ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 29 (2006) 396 – 408

Application of artificial neural network to fMRI regression analysis

Masaya Misaki* and Satoru Miyauchi

Brain Information Group, Kansai Advanced Research Center, National Institute of Information and Communications Technology, 588-2 Iwaoka, Iwaoka-cho, Nishi-ku, Kobe-shi, Hyogo 651-2429, Japan

Received 18 January 2005; revised 29 July 2005; accepted 1 August 2005
Available online 6 September 2005

We used an artificial neural network (ANN) to detect correlations between event sequences and fMRI (functional magnetic resonance imaging) signals. The layered feed-forward neural network, given a series of events as inputs and the fMRI signal as a supervised signal, performed a non-linear regression analysis. This type of ANN is capable of approximating any continuous function, and thus this analysis method can detect any fMRI signals that correlated with corresponding events. Because of the flexible nature of ANNs, fitting to autocorrelation noise is a problem in fMRI analyses. We avoided this problem by using cross-validation and an early stopping procedure. The results showed that the ANN could detect various responses with different time courses. The simulation analysis also indicated an additional advantage of ANN over non-parametric methods in detecting parametrically modulated responses, i.e., it can detect various types of parametric modulations without a priori assumptions. The ANN regression analysis is therefore beneficial for exploratory fMRI analyses in detecting continuous changes in responses modulated by changes in input values.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Artificial neural network; fMRI; Semi-parametric analysis; Autocorrelation noise; Paramedic modulation

Introduction

In functional magnetic resonance imaging (fMRI) studies, searches for the occurrence of signal changes that correlated with certain events are conducted to detect brain activations. In the general linear model (GLM) approach (Friston et al., 1995), a regressor representing the canonical hemodynamic response function (HRF) is used to detect such correlations. However, if the shape of the hemodynamic response differs greatly from the pre-assumed shape (Aguirre et al., 1998; Miezin et al., 2000) or an unknown process mediates such correlations, we cannot detect those correlations. In some cases, a combination of regressors, the canonical HRF, its temporal derivative, and a dispersion derivative, is used to absorb this diversity of response shape (Friston et al., 1998a,b). Although this method can absorb minor changes in the canonical HRF, it cannot absorb all diversity and the response variability is still a problem.

Various other approaches, which do not assume the shape of the HRF a priori, have been proposed including selective averaging (Dale and Buckner, 1997), smooth FIR filters (Goutte et al., 2000), and non-parametric Bayesian estimation of the HRF (Marrelec et al., 2003). The primary advantage of these methodologies, which are called non-parametric methods because no parametric models of the HRF are used, is that even when the response functions diverge from region to region or subject to subject, correlated responses can be detected.

In this study, we proposed another 'semi-parametric' method for fMRI analysis: the use of a very general class of functional forms to build more flexible models (Bishop, 1995). The proposed method uses a feed-forward layered artificial neural network (ANN) to describe a non-linear dynamic system of hemodynamic response. Some event over its recent history is used as an input and the BOLD signal from a particular voxel is used as a supervised signal. This type of artificial neural network is called a multi-layer perceptron (MLP). It is known that for an infinite number of hidden units, an MLP with one or more hidden layers whose output functions are sigmoid functions can approximate any continuous function to any degree of accuracy (Funahashi, 1989; Hornik et al., 1989). Thus this method can perform a non-linear regression analysis between BOLD signals and events without explicit modeling of the response function.

In the MLP, hidden units receive input vector $\boldsymbol{x}_{\text{in}}$ multiplied by an adjustable weight matrix $\mathbf{W}_{\text{hidden-in}}$ and bias value $\boldsymbol{b}_{\text{hidden}}$. Hidden units have a general class of transfer function, sigmoid function tanh, and transfer the input value to the output layer.

$$\boldsymbol{x}_{\text{hidden}} = \tanh(\mathbf{W}_{\text{hidden}-\text{in}}\boldsymbol{x}_{\text{in}} + \boldsymbol{b}_{\text{hidden}}) \quad (1)$$

At the output layer, the outputs of the hidden layer $\boldsymbol{x}_{\text{hidden}}$ are multiplied by an adjustable weight matrix $\mathbf{W}_{\text{out-hidden}}$ and a bias value $\boldsymbol{b}_{\text{out}}$ is added.

$$y = \mathbf{W}_{\text{out}-\text{hidden}}\boldsymbol{x}_{\text{hidden}} + \boldsymbol{b}_{\text{out}} \quad (2)$$

* Corresponding author. Fax: +81 78 969 2279.
E-mail address: misaki@po.nict.go.jp (M. Misaki).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2005.08.002

y is a network output. Thus, an MLP approximates the regression function $y = f(\boldsymbol{x}_{\text{in}})$, by a convolution of sigmoid functions with adjustable input weights. Hereafter, we call this regression the ANN regression.

The ANN regression produces results known to be equivalent to the Volterra kernel method (Volterra, 1959) without explicit definition of the kernel functions (Wray and Green, 1994). Friston et al. (1998b, 2000) have already applied the Volterra series approach to fMRI analysis. In their approach, the kernel functions are explicitly modeled (Friston et al., 2000). From the viewpoint of MLP, this approach can be seen as using fewer pre-assumed transfer functions at the hidden layer and only a restricted number of input-hidden connections. In contrast, the proposed ANN regression has no explicit form of response functions and uses fully adjustable connections between the input and hidden layers. Because an adequately trained network is equivalent to the Volterra series expansion with all the dimensions of that system (Wray and Green, 1994), the ANN regression provides more flexibility than using only a restricted set of Volterra kernels. Furthermore, such a network can describe any relations between input and output series.

The ANN regression analysis has another advantage over non-parametric methods. ANN regression can detect any type of response modulation by the input values. This type of analysis, which is known as parametric modulation analysis, deals not only with the existence of activation, but also with how that activation is modulated by the parameters of the event. To detect parametric modulation using non-parametric methods, we have to explicitly model the shape of the modulations; otherwise we can detect only linear modulations. The ANN regression, in contrast, does not require explicitly modeling of the modulation shape. Furthermore, any continuous modulation can be detected, owing to the ability of the ANN regression to adjust to any continuous functions.

This paper presents a method of ANN regression that can be applied to fMRI studies. In fMRI analyses, autocorrelation noise involved in the BOLD signal time course (Bullmore et al., 1996; Purdon and Weisskoff, 1998; Woolrich et al., 2001) may be a problem for the ANN regression, because it can fit any relations between events and BOLD signals. We considered this problem in analyzing null-task fMRI data and synthetic white noise data. In the following section, we describe the application of the ANN method to a practical case: a memory-guided saccade task. Because this task includes various responses with different shapes, it is a good example of how the ANN method can fit various shapes of responses. Finally, to explore the potential of the ANN method, a parametric modulation analysis was performed. Using a synthetic data set, we compared the ANN method and a non-parametric method and their ability to detect non-linear parametric modulations.

Analysis methods

Our analysis used a multi-layer ANN to regress fMRI signals using a history of recent events. The ANN is a network of simple processing nodes connected by a certain weights (Hertz et al., 1991). The model was inspired by biological processes, i.e., modeling biological neural network architectures such as parallel processing, with the node transfer function providing simple modeling of neural activation. Although it has been used for modeling various perceptual and cognitive processes in psychology and neuroscience, we used it here as an analytical tool. ANNs have been studied in-depth as multivariate analysis methods. In particular, feed-forward neural networks, trained using a back propagation algorithm (Rumelhart et al., 1996), can be used in a multivariate non-linear regression analysis by feeding explanatory variables as input and a dependent variable as the supervised signal. Back propagation is a network weight update method that minimizes errors between network output and the supervised signal. Here, we used a feed-forward artificial neural network (multi-layer perceptron) for regression of BOLD signals by event sequences using a learning algorithm that was a variant of standard back propagation.

Materials and methods

Fig. 1 shows a schematic diagram of the network architecture used in this analysis. The ANN has three layers: input, hidden, and output. Each unit in the input layer is connected to all the units in the hidden layer, and each unit in the hidden layer is connected to the sole output unit. The number of input units is set to be equal to the length of the event history. The number of hidden units should be set appropriately according to how non-linear the regressed function is, the size of the input space, etc. As described later, we used an excessive number of hidden units and an early stopping procedure to avoid over-fitting.

Network training

The analysis was conducted for each voxel independently. First, we determined the network size, length of the input event time

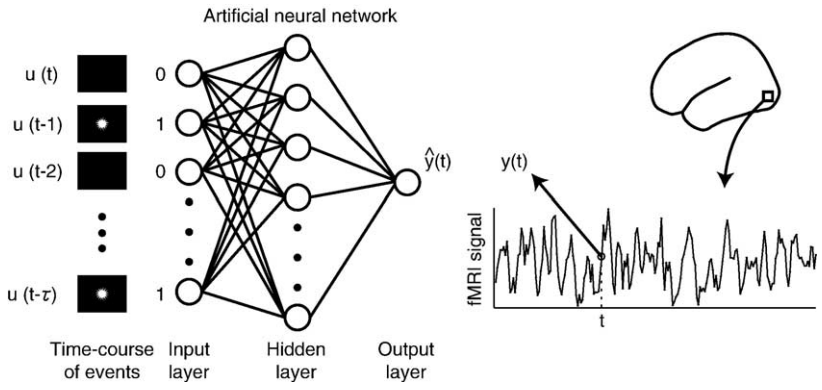

Fig. 1. Schematic illustration of artificial neural network used in fMRI analysis. Detailed procedures of the method are in the text.

course, temporal resolution of the input sequence, and number of hidden units. The temporal resolution of the input time course can be set higher than the MRI scanning interval, TR, when a jittered event-related design is used.

Then, the network was trained to output a BOLD response with the input of the event time course by adjusting its connection weights. For this training, RPROP (Riedmiller, 1994; Riedmiller and Braun, 1996), a variant of back propagation, was used. The RPROP (i.e., resilient propagation) is faster at converging and more stable without requiring adjustment of the learning parameters as opposed to standard back propagation. In this training, the BOLD signal at a certain time was used as the supervised signal, and an event time course before the sample time of the BOLD signal was used as the input. Through the RPROP procedure, network weights were modified to reduce the error between the network output and the BOLD signal. Appendix A contains a detailed explanation of the RPROP procedure with the specific parameters used in this paper.

Early stopping with cross-validation

It should be noted that the training should not be repeated until convergence. A critical problem of neural network learning is over-fitting or the bias-variance problem (Geman et al., 1992; Bishop, 1995). Because of their adjustability, artificial neural networks sometimes fit even a specific trend of a learning sample. The results of this kind of over-fitted network cannot be generalized to other samples. Various methods of regularization have been proposed to avoid over-fitting, such as jittering input, early stopping, weight decay, and Bayesian learning. Of these, we chose cross-validation and an early stopping procedure (Nelson and Illingworth, 1991; Wang et al., 1994). Bayesian regularization is also often used in ANN regression (Thodberg, 1996; MacKay, 1995); however, it is a slower method than early stopping because the training is repeated until convergence with a weight decay parameter being used to control regularization. In the early stopping method, network training is aborted before convergence, making it a faster method of regularization compared to the others. This is important since computation time is critical in fMRI analysis, which involves a huge amount of data.

In this procedure, a large number of hidden units were used and the training data were divided into a training set and a validation set. Only the training set was used for network training. After updating the weights using the training set, generalization error (validation error) was evaluated using the validation set. This procedure was repeated, changing out the training set for the validation set. We divided the samples (fMRI scans) into 10 sets of equal size, using 9 sets for training and 1 set for validation. One epoch of training was defined as 10 training repetitions, with the validation set being changed for each repetition. The division of sets was changed randomly in each epoch.

The network training continued as validation errors decreased. However, in reality, validation errors do not decline smoothly but rather oscillates, so it is difficult to determine when training should stop. Prechelt (1998) identified several criteria for early stopping to resolve this difficulty. Here, we used his PQ_1 criterion: the ratio of the increase in validation errors to the learning progress (training errors). Using this criterion, training was continued while training errors were declining rapidly and was stopped when error reduction slowed down and validation errors increased. A detailed description of PQ_1 and the stopping procedures are provided in Appendix B.

Reducing computation time

Computation time is another problem with this method. We had to perform time-consuming repeated calculations for each of the vast number of voxels in the fMRI data. The early stopping described above helps by shortening the computation time. We also took the following steps to further shorten the computation time. We did not analyze all the voxels, but we restricted the analysis only the brain section that was extracted using the BET algorithm (Smith, 2002) and we used the fast approximation of tanh (Anguita et al., 1993) in calculating the unit transfer function in the hidden layer.

Summary

Using the above procedures, the network learnt to output a BOLD signal from the input of event sequence. This is equivalent to a regression of the BOLD signal by the event sequence. In the following sections, we described the application of the ANN analysis method to fMRI data (null data, memory-guided saccade task, and parametric modulation) to evaluate ANN's capabilities and limitations.

ANN fitting to white noise and null fMRI data

When the ANN regression is applied to fMRI data, autocorrelation noise, which is often involved in a BOLD signal time course (Bullmore et al., 1996), may become a problem. Although we were able to avoid over-fitting by using the cross-validation procedure described above, an ANN may fit autocorrelation noise because it is not white noise. In this section, we examine how an ANN fits to noise signals, i.e., synthetic white noise and a null fMRI signal (no task was applied and the subject rested during the scan).

Autocorrelation noise is an important consideration in fMRI analyses (Purdon and Weisskoff, 1998; Woolrich et al., 2001), and several approaches have been used to solve the problem. One approach is 'coloring,' in which a BOLD signal time course is filtered to pass only the frequencies where there is a hemodynamic response. However, we did not want to restrict the shape of the hemodynamic response function, because one advantage of the ANN method is that any shape of response can be detected. Another approach is 'whitening,' in which autocorrelation noise is estimated and then removed to whiten the residual signal. This method requires repeated cycles of regression and noise estimation, which is impractical in an ANN regression analysis because the regression itself is time consuming in ANN regression.

A detrend procedure is also used to relax the autocorrelation noise (Tanabe et al., 2002). This method was used to fit a curve to the signal, thus enabling the removal of fitted low-frequency fluctuations from the signals. Because autocorrelation noise often appears as low-frequency fluctuations in BOLD signals, this method can reduce the autocorrelation noise. It has been previously shown that activation detectability was increased by detrending in an fMRI study (Marchini and Ripley, 2000; Tanabe et al., 2002).

We therefore used a detrend procedure to evaluate the effect of autocorrelation noise on an ANN regression analysis and compared the results of ANN fittings to synthetic white noise and null fMRI data with and without detrending.

Materials and methods

The ANN was fitted to the three noise signals: a synthetic white noise signal, a null-task fMRI signal, and a null-task fMRI signal with detrend. It was thought that the ANN fitting would depend on regressors representing the task design. We used three different designs as regressors: blocked, fixed-interval event-related, and jittered-interval event-related.

The blocked design consisted of a 30-s task period and a 30-s rest period, which was repeated eight times. The task period started 20 s after sequence onset. The fixed-interval event-related design consisted of a sequence of 25 events occurring at 20-s intervals. The jittered-interval event-related design was a sequence of 41 events occurring at one of seven different intervals (5–20 s by 2.5 s units). The interval during which an event occurred was random. From these sequences, regressors for the ANN analysis were established by setting 1 as the task period or event onset and 0 for all other intervals. The length of input event history was 30 s in the blocked design, 20 s in the fixed-interval event-related design, and 25 s in the jittered-interval event-related design. The temporal resolution of the input sequence was 2.5 s for all designs. Then, the number of input units was 12 in the blocked design, 8 in the fixed-interval event-related design, and 10 in the jittered-interval event-related design. The number of hidden units was 50 in all cases. We also examined a two-hidden-layer ANN; two hidden layers of 25 units were connected hierarchically between the input and output layers.

Null fMRI images were collected using a 1.5-T Siemens Vision (Erlangen, Germany). The functional images were 22 contiguous T2*-weighted axial echo-planar images (TR = 2500 ms, TE = 55.24 ms, flip angle = 90°, FOV = 256 × 256 mm, matrix = 64 × 64, slice thickness = 4 mm with no gap) of a male subject, aged 31. Written informed consent was obtained from the subject in accordance with institutional guidelines approved by the Ethics Committee for Human and Animal Research of the National Institute of Information and Communications Technology (NICT). Two hundred and five volumes were acquired with the first 5 volumes being discarded for the stability of the magnetic resonance state. The analysis region was restricted to brain sections and 18,810 voxels were analyzed. A synthetic white noise time course was made up of 200 random samplings from a zero-mean unit-variance normal distribution. The number of white noise time courses was also 18,810.

In the analysis of the detrended null fMRI data, we used a running-lines smoother (Marchini and Ripley, 2000). The running-lines smoother fits a linear regression of the nearest neighbors of a given point and uses the line to predict the value of the fitting curve at that point. The nearest span was set to 70 s, i.e., 28 scans. The ANN regression was performed with early stopping by cross-validation as described in the previous section.

Results and discussion

The ANN fitting was evaluated by its goodness of fit (GOF), which is the ratio of the regressed variance to the total variance. The GOF is same as the square of the correlation coefficient R^2. F statistics also have a one-to-one relationship with GOF and are widely used in neuroimaging studies. However, in the case of ANN, we could not define the degrees of freedom, which are needed to define the F distribution. In the ANN, there are numerous adjustable parameters, but not all these degrees of freedom were used because of early stopping. Thus, we did not use F statistics here.

Fig. 2a shows the distributions of GOF for each data set and each design with a one-hidden-layer ANN. This shows that the distribution of the GOF for a null-fMRI signal was similar to that for synthetic white noise. In every case, the results of fitting to detrended null data were better than fitting to non-detrended null data. The results of fitting to non-detrended null data were even worse than fitting to white noise in the blocked and the fixed-interval event-related designs. If the ANN fitted to low-frequency fluctuations, fitting to non-detrended null data would be better than fitting to white noise data and detrended null data. Thus, these results indicated that the ANN did not fit low-frequency fluctuation or autocorrelation noise in fMRI signals in these cases.

Though the one-hidden-layer ANN did not fit the low-frequency fluctuation noise, this could be due to the insufficient fitting ability of the ANN with one hidden layer. A feed-forward neural network with one hidden layer sometimes cannot fit functions with large non-linearity. To examine whether or not the lack of fitting to the low-frequency fluctuation noise was due to the network's inability to fit, we used the two-hidden-layer ANN. It is known that a two-hidden-layer ANN can more easily fit the non-linearity of input–output relations (Sontag, 1992; Bishop, 1995). Here, we used a two-hidden-layer ANN with 25 units for each layer. The training procedure was the same as for a one-hidden-layer ANN. Fig. 2b shows the distribution of the GOF for a two-hidden-layer ANN, which showed the same trend as for the one-hidden-layer ANN. Thus, even when the ANN's fitting capability was increased, it did not fit low-frequency noise, and the cross-validation procedure acted to avoid over-fitting to this noise.

The GOF was better for the jittered-interval event-related designs than it was for the other designs. In this case, the effect of detrending was small: In both the detrended and non-detrended cases, the GOF was better than in the white noise case. However, the results of fitting to non-detrended null data were at most the same as the fitting to detrended null data. This shows that low-frequency fluctuations did not supply additional fit for the ANN regression. Thus, the ANN did not fit this noise in this case either.

From these results, we conclude that autocorrelation noise is not as problematic as we thought in an ANN analysis. However, these results revealed another problem with the ANN regression: It is still difficult to predict the null distribution of the GOF, which is necessary for statistical inference. As shown in Fig. 2, the null distribution changed dramatically from design to design and was sometimes irregularly distributed as in the jittered-interval event-related design with a one-hidden-layer ANN. It is thus difficult to model these null distributions.

Instead of making a statistical inferences by modeling the null distributions, we took another approach. We took null fMRI data after taking task fMRI data with the same parameters. From this null data set, we created an empirical null distribution and used it for statistical inference. Though this procedure requires additional

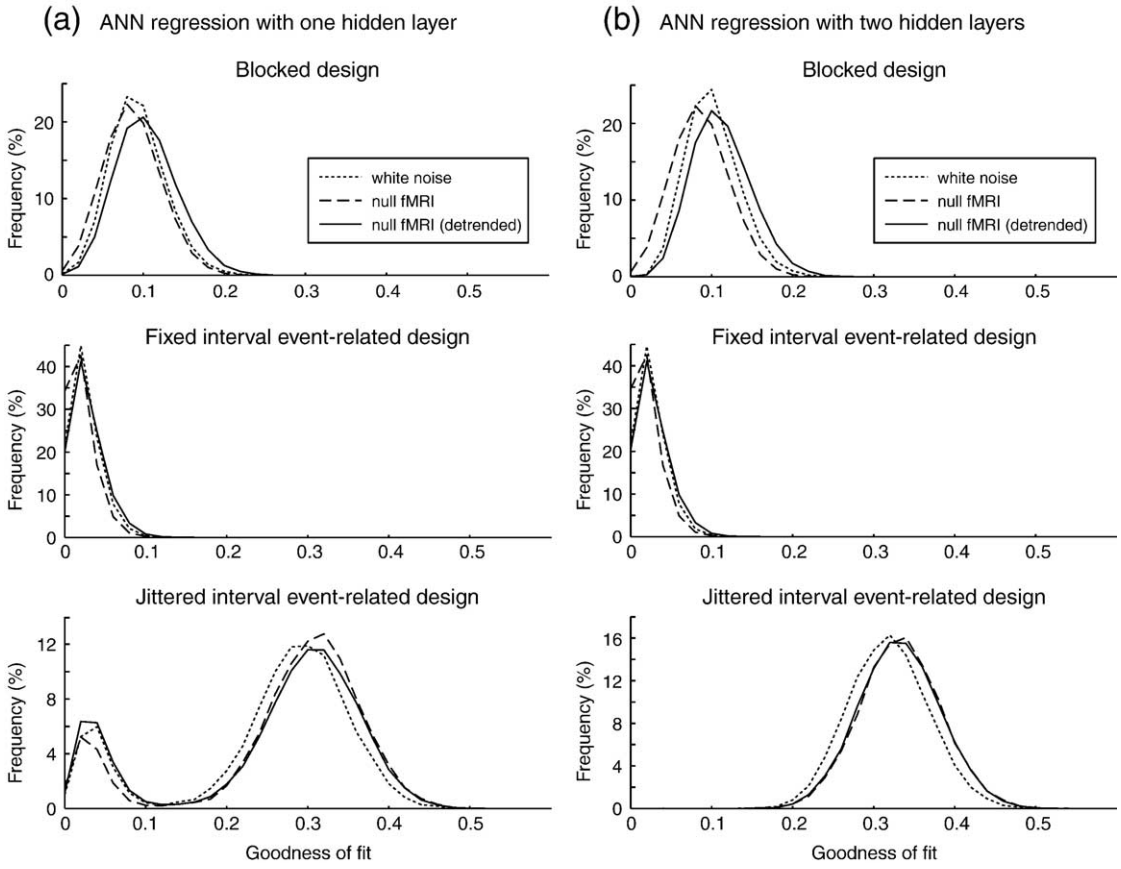

Fig. 2. Goodness of fit (GOF) distributions for each design and data regressed by one-hidden-layer ANN (a) and two-hidden-layer ANN (b). Horizontal axis shows GOF. Vertical axis shows percent frequency of each bin (width = 0.02).

fMRI measurement, no explicit null model is needed. We used this procedure in the next section.

Application to practical data: memory-guided saccade task

In an earlier section, we showed that the ANN did not regress the autocorrelation noise relative to the white noise. However, the ANN may fit noises as well as activation signals. In this section, we describe the use of a practical task experiment to examine whether we could discriminate activation from noise in the ANN regression result.

In the experiment, a memory-guided saccade task (Funahashi et al., 1989, 1991; Hikosaka and Wurtz, 1983) was performed. The subject was required to remember the target position and saccade to the target after a delay. This task required the subject to utilize various cognitive components with different time courses, i.e., the transient component of target detection, the sustained component of working memory, and the delayed component of saccade execution. We also examined whether the different time courses of the hemodynamic responses could be detected through the ANN regression.

Materials and methods

Subject and experimental task

One male subject, aged 31, participated in the experiment. Written informed consent was obtained from the subject in accordance with institutional guidelines approved by the Ethics Committee for Human and Animal Research of NICT.

Fig. 3 shows a schematic diagram of the experimental procedures (Kato and Miyauchi, 2003). On a screen with a gray background, a fixation square and six target squares were presented. The size of each square was 0.5 × 0.5° of a visual angle. The distance from the center to the target was 4° of the

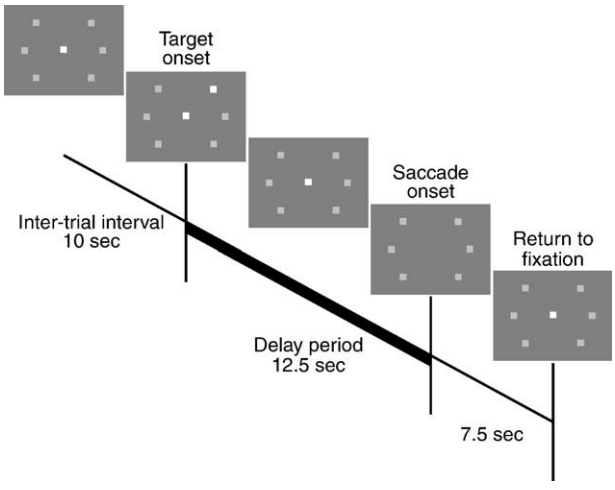

Fig. 3. Diagram of memory-guided saccade task. Refer to text for detailed procedure.

visual angle. Ten seconds after the end of the preceding trial, one of the target squares was brightly lit for 500 ms. The target position of the saccade was in the opposite direction to the lighted position. The subject was required to remember the target position and fixate on the center during a 12.5-s delay period. When the fixation square disappeared, the subject was to saccade to the target position. After 7.5 s, the fixation square reappeared and subject returned his eyes to the fixation position. Thirty trials were performed, with the order of the target direction changing randomly. The subject's eye movement was monitored by an infrared-video eye-monitoring system (NAC Image Technology, Tokyo, Japan).

Image acquisition

The fMRI images were collected using a 1.5-T Siemens Vision (Erlangen, Germany). The functional images were 20 contiguous T2*-weighted axial echo-planar images (TR = 2500 ms, TE = 55.24 ms, flip angle = 90°, FOV = 256 × 256 mm, matrix = 64 × 64, slice thickness = 5 mm with no gap) for the measurement of the BOLD effect. We acquired 370 volumes of functional images and discarded the first 5 volumes. We also acquired null-task fMRI data with the same parameters.

Analysis of fMRI images

The functional images were motion corrected using the first remaining image as the reference, and spatially smoothed using an 8 × 8 × 10 mm Gaussian kernel. We used SPM2 (Wellcome Department of Cognitive Neurology, London, UK) for these processing. The time course of the BOLD signal in each voxel was normalized to zero mean and unit variance, and detrend was applied using the running-lines smoother with a 70-s window. Before the ANN regression, the linear correlations between the time course of the BOLD signal and the realignment parameters were calculated, and then the correlated components were removed from the signal as nuisance effects.

In the ANN regression, the number of input units was set to 12. Each input unit represented the event in each TR before the scan onset time of the regressed signal (0–27.5 s before the scan onset time). The event sequence was composed of binary values representing the target onset. In the memory-guided saccade task, all the task-related components were correlated with the target onset. If the ANN detected any correlations, it was sufficient to use only one regressor for finding various task-related responses. Thus, this experiment was a fixed-interval event-related design and only the task onset was used as an event. We used a one-hidden-layer network with 50 hidden units. The network was trained using RPROP and an early stopping with cross-validation procedure was used. The learning parameters are as given in Appendix A. The analysis was applied only to the voxels in the brain sections.

The same procedure was applied to both the task fMRI data and the null fMRI data. We got the empirical null distribution of the GOF from the null fMRI data and used this distribution to evaluate the false-positive rate (P value) of the GOF.

Results and discussion

Fig. 4 shows the distributions of the GOF for the null and task data fitted by the ANN regression. The distribution of the task data set had a longer upper tail than that of the null data set. The center of Fig. 5 shows the regions in which a significantly higher fitting was obtained. The threshold of the map was P corrected <0.01 (GOF > 0.09) by FDR correction (Genovese et al., 2002; Benjamini and Hochberg, 1995). This activation map was normalized to the MNI brain template and projected onto a rendered view of it. The surrounding graphs show the hemodynamic responses in each region of interest (ROI). Each ROI was an 8 × 8 × 10 mm rectangle. The fitted response curve (solid line) and mean time course of the adjusted response with standard error (dotted line) were plotted in the graph. Theses lines were in good agreement with one another, indicating that the fitting was successful.

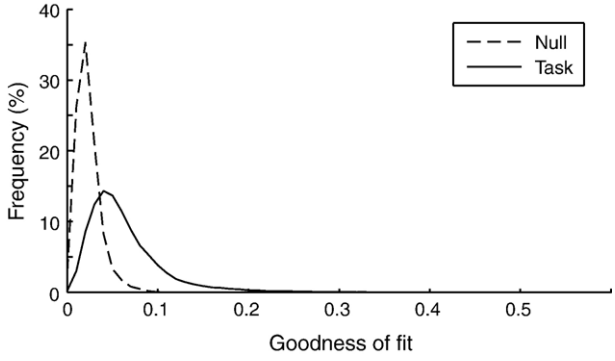

Fig. 4. Goodness of fit (GOF) distributions for null data and task data in memory-guided saccade task. Format is identical to that in Fig. 2.

As shown in Fig. 5, the activated regions had various time courses with regard to the hemodynamic response. In the occipital visual areas, the response was delayed until saccade onset. The inferior parietal and the ventral premotor regions responded to both the target onset and saccade onset. The superior parietal and dorsal premotor regions also responded to both the target and saccade onset, but the response was stronger for the target onset than for the saccade onset. The prefrontal regions responded to the target onset and this activation was sustained until the saccade onset.

The results of this experiment demonstrated that an ANN analysis could discriminate activation signals from noise and that the ANN could regress various shapes of hemodynamic responses. Various responses were regressed using only one event regressor (target onset). This has the advantage of avoiding the problem of colinearity between regressors, which may arise when multiple regressors are used. Correlation between model regressors often leads to difficulties in interpreting of the results (Andrade et al., 1999; Poline et al., 2003).

This analysis might be possible using a standard GLM approach if we modeled each event type, such as target onset, saccade onset, and returns to the center, and used a canonical HRF with temporal and dispersion derivatives to form the regressors (Friston et al., 1998a). However, a GLM analysis would only be possible if we knew that these events affect brain responses. With an ANN analysis, even if unknown or unpredicted events affect brain responses, we can detect them in as far as these events correlate with the modeled event. After detecting them, we can interpret the estimated hemodynamic responses if we know about these experimental events.

Thus far, we have described the validation of the ANN regression method as a viable method for detecting activations with different shapes of hemodynamic responses without any a priori assumptions. Though this is one of the advantages of this method, non-parametric methods can do the same thing. The ANN

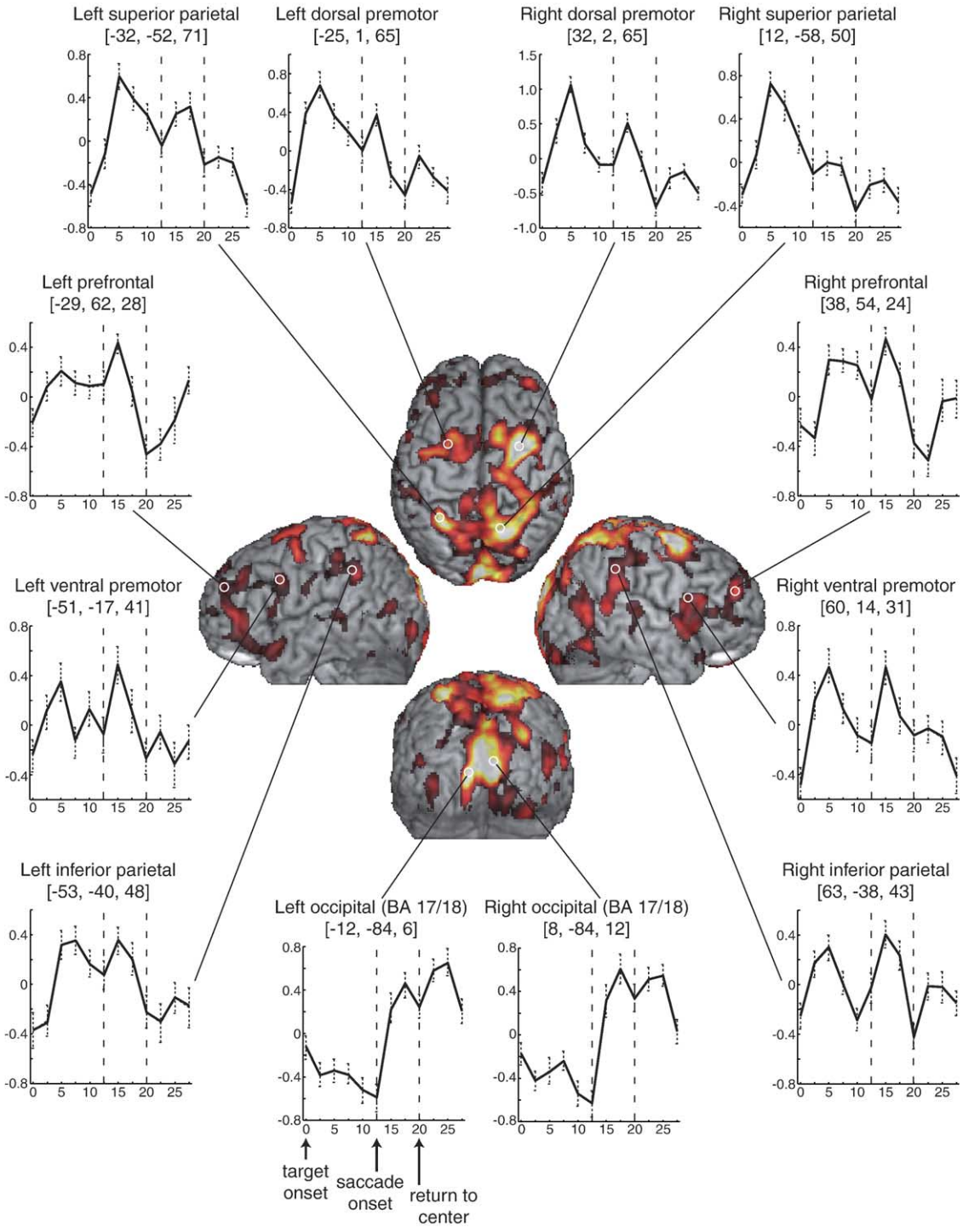

Fig. 5. Activation map (GOF statistical parametric map) and hemodynamic responses in each region of interest (ROI) for memory-guided saccade task. Map was projected onto rendered view of canonical MNI brain template. Map threshold was P corrected <0.01. Surrounding graphs show fitted response (solid line) and mean time course of adjusted response with its standard error (dotted line). ROIs were $8 \times 8 \times 10$ mm rectangle. Values above each graph show ROI center in MNI coordinates (mm). Horizontal 0 point in each graph shows target onset time, and broken vertical lines show saccade onset time and return to fixation time, respectively. Vertical axis value was normalized BOLD signal.

is more time consuming and it is more difficult to make statistical inferences than it is using non-parametric methods. What then is the advantage of applying the ANN regression to fMRI analysis? In the next section, we explore the potential advantages of an ANN regression over non-parametric methods.

Parametric modulation analysis using an ANN regression and non-parametric methods

One advantage of the ANN analysis over non-parametric methods is in finding modulated responses by the stimulus

parameters in a parametric modulation analysis (e.g., Pinel et al., 2001; Riecker et al., 2003; Coull et al., 2004). This analysis tells us not only 'which area' of the brain was activated by the stimulus, but also 'how' the brain region was activated by changes in the stimulus parameters. Though conventional non-parametric methods can describe response modulations, these methods require explicit definition of the types of modulations, such as linear or non-linear, monotonic or non-monotonic. In contrast, an ANN can represent arbitrarily shaped modulations.

Here, we describe the results of comparing the use of an ANN and a non-parametric method to detect various response modulations. The non-parametric method used here was a robust Bayesian estimation of the hemodynamic response function (Marrelec et al., 2003). This method fits the hemodynamic response non-parametrically and the smoothness prior of the response function was imported in the Bayesian framework. Marrelec et al. (2003) showed that this method could robustly estimate hemodynamic responses even when the data were noisy. Hereafter, we refer to this method as the non-parametric method.

To validate the result of the analysis, we used a synthetic data set. Simulation analysis using a synthetic data set is helpful for validating an analysis method because we know the correct result and can compare it with the estimated results. The synthetic data sets included four types of modulation: linear, monotonic non-linear, non-monotonic non-linear, and non-smooth (Fig. 6b). For these data sets, we evaluated how accurately and robustly the ANN and non-parametric methods could estimate these modulations.

It was presumed that a non-parametric method in its original form could not describe the modulations other than in their linear form. To overcome this limitation, one modification was applied to the non-parametric analysis, i.e., different parameters were handled as discrete events. After applying this modification, the modulation pattern was irrelevant to the analysis because each modulation of each parameter was evaluated independently.

Thus, we evaluated three analysis methods (ANN, non-parametric, and non-parametric with discrete handling) by applying them to four types of modulation (linear, monotonic non-linear, non-monotonic non-linear, and non-smooth).

Materials and methods

Synthetic data set

The synthetic data set was made from convolution of the pulse input representing the event onset with a hemodynamic response

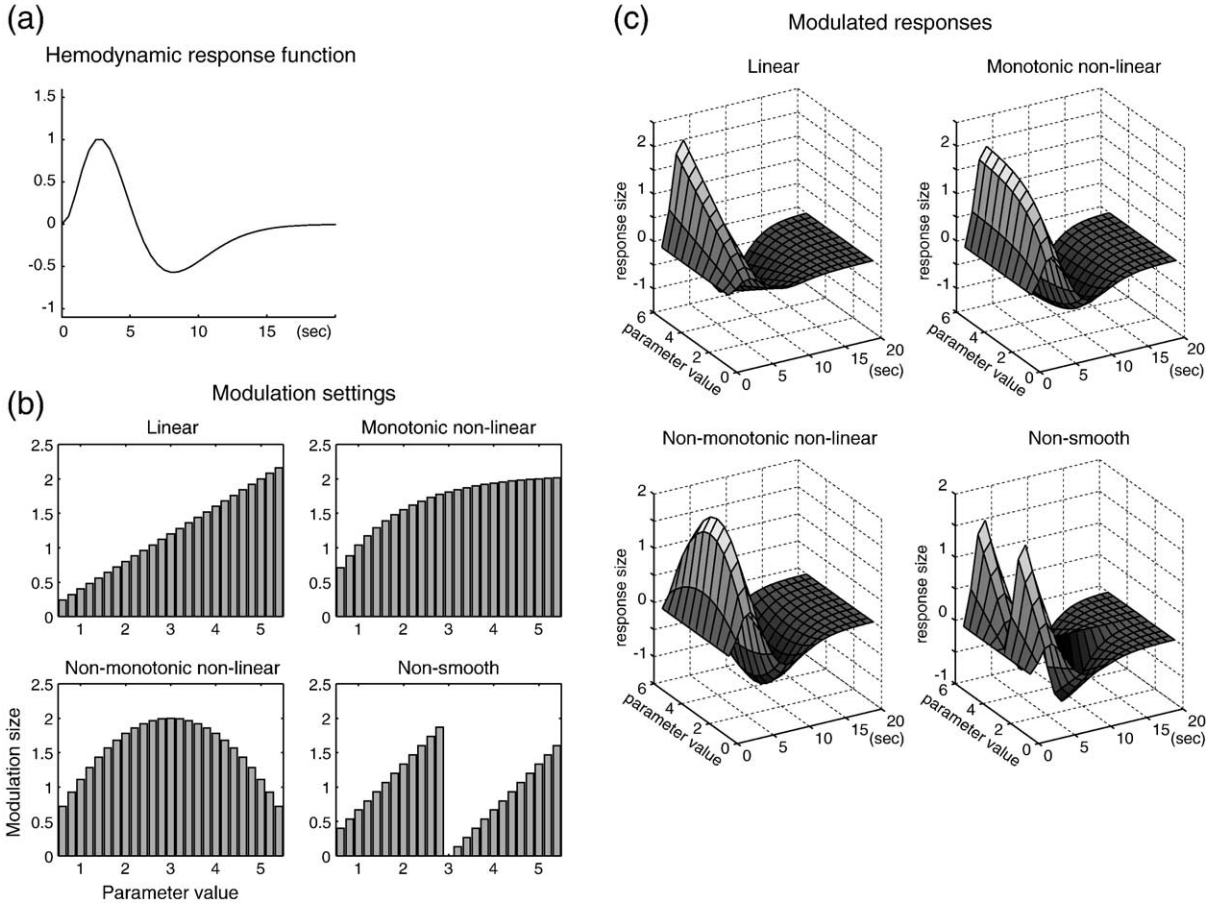

Fig. 6. Response function and modulation shapes in parametric modulation simulation. (a) Response function was modification of canonical shape in SPM2. Response was modulated by parameters as shown in panel b. (b) Modulation sizes at each parameter in each modulation type. Modulation shapes were $y = 0.4x$ in linear, $y = 2.062*(1 - \exp(-0.7x))$ in monotonic non-linear, $y = (9 - x^2) / 4.5$ in non-monotonic non-linear, and $y = (x\%3)*2 / 3$ in non-smooth, where x is parameter, y is modulation size, and % is modulo operator, which returns remainder when left value was divided by right value. (c) Modulated responses for each type.

function. The shape of the response function is shown in Fig. 6a. This shape was a modification of the canonical hemodynamic response function implemented in SPM2 (Wellcome Department of Cognitive Neurology, London, UK). This response has an early positive peak and a large undershoot, as is sometimes observed in actual fMRI data (Aguirre et al., 1998). This response was modulated by the parameter values in the four types of modulation shown in Fig. 6b. Fig. 6c shows the modulated response functions for each modulation type.

The linear modulation represented a linear relationship between the input parameters and modulation sizes. The monotonic non-linear modulation showed that the modulation sizes were saturated at higher parameter values, though the input–output relation was monotonic, i.e., the larger the input parameter, the larger the output modulation. The non-monotonic non-linear modulation showed that the modulation size peaked in the middle of the parameter value, indicating that the input–output relation was not monotonic. The non-smooth modulation demonstrated that there was discontinuity in the input–output relationship.

We made synthetic signals simulating the following fMRI design. TR was set to 3.0 s and each event occurred at a fixed interval of 16.5 s. The parameter set had 25 parameters from 0.6 to 5.4 in units of 0.2, and each parameter event occurred only twice. The input values and modulation size are shown in Fig. 6b. The total number of events was 50, and the order of parameter occurrence was randomized. To simulate the fMRI signals, an event sequence was convolved with the hemodynamic response function shown in Fig. 6a, and then Gaussian noise was added. The SNR of the synthetic signals was set to 1 ($SNR = 20 \log_{10}(Var_{signal} / Var_{noise})$). The event sequence consisted of pulses at the event onset. The pulses were the same size as the modulation and the length of the signal was 275. This simulated signal time course was made 1000 times for each modulation type. Three different analyses were performed on these data to compare their estimating accuracy and stability.

Analysis method

In the ANN regression, the number of input units was 11 and the temporal resolution of these units was 1.5 s. Thus, each input unit represented the event in the 0–15 s before the measurement time of a regressed fMRI signal. The input values at event onset were the parameter values of the event and the values at other times were zero. The network had one hidden layer with 50 units and was trained using the procedures described in the previous sections.

In the non-parametric analyses, the robust Bayesian estimation of hemodynamic response function (Marrelec et al., 2003) was used. Here, the signal time course y was modeled by $y = Xh + D\lambda + e$ (X: design matrix; h: hemodynamic response function (HRF); D: nuisance effects; λ: nuisance coefficient; e: residual). The HRF was estimated by the maximum likelihood method in a Bayesian frame with two priors: the HRF starts and ends at 0, and the HRF is smooth (refer to Marrelec et al., 2003, for detailed procedures). The length of h was set to 11, which is 15 s in units of 1.5 s. Only a constant term was modeled as a nuisance effect because only white noise was added to the synthetic data. This method was used in two different types of parameter handling. The first type dealt with all parameters as the same event, i.e., it was assumed that the same response was modulated by different parameters. The second type dealt with different parameters as different events, which allowed different shapes for different parameters. The first type of handling could adopt only linear modulation because no modulation model was implemented. In contrast, the latter type could adopt any modulation shape because it adopted each response independently. However, this type of handling reduced the sample size for each estimation, so it may be a disadvantage in terms of robustness.

Results and discussion

Fig. 7 shows the results of the estimated modulation sizes for each method and modulation. The peak value of the estimated response for each parameter was taken as the modulation size. The ANN performed well in estimating the modulations for the linear, monotonic non-linear, and non-monotonic non-linear modulations, but less well in estimating the non-smooth modulation. The non-parametric method was able to adopt the linear modulation well but not the other modulation shapes instead adopting a linear shape for all modulation types. The non-parametric method with discrete handling was able to estimate all modulation shapes but with a large standard error.

We also evaluated how accurately and robustly the methods fitted the responses. Fig. 8 shows the mean squared error of fitting and the variances of the fitted responses for each method and modulation. The ANN fitted well in all cases except non-smooth modulation. Its estimation variance did not change across the first three modulation types. This indicated that the ANN analysis was capable of robustly estimating the various types of modulations, though it did not fit the non-smooth modulation. The non-parametric method performed best when estimating linear modulation but its estimation accuracy worsened as the modulation became increasingly less linear. Though its variance was very small in all cases, it could only detect linear modulations. The non-parametric method with discrete handling was the worst-performed method with the largest errors and the largest variances for every modulation. These results demonstrated that the ANN has advantages over non-parametric methods in robustly estimating various modulations.

If we knew or could hypothesize the type of modulation in advance, we could use a non-parametric method built into a modulation model. However, the need for an a priori model of the response function reduces the benefit of the non-parametric approach; no explicit model of response shape was used. The failure of robust fitting in non-parametric analysis with discrete handling was due to the low number of event occurrences for each parameter. If a restricted number of parameters were used and each parameter occurred more frequently, this method might be more useful. Though in cases where the input parameters changes continuously and the same parameter values merely occur, this method cannot be used.

The ANN regression has no need for a priori assumptions of the modulation type and response shape and can be used for continuous input parameters. While we should remember that the ANN cannot adapt to non-smooth or non-continuous modulation shapes, it can be applied in broader areas of study, enabling us to perform exploratory analysis.

General discussion

This paper examined the application of an artificial neural network (ANN) regression analysis to fMRI data. To reduce

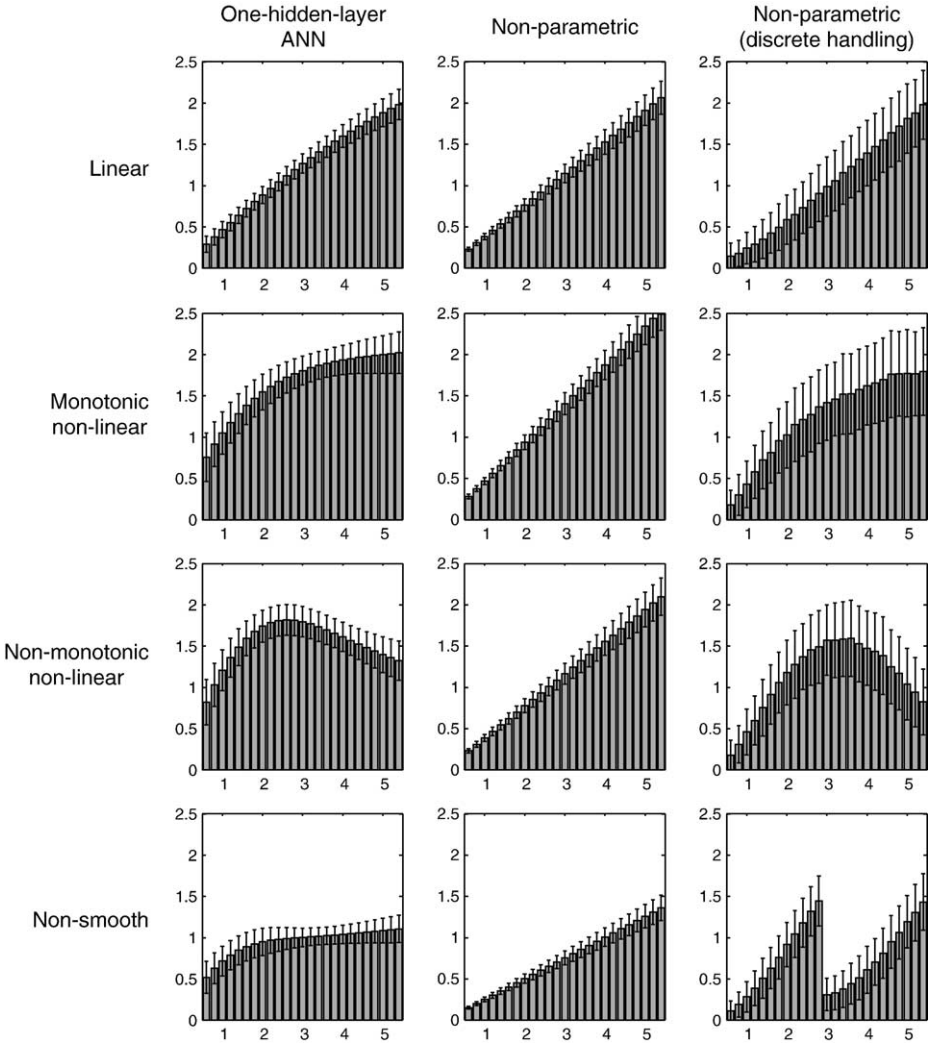

Fig. 7. Estimated modulation sizes for each analysis and modulation type. Modulation size (peak value in estimated response) of each parameter and its standard deviation are presented.

the computational time, we used an RPROP algorithm (Riedmiller, 1994; Riedmiller and Braun, 1996), and to avoid over-fitting we used early stopping with cross-validation. The result of fitting to synthetic white noise and null fMRI data indicated that the autocorrelation noise intrinsically included in most fMRI time course data was less problematic in the ANN regression. Applying the ANN regression analysis to data for a practical task (memory-guided saccade task) demonstrated that the ANN could discriminate activation from noise and extract various responses with different time courses. These results showed that the ANN regression could be used for fMRI data analysis.

The ANN regression analysis can extract arbitrary changes in BOLD responses that correlated with task sequences. This feature is similar to non-parametric analysis methods. The ANN regression analysis has another advantage over non-parametric methods in that it can detect continuous changes in responses modulated by input values. Though non-parametric methods can fit a linear modulation, they cannot detect other modulations except those that are explicitly modeled. Some parametric methods of fMRI analysis are able to detect various modulations by using polynomial basis functions as regressors in GLM (Buchel et al., 1998; Clark, 2002). However, these methods still have to hypothesize the shape of the response function. The ANN regression combines the advantages of non-parametric analysis (no need for assumptions about the shape of the response function) and parametric analysis (various continuous relations between input and output can be fitted).

We demonstrated the advantage of the ANN analysis in detecting various parametric modulations using synthetic data sets. Though a simulation analysis with synthetic data sets was helpful for validating the method, further research in which the method is applied to practical data is needed to test its real capability. The advantage of the ANN method is its ability to detect unknown or unassumed modulations. Thus, in future work, we may find previously unknown brain responses using this method.

In the above examples, we used sparse event-related design in which events were widely separated in time. We used this design to demonstrate the ability of ANN analysis to estimate the HRF. Sparse event-related design is the simplest design for estimating the shape of the HRF. As shown in the example of a jittered event-related design in the section of ANN fitting to white noise

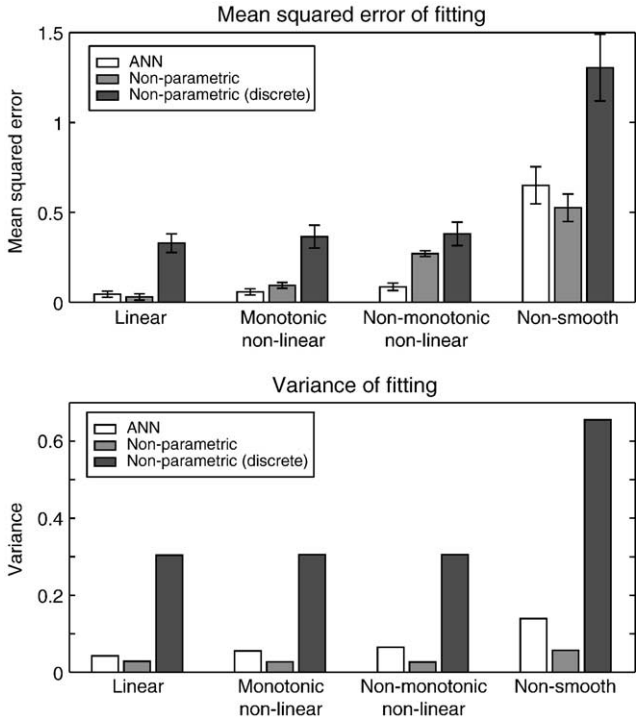

Fig. 8. Estimation errors and variances for each method. Upper panels show mean squared error from true signal and its standard deviation. Lower panel shows estimation variance: mean variance of each time point of estimated response.

and null fMRI data, the ANN analysis is also applicable to other designs, such as rapid event-related design, in which much shorter inter-trial intervals are used. In this case, it would be advisable to optimize the design for estimation efficiency, which is a measure of the variance in estimates of the HRF (Liu and Frank, 2004; Liu, 2004; Josephs and Henson, 1999). If we are unable to correctly estimate the shape of the HRF, we cannot estimate the parametric modulations, and thus the advantage of the ANN over the non-parametric methods decreases.

The limitations of the ANN regression derived from the difficulty of performing statistical tests and the computational time needed for repeated calculation in the ANN training. As a statistical test of regression, we proposed using null data to create an empirical null distribution. This imposes additional costs and time on an fMRI experiment. Furthermore, this method assumes that noise processes are the same for all brain regions; each voxel in different regions was assumed to be independently and identically distributed (i.i.d.). Though this method of statistical analysis is useful in practice, we should be aware that the P value is defined in this approximated null distribution.

To reduce the computational time, we restricted the analyzed regions to the voxels within the brain sections, and early stopping also saved time. In the memory-guided saccade task experiment, it took about 4.5 h for 19,976 voxels using a common Linux PC workstation (Dual Xeon 2.8 GHz, 2 GB memory). Though it took longer than other methods such as SPM analysis or non-parametric estimation, this is still within an acceptable range. If the analysis regions are further restricted to only gray matter, we could save more time, though extracting gray matter is also time consuming.

The ANN regression analysis can detect various fMRI signal changes that correlate with a task without a priori assumptions. This method will enable us to carry out exploratory searches for brain responses that have an unknown response shape and unknown parametric relation with the stimulus parameters.

Acknowledgment

We would like to thank to Dr. Makoto Kato for his helpful comments and technical assistance with the memory-guided saccade task.

Appendix A. RPROP training procedure

Feed-forward artificial neural networks function by propagating input to output by applying weighted sum and transfer function operations. The propagation from input to output is described by the equations below.

$$\begin{aligned} \boldsymbol{x}_{\text{in}} &= [u(t), u(t-1), \cdots, u(t-\tau)] \\ \boldsymbol{x}_{\text{hidden}} &= \tanh(\mathbf{W}_{\text{hidden-in}}\boldsymbol{x}_{\text{in}} + \boldsymbol{b}_{\text{hidden}}) \\ \hat{y} &= \mathbf{W}_{\text{out-hidden}}\boldsymbol{x}_{\text{hidden}} + \boldsymbol{b}_{\text{out}} \end{aligned} \tag{A.1}$$

The input vector $[u(t),u(t-1),\ldots,u(t-\tau)]$ of the event time course was multiplied by the weight matrix $\mathbf{W}_{\text{hidden-in}}$. The hidden units received this input with the bias values $\boldsymbol{b}_{\text{hidden}}$ of each hidden unit and applied the transfer function tanh to output $\boldsymbol{x}_{\text{hidden}}$. The output unit received the sum of $\mathbf{W}_{\text{out-hidden}}\boldsymbol{x}_{\text{hidden}}$ with the bias value, $\boldsymbol{b}_{\text{out}}$. This sum, $\hat{y}(t)$, was an output of the network. To reduce the calculation time, tanh was approximated using the following (Anguita et al., 1993):

$$\tanh(x) = \begin{cases} 0.96016 & \text{if} \quad x > 1.92033 \\ 0.96016 - 0.26037(x-1.92033)^2 & \text{if} \quad 0 < x \le 1.92033 \\ 0.26037^*(x+1.92033)^2 - 0.96016 & \text{if} \quad -1.92033 < x < 0 \\ -0.96016 & \text{if} \quad x \le -1.92033 \end{cases} \tag{A.2}$$

The purpose of network training is to minimize error between the network output and the supervised signal by adjusting the weights of the network connection. Error E was defined as

$$E = \frac{1}{2}(\hat{y}(t) - y)^2 \tag{A.3}$$

where y is a supervised signal and $\hat{y}(t)$ is a network output at time t. In a standard back propagation algorithm (Rumelhart et al., 1996), network weights are updated by (∂E) / (∂w_{ij}), a partial derivative of E for each weight. The RPROP procedure (Riedmiller, 1994; Riedmiller and Braun, 1996) introduced an individual update value Δ_{ij} for each weight, which determines the

size of the weight update to enhance the training. The procedures for weight updating in RPROP are as follows (Riedmiller and Braun, 1996):

$$
\begin{aligned}
&\text{if}\left(\frac{\partial E}{\partial w_{ij}}(t-1)*\frac{\partial E}{\partial w_{ij}}(t)>0\right)\text{then }\{\\
&\Delta_{ij}(t)=\mathbf{min}\left(\Delta_{ij}(t-1)*\eta^{+},\Delta_{\max}\right)\\
&\Delta w_{ij}(t)=-\mathbf{sign}\left(\frac{\partial E}{\partial w_{ij}}(t)\right)*\Delta_{ij}(t)\\
&w_{ij}(t+1)=w_{ij}(t)+\Delta w_{ij}(t)\\
&\}\text{else if }\left(\frac{\partial E}{\partial w_{ij}}(t-1)*\frac{\partial E}{\partial w_{ij}}(t)<0\right)\text{then }\{\\
&\Delta_{ij}(t)=\mathbf{max}\left(\Delta_{ij}(t-1)*\eta^{-},\Delta_{\min}\right)\\
&w_{ij}(t+1)=w_{ij}(t)-\Delta w_{ij}(t-1)\\
&\frac{\partial E}{\partial w_{ij}}(t)=0\\
&\}\text{else }\{\\
&\Delta w_{ij}(t)=-\mathbf{sign}\left(\frac{\partial E}{\partial w_{ij}}(t)\right)*\Delta_{ij}(t)\\
&w_{ij}(t+1)=w_{ij}(t)+\Delta w_{ij}(t)\ \}
\end{aligned}
\tag{A.4}
$$

where min and max operators return minimum and maximum value of two numbers, respectively. The sign operator returns +1 if the argument is positive, −1 if the argument is negative, and otherwise, 0. We used the parameters $\eta^{-}=0.5$, $\eta^{+}=1.4$, $\Delta_{\min}=1e^{-6}$, and $\Delta_{\max}=50.0$. The initial values were $(\partial E)/(\partial w_{ij})(0)=0.0$, $\Delta_{ij}(0)=0.1$, and $w_{ij}(0)$: random values sampled from a Gaussian distribution with 0 mean and 0.01 standard deviation. This weight update procedure was repeatedly applied to all weights until the stopping criterion described in Appendix B was met.

Appendix B. Criterion for early stopping

To avoid over-fitting, a procedure that ensured early stopping with cross-validation was used. The error for the validation set was a criterion for stopping. Because validation error changes are not smooth, it is difficult to determine when training should be stopped. Prechelt (1998) showed some criteria for early stopping for resolution of this difficulty. Here we used his PQ_1 criterion: ratio of increase in validation error to learning progress (training error). Using this criterion, training was continued while training errors were reducing rapidly. It is thought that even when the validation error reaches a local minimum, there may be a more optimal value while the weight update is active. When the error reduction slows down and validation errors start to increase, the training is stopped. This concept was implemented as follows.

Network training was continued until the ratio of generalization loss to training progress (training error) exceeded 1. Generalization loss was defined by

$$
\mathrm{GL}(t)=100\left(\frac{E_{\mathrm{va}}(t)}{E_{\mathrm{opt}}(t)}-1\right) \tag{B.1}
$$

where $E_{\mathrm{va}}(t)$ is the validation error at i, and $E_{\mathrm{opt}}(t)$ is the minimum validation error until t. Training progress was defined by

$$
P_k(t)=1000\left(\frac{\sum_{t-k+1}^{t}E_{tr}(t')}{k\mathbf{min}_{t-k+1}^{t}E_{tr}(t')}-1\right) \tag{B.2}
$$

where $E_{\mathrm{tr}}(t)$ is the training error at t and $\min_{t-k+1}^{t}E_{tr}(t)$ is the minimum training error from $t-k+1$ to t (here we used $k=5$). Then PQ_1 was defined as

$$
\text{continue training until }\frac{\mathrm{GL}(t)}{P_k(t)}>1. \tag{B.3}
$$

References

Aguirre, G.K., Zarahn, E., D'esposito, M., 1998. The variability of human, BOLD hemodynamic responses. NeuroImage 8, 360–369.

Andrade, A., Paradis, A.L., Rouquette, S., Poline, J.B., 1999. Ambiguous results in functional neuroimaging data analysis due to covariate correlation. NeuroImage 10, 483–486.

Anguita, D., Parodi, G., Zunino, R., 1993. Speed Improvement of the back-propagation on current generation workstations. Proceedings of the World Congress on Neural Networking. Lawrence Erlbaum/INNS Press, Portland, Oregon, pp. 165–168.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J.R. Stat. Soc. B57, 289–300.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, New York.

Buchel, C., Holmes, A.P., Rees, G., Friston, K.J., 1998. Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. NeuroImage 8, 140–148.

Bullmore, E., Brammer, M., Williams, S.C., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., Sham, P., 1996. Statistical methods of estimation and inference for functional MR image analysis. Magn. Reson. Med. 35, 261–277.

Clark, V.P., 2002. Orthogonal polynomial regression for the detection of response variability in event-related fMRI. NeuroImage 17, 344–363.

Coull, J.T., Vidal, F., Nazarian, B., Macar, F., 2004. Functional anatomy of the attentional modulation of time estimation. Science 303, 1506–1508.

Dale, A.M., Buckner, R.L., 1997. Selective averaging of rapidly presented individual trials using fMRI. Hum. Brain Mapp. 5, 329–340.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. NeuroImage 7, 30–40.

Friston, K.J., Josephs, O., Rees, G., Turner, R., 1998. Nonlinear event-related responses in fMRI. Magn. Reson. Med. 39, 41–52.

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12, 466–477.

Funahashi, K., 1989. On the approximate realization of continuous mappings by neural networks. Neural Netw. 2, 183–192.

Funahashi, S., Bruce, C.J., Goldman-Rakic, P.S., 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. 61, 331–349.

Funahashi, S., Bruce, C.J., Goldman-Rakic, P.S., 1991. Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. 65, 1464–1483.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Comput. 4, 1–58.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical

maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.
Goutte, C., Nielsen, F.A., Hansen, L.K., 2000. Modeling the haemodynamic response in fMRI using smooth FIR filters. IEEE Trans. Med. Imag. 19, 1188–1201.
Hertz, J., Krogh, A., Palmar, R.G., 1991. Introduction to the Theory of Neural Computation. Perseus Books, Reading, MA.
Hikosaka, O., Wurtz, R.H., 1983. Visual and oculomotor functions of monkey substantia nigra pars reticulata: III. Memory-contingent visual and saccade responses. J. Neurophysiol. 49, 1268–1284.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366.
Josephs, O., Henson, R.N., 1999. Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philos. Trans. R. Soc. Lond., B Biol. Sci. 354, 1215–1228.
Kato, M., Miyauchi, S., 2003. Human precentral cortical activation patterns during saccade tasks: an fMRI comparison with activation during intentional eyeblink tasks. NeuroImage 19, 1260–1272.
Liu, T.T., 2004. Efficiency, power, and entropy in event-related fMRI with multiple trial types Part II: design of experiments. NeuroImage 21, 401–413.
Liu, T.T., Frank, L.R., 2004. Efficiency, power, and entropy in event-related FMRI with multiple trial types Part I: theory. NeuroImage 21, 387–400.
MacKay, D.J.C. 1995. Probable networks and plausible predictions—A review of practical Bayesian methods for supervised neural networks. Network: Computation in Neural Systems 6, 469-505.
Marchini, J.L., Ripley, B.D., 2000. A new statistical approach to detecting significant activation in functional MRI. NeuroImage 12, 366–380.
Marrelec, G., Benali, H., Ciuciu, P., Pelegrini-Issac, M., Poline, J., 2003. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. Hum. Brain Mapp. 19, 1–17.
Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E., Buckner, R.L., 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. NeuroImage 11, 735–759.
Nelson, M.C., Illingworth, W.T., 1991. A Practical Guide to Neural Nets. Addison-Wesley, Reading, MA.
Pinel, P., Dehaene, S., Riviere, D., LeBihan, D., 2001. Modulation of parietal activation by semantic distance in a number comparison task. NeuroImage 14, 1013–1026.
Poline, J.B., Kherif, F., Penny, W., 2003. Contrasts and Classical Inference. In: Frackowiak, R.S.J., Friston, K.J., Frith, C., Dolan, R., Price, C.J., Zeki, S., Ashburner, J., Penny, W.D. (Eds.), Human Brain Function, 2nd edition. Academic Press.
Prechelt, L., 1998. Automatic early stopping using cross validation: quantifying the criteria. Neural Netw. 11, 761–767.
Purdon, P.L., Weisskoff, R.M., 1998. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. Hum. Brain Mapp. 6, 239–249.
Riecker, A., Wildgruber, D., Mathiak, K., Grodd, W., Ackermann, H., 2003. Parametric analysis of rate-dependent hemodynamic response functions of cortical and subcortical brain structures during auditorily cued finger tapping: a fMRI study. NeuroImage 18, 731–739.
Riedmiller, M., 1994. Rprop—Implementation Details. Technical Report. University of Karlsruhe.
Riedmiller, M., Braun, H., 1996. A direct adaptive method for faster backpropagation learning: the Rprop algorithm. Proceedings of the ICNN (San Francisco).
Rumelhart, D.E., McClelland, J.L., the P.R.G. (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations 1. MIT Press, Cambridge, MA.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.
Sontag, E.D., 1992. Feedback stabilization using two-hidden-layer nets. IEEE Trans. Neural Netw. 3, 981–990.
Tanabe, J., Miller, D., Tregellas, J., Freedman, R., Meyer, F.G., 2002. Comparison of detrending methods for optimal fMRI preprocessing. NeuroImage 15, 902–907.
Thodberg, H.H., 1996. A review of Bayesian neural networks with an application to near infrared spectroscopy. IEEE Trans. Neural Netw. 7, 56–72.
Volterra, V., 1959. Theory of Functionals and of Integral and Integro-Differential Equations. Dover Publications, New York.
Wang, C., Venkatesh, S.S., Judd, J.S., 1994. Optimal stopping and effective machine complexity in learning. In: Cowan, J.D., Tesauro, G., Alspector, J. (Eds.), Advances in Neural Information Processing Systems 6. Morgan Kaufmann Publishers, Inc, pp. 303–310.
Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. NeuroImage 14, 1370–1386.
Wray, J., Green, G.G.R., 1994. Calculation of the Volterra kernels of non-linear dynamic systems using an artificial neural network. Biol. Cybern. 71, 187–195.