LETTER

Artificial Neural Networks as Analytic Tools in an ERP Study of Face Memory

Reiko Graham[1] and Michael R.W. Dawson[2]

[1] Center for Cognitive Neuroscience, Duke University,
Durham, NC, U.S.A. 27708
E-mail: rgraham@duke.edu

[2]Department of Psychology, University of Alberta
Edmonton, Alberta Canada T6G 2E9
E-mail: mdawson@ualberta.ca

(Submitted on June 17, 2003; Accepted on October 24, 2003)

Abstract - Despite inquiry, the existence of early event-related potential (ERP) correlates of face memory has yet to be confirmed. We investigated the possibility that such correlates exist but cannot be reliably detected by linear analysis. We compared the abilities of artificial neural networks (ANN's) and ANOVA in classifying ERP's from right temporal areas elicited by recognized and novel faces. ANOVA's were unable to distinguish between ERP types; however, ANN's were. Results suggest that early ERP's recorded over right areas do index memory related activity, but that this activity is in the form of higher-order relationships between voltage and time. Wiretapping revealed that classification was achieved through coarse coding in the hidden units and that a subset of timepoints seemed to be driving their activity. Although much remains to be resolved, these preliminary results provide support for early face memory effects and attest to the utility of ANN's in ERP analysis.

Keywords **-** Artificial neural networks, Early latency activity, Event-related potentials, Face memory, Temporal lobe

1. Introduction

Memory can be conceptualized as the result of processing in two cortical areas: unimodal and transmodal association areas [1]. Unimodal areas are modality specific and receive projections from primary sensory cortex, whereas transmodal areas receive inputs from more than one modality. Within a modality, it is also possible to further conceptualize memory as being the result of general and stimulus specific processes. Event-related potential (ERP) studies have associated visual memory with two effects that likely reflect the activity of areas involved in general memory because they are elicited by different stimuli and experimental contexts. First, ERP's to remembered items show a positivity relative to new items that occurs after about 400ms over parietal regions. This has been found with both words [2-5] and faces [6, 7]. The second effect is a sustained positivity over frontal areas, which occurs later than the parietal effect and has also been found with words [3-5] and faces [6, 7].

These results suggest that general visual memory effects occur relatively late after stimulus presentation. However, the timecourse of early face specific activity remains uncertain, especially as it pertains to memory. Neuroimaging [e.g., 8] and intracranial [e.g., 9] studies have identified right temporal lobe areas, particularly the right fusiform gyrus, as putative face sensitive areas. Although neurons in the macaque temporal lobe show early latency changes as a face becomes familiar [11], ERP evidence from humans is mixed. Electrophysiologically, fusiform activity is correlated with the N200, a negative deflection that occurs 200ms after face presentation and is

maximal over right occipitotemporal areas [10]. Some studies have shown evidence of face memory effects over right temporal areas during the timeframe of the N200 [e.g.,12, 13], while another study reports memory effects as early as 50ms [14]. However, other studies of face memory have not replicated these effects [e.g., 6, 7, 15-17].

There could be several explanations for these inconsistencies. We were interested in the possibility that memory-related activity could be represented in early latency ERP's from right temporal sites that are believed to exhibit face specific activity, but conventional methods of statistical analysis may not consistently detect them. The current study explored this possibility by comparing the ability of linear techniques and artificial neural networks (ANN's) to discriminate between ERP's elicited by remembered and novel faces recorded from right temporal sites. ANN's offer advantages over analyses based on the general linear model, particularly when they are employed for pattern recognition [18]. Because ANN's are taught by providing them feedback about their performance, they can tolerate noisy data [19] and detect complex, nonlinear interactions between input features that are not normally detected by conventional techniques. Therefore, they have the potential to be powerful tools for ERP analysis.

This study explored this potential by determining whether ANN's could distinguish between early latency ERP's recorded from right temporal sites to correctly recognized old and new faces, and comparing these results to those obtained using analysis of variance (ANOVA). Furthermore, if ANN's were successful, we were interested if they could elucidate the timecourse of these ERP memory effects recorded over face-sensitive cortical areas.

2. Methods

Subjects

The sample consisted of 54 right-handers with normal or corrected-to-normal vision participated for course credit or paid participation. After obtaining written consent, participants were seated in a soundproof, shielded chamber and EEG recording equipment was applied. Four subjects were rejected because of equipment problems, and 3 were rejected due to excessive artifact. The final data set consisted of 47 subjects (19 male, 28 female).

Materials and Methods

The stimuli consisted of 240, 3.5 x 2.7 inch, black-and-white photos of unfamiliar faces taken from the Purdue University, University of Stirling, and University of Northern British Columbia face databases. In each of 4 blocks, participants studied 30 faces in random order, and then performed a recognition test that included 30 studied and 30 new faces. The assignment of faces to the old and new conditions was counterbalanced. Trials consisted of a fixation (500 ms), a face (400 ms), a fixation (1600 ms), a response selection screen until a response was made, and a blank screen (1000 ms). The fixation was a 3.5 by 2.7 inch gray rectangle. At study, subjects were asked to study the faces for a subsequent memory test. At test, a response screen prompted subjects to make an old/new decision. Subjects made their responses by pressing keys with different hands. Hand use was counterbalanced.

ERP methods

EEG was collected from 32 Ag/AgCl electrodes embedded in a Quikcap (Neurosoft Inc., Sterling, Virginia). Sites included frontopolar (FPI, FP2), frontal (F7, F3, FZ, F4, F8), frontocentral (FC3, FZ, FC4), central (C3, CZ, C4), centroparietal (CP3, CPZ, CP4), parietal (P3, PZ, P4), frontotemporal (FT7, FT8), temporal (T7, T8), temporoparietal (TP7, TP8), and occipital (O1, OZ, O2) electrodes with linked mastoids as a reference. Horizontal eye movements were monitored with bipolar electrodes on the outer canthus of each eye; vertical movements, from electrodes placed above and below the left eye. EEG was recorded with a sampling rate of 500 Hz for 1700 ms starting 100 ms prior to face onset. Channels were amplified with a filter bandwidth of 0.03 - 50 Hz. Trials with values above 100 or below -100 μV were excluded. ERP's were averaged according to 2 categories: correctly classified old (hits) and new faces (CR's). The grand-averaged ERP's for the first 500 ms after face onset are shown in Figure 1 for the electrodes of interest (FT8, T8 and TP8).

3. Results

Two types of analyses were performed on the three electrode channels of interest: FT8, T8, and TP8. The first was a linear analysis of the data via repeated measures ANOVA; the second, analysis with ANN's.

Repeated measures ANOVA's were performed for each of the three channels on the first 500 ms of data after face onset from each of the temporal electrode channels. Data from each channel was parsed into 25 contiguous

20ms epochs that were also used as ANN inputs. Each electrode site was analysed with trial type (hit vs. CR's) and time (25 x 20ms epochs) as within-subjects factors. None of the ANOVA's was able to detect reliable differences between ERP's to hits and CR's during the first 500ms at any electrode site (main effect of type, type x time interaction, all F's < 1, p's>.05). All ANOVA's did find a significant effect of time, indicating that voltages were changing over time. In order to keep results as similar as possible to the ANN analysis, ANOVA's were repeated using within-subject standardized averaged voltage values for a 20ms epoch. Although it was not possible to assess main effects because of standardization, no time by type interactions were significant (all F's < 1, p's>.05). Three ANN's (one for each channel) were trained to discriminate between the two ERP types. To promote training and reduce network complexity, an input representation scheme based on [20] was used. Each ANN had 25 input units, each corresponding to the within-subject standardized averaged voltage values for a 20ms epoch, three hidden units and one output unit. All processing units used a logistic activation function. This architecture is shown in Figure 2 and was adopted because pilot studies indicated that ANN's with these specifications could discriminate between hits and CR's with the fewest number of hidden units and could generalize well to new cases. The output unit was trained to generate a response of 1 to ERP's corresponding to hits, and a response of 0 to ERP's to CR's.

The data from 47 subjects was used to construct two data sets; a training and a test set. The training set consisted of data from 42 subjects (42 hits and 42 CR's), while the test set contained data from 5 randomly selected subjects (10 patterns, 5 of each type). Data from the test set was not used for training. The ANN's were trained using the generalized delta rule [21]. Initially, all connection weights and processing unit biases were randomly assigned values between 1.0 and -1.0. The networks were trained with a learning rate of 0.075 and zero momentum. Weights and biases were updated after each training pattern presentation. Each pattern was presented once during a "sweep" of training; the order of pattern presentation was randomized before each sweep. The networks were trained until they converged (generated a hit for every pattern), where a hit was defined as response of .99 or higher when the desired response was 1, or .01 or lower when the desired response was 0. All ANN's were able to differentiate between ERP types. The first ANN

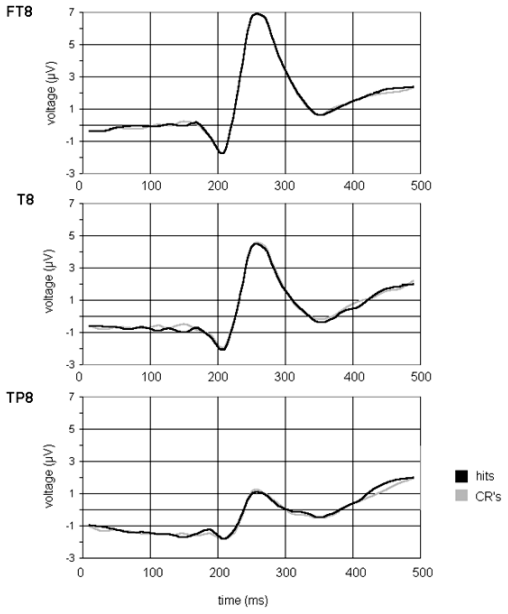

Figure 1. Grand-averaged voltages (μV) obtained from selected electrodes during face presentation.

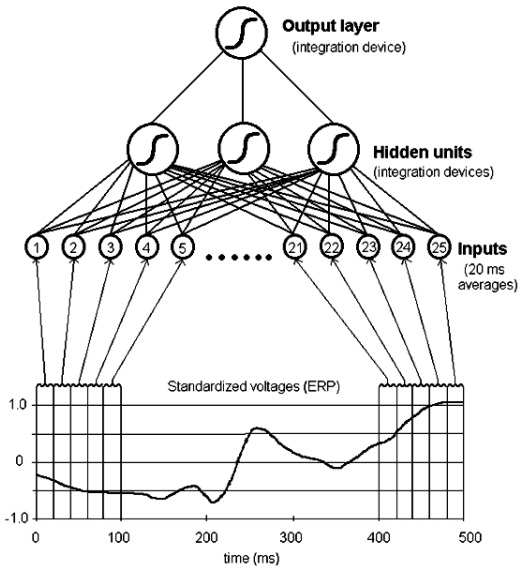

Figure 2. The network architecture used in this study. Average standardized voltage values represent averaged standardized ERP's obtained from each subject for each ERP type

(corresponding to the electrode FT8) converged after 3,640 sweeps; the second (corresponding to the electrode T8) after 2,263 sweeps, and the third (corresponding to the electrode TP8) after 2,828 sweeps. The prediction accuracy of each ANN was evaluated by exposing each network to the 10 test patterns. Because of the small number of testing patterns, stringent criteria identical to those used during training were employed. The first network (FT8) generalized to 80% of the new test patterns (3/5 CR's, 5/5 hits); the second (T8), to 90% (4/5 CR's, 5/5 hits); and the third (TP8), generalized to 80% (3/5 CR's, 5/5 hits).

To understand how the ANN's discriminated between hits and CR's, the activities of each hidden unit to each individual input pattern were examined. Alone, each unit was a poor discriminator of ERP type. However, when the activities of all three units were examined simultaneously, it was possible to see how discrimination was achieved. As can be seen in the scatterplots of hidden unit activity in Figure 3, the ANN's were able to classify ERP's by representing inputs in 3-D space that can be divided by a plane. Inputs representing hits fall on one side of this hypothetical plane; inputs corresponding to CR's fall on the other side.

Figure 3 demonstrates that hidden unit activity discriminates the two pattern types. Hidden unit activity is a nonlinear function of net input, which in this study is the sum of the signals traveling from the input units through the weighted connections to the hidden unit. In order to determine if some inputs (timepoints) were more influential than others in classifying hits and CR's, we used multiple regression to determine if a subset of timepoints could account for most of the variance in the input to each hidden unit. A linear approach seemed appropriate since no nonlinear transformations occur at the level of net input. This analysis was chosen over a straight interpretation of network weights because examination of connection weights revealed that inputs were not mapping onto only one hidden unit. Furthermore, due to large differences in voltage over time, the importance of a particular input in determining hidden unit activity is not only based on the magnitude of its connection weight to the hidden unit, but also on its magnitude before weighting. Regressions with net input as the predicted variable and weighted inputs as predictors permit the examination of both the connection weights and the initial values of the inputs.

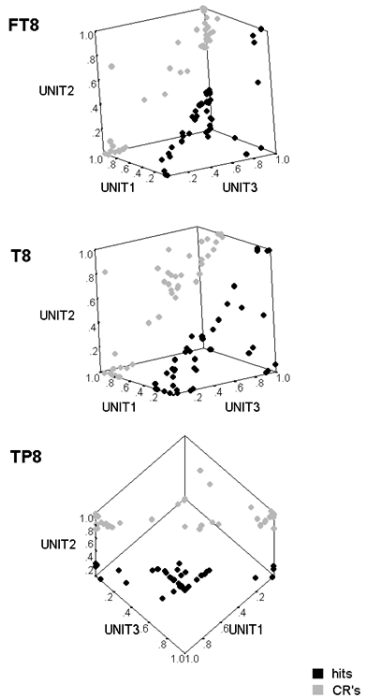

Figure 3. 3-D scatterplots of hidden unit activity for each of the ANN's. Plots correspond to ANN's trained on data from right temporal electrode sites (FT8, T8 and TP8, respectively).

Three regressions, one for each hidden unit, were performed on each of the three networks, with net input as the predicted variable. Timepoints entered as predictors were chosen on the basis of their standard deviations after weighting by the appropriate hidden unit, the rationale being that only timepoints with fairly large variance would contribute significantly to hidden unit activity. Inputs with the highest standard deviations after weighting were entered into each regression to yield the final regression equations.

A summary of these regressions is shown in Table 1, which shows the timepoints entered into the regression for each hidden unit and each channel, as well as measures of fit. Few inputs were necessary to account for variance in net input (one-third of the original inputs). All had high standard deviations after weighting, having the top third of all standard deviations for their respective hidden unit. There is also similarity between the timepoints entered, both within and across channels. These timepoints cover the entire range of the 500ms epoch; however, Table 1 shows some global characteristics of these inputs. Inputs between 100-200ms, as well as those between 400-500ms, appear to dominate each regression, with very few timepoints lying between 300-400ms. Early timepoints (<100ms) are also included in each regression. Additionally, timepoints between 200-300ms become more dominant in each solution as you move from the frontal channel (FT8) to the posterior channel (TP8).

4. Discussion

While ANOVA was unable to distinguish between ERP's, ANN's were able to differentiate between ERP's recorded over right temporal areas thought to be involved in the early stages of face processing. ANN's were able to discriminate between ERP's to correctly recognized old and new faces and generalized well to new test cases. Due to the small number of testing patterns, generalizations beyond this particular data set must be made cautiously. However, results suggest that memory-related activity is represented in early latency ERP's from temporal areas. This information may be in the form of higher-order voltage/time relationships that cannot be consistently detected with linear methods. The success of the ANN's was not particularly surprising: ANN's have been used to differentiate between ERP's produced by individuals with brain pathology and controls [e.g., 20, 21, 23]. This study provides evidence that ANN's can also be useful in differentiating between cognitive states.

Table 1. Timepoints identified as predictors of net input for each hidden unit of each network. Numbers in the table represent the input number, and r and R^2 values for each regression; time is indicated along to bottom of the table.

Inputs entered							r	R^2
FT8								
	HU1:	4	7 8 9 10	15		22 24	.92	.85
	HU2:	1 2	6 7 8		19 20	22	.83	.69
	HU3:	1	7	11	16 19	23	.96	.92
T8								
	HU1:		7 8	11 12 13		21 22 25	.91	.83
	HU2:	1 2 3	6 9	12		22 25	.97	.93
	HU3:	1	8	11 13	20	21 23 25	.98	.97
TP8								
	HU1:	3	7 10	12 13 14		23 25	.94	.88
	HU2:	1	7 8	12 13 14 15		25	.93	.86
	HU3:	2	8 9	11 13 14	19	23	.96	.93
time (ms)		0	100	200	300	400	500	

One reason for the inconsistencies between previous studies could be that early memory effects recruit relatively small cortical areas whose location and activity vary across individuals. This sample dependent variability could obstruct the detection of mean differences. ANN's, on the other hand, can distinguish between groups with overlapping distributions and variances [20] and classification occurs by comparing the entire timecourse of an ERP against those which belong to known types, without the aid of any a priori models. Effects in previous studies may also have been attenuated, either by virtue of their cortical generators or by the use of mastoid references. Since ANN's can tolerate noisy, incomplete data, they should be more robust to signal attenuation, as evidenced by the successful training of ANN's in this study despite the use of mastoid references.

Examination of hidden unit activity revealed that the ANN's were able to differentiate between hits and CR's by transforming inputs into three-dimensional pattern space. Furthermore, timepoints that could account for a significant amount of variability in the net input to each hidden unit had high variability after weighting. This suggests that standard deviation after weighting could be useful for the interpretation of trained networks.

Timepoints identified as predictors of net input via multiple regression were surprisingly consistent between and within channels, but were not circumscribed to any particular epoch, suggesting that voltage values over the entire timecourse were necessary for correct classification. However, Table 1 shows an almost bimodal distribution of these timepoints, with inputs between 100-300ms and 400-500ms dominating each solution, as if the ANN's were detecting two features in the data. The first feature (100-300ms) may be a memory effect corresponding to the N200 complex, while the second feature (400-500ms) may be the onset of the parietal effect, which was prominent after 500ms. Further examination of this first epoch is warranted; in particular, it would be interesting to see if an ANN can be trained to discriminate between hits and CR's when presented with data from the first 300ms after face presentation. While the second feature may be the onset of the parietal effect [see 2-7] and could be interpreted negatively, these results are actually quite heartening at this preliminary stage, as they do

provide some validation for network interpretation. The fact that the ANN's could detect this regularity before it became statistically significant is one reason to be optimistic about the use of ANN's for ERP analysis.

An interesting result of the regressions was the predominance of adjacent timepoints in their solutions. The fact that these timepoints were prominent in each solution was surprising given that adjacent timepoints should be highly correlated and therefore should not contribute much unique variance before weighting. The inclusion of adjacent timepoints as predictors in each regression, particularly between 100-300ms and 400-500ms, could signify that the ANN's are picking up differences in the onset latencies of both early and late ERP components. This suggestion can only be confirmed through topographic analysis of these epochs.

The purpose of this study was to extend the ERP literature by re-examining ERP memory effects with a different analytic tool, ANN's. Much remains to be revealed about how ANN's classify ERP's and a framework for future analyses remains to be established, particularly with regard to network interpretation. In addition, future work must examine the ability of these networks to generalize to new cases. Nevertheless, preliminary results are encouraging and provide support for early face memory effects and the continued use of ANN's in ERP analysis.

References

[1] M-M. Mesulam, "From sensation to cognition," Brain, Vol. 121, pp 1013-1052, 1998.
[2] A.J. Senkfor & C.VanPetten, "Who said what? An event-related potential investigation of source and item memory," Journal of Experimental Psychology: Learning, Memory and Cognition, Vol. 24, pp, 1005-1025, 1998.
[3] E.L. Wilding, M.C. Doyle & M.D. Rugg, "Recognition memory with and without retrieval of context: An event-related potential study," Neuropsychologia, Vol. 33, pp.743-767, 1995.
[4] E.L. Wilding & M.D. Rugg MD, "An event-related potential study of recognition memory with and without retrieval of source," Brain, Vol. 119, pp. 889-905, 1996.
[5] E.L. Wilding, "Separating retrieval strategies from retrieval success: An event-related potential study of source memory," Neuropsychologia, Vol. 37, pp. 441-454, 1999.
[6] R. Graham & R. Cabeza, "Event-related potentials of recognizing happy and neutral faces," Neuroreport, Vol.12, pp. 245-248, 2001.
[7] R. Graham & R. Cabeza, "Dissociating the neural correlates of item and context memory: An ERP study of face recognition," Canadian Journal of Experimental Psychology, Vol. 55, pp. 154-161, 2001.
[8] N. Kanwisher, F. Tong & K. Nakamura, "The effect of face inversion on the human fusiform face area," Cognition, Vol. 68, pp. B1-11, 1998.
[9] G. McCarthy, A. Puce, J.S. Gore & T. Allison, "Face-specific processing in the human fusiform," Journal of Cognitive Neuroscience, Vol. 9, pp. 605-610, 1997.
[10] S. Bentin, T. Allison, A. Puce, E. Penez & G. McCarthy, "Electrophysiological studies of face perception in humans," Journal of Cognitive Neuroscience, Vol. 8, pp. 551-565, 1996.
[11] E.T. Rolls, C.G.Baylis, M.E. Hasselmo & V. Nalwa, "The effect of learning on the face selective responses of neurons in the cortex of the superior temporal sulcus of the monkey," Experimental Brain Research, Vol. 76, pp. 153-164, 1989.
[12] W. Endl, P. Walla, G. Lindinger, W. Lalouschek, F.G. Barth, L. Deecke & W. Wang, "Early cortical activation indicates preparation for retrieval of memory for faces: An event-related potential study," Neuroscience Letters, Vol. 240, pp. 58-60, 1998.
[13] S. Campanella, C. Hanoteau, D. Dépy, B. Roisson, R. Bruyer, M. Crommelinck & J-M. Guérit, "Right N170 modulation in a face discrimination task: An account for categorical perception of familiar faces," Psychophysiology, Vol. 37, pp. 796-806, 2000.
[14] M. Seeck, C.M. Michel, N. Mainwaring, R. Cosgrove, H. Blume, J. Ives, T. Landis & D.L. Schomer, "Evidence for rapid face recognition from human scalp and intracranial electrodes,", Neuroreport, Vol. 8, pp. 2749-2754, 1997.
[15] M. Eimer, "Event-related brain potentials distinguish processing stages involved in face perception and recognition," Clinical Neurophysiology, Vol. 111, pp. 694-705, 2000.
[16] M. Eimer, "Effects of face inversion on the structural encoding and recognition of faces: Evidence from event-related brain potentials," Cognitive Brain Research, Vol. 10, pp. 145-158, 2000.

[17] W. Sommer, A. Heinz, H. Leuthold, J. Matt & S.R. Schweinberger, "Metamemory, distinctiveness and event-related potentials in recognition memory for faces," Memory and Cognition, Vol. 23, pp. 1-11, 1995.
[18] M.R.W. Dawson, A. Dobbs, H.R. Hooper, A.J.B. McEwan, J. Triscott & J. Clooney, "Artificial neural networks that use single-photon emission tomography to identify patients with probable Alzheimer's disease," European Journal of Nuclear Medicine, Vol. 21, pp. 1303-1311, 1994.
[19] L. Gupta, D.L. Molfese & R. Tammana, "An artificial neural network approach to ERP classification," Brain and Cognition, Vol. 27, pp. 311-330, 1995.
[20] J.D. Slater, F.Y. Wu, L.S. Honig, R.E. Ramsay & R. Morgan, "Neural network analysis of the P300 event-related potential in multiple sclerosis," Electroencephalography and Clinical Neurophysiology, Vol. 90, pp. 114-122, 1994.
[21] D.E. Rumelhart, G.E. Hinton & R.J. Williams, "Learning internal representations by error propagation," in D.E. Rumelhart and J. McLelland (eds.), *Parallel distributed processing, Vol 1*, MIT Press, pp. 319-362, 1986.
[22] B. Klöppel, "Classification by neural networks of evoked potentials. A first case study," Neuropsychobiology, Vol. 29, pp. 47-52. 1994.

Reiko Graham received her Ph.D. in psychology from the University of Alberta in 2002, and is currently a postdoctoral researcher at the Center for Cognitive Neuroscience at Duke University. Her research interests are focused on face perception and recognition and the functional neuroimaging of the various aspects of these processes. (Homepage: http://www.bcp.psych.ualberta.ca/~reiko/)

Michael R.W. Dawson received his Ph.D. in psychology from the University of Western Ontario in 1986, and is currently a full professor in the Psychology Department at the University of Alberta. His research interests include pure and applied research on artificial neural networks and the relationship of this research to empirical and theoretical issues in Cognitive Science. (Homepage: http://www.bcp.psych.ualberta.ca/~mike/)