

A model constructed from a particular data set may be so overfit that the deviance between that model and another data set generated from an identical source is significant ("variance" is large). On the other hand, if the model is too underfit, then the model cannot reflect the data set at all ("bias" is large). It is therefore necessary to balance the variance which a model has towards data sets against the bias of the model. One of the standard techniques to find a balanced model is called cross-validation. Ten-fold cross-validation (constructing a model from 90% of the data set and validating the model with the remaining 10%, then constructing another model from another 90% of the data set, and so on) is employed for evaluation of the appropriateness of the model. By changing the model size, one can see the relationship between the model size and the deviance defined by equation 1.

Suppose that as the model size increases, the deviance decreases monotonically. This means that there are no simple models which can be used to predict the temporal patterns of spikes. On the other hand, if there are some local minima in the deviance, then there are simple models. Figure 2 shows how the size of the model is related to the deviance for neuronal spikes recorded during and after presentation of various stimuli.



ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 33-38

Fig. 2
The graphs represent the relationship between effective model size (the lower abscissa, see below for definition), deviance (the ordinate) and α (the upper abscissa, see below). The curves were computed for two neurons; the temporal pattern of neuron a (ge1112.01) is fairly regular (the time bin width is 25 ms) whereas that of neuron b (ge1119.11) is random (the time bin width is 14 ms). For each point on the curve, 10-fold cross validation was performed.

In the case of neuron a, when the size of the model is equal to 3, the deviance is the minimum whereas there is no clear minimum for neuron b, which implies that the temporal patterns of neuron b have no clear structure.

It is obvious that as the time bin is widened, the temporal patterns of responses tend to become weaker. While a simple model exists for narrow time bins, there are no clear minima for wide time bins (Fig. 3).

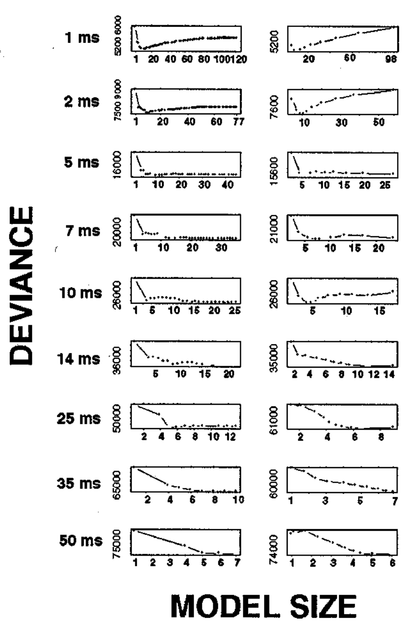


Fig. 3
The relationships among the time bin width (the first column), the deviance (the ordinate) and the model size (the abscissa) are computed from the spike trains of neuron ge1119.01. Shrinking (the second column) and pruning [2] (the third column, see below for definition) were employed to reduce the model size. As the time bin width increases, the deviance tends to become a monotonically decreasing function of the model size.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 33-38

Although the deviance at the local minimum and the model size which gives the local minimum are dependent on how the trees are truncated (for one neuron, pruning and shrinking give very different values, 1 ms and 50 ms), for all the neurons we analyzed, it holds true that if the time bin width is larger than 50 ms, the deviance is a monotonically decreasing function of the model size regardless of how trees are truncated. This means that the time scale in the primary auditory cortex is not larger than 50 ms [8].

References

1. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., Classification and Regression Trees, Wadsworth, Belmont, CA, 1984, pp. 18-58.
2. Clark, L.A. and Pregibon, D., Tree based model, In J.M. Chambers and T.J. Hastie (Eds), Statistical Models in S, Wadsworth, Pacific Grove, CA, 1992, pp. 377-419.
3. Hastie, T. and Pregibon, D., Shrinking Trees. AT&T Bell Laboratories Tech. Report. Murray Hill, NJ, 1990, pp. 11-12.
4. Quinlan, J.R., C4.5: Programs for machine learning, Morgan Kaufmann, San Francisco, CA, 1993, pp. 17-26.
5. Riquimaroux, H., Takahashi, T. and Hashikawa, T., Neuronal responses of the primary auditory cortex in the Japanese monkey: Tonotopicity revisited, J. Acoust. Soc. Jpn, 15 (1994) 171-177.
6. Takahashi, T., Time-variant structure of auditory cortex of the macaque monkey, Neurosci. Lett., 186 (1995) 1-4.
7. Takahashi, T., Two statistical methods for analyzing responses of multiple neurons, Intl. J. Neurosci., 88 (1996) 11-26.
8. Takahashi, T., Analyses on the temporal patterns of spikes of auditory neurons of the macaque monkey by means of an artificial neural network and tree-based models", Neurosci. Lett. 233 (1997) 17-20.