Introduction of a very important neuroscience research article

for the non-neuroscientist reader

Prolonged exposure to unremitting stress damages a person's health. The research is unequivocal (read the science in the WBI Research Library). Mental health impact begins with anxiety. In worst cases, trauma can result. The diagnosis can be elusive because of the strict definition in the DSM-IV-TR (the diagnostic bible) and the reluctance of clinicians to admit what Heinz Leymann knew back in the late 1980's — work trauma is real.

Now comes a potential new neuroscience tool to complement the diagnostic toolkit — MEG. MEG stands for magnetoencephalography. PTSD can be detected with 97% accuracy using this non-invasive, but still experimental, procedure.

MEG measures the magnetic signals produced by the activity of the brain. Signals derive from the net effect of ionic currents flowing in the dendrites of neurons during synaptic transmission (EEG tests also measure these currents though slightly differently). These signals are very small. By comparison, the heartbeat produces a stronger signal. Magnetic resonance imaging (MRI) uses magnetic fields with a signal 3,000,000,000,000,000 stronger than the signal produced by the brain. In order to generate a signal that is detectable, approximately 50,000 active neurons are needed.

The essence of the MEG test is the measurement of the dynamic synchronous neural (bundled) interactions, an essential aspect of the brain function. MEG Dewars (caps) are helmet-shaped and contain as many as 300 sensors, covering most of the head. Then, complex statistical analyses of the data are required to differentiate activity across various areas of the brain to identify specific patterns.

MEG can detect neuronal events with a precision of 10 milliseconds or less, while fMRI, which depends on changes in blood flow, has a lower precision of several hundred milliseconds. MEG also accurately pinpoints sources in primary auditory, somatosensory and motor areas.

Research on brain–machine interfaces has been ongoing for at least a decade. During this period, simultaneous recordings of the extracellular electrical activity of hundreds of individual neurons have been used for direct, real-time control of various artificial devices. Thinking about moving an arm is converted to moving an artificial limb. Neuroprosthetics restores mobility in severely paralyzed patients.

MEG has been used to diagnose Alzheimer's, Parkinson's disease, epilepsy, schizophrenia, Sjögren's syndrome, chronic alcoholism, facial pain, and multiple sclerosis. "Communication patterns are very different from disease to disease," says Dr. Apostolos Georgopoulos from the Brain Sciences Center at the Minneapolis VA Medical Center and University of Minnesota. "So the different diseases create disturbances in the communication that can be used as a fingerprint, a signature, for the disease."

For years, the diagnosis of PTSD, has been a subjective process involving mental-health professionals conducting structured interviews with patients suffering PTSD-like symptoms.

In the attached article published in the Jan. 2010 issue of the *Journal of Neural Engineering,* Georgopoulos and his research team reported the successful diagnosis of PTSD using MEG. The Minnesota researchers used MEG to assess 74 U.S. military veterans believed to be suffering from PTSD, along with 250 subjects not thought to be suffering from the condition. Distinctive brain patterns indicating PTSD were found in 72 — or 97.3% — of the 74 people diagnosed with PTSD through the traditional interview process; false positives turned up in 31 of the 250 subjects (12.4%) without PTSD. The findings counter the popular notion that PTSD is not a real disease but a fabricated disorder. The neuronal patterns revealed a distinctive communication pattern, the "PTSD fingerprint."

Georgopoulos likens the MEG test for PTSD to diabetic blood-glucose monitoring tests to keep the disease under control. "The test is totally safe — there are no magnets, no isotopes — you can do it as frequently as you want," Georgopoulos says, adding that it also doesn't require dredging up the traumatic events that generate PTSD. "The whole thing takes literally a minute."

This most recent application of neuroscience to the world of stressed and traumatized individuals seems profound. However, the question remains about how to distribute the technology and methods to medical practitioners for use in the field.

Thanks to Wikipedia, *Time* magazine, and the *J. Neural Engineering* for help.

J. Neural Eng. **7** (2010) 016011 (7pp) doi:10.1088/1741-2560/7/1/016011

The synchronous neural interactions test as a functional neuromarker for post-traumatic stress disorder (PTSD): a robust classification method based on the bootstrap

A P Georgopoulos[1,2,4,5,6,9], H-R M Tan[1,2,8], S M Lewis[1,3], A C Leuthold[1,2], A M Winskowski[1,6], J K Lynch[1,2] and B Engdahl[6,7]

[1] Brain Sciences Center, US Department of Veterans Affairs Medical Center (11B), Minneapolis, MN 55417, USA
[2] Department of Neuroscience, University of Minnesota Medical School, Minneapolis, MN 55455, USA
[3] Department of Neurology, University of Minnesota Medical School, Minneapolis, MN 55455, USA
[4] Department of Psychiatry, University of Minnesota Medical School, Minneapolis, MN 55455, USA
[5] Center for Cognitive Sciences, University of Minnesota, Minneapolis, MN 55455, USA
[6] Psychology Section, US Department of Veterans Affairs Medical Center (11B), Minneapolis, MN 55417, USA
[7] Department of Psychology, University of Minnesota, Minneapolis, MN 5545, USA
[8] Centre for Cognitive Neuroimaging, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK

E-mail: omega@umn.edu

Received 31 August 2009
Accepted for publication 25 November 2009
Published 20 January 2010
Online at stacks.iop.org/JNE/7/016011

Abstract
Traumatic experiences can produce post-traumatic stress disorder (PTSD) which is a debilitating condition and for which no biomarker currently exists (Institute of Medicine (US) 2006 *Posttraumatic Stress Disorder: Diagnosis and Assessment* (Washington, DC: National Academies)). Here we show that the synchronous neural interactions (SNI) test which assesses the functional interactions among neural populations derived from magnetoencephalographic (MEG) recordings (Georgopoulos A P *et al* 2007 *J. Neural Eng.* **4** 349–55) can successfully differentiate PTSD patients from healthy control subjects. Externally cross-validated, bootstrap-based analyses yielded >90% overall accuracy of classification. In addition, all but one of 18 patients who were not receiving medications for their disease were correctly classified. Altogether, these findings document robust differences in brain function between the PTSD and control groups that can be used for differential diagnosis and which possess the potential for assessing and monitoring disease progression and effects of therapy.

1. Introduction

In a previous paper [2], we reported on the power of SNI to discriminate various brain disorders. In that study, using a resting state paradigm, neuromagnetic signals were recorded using MEG while subjects fixated on a spot of light for 60 s. After fitting an autoregressive integrative moving average (ARIMA) model and taking the stationary and non-autocorrelated residuals, all pairwise, zero-lag, partial crosscorrelations $\left(PCC_{ij}^{0}\right)$ and their z-transforms $\left(z_{ij}^{0}\right)$ between i and j sensors were calculated, yielding estimates of the strength and sign of direct synchronous coupling at

[9] Author to whom any correspondence should be addressed.

1 ms temporal resolution. Small subsets of z^0_{ij} reliably classified subjects to seven groups (six brain disorders and a control group). A key feature of that work was the use of genetic algorithms to successfully screen for and identify those subsets of z^0_{ij} that would be good classifiers. The excellent results obtained documented the presence of significant information in z^0_{ij} but left unclear the issue of how to find those subsets in a systematic fashion. In the present study, we had four major objectives, as follows. First, we applied the SNI test to a new (with respect to [2]) brain disorder, namely PTSD, for which no biomarkers are currently available [1]; second, we developed a robust classification method without using genetic algorithms; third, we used extensively the bootstrap (resampling with replacement) [3] to take into account practically all the variability in the subject populations; and fourth, we applied this new approach to simulate an 'externally validated' clinical trial and derive expected long-term, overall group classification rates as well as confidence intervals in classifying individual subjects.

2. Materials and methods

2.1. Subjects

All subjects participated in the study after providing informed consent, in adherence to the Declaration of Helsinki, and were financially compensated for their time. All study protocols were approved by the respective Institutional Review Boards.

2.1.1. PTSD group. Subjects were recruited from a database of veterans living in Minnesota and Wisconsin who had one or more contacts with the Minneapolis VA Medical Center. Their clinical records were reviewed to identify those with a likely current PTSD diagnosis, determined using a structured clinical diagnostic interview. These mental health assessments were based either on the CAPS [4] or DSM-IV-TR SCID PTSD module [5]. Recruitment letters were then sent to the subjects identified and the study coordinator followed up with phone calls to inquire whether they would be interested in participating. Interested veterans were invited to the VA Medical Center for an interview and to possibly participate in the MEG study. Ultimately, we included in this study only patients ($N = 74$) with confirmed PTSD as their primary diagnosis based on standard structured clinical interviews appropriate for PTSD. To minimize subject burden, we did not contact veterans with indicators of instability within the last 6 months (e.g. inpatient medical or mental health treatment, significant changes in health or medications, etc) or those with psychotic disorders based on medical record review. We also excluded veterans with active substance use disorders, serious chronic pain and other CNS disorders (e.g. Parkinson's disease, dementia, cerebral vascular accidents, etc).

Of the 74 PTSD patients studied, 69 were men (52.2 ± 1.77 years, mean ± SEM) and 5 were women (46.2 ± 4.12 years). Several patients had PTSD linked to childhood abuse or non-military trauma experienced as an adult (e.g. sexual assault). Those with combat trauma served in various wars, including World War II and the wars in Iraq and Afghanistan; most had fought in the Vietnam War. With respect to medications, 56 patients received medications related to PTSD whereas the remaining 18 were not receiving any such medications. Finally, for 50 patients, it was possible to derive an index of PTSD symptoms severity [6] through summation of 17 symptom scores from the SCID-PTSD module. Each of these symptoms was scored 1 (absent), 2 (subthreshold) or 3 (threshold); scores ranged from 17 to 51.

2.1.2. Control subjects. Healthy subjects ($N = 250$) within the age range of the PTSD patients were recruited from the general public; 151 were men (52.2 ± 1.3 years, mean ± SEM) and 99 were women (49.9 ± 1.6 years). Health was assessed by clinical interview of the subject upon initial contact and again in more detail at the time of consent. This included a general medical history, medication use and a detailed review of neurologic and psychiatric history.

2.2. Task and data acquisition

As with previous work [2], a simple fixation task was employed to engage the brain in a stable condition. Subjects lay supine within the electromagnetically shielded chamber and fixated their eyes on a spot ∼65 cm in front of them, for 60 s. MEG data were acquired using a 248-channel axial gradiometer system (Magnes 3600WH, 4-D Neuroimaging, San Diego, CA), sampled at 1017.25 Hz and band filtered between 0.1and 400 Hz.

2.3. Data pre-processing

Prior to the main analyses, cardiac correction of the MEG signals was performed using synchronous event subtraction [7, 8]. Subsequently, single trial MEG data from all sensors underwent 'prewhitening' [9, 10] using an (25,1,1) ARIMA model [2] to yield practically stationary and non-autocorrelated results. Residuals were estimated using the SPSS statistical package (SPSS for Windows, version 15, SPSS Inc., Chicago, IL, 2006).

2.4. Crosscorrelations

All possible pairwise zero-lag crosscorrelations ($N = 30\,628$, given 248 sensors) were computed using the DDCCF routine of the IMSL statistical library (Compaq Visual Fortran Professional edition version 6.6B). Next, the partial zero-lag crosscorrelations PCC^0_{ij} between i and j sensors was computed for all sensor pairs; thus, for any given pair of sensors (from a total of 248), the effects of the remaining 246 sensors were partialled out. The PCC^0_{ij} was then transformed to z^0_{ij} using Fisher's z-transformation [11] to normalize its distribution:

$$z^0_{ij} = 0.5\left[\ln\left(1 + PCC^0_{ij}\right) - \ln\left(1 - PCC^0_{ij}\right)\right]. \qquad (1)$$

In total, there were 30 628 such Z predictors; we denote by z^0_{ijs} the specific value of Z for a particular subject.

2.5. Group discrimination and intermediate classification

Classification of individual subjects to the PTSD or control group was performed using the bootstrap [3] as the key approach. This approach was motivated by the common practical situation where a 'new' subject S needs to be classified, given subjects of known condition, i.e. belonging to a control (i.e. non-PTSD) group C or to the PTSD group D. In classical discriminant analyses, S is assigned to the group (C or D) which yields the greater classification score in a single run. However, a major objective in clinical practice is to evaluate, if possible, S's assignment with respect to the larger populations C' and D', of which C and D are samples. Classification rates for C' and D' can be approximated using the bootstrap, by calculating S's assignment R times with respect to bootstrap samples of equal (between groups) and sufficiently large size B drawn with replacement from C and D. In this way, R assignments (i.e. classification outcomes and corresponding posterior probabilities) are obtained from which S's classification is then computed (based on the outcomes of R bootstraps), together with a confidence interval on the assignment. A sufficiently large number G of subjects (per group) can thus be classified to obtain long-term estimates of the accuracy of classification. This approach provides a more realistic and practically more useful evaluation of S's assignment than a single run, since it takes into account the variability of C and D. This plan was implemented as described below and illustrated in figure 1.

Let C and D denote the control and disease (PTSD) samples, respectively, where $N_C = 250$ and $N_D = 74$ subjects ($N_{\text{TOTAL}} = N_C + N_D = 324$ subjects). The analysis proceeded in the following steps.

Step 1: subject to be classified. This 'external' subject S to be classified (figure 1, top left) is left out of the calculations in steps 2–5 below.

Step 2: first-stage bootstrap samples for classification coefficients. From the remaining subjects, compose two bootstrap samples C^* and D^* of equal size $B = N_{C^*} = N_{D^*}$, by resampling with replacement from C and D, respectively. These samples will be used to generate classification coefficients for an intermediate classification of S. This step is shown in figure 1 as the parallel arrows from the green box loop leading into the blue box loop, wherein the two bootstrap samples are composed.

Step 3: second-stage bootstrap samples for predictor selection based on a distance measure. Compose two new bootstrap samples, C^{**} and D^{**} of equal size $B' = N_{C^{**}} = N_{D^{**}}$, by resampling with replacement from C^* and D^*. Now, for each Z predictor, calculate a distance measure Q_Z between C^{**} and D^{**} based on some attribute ξ of z_{ij}^0 (e.g. mean, median, variance, etc; see the appendix below) and keep a running sum of Q_Z, $\sum Q_Z$, over a number of M repetitions of this step. This is illustrated in figure 1 as the procedures within the

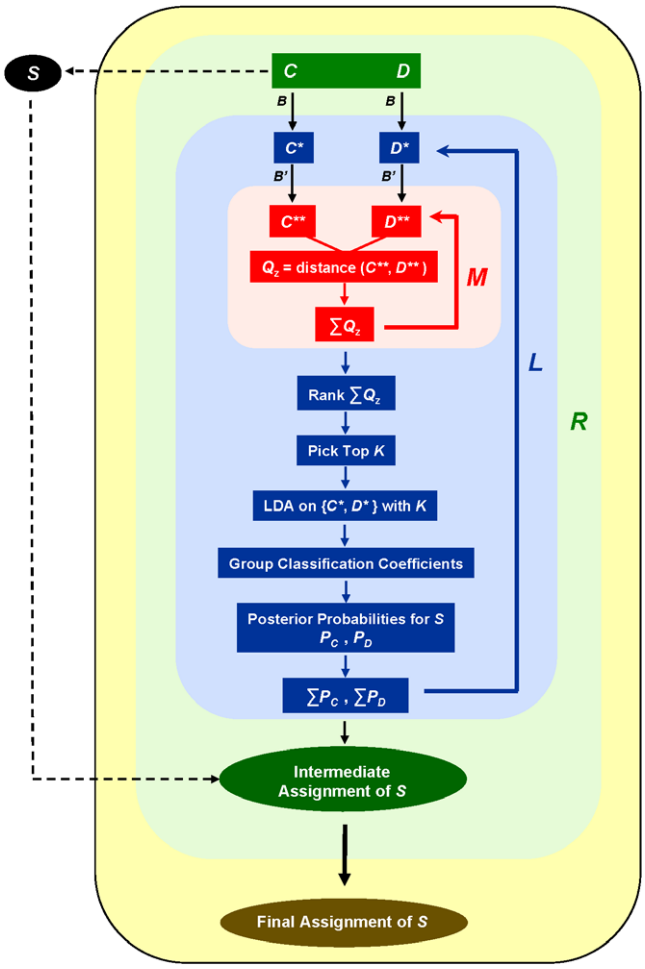

Figure 1. The classification procedure flowchart. Three distinct bootstrap loops that are integrated into the whole classification procedure are highlighted as (i) red box; (ii) blue box and (iii) green box loops. The classification of an external subject (S) is the first step (step 1 in the text) and begins at the top, within the green box loop. Thereafter, two equal-size bootstrap samples ($B = N_{C^*} = N_{D^*}$) are composed from the remaining pool of subjects ($N_C + N_D - 1$) and this (step 2 in the text) is indicated in the figure as the parallel arrows stemming from the green box loop into the blue box loop toward the two bootstrapped samples. From C^* and D^*, another two new sets of equal-size bootstrap samples are composed ($B' = B = N_{C^{**}} = N_{D^{**}}$) as the classification procedure enters the red box loop. Within the red box loop, a measure of distance Q between the two groups for a given predictor Z of the two groups is determined and summed cumulatively over M repetitions (step 3 in the text). This cumulative sum of the distance measure ($\sum Q_Z$) is ranked for all 30 628 predictors, and the top-ranking K predictors are used to derive group classification coefficients using linear discriminant analysis (LDA) on C^* and D^*. Next, the posterior probabilities for the assignment of S into either the control or diseased group are calculated and cumulatively summed over L repetitions. The blue box loop represents the series of calculations (steps 4, 5, 6 in the text) for each of the L repetitions, each of which also includes the red box loop of M repetitions. From these cumulative posterior probabilities ($\sum p_C$ and $\sum p_D$), the external subject S is classified based on the larger cumulative probability of the two. This intermediate classification process (steps 2 through 6) is repeated R times. This is represented by the green box loop, which encompasses both the blue box and red box loops, as well. Finally, the outer yellow box loop describes the full classification procedure (refer to section 2.6).

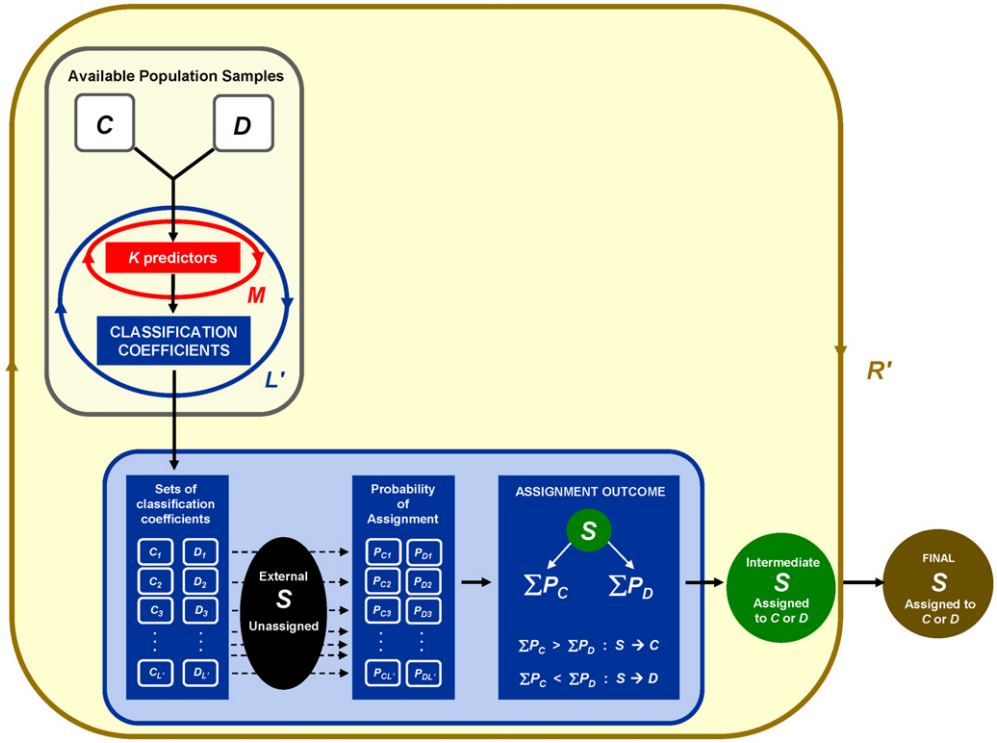

Figure 2. An overview of the procedure to classify a new subject. Existing population samples of control (C) and disease groups (D) contribute to the derivation of top-ranking K predictors (over M repetitions in the predictor reduction loop; figure 1), with which classification coefficients are calculated (over L' repetitions in the assignment of classification functions loop; figure 1). A set of classification coefficients for the assignment of S to either group is derived for each L' repetition, and the corresponding probabilities of assignments (p_C;p_D) are calculated. An intermediate assignment outcome is based on the cumulative probability of assignment of S to either group ($\sum p_C$;$\sum p_D$) at the end of L' repetitions, the larger of which determines the assignment of S. The final assignment is based on the mean of the R' repeated bootstrapped classification outcomes for the external validation of S (refer to section 3.3).

red box loop, which culminates in the cumulative sum of the distance measure between the two samples' predictors $\left(z_{ij}^0\right)$.

Step 4: predictor selection. Rank-order $\sum Q_Z$ and retain those K number of Z predictors with the top-ranked $\sum Q_Z$. These are the predictors with the highest potential for separating the two groups. (Note that this procedure does not rely on, and is independent of, any classification analysis.)

Step 5: classification coefficients from first-stage bootstrap samples. Use these K predictors above to derive classification functions for the two groups (i.e. control and PTSD) using the C^* and D^* samples as input subjects in a linear discriminant analysis (LDA). (For that purpose, the DDSCRM routine of the IMSL statistical library was used, with each observation weighted by the inverse of variance of that observation with respect to the K predictors.)

Step 6: intermediate classification of S.

6.1. Use these functions to compute pairs of posterior probabilities p_C^S and p_D^S that S belongs to the control or disease group, respectively.

6.2. Keep running sums of $\sum p_C^S$ and $\sum p_D^S$ over a number of L repetitions of steps 2 through 5. (These calculations are shown in figure 2, as the analysis moves from the red box loop to the blue box loop where the posterior probabilities are cumulatively summed over the L repetitions.)

6.3. Intermediate classification: classify subject S (at the end of L repetitions) based on the greater sum: if $\sum p_C^S > \sum p_D^S$, subject S is assigned to the control C group, whereas if $\sum p_C^S < \sum p_D^S$, subject S is assigned to the disease D (PTSD) group. (If $\sum p_C^S = \sum p_D^S$, repeat the analysis; this never arose in this study.) This step is depicted in figure 1 as the analysis moves from the blue box loop into the green box loop. (Very similar results are obtained if the binary outcomes of the intermediate classification are used instead of the posterior probabilities; we prefer the latter due to their finer grain.)

6.4. Repeat steps 1 through 6.3 R times to generate a large sample of intermediate classifications. This is illustrated as the procedures within the green box loop, which encompasses the blue and red box loops in figure 1.

Step 7: systematic evaluation of classification parameters. For any number of R intermediate classifications, compute the following measures:

$$\text{Sensitivity (\%)} = \frac{\text{Number of patients correctly classified}}{\text{Total number of patients}} \times 100 \qquad (2)$$

Table 1. Intermediate classification results for different measures ξ and combinations of the parameters L, M, using $B = B' = 100$ and $R = 1$. See the text for explanation of abbreviations.

ξ	K	L	M	Classification accuracy
1	40	7	3	89.3
2	40	7	3	88.5
3	40	7	3	90.2
4	40	7	3	90.8
4	40	7	7	92.8
4	40	11	7	92.0

Specificity (%)

$$= \frac{\text{Number of control subjects correctly classified}}{\text{Total number of control subjects}} \times 100 \tag{3}$$

$$\text{Overall accuracy (\%)} = \frac{\text{Sensitivity} + \text{Specificity}}{2}. \tag{4}$$

These measures allow for a quick way (for small L and R) to assess key parameters employed, namely, the sizes of bootstrap samples (B, B'), the number of predictors (K), the number of bootstrap iterations (M, L, R) and the ξ attribute of z^0_{ij}. This is shown as the external yellow box loop in figure 1.

2.6. Final classification of an external subject, including confidence intervals

The preceding section outlined the bootstrap-based classification procedure and provided the means by which to assess the performance of the classifier for different parameter values based on intermediate classifications of S. In this section, we present the procedure by which an external subject is ultimately classified (figure 2) and by which confidence intervals are generated on the classification outcome. Let r^S_C and r^S_D be the running counts of intermediate S assignments (based on L′) to group C or D, respectively, over a suitably large number of R' repetitions of steps 2–6. Then, the resulting proportions of classification to each one of the two groups, $h^S_C = \frac{r^S_C}{R'}$ and $h^S_D = \frac{r^S_D}{R'}$, can be regarded as long-term, expected probabilities for classifying S to the control or the PTSD group, respectively, at a 0.5 probability threshold (for a binary outcome); corresponding confidence intervals are obtained by bootstrapping h^S_C and h^S_D over a W number of R'. Then, h^S_C and h^S_D (and their confidence intervals) denote the strength (and associated uncertainty) of single subject classification, based on the existing C and D samples.

3. Results

3.1. General classification results

The results obtained using $K = 40$ top-ranked predictors and different distance measures and combinations of parameters are given in table 1. It can be seen that the overall results were excellent with high rates of accuracy of classification. The results were very similar for various ξ attributes of z^0_{ij}. We chose $K = 40$ as an upper bound for detailed testing for three reasons. First, for simplicity, we wanted to keep K small. Second, the summed distance measure $\sum Q_Z$ used to select the top K predictors was steeply exponentially distributed (data not shown), which means that only the initial part of the distribution carried the most useful information for separating C^{**} and D^{**}. And third, the accuracy of prediction increased with K, tending to a plateau with $K > 20$ (refer to table 2).

Table 2. Intermediate classification results for different numbers of top-ranked predictors K using $B = B' = 100$, $G = 100$, $L = 7$, $M = 3$, $\xi = 4$.

K	Classification accuracy
5	87.3
10	90.0
15	91.2
20	91.6
25	92.2
30	93.0
35	90.3
40	90.8

Table 3. Final classification results and associated statistics for the following parameters: $B = B' = 100$, $\xi = 4$, $K = 40$, $R' = 101$, $L' = 1000$, $M = 3$. χ^2 is the chi-square statistic, φ is the phi coefficient and ω is the odds ratio.

		Actual		
		Control	PTSD	Total
Predicted	Control	219	2	221
	PTSD	31	72	103
	Total	250	74	324

Two-way table statistics

	%			*P*-value
Sensitivity	97.3	χ^2	189.8	<0.001
Specificity	87.6	φ	0.765	
Accuracy	92.4	ω	254.3	

3.2. Single subject classification

Each one of the 324 subjects in this study was classified based on 12 000 pairs of $\{p^S_C, p^S_D\}$ per subject, generated using $K = 40$, $B = B' = 100$ and $M = 3$. Intermediate classifications of S were obtained from these data by summing a suitably large number of L' pairs $\{p^S_C, p^S_D\}$, randomly drawn with replacement (out of the 12 000 available), and counting outcomes over $R' = 101$ repetitions to yield h^S_C and h^S_D (and confidence intervals of h^S_C and h^S_D over a W number of R', with $W = 100$). Table 3 shows the results obtained using $B = B' = 100$, $\xi = 4$, $K = 40$, $R' = 101$, $L' = 1000$ and $M = 3$. It can be seen that excellent classification was achieved, with a high (>90%) rate of classification accuracy. With respect to the two-way table itself, there was a highly statistically significant association between actual diagnosis and predicted classification. In addition, the high value of the φ coefficient (0.765) validates the high accuracy of the prediction of diagnosis. The distributions of h^S_C and h^S_D are shown in figure 3.

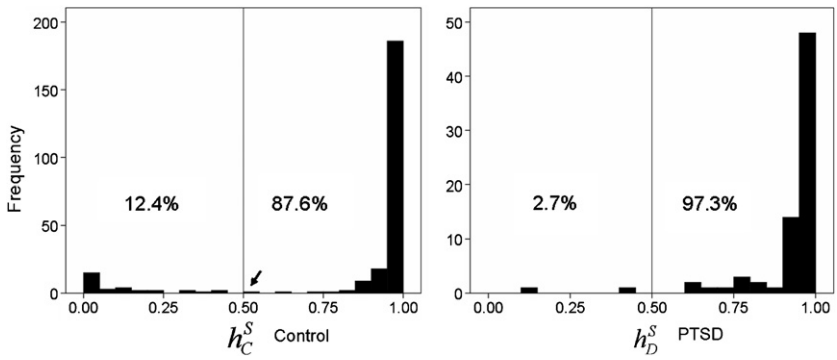

Figure 3. Distributions of classification probabilities h_C^S and h_D^S ($R' = 101$, $L' = 1000$, $K = 40$, $M = 3$) for the control (left panel) and PTSD groups (right panel).

Table 4. Descriptive statistics for a subject with the h^S value near 0.5. Results were obtained by running $W = 100$ times the classification procedure mentioned in table 3 ($B = B' = 100$, $\xi = 4$, $K = 40$, $R' = 101$, $L' = 1000$, $M = 3$). C/W denotes the ratio of *correct*/*wrong* predicted diagnosis; no cases with $h^S = 0.5$ were encountered.

			h^S from $W = 100$ repetitions					
Actual diagnosis	Predicted diagnosis	h_C^S	Median	Mean	SEM	Lower 95% CI	Upper 95% CI	C/W
C	C	0.545	0.574	0.576	0.004 39	0.568	0.585	95/5

Finally, of 18 non-medicated PTSD patients, only one was misclassified and $17/18 = 94.4\%$ were correctly classified. Similarly, of the 56 medicated patients, only one was misclassified and $55/56 = 98.2\%$ were correctly classified. These two classification proportions between the non-medicated and medicated PTSD patients did not differ statistically significantly ($P = 0.98$; the test of two proportions, see [11]).

3.3. Strength of single subject classification

The estimates of the probabilities of classification h_C^S or h_D^S can be regarded as estimates of the strength of classification, irrespective of whether the classification is correct or not, where a value of $h^S = 1$ (h_C^S or h_D^S) indicates highest certainty of prediction. Indeed, such an unequivocal result was obtained in $189/324$ (58.3%) subjects; the median h^S was 1.0 and the interquartile range 0.08. For the control group, h^S was 1.0 in $155/250 = 62\%$ subjects, with the median $=$ 1.0 and the interquartile range $= 0.06$; for the PTSD group, h^S was 1.0 in $34/74 = 45.9\%$ subjects, with the median $=$ 0.98 and the interquartile range $= 0.12$. These results attest to the ability of our classification procedure to yield clear outcomes. In addition, such outcomes remained robust over many repetitions of the procedure (e.g. $W = 100$). Now, in those cases where $h^S < 1$, confidence intervals on h^S can be generated based on W repetitions to further assess the robustness of the classification outcome. When h^S is high enough (e.g. >0.8), there is no issue, whereas for cases with h^S near 0.5 further analysis is warranted. In the present study, there was one such case (indicated by an arrow in figure 3, control group). For that subject, the values $h_C^S = 0.545$ and $h_D^S = 0.455$ were obtained; hence, the subject was assigned to the control group. To confirm this assignment, we then ran the procedure $W = 100$ times. The results are shown in table 4. It can be seen that the assignment of this subject to the control group was justified, based on more extensive testing. In fact, the results of this more extensive testing gave even stronger evidence (0.576) for the correct classification than the original procedure (0.545). The point is that, in general, this procedure can lead to specific, quantitative criteria for classification (e.g. that a lower 95% CI value be on the same side as the mean) or for declaring a subject unclassifiable, based on inconclusive results.

3.4. Relation to PTSD symptom severity

As mentioned above, a quantitative index of PTSD symptom severity was available for 50 patients. The magnitude of this index was compared between two subgroups of patients, namely those for whom $h_D^S = 1$ (most certain classification; $N = 21$) and those for whom $h_D^S < 1$ ($N = 29$). We found that the mean PTSD severity index was significantly higher in the former than the latter group ($P = 0.037$, independent samples two-tailed t-test; mean difference $= 3.46$). This result indicates a significant positive association between the certainty of prediction and the severity of symptoms.

4. Discussion

The results of this study extend the power of the SNI test as a functional neuromarker for PTSD, a disorder for which no useful biomarker currently exists [1]. Remarkably, a high correct classification rate was obtained for PTSD patients who were not medicated for their disease. This finding in particular, and the high overall accuracy of classification in general, indicate a distinct difference in synchronous neural interactions between PTSD and control subjects. This

distinction is robust, for it was observed consistently across many random bootstrap samples from the whole population. The excellent results obtained offer major promise for the usefulness of the SNI test for differential diagnosis as well as for monitoring disease progression and for evaluating the effects of psychological and/or drug treatment. This prospect, and a specific link of SNI to PTSD, is strengthened by the significant association found between the certainty of PTSD patient classification based on the SNI and their symptom severity.

The bootstrap-based approach used in this study is in keeping with the approach advocated recently for biomarker research in general [12]. Essentially, our approach simulates external cross-validation where a subject is classified without feedback of classification information to improve the procedure. Instead, the subject is first classified many times against random samples from the existing populations ('intermediate' classification), and then the bias in these classification outcomes toward one or the other group determines the 'final' assignment. It is worth noting that the only use of the linear discriminant analysis in this approach is to derive classification coefficients from bootstrap samples (to apply to the external subject) without using, or paying attention to, the performance of the linear discriminant analysis itself with respect to the bootstrap samples. Actually, a crucial and innovative aspect of our approach is the procedure for the selection of a very small fraction of predictors (40 out of 30 628 available, i.e. 0.13%) to be used in the discrimination. This procedure is based on accumulating information concerning the distance between the two groups' distributions from a number of bootstrap samples. The conceptual hallmark of the whole approach lies in the extensive resampling of the existing populations in an open-loop way, thus utilizing most of the information available. This approach renders further credence and robustness to the present results and lends itself as a general-purpose procedure for assessing the usefulness of potential biomarkers in other applications.

Acknowledgments

We thank Nancy Tabaka and Cynthia Marceau for their help in patient consent and various aspects of database entry. This work was partially supported by a grant from the United States Department of Veterans Affairs VISN23 and the American Legion Brain Sciences Chair. A P Georgopoulos has a financial interest in the technology described in this publication.

Appendix A

Attributes ξ of z_{ij}^0 in C^{**} and D^{**} samples used in the analysis.
(1) Absolute difference between means $m_{C^{**}}$ and $m_{D^{**}}$:

$$\xi = |m_{C^{**}} - m_{D^{**}}|. \tag{A.1}$$

(2) Absolute difference between medians $m'_{C^{**}}$ and $m'_{D^{**}}$:

$$\xi = |m'_{C^{**}} - m'_{D^{**}}|. \tag{A.2}$$

(3) Standard normal score of the Wilcoxon rank-sum statistic W (based upon a variance that has been adjusted for ties) [13]:

$$\xi = W_{\text{normal}}. \tag{A.3}$$

(4) Signal-to-noise ratio (SNR):

$$\xi = \text{SNR} = \frac{|m_{C^{**}} - m_{D^{**}}|}{\sqrt{\frac{v_{C^{**}}}{N_{C^{**}}} + \frac{v_{D^{**}}}{N_{D^{**}}}}}, \tag{A.4}$$

where $v_{C^{**}}$ and $v_{D^{**}}$ are the variances of z_{ij}^0 in C^{**} and D^{**}.

References

[1] Institute of Medicine (US) 2006 Subcommittee on posttraumatic stress disorder of the committee on gulf war and health: physiologic, psychologic, and psychosocial effects of deployment-related stress *Posttraumatic Stress Disorder: Diagnosis and Assessment* (Washington, DC: National Academies Press)
[2] Georgopoulos A P *et al* 2007 Synchronous neural interactions assessed by magnetoencephalography: a functional biomarker for brain disorders *J. Neural Eng.* **4** 349–55
[3] Efron B and Tibshirani R 1993 *An Introduction to the Bootstrap* (New York: Chapman & Hall)
[4] Blake D D *et al* 1995 The development of a clinician-administered PTSD scale *J. Trauma Stress* **8** 75–90
[5] First M B, Spitzer R L, Williams J B W and Gibbon M 1997 *Structured Clinical Interview for DSM-IV Axis I Disorders SCID-I: Clinician Version, Administration Booklet* (Washington, DC: American Psychiatric Press)
[6] Dikel T N, Engdahl B E and Eberly R 2005 PTSD in former prisoners of war: prewar, wartime, and postwar factors *J. Trauma Stress* **18** 69–77
[7] Strobach P, Abraham-Fuchs K and Harer W 1994 Event-synchronous cancellation of the heart interference in biomedical signals *IEEE Trans. Biomed. Eng.* **41** 343–50
[8] Leuthold A C 2003 Subtraction of heart artifact from MEG data: the matched filter revisited *33rd Ann. Meeting of the Society for Neuroscience (New Orleans, LA, 8–12 November 2003)* 863.15
[9] Box G E P and Jenkins G M 1976 *Time Series Analysis: Forecasting and Control* (San Francisco, CA: Holden-Day)
[10] Leuthold A C, Langheim F J, Lewis S M and Georgopoulos A P 2005 Time series analysis of magnetoencephalographic data during copying *Exp. Brain Res.* **164** 411–22
[11] Snedecor G W and Cochran W G 1989 *Statistical Methods* (Ames: Iowa State University Press)
[12] Pepe M S, Feng Z, Janes H, Bossuyt P M and Potter J D 2008 Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design *J. Natl Cancer Inst.* **100** 1432–8
[13] Conover W J 1980 *Practical Nonparametric Statistics* (New York: Wiley)