

# A Kronecker Product Structured EEG Covariance Estimator for a Language Model Assisted-BCI

Paula Gonzalez-Navarro<sup>1</sup>(✉), Mohammad Moghadamfalahi<sup>1</sup>,  
Murat Akcakaya<sup>2</sup>, and Deniz Erdogmus<sup>1</sup>

<sup>1</sup> Northeastern University, Boston, USA  
gonzaleznavarro@ece.neu.edu

<sup>2</sup> University of Pittsburgh, Pittsburgh, USA

**Abstract.** Electroencephalography (EEG) recorded from multiple channels is typically used in many non-invasive brain computer interfaces (BCIs) for inference. Usually, EEG is assumed to be a Gaussian process with unknown mean and covariance, and the estimation of these parameters are required for BCI inference. However, relatively high dimensionality of the feature vectors extracted from the recorded EEG with respect to the number of supervised observations usually leads to a rank deficient covariance matrix estimator. In our typing BCI, RSVP Keyboard<sup>TM</sup>, we solve this problem by applying regularization on the maximum likelihood covariance matrix estimators. Alternatively, in this manuscript we propose a Kronecker product structure for covariance matrices. Our underlying hypothesis is that the a structure imposed on the covariance matrices will improve the estimation accuracy and accordingly will result in typing performance improvements. Through an offline analysis we assess the classification accuracy of the proposed model. The results represent a significant improvement in classification accuracy compared to an RDA approach which does not assume any structure on the covariance.

**Keywords:** Structured covariances kronecker · Brain-Computer Interface (BCI) · Spatial temporal discriminant analysis · Event-Related Potential (ERP) · Multichannel Electroencephalogram (EEG)

## 1 Introduction

Non-invasive electroencephalography (EEG) based brain computer interfaces (BCIs) are designed as assistive technologies for people with severe speech and muscle impairments providing means for them to communicate with their caretakers and families [2]. Event relate potentials (ERPs) are commonly employed by the EEG-based BCIs to detect the user intend [1–3, 5, 7]. Donchin and Farewell demonstrated that ERPs can be used to design a letter by letter typing BCI [3]. The matrix based presentation paradigm used in their design is shown to be highly gaze dependent [10]. On the other hand, rapid serial visual presentation

(RSVP) paradigm is a gaze-independent alternative for matrix presentation paradigms. In RSVP, the symbols are rapidly presented as a time series on a prefixed location on the screen in a pseudo-random order [1, 5, 7].

RSVP Keyboard<sup>TM</sup> is a non-invasive EEG-based language-model-assisted BCI for typing which utilizes ERPs for intent detection. Inference module of the RSVP Keyboard<sup>TM</sup> probabilistically fuses the evidence extracted from the recorded multiple EEG channels with the probabilistic context information provided by a 6-gram language model [5–7]. This BCI system currently can employ both matrix-based presentation and RSVP paradigms. The EEG evidence is extracted using regularized discriminant analysis (RDA [6, 7]). RDA is a generalization of the quadratic discriminant analysis (QDA) which applies regularization and shrinkage on the maximum likelihood class covariance matrix estimators to remedy rank deficiencies [4]. RSVP Keyboard<sup>TM</sup> utilizes RDA because the dimensionality of the extracted EEG feature vectors is relatively higher than the number of measurements collected for supervised learning.

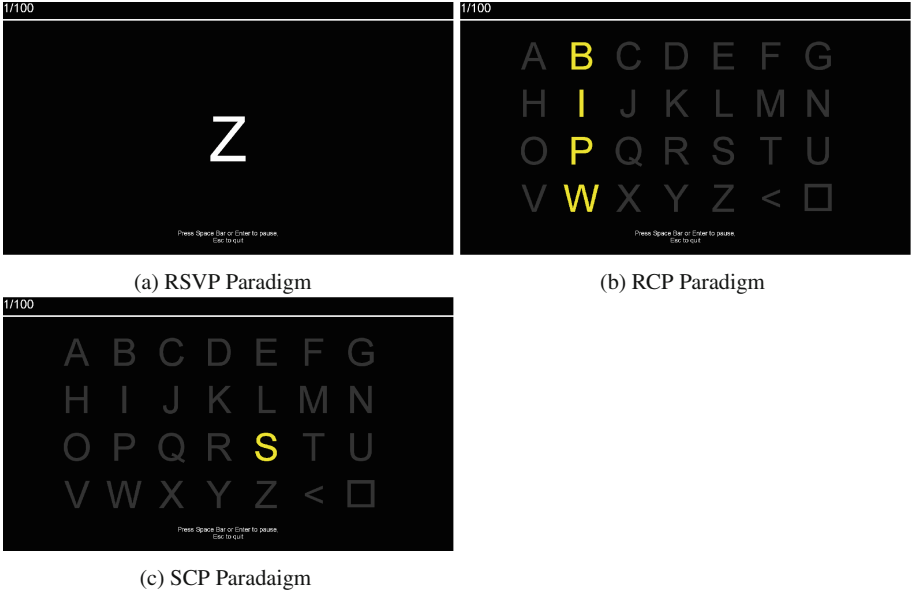
Alternative to the RDA method, in this manuscript, we propose a Kronecker product structure for the covariance matrices. We show that modeling multichannel EEG using an auto-regressive moving average (ARMA) model under certain assumptions leads to a covariance matrix with a Kronecker product structure. In this structure the number of parameters is significantly lower than RDA. The maximum likelihood estimation of the proposed parametric model of covariance matrix along with regularization lead to significant improvement in classification performance. Our offline analysis shows that the median of the percentage of improvement for different subjects across different presentation paradigms is 1.111%.

## 2 Inference in RSVP Keyboard<sup>TM</sup>

RSVP Keyboard<sup>TM</sup> utilizes a visual presentation module to detect the user intent. The EEG collected during the visual stimulation is then employed in decision making procedure.

### 2.1 Visual Presentation

In letter by letter typing task we assume a dictionary set  $\mathcal{D}$  of 26 letters in English alphabet, a space symbol “\_” and a backspace symbol “<” as the set of all possible choices. Our system utilizes both matrix-based and rapid serial visual presentation paradigms. The different presentation paradigms are shown in Fig. 1a, b and c. Generally for all matrix-based presentation paradigms the dictionary members are arranged on a matrix shaped layout on the screen in gray color. In row and column presentation (RCP) paradigm the elements on each row or column of the matrix are assumed as a “trial” which are then flashed rapidly and in a pseudo-random order. One sequence for this presentation paradigm contains the presentation of all the rows and columns.



**Fig. 1.** Rapid serial visual and matrix-based presentation paradigms

After each sequence the system will attempt to make a decision however if a predefined confidence threshold is not satisfied the system will capture more EEG in response to more sequences to improve the decision confidence. Accordingly the set of sequences which lead to a decision is called an “epoch”. Similarly, a trial in single character presentation (SCP) paradigm consists of only one element of dictionary and a sequence is defined as flashes of a subset of dictionary. For RSVP paradigm there exist no background matrix of characters but all the characters are presented on a prefixed location of the screen in a pseudo-random order and rapidly in time. In this presentation scheme each flashing letter is a trial and in each sequence a subset of dictionary is presented. The definition of epoch is the same among all presentation paradigms.

## 2.2 Decision Making

The decision making process in RSVP Keyboard<sup>TM</sup> utilizes a maximum a posteriori (MAP) inference mechanism. During this procedure the context information from a language model (LM) is probabilistically fused with EEG evidence to produce a more accurate decision. The inference mechanism at epoch  $k$  and after observing sequence  $l$  is defined as follows:

$$\hat{s}_k^* = \arg \max_{s \in \mathcal{D}} P(s_k^* = s | \mathcal{E}^l; C) \quad (1)$$

where  $s_k^*$  is a random variable which represents the user intent in epoch  $k$ ,  $\hat{s}_k^*$  is the estimated user intent,  $\mathcal{E}^l$  is the EEG evidences for all the observed  $l$

sequences in epoch  $k$ . Assuming that conditioned on the unknown symbol, the EEG evidence and context information are independent from each other, and again conditioned on the unknown symbol all EEG evidence from different trials are independent, we can simplify Eq. (1) as:

$$\widehat{s}_k^* = \arg \max_{s \in \mathcal{D}} P(s_k^* = s|C) \prod_{\substack{i=1, \dots, l \\ \{j \mid s \in \mathbf{s}_j^i\}}} \frac{p(e_j^i|1)}{p(e_j^i|0)} \quad (2)$$

Here in (2),  $\mathbf{s}_j^i$  is the  $j^{\text{th}}$  trial of the sequence  $i^{\text{th}}$  of epoch  $k$  and  $e_j^i$  represents the EEG evidence associated with  $\mathbf{s}_j^i$ .

As in (2), one needs to define  $P(s_k^* = s|C)$  and class conditional distributions  $p(e|1)$ ,  $p(e|0)$  to be able to perform an inference.

**Context Information.** To define  $P(s_k^* = s|C)$  we utilize a letter  $n$ -gram LM which provides a prior probability mass function (PMF) over the dictionary. We have shown that context information fused with EEG evidence improves system performance effectively [5,6]. An  $n$ -gram LM, mimics a Markov model of order  $n - 1$ , through which it estimates the conditional PMF over the dictionary set based on  $n - 1$  previously typed letters. Let  $C = \{s_m^*\}_{m=n-1, \dots, 1}$ , where  $s_m^*$  is the  $m^{\text{th}}$  previously typed character, then

$$P(s|C) = P(s|\{s_m^*\}_{m=n-1, \dots, 1}) = \frac{P(s, s_{n-1}^*, \dots, s_1^*)}{P(s_{n-1}^*, \dots, s_1^*)} \quad (3)$$

In our system, we use a 6-gram letter model, which is trained on the NY Times portion of the English Gigaword corpus [8].

**Preprocessing and Feature Extraction.** The class conditional distributions  $p(e|1)$ ,  $p(e|0)$  in RSVP Keyboard<sup>TM</sup> are estimated over the EEG evidences. To extract the EEG evidence from the EEG time signals, we begin with applying a two step dimensionality reduction following a preprocessing of recorded EEG. We use g.USBamp bio-signal amplifier with the sampling frequency of 256 Hz to acquire the data. A bandpass linear-phase finite impulse response (FIR) filter with bandpass of [1.5, 42] Hz is then applied on the EEG data in order to improve the signal to noise ratio (SNR) and eliminate DC drifts. We down-sample the preprocessed data by order of 2. We concatenate the data from every channel in a time window of [0, 500] ms, time locked to onset of  $i^{\text{th}}$  trial, to form the feature vector  $\mathbf{x}_i$  for that trial.

The supervised data required for estimating the class conditional distributions is recorded during “calibration” mode of the system [5]. Each calibration task of RSVP Keyboard<sup>TM</sup> consists of 100 sequences. Before each sequence the user is presented with a target character which she/he is supposed to locate during that sequence. For RSVP and SCP paradigms the number of trials in each sequence is set to 10, and for RCP it is equal to number of all rows and columns

in the matrix (for instance, here we are using a  $4 \times 7$  matrix which leads to 11 trials in a sequence).

The labels for the feature vectors is assigned as 0 if the trial in a non-target, or 1 if the trial contains the target character. To increase the EEG ERP detection accuracy, we further quadratically project the feature vectors on to a space which maximizes the distance between two classes. In RSVPCKeyboard<sup>TM</sup> we use regularized discriminant analysis (RDA) which is a generalization of quadratic discriminant analysis (QDA) to perform this projection. In our system the dimensionality of feature vectors is relatively higher than the number of observation during a calibration session, hence we mainly utilize RDA to be able to estimate invertible covariance matrices. The maximum likelihood class conditional mean and covariance matrices are computed as follows:

$$\begin{aligned} \boldsymbol{\mu}_h &= \frac{1}{N_h} \sum_{i=1}^N \mathbf{x}_i \delta(y_i, h) \\ \boldsymbol{\Sigma}_h &= \frac{1}{N_h} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_h)(\mathbf{x}_i - \boldsymbol{\mu}_h)^T \delta(y_i, h) \end{aligned} \tag{4}$$

where  $y_i \in \{0, 1\}$  is the label of  $\mathbf{x}_i$ ,  $h \in \{0, 1\}$  is the class for which we are performing the estimation,  $N_h$  is the number of observations in class  $h$  and  $N = N_0 + N_1$ . RDA makes the estimated covariance matrices invertible by applying regularization and shrinkage.

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_h(\lambda) &= \frac{(1 - \lambda)N_h \boldsymbol{\Sigma}_h + (\lambda) \sum_{h=0}^1 N_h \boldsymbol{\Sigma}_h}{(1 - \lambda)N_h + (\lambda) \sum_{k=0}^1 N_k} \\ \widehat{\boldsymbol{\Sigma}}_h(\lambda, \gamma) &= (1 - \gamma) \widehat{\boldsymbol{\Sigma}}_h(\lambda) + (\gamma) \frac{1}{p} \text{tr}[\widehat{\boldsymbol{\Sigma}}_h(\lambda)] \mathbf{I}_p \end{aligned} \tag{5}$$

Here,  $\lambda, \gamma \in [0, 1]$  are the shrinkage and regularization parameters,  $\text{tr}[\cdot]$  is the trace operator and  $\mathbf{I}_p$  is an identity matrix of size  $p \times p$ . RSVPCKeyboard<sup>TM</sup> optimizes the  $\lambda$  and  $\gamma$  for the maximum area under the receiver operating characteristics (ROC) curve (AUC) in a 10-fold cross validation framework. The RDA score for  $e_i$ , is then referred to as EEG evidence for trial  $\mathbf{s}_i$ .

$$e_i = \log \left( \frac{f_{\mathcal{N}}(\mathbf{x}_i; \boldsymbol{\mu}_1, \widehat{\boldsymbol{\Sigma}}_1(\lambda, \gamma))}{f_{\mathcal{N}}(\mathbf{x}_i; \boldsymbol{\mu}_0, \widehat{\boldsymbol{\Sigma}}_0(\lambda, \gamma))} \right) \tag{6}$$

where  $f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the Gaussian probability density function with mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ .

Consequently we use these EEG evidences in kernel density estimation (KDE) framework to define class conditional distributions. In our system we use Silverman rule of thumb to define the kernel width for KDE [9].

### 3 Signal Modeling and Covariance Estimation

Currently in our system we employ RDA to estimate full-ranked class conditional covariance estimates. But for a non-structured maximum likelihood estimation

of covariance matrix one needs to estimate many parameters (i.e. elements of covariance matrix). But due to lack of enough observation in a calibration session, this estimation might be prone to errors. We propose to use a Kronecker product structure for the covariance matrices. This structure reduces the number of the covariance parameters to be estimated using the assumption of stationarity in time and space. We show that defining an auto-regressive moving average (ARMA) (p,q) model for the multi-channel EEG recordings leads to Kronecker product structure under certain assumptions.

Define  $\mathbf{v}[n]$  as the spatial feature vector of EEG recorded from  $N_{ch}$  EEG channels at time instant  $n$ :

$$\mathbf{v}[n] = \begin{bmatrix} v_1[n] \\ v_j[n] \\ \vdots \\ v_{N_{ch}}[n] \end{bmatrix}_{N_{ch} \times 1} \quad (7)$$

where  $v_i[n]$  is the  $n^{\text{th}}$  time sample recorded at channel  $i$ . Then define the feature vectors as:

$$\mathbf{x}[i] = \begin{bmatrix} \mathbf{v}[1] \\ \mathbf{v}[2] \\ \vdots \\ \mathbf{v}[N_t] \end{bmatrix}_{(N_{ch}N_t) \times 1} \quad (8)$$

where  $\mathbf{v}[\cdot]$  and accordingly  $\mathbf{x}$  follows a multivariate Gaussian distribution and  $N_t$  is the number of time samples. We define an ARMA model for EEG signal as follows:

$$\mathbf{v}[n] = \sum_{k=1}^p A_k \mathbf{v}[n-k] + \sum_{j=0}^q b_j \mathbf{w}[n-j] \quad (9)$$

In (9),  $A_k$  represents the  $N_{ch} \times N_{ch}$  signal weight matrix at lag  $k$ ,  $b_j$  is an scalar weight for noise at lag  $j$  and  $\mathbf{w}[n]$  represents multivariate wide sense stationary Gaussian noise for the  $n^{\text{th}}$  time sample. Let us assume that the EEG signals among the channels is stationary. Then one can write (9) as:

$$\mathbf{v}[n] = \sum_{k=1}^p c_k \cdot \mathbf{v}[n-k] + \sum_{j=0}^q b_j \mathbf{w}[n-j] \quad (10)$$

in which  $c_k$  is an scalar weight of time for the signal at time lag  $k$ . Now lets further assume  $p = 1$  and  $b_j = 0 \forall j = 0 \dots q$  then we can write:

$$\mathbf{v}_n = c_1 \cdot \mathbf{v}[n-1] \quad (11)$$

Then define:

1. Initial state  $\mathbf{v}[0] \sim \mathcal{N}_{N_{ch}}(\boldsymbol{\mu}_{\mathbf{v}[0]}, \boldsymbol{\Sigma}_{\mathbf{v}}[0, 0])$
2.  $E[\mathbf{v}[n]] = \boldsymbol{\mu}_{\mathbf{v}[n]}$
3.  $\boldsymbol{\Sigma}_{\mathbf{v}}[m, n] = \text{Cov}[\mathbf{v}[m], \mathbf{v}[n]]$   
 $= E[(\mathbf{v}[m] - \boldsymbol{\mu}_{\mathbf{v}[m]})(\mathbf{v}[n] - \boldsymbol{\mu}_{\mathbf{v}[n]})^T]$ .
4.  $\mathbf{x} = \begin{bmatrix} \mathbf{v}[1] \\ \mathbf{v}[2] \\ \vdots \\ \mathbf{v}[N_t] \end{bmatrix}_{(N_{ch}N_t) \times 1}$

Here note that based on the above definition we have  $\boldsymbol{\Sigma}_{\mathbf{v}}[m, n] = \boldsymbol{\Sigma}_{\mathbf{v}}[n, m]$ . Also we have:

$$\mathbf{v}[n] = (c_1)^n \cdot \mathbf{v}[0] \Rightarrow E[\mathbf{v}[n]] = (c_1)^n \cdot \boldsymbol{\mu}_{\mathbf{v}[0]} \quad (12)$$

We further assume that the EEG signal is stationary in time. Lets assume  $m < n$  hence we have:

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{v}}[m, n] &= E\{\mathbf{v}[n]\mathbf{v}[m]^T\} - E\{\mathbf{v}[n]\}E\{\mathbf{v}[m]\}^T \\ &= c_1^{(n-m)} \cdot (E\{\mathbf{v}[m]\mathbf{v}[m]^T\} - \boldsymbol{\mu}_{\mathbf{v}[m]}\boldsymbol{\mu}_{\mathbf{v}[m]}^T) \\ &= c_1^{(n-m)} \cdot \boldsymbol{\Sigma}_{\mathbf{v}}[m, m] \\ &= c[|n - m|] \cdot \boldsymbol{\Sigma}_{\mathbf{v}}[0, 0] \\ &\text{where } c[|n - m|] = c_1^{(|n-m|)} \end{aligned} \quad (13)$$

According to definition of  $\mathbf{x}$  one can define the covariance matrix of  $\mathbf{x}$  as:

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{x}} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{v}}[1, 1] & \dots & \boldsymbol{\Sigma}_{\mathbf{v}}[1, N_t] \\ \boldsymbol{\Sigma}_{\mathbf{v}}[2, 1] & & \boldsymbol{\Sigma}_{\mathbf{v}}[2, N_t] \\ \vdots & & \vdots \\ \boldsymbol{\Sigma}_{\mathbf{v}}[N_t, 1] & \dots & \boldsymbol{\Sigma}_{\mathbf{v}}[N_t, N_t] \end{bmatrix} \\ &= \begin{bmatrix} c[0] & \dots & c[N_t - 1] \\ c[1] & & c[N_t - 2] \\ \vdots & & \vdots \\ c[N_t - 1] & \dots & c[0] \end{bmatrix} \otimes \boldsymbol{\Sigma}_{\mathbf{v}}[0, 0] \end{aligned} \quad (14)$$

Finally, we assume the EEG signal is independent in time which means that  $c[|n - m|] = 0$  for all  $m \neq n$ . Then we have:

$$\boldsymbol{\Sigma}_{\mathbf{x}} = c[0] \cdot \mathbf{I}_{N_t} \otimes \boldsymbol{\Sigma}_{\mathbf{v}}[0, 0] \quad (15)$$

Here  $\mathbf{I}_{N_t}$  is an  $N_t \times N_t$  identity matrix.

Through a maximum likelihood framework we can estimate the parameter values of the structured covariance matrices. We specifically utilize a flipflop algorithm presented by Karl Werner in [11] for which we fix the time covariance matrix to identity and perform a one time estimation on channel covariance matrix.

## 4 Results

### 4.1 Participants

In this manuscript we utilized the calibration data collected from 9 healthy users who had consented to participate in our study according to the IRB-approved protocol (IRB130107) [5]. In our study, each user performed 12 calibration sessions for all possible combinations of 4 inter trial interval (ITI) values ( $\{200; 150; 100; 85\}$  ms) and 3 presentation paradigms (RCP, SCP and RSVP). According to the International 10/20 configuration, data recorded from 16 EEG locations: Fp1, Fp2, F3, F4, Fz, Fc1, Fc2, Cz, P1, P2, C1, C2, Cp3, Cp4, P5 and P6.

### 4.2 Data Analysis and Results

We calculated the area under the receiver operating characteristics (ROC) curve (AUC) values, for every calibration data using a 10-fold cross validation. The goal of this analysis is to assess the changes in classification accuracy under the proposed signal model.

For each particular ITI and presentation paradigm (PP) combination, we compared the median of AUC values for RDA and the proposed model in Table 1, and also we show the number of participants who demonstrate improvement under the proposed model in Table 2. In Table 1 we can see an improvement for RSVP at ITI = 150 ms which is the optimal speed for this presentation paradigm [5]. Also, the proposed model seems to be most effective in RCP paradigm. However, we cannot observe any significant improvement for SCP at any ITI. As shown in Table 2 most of the population could benefit from the proposed model at every PP and ITI combination. Among all the users at every ITI and PP combinations, half of the AUC values fall below .811. We utilized this value to define a threshold for high AUCs and low AUCs. The Table 3 represents the median AUC values for regular RDA and proposed covariance estimation technique. As one can clearly see in this table the participants with low AUCs can benefit more from the new signal modeling scheme.

We also compute the number of participants who demonstrate a classification improvement regardless of particular ITI value and PP. We assumed a participant can benefit from this signal modeling scheme if the median of all 12 ITI and PP combination AUCs improves. The corresponding results are shown in Table 4.



**Table 1.** The median of changes in AUC for each PP and ITI combination among nine users

	RSVP	SCP	RCP
85 ms	2.867	-1.178	-0.383
100 ms	-2.089	0.956	2.156
150 ms	2.000	0.756	1.206
200 ms	-2.189	-1.633	3.022

**Table 2.** The number of participant for whom the proposed model improved the classification AUC, for each PP and ITI combination and among nine users

	RSVP	SCP	RCP
85 ms	5	6	7
100 ms	5	7	7
150 ms	4	4	6
200 ms	5	4	5

**Table 3.** Median of AUCs lower than 0.811 for the nine subjects when we use the signal modeling (SM) versus RDA for all PP and ITIs.

	RSVP		SCP		RCP	
	SM	RDA	SM	RDA	SM	RDA
85 ms	0.680	0.656	0.705	0.706	0.786	0.776
100 ms	0.722	0.698	0.788	0.754	0.781	0.754
150 ms	0.721	0.736	0.756	0.722	0.777	0.777
200 ms	0.725	0.736	0.790	0.780	0.795	0.786

**Table 4.** Improvement in median of AUCs among all 12 ITI and PP combination for each user.

US1	US2	US3	US4	US5	US6	US7	US8	US9
-0.402	1.54	1.269	-0.772	1.111	1.95	-0.181	2.750	-0.108

Table 4 shows that most of the population, 5 out of 9, demonstrate an improvement in classification AUC. Besides the amount of improvements is generally higher than 1% while the performance degradation is less than 0.5% for other users.

## 5 Discussions and Future Work

In this manuscript, we considered the EEG as a structured multivariate Gaussian data, and under certain assumptions, we modeled the covariance matrix of this

signal to have a Kronecker product of a channel covariance matrix and an identity time covariance matrix. With this assumption on the covariance matrix, we reduced the number of parameters that are needed to be estimated. Correspondingly this decrease in the number of parameters to be estimated led to an increase in classification performance.

In this study at every presentation paradigm and inter trial interval combination, we compared the classification performances of two methods when the covariance matrix is estimated under the new structure versus the covariance is estimated without a specific structure using typical RDA. Results suggested that considering a structured EEG signal can significantly improve the ERP-detection specially when the RDA AUC is below 80 %. Future work will analyze and optimize additional structures such as Toeplitz or AR(p) structures for the covariance of the multichannel EEG signal.

**Acknowledgment.** This work is supported by NIH 2R01DC009834, NIDRR H133E140026, NSF CNS-1136027, IIS-1118061, IIS-1149570, CNS-1544895, SMA-0835976. For supplemental materials, please visit <http://hdl.handle.net/2047/D20199232> for the CSL Collection in the Northeastern University Digital Repository System.

## References

1. Acqualagna, L., Treder, M.S., Schreuder, M., Blankertz, B.: A novel brain-computer interface based on the rapid serial visual presentation paradigm. *Proceed. EMBC* **1**, 2686–2689 (2010)
2. Akcakaya, M., Peters, B., Moghadamfalahi, M., Mooney, A., Orhan, U., Oken, B., Erdogmus, D., Fried-Oken, M.: Noninvasive brain computer interfaces for augmentative and alternative communication. *IEEE Rev. Biomed. Eng.* **7**(1), 31–49 (2014)
3. Farwell, L.A., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* **70**(6), 510–523 (1988)
4. Friedman, J.H.: Regularized discriminant analysis. *J. Am. Stat. Assoc.* **84**(405), 165–175 (1989)
5. Moghadamfalahi, M., Orhan, U., Akcakaya, M., Nezamfar, H., Fried-Oken, M., Erdogmus, D.: Language-model assisted brain computer interface for typing: a comparison of matrix and rapid serial visual presentation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **23**(5), 910–920 (2015)
6. Orhan, U., Erdogmus, D., Roark, B., Oken, B., Fried-Oken, M.: Offline analysis of context contribution to ERP-based typing BCI performance. *J. Neural Eng.* **10**(6), 066003 (2013)
7. Orhan, U., Hild, K.E., Erdogmus, D., Roark, B., Oken, B., Fried-Oken, M.: RSVP keyboard: an EEG based typing interface. In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 645–648. IEEE (2012)
8. Roark, B., De Villiers, J., Gibbons, C., Fried-Oken, M.: Scanning methods and language modeling for binary switch typing. In: *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pp. 28–36. Association for Computational Linguistics (2010)

9. Silverman, B.W.: Density Estimation for Statistics and Data Analysis, vol. 26. CRC Press, Boca Raton (1986)
10. Treder, M.S., Blankertz, B.: Research (c)overt attention and visual speller design in an ERP-based brain-computer interface. *Behav. Brain Funct.* **6**, 28 (2010)
11. Werner, K., Jansson, M., Stoica, P.: On estimation of covariance matrices with kronecker product. *Structure* **56**(2), 478–491 (2008)