RSVP KEYBOARD: AN EEG BASED TYPING INTERFACE

Umut Orhan[†], Kenneth E. Hild II[°], Deniz Erdogmus[†], Brian Roark[°], Barry Oken[°], Melanie Fried-Oken[°]

Abstract— **Humans need communication. The desire to communicate remains one of the primary issues for people with locked-in syndrome (LIS). While many assistive and augmentative communication systems that use various physiological signals are available commercially, the need is not satisfactorily met. Brain interfaces, in particular, those that utilize event related potentials (ERP) in electroencephalography (EEG) to detect the intent of a person noninvasively, are emerging as a promising communication interface to meet this need where existing options are insufficient. Existing brain interfaces for typing use many repetitions of the visual stimuli in order to increase accuracy at the cost of speed. However, speed is also crucial and is an integral portion of peer-to-peer communication; a message that is not delivered timely often looses its importance. Consequently, we utilize rapid serial visual presentation (RSVP) in conjunction with language models in order to assist letter selection during the brain-typing process with the final goal of developing a system that achieves high accuracy and speed simultaneously. This paper presents initial results from the RSVP Keyboard system that is under development. These initial results on healthy and locked-in subjects show that single-trial or few-trial accurate letter selection may be possible with the RSVP Keyboard paradigm.**

Index Terms— **Brain computer interface, language model, event related potential, Bayesian fusion**

I. INTRODUCTION

Brain computer interfaces (BCI) are seen as the future of human computer interaction. In particular, BCI-enabled devices that will allow people with severe speech and motor disabilities to communicate and interact with their personal networks and environments have received significant interest by the research community in the last decades [1], [2]. Noninvasive BCI design paradigms, particularly those using electroencephalography (EEG), are increasingly popular due to their portability, noninvasiveness, cost-effectiveness, and reliability [1], [2], [3]. A large portion of the functionally locked-in population may lack precise gaze control to be able to use the commercially available eye-tracking based systems and the matrix layout based P300-speller [1] paradigms. Therefore, by also considering the possible heavier cognitive load of matrix based paradigms, we pursue the rapid serial visual presentation (RSVP) paradigm which does not require precise gaze control to discriminate intent between various symbols that one could choose from [4], [5]. Our approach, referred to as the RSVP Keyboard, allows user to sequentially scan the options until the desired symbol is selected as opposed to the two-tiered selection mechanism of Hexo-Spell by the Berlin BCI group [3]. In the RSVP paradigm, each candidate letter is shown at the same place on the screen in a temporally ordered sequence at a comfortably high presentation rate. In the current prototype system, EEG responses corresponding to each visual stimuli are assessed using regularized discriminant analysis (RDA).

Due to the low signal-to-noise ratio in EEG (where noise is primarily due to irrelevant brain activity), current BCI typing systems require a high number of stimulus repetitions, resulting in low symbol rates. Hierarchical symbol trees may lead to some speed increase [3], [6], [7]. Our simulations indicate that with error-prone binary-intent-detectors (such as ERP-detectors), for the English language, the expected bits-per-symbol is negligibly degraded in the RSVP paradigm (to about 3bits/symbol at no error in intent detection) compared to the Huffman tree (about 2.5bits/symbol with no errors in binary intent detection), which minimizes the expected bits to type a symbol [8]. Probabilistic incorporation of a language model in decision making (as opposed to simple word completion) could help improve EEG classification accuracy, therefore increase speed by reducing the number of stimulus repetitions required to achieve a certain level of accuracy.

In this paper, we present the design and initial results of the RSVP Keyboard system, a novel RSVP and EEG based BCI typing system, which tightly incorporates language models into the decision mechanism. The fusion of language and EEG evidence is achieved using a probabilistic framework, assuming that these two pieces of evidence are conditionally independent given class labels (a symbol is desired or not). An extensive off-line analysis of the proposed fusion approach was done in a previous study [9]. An initial demonstration of RSVP keyboard was done in [10]. Here we focus on the first analysis of real-time closed-loop typing results using the designed prototype, using number of stimulus repetitions per symbol as the primary performance measure.

II. METHODS

The RSVP Keyboard consists of three main components: (1) visual presentation, (2) EEG data acquisition, and (3) the decision mechanism to select a symbol to type. During the typing of each symbol, EEG responses to all symbols are collected and the selection of the symbol to be typed is based on this EEG evidence and the language model assessment of each candidate symbol. This tight incorporation of language evidence in EEG response evaluation sharply separates the RSVP Keyboard from existing prototypes of BCI spellers. While the fusion of language information helps improve speed and accuracy, the use of RSVP eliminates precise gaze control to focus one's attention on a relatively narrow region on the display - a challenging feat for most members of the target population, and a crucial requirement for these BCI systems to work properly [11].

A. RSVP: Rapid Serial Visual Presentation

RSVP is a presentation technique in which visual stimuli are displayed as a temporal sequence at a fixed location on the screen and with arbitrarily large sizes if needed. In contrast, the Matrix-P300-Speller places all symbols on the display in a matrix and highlights a subset (for instance, a column, a row, checkerboard, or an individual letter) [1], [12]. Berlin BCI's recent variation of their

ERP-based Hexo-Spell uses an RSVP like presentation at the central area, while options (sets of symbols) are displayed around it [3]. RSVP is particularly useful for most users with weak or no gaze control and for those whose cognitive skills do not allow processing matrix presentation of letters. An example screen snapshot from the current RSVP Keyboard prototype is given in Fig. 1. The optimal layout of the screen is a point of research in itself and we will not discuss that issue in detail here.

Fig. 1. RSVP keyboard interface

In the current study, RSVP contains random permutations of the 26 letters in English alphabet, a space symbol and a backspace symbol (a total of 28 symbols to choose from). If repetition is needed, all symbols are repeated multiple times to improve classification accuracy until a preset desired confidence level is reached[1]. In the RSVP Keyboard, the user is assumed to show positive intent exactly for one symbol per epoch (section in which user types a symbol of the text). Each epoch contains a block of sequences, currently containing all 28 symbols, of which one is the target symbol. During the calibration phase (classifier training session), a designated target symbol is shown before each epoch, so that the user focuses his intent on this known given symbol in order for the classifier calibration to have labeled positive and negative intent examples of EEG responses.

B. EEG feature extraction via regularized discriminant analysis

As a response to the infrequent target stimulus shown in RSVP sequences, the brain generates event related potentials including the P300 wave - a positive deflection in the scalp voltage mainly in the centro-parietal areas with an average latency just over 300 ms. This natural novelty detection or target matching response allows us to detect intent using EEG signals. This process starts with extracting stimulus-time-locked bandpass filtered EEG signals for each stimulus in the sequence. Since physiologically, the most relevant signal components are expected to occur within the first 500ms following the stimuli, the $[0,500)$ms portion of the EEG following each stimulus is extracted. At this stage it is important to design bandpass filters whose group delay does not shift the physiological response to outside this interval. A linear dimension reduction is applied on the temporal signals using Principal Component Analysis in order to to remove zero variance directions (i.e. zero-power bands based on the estimated covariance). The final feature vector to be classified is obtained as a concatenation of the PCA-projected temporal signals for each channel. Regularized Discriminant Analysis (RDA) [13] is used to further project the

[1]In future designs, we are contemplating adaptive and optimal sequencing of a subset of candidate symbols to save time and increase speed without negatively affecting accuracy - initial results are encouraging.

EEG evidence into scalar-feature for use in fusion with language model evidence.

RDA is a modification of quadratic discriminant analysis (QDA). QDA yields the optimal minimum-expected-risk Bayes classifier under the assumption of multivariate Gaussian class distributions. This classifier depends on the inverses of covariance matrices for each class, which are estimated from training data. In BCI, to keep the calibration phase short, few training samples are acquired - especially for the positive intent class. Therefore, the sample covariance estimates may become singular or ill-conditioned for high-dimensional feature vectors, which is the case here. RDA applies shrinkage and regularization on class covariance estimates. Shrinkage forces class covariances closer towards the overall data covariance as

$$\hat{\mathbf{\Sigma}}_c(\lambda) = (1-\lambda)\hat{\mathbf{\Sigma}}_c + \lambda\hat{\mathbf{\Sigma}}, \tag{6}$$

where λ is the shrinkage parameter; $\hat{\mathbf{\Sigma}}_c$ is the class covariance matrix estimated for class $c \in \{0,1\}$ with $c = 0$ for non-target class and $c = 1$ for target class; $\hat{\mathbf{\Sigma}}$ is the weighted average of class covariance matrices. Regularization is administered as

$$\hat{\mathbf{\Sigma}}_c(\lambda,\gamma) = (1-\gamma)\hat{\mathbf{\Sigma}}_c(\lambda) + \frac{\gamma}{d}\text{tr}[\hat{\mathbf{\Sigma}}_c(\lambda)]\mathbf{I}, \tag{7}$$

where γ is the regularization parameter, $\text{tr}[\cdot]$ is the trace function and d is the dimension of the data vector. We employ the Nelder-Mead simplex-reflection method [14] to optimize the free parameters λ and γ such that a local maximizer of the area-under-the-ROC-curve (AUC) estimated using 10-fold cross-validation is achieved.

After regularization and shrinkage, the covariance and mean estimates for each class are used in generating a scalar feature that minimizes expected risk under the Gaussianity assumption of class distributions. This is the log-likelihood ratio

$$\delta_{\text{RDA}}(\mathbf{x}) = \log \frac{f_{\mathcal{N}}(\mathbf{x};\hat{\boldsymbol{\mu}}_1,\hat{\mathbf{\Sigma}}_1(\lambda,\gamma))\hat{\pi}_1}{f_{\mathcal{N}}(\mathbf{x};\hat{\boldsymbol{\mu}}_0,\hat{\mathbf{\Sigma}}_0(\lambda,\gamma))\hat{\pi}_0}, \tag{8}$$

where $\boldsymbol{\mu}_c$, $\hat{\pi}_c$ are estimates of class means and priors respectively; $\mathbf{x}$ is the data vector to be classified and $f_{\mathcal{N}}(\mathbf{x};\boldsymbol{\mu},\mathbf{\Sigma})$ is the pdf of a multivariate Gaussian (normal) distribution.

C. Language modeling

For many text processing applications language modeling constitutes an important part; likewise, it is significant for BCI typing systems as demonstrated in our earlier analysis of various letter scanning options used in BCI speller designs [8]. Typically, previously written text is used to predict upcoming symbols, which might become very predictable in some contexts and symbol locations. Especially in situations where symbol output rate is low, language modeling can have a significant impact on speed and accuracy.

In letter-by-letter typing, we adopt n-gram language models at the symbol level. These models estimate the conditional probability of a letter given $n-1$ previously typed letters. Let W be a sequence of letters where W_i is the ith letter and $\mathcal{S}$ be the set of candidate symbols. For an n-gram model, the conditional probability of W_i given previously written symbols, W'_j, is obtained from (1), $s \in \mathcal{S}$ and the joint probabilities are estimated by regularized relative frequency estimation from a large text corpus. The backspace symbol is assumed to have a constant conditional probability of 0.1 and the conditional probabilities of the other symbols are

$$P(W_i = s | W'_{i-1}, W'_{i-2}, \cdots, W'_{i-n+1}) = \frac{P(W_i = s, W'_{i-1}, \cdots, W'_{i-n+1})}{P(W'_{i-1}, \cdots, W'_{i-n+1})} \tag{1}$$

$$P(c_s = c | \boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s), \mathbf{W}'_{i-1}) = P(c_s = c | \delta_{\text{RDA}}(\mathbf{x}_{s,1}), \delta_{\text{RDA}}(\mathbf{x}_{s,2}), \cdots, \delta_{\text{RDA}}(\mathbf{x}_{s,N_S}), \mathbf{W}'_{i-1}) \tag{2}$$

$$= \frac{(\prod_{n_s=1}^{N_S} P(\delta_{\text{RDA}}(\mathbf{x}_{s,n_s}) | c_s = c)) P(c_s = c | \mathbf{W}'_{i-1}) P(\mathbf{W}'_{i-1})}{P(\delta_{\text{RDA}}(\mathbf{x}_{s,1}), \delta_{\text{RDA}}(\mathbf{x}_{s,2}), \cdots, \delta_{\text{RDA}}(\mathbf{x}_{s,N_S}), \mathbf{W}'_{i-1})} \tag{3}$$

$$P(W_i = s | \boldsymbol{\delta}_{\text{RDA}}, \mathbf{W}'_{i-1}) = P\left(c_s = 1 | \boldsymbol{\delta}_{\text{RDA}}, \mathbf{W}'_{i-1}, \sum_{t \in \mathcal{S}} c_t = 1\right) \tag{4}$$

$$= \frac{P(c_s = 1 | \boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s), \mathbf{W}'_{i-1}) / P(c_s = 0 | \boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s), \mathbf{W}'_{i-1})}{\sum_{t \in \mathcal{S}} P(c_t = 1 | \boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_t), \mathbf{W}'_{i-1}) / P(c_t = 0 | \boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_t), \mathbf{W}'_{i-1})} \tag{5}$$

normalized accordingly.[2]

In this study, a 6-gram model that is trained using a one-million-sentence (210M character) sample of the NY Times portion of the English Gigaword corpus. Corpus normalization and smoothing methods are described in [8]. Most importantly for this work, the corpus was case normalized, and Witten-Bell smoothing was used for regularization [15].

D. Joint target decision

The evidence obtained from EEG and the language model is used collaboratively to make a more informative symbol decision. For each epoch and a number of sequences shown, N_S, a decision will be made using the previously written symbols and EEG classification scores corresponding to N_S sequences. Let $\delta_{\text{RDA}}(\mathbf{x}_{s,n_s})$ be the corresponding posterior ratio scores obtained from RDA for letter $s \in \mathcal{S}$, where $n_s \in \{1, 2, \cdots, N_S\}$. Then the posterior probability of letter s to be in class c given the classification scores for letter s trials in each sequence and the previous letters is given in (2), where c_s is the candidate class label of letter s, n_{LM} is the order of the language model, $\mathbf{W}'_{i-1} = [W'_{i-1}, W'_{i-2}, \cdots, W'_{i-n_{LM}+1}]$ and $\boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s) = [\delta_{\text{RDA}}(\mathbf{x}_{s,1}), \delta_{\text{RDA}}(\mathbf{x}_{s,2}), \cdots, \delta_{\text{RDA}}(\mathbf{x}_{s,N_S})]$. If we further assume that the scores obtained from RDA for the stimuli corresponding to the current letter and previously written letters are conditionally independent given class label, i.e $\boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s) \perp\!\!\!\perp \mathbf{W}'_{i-1} | c$, and the RDA scores corresponding to EEG responses for different trials of the same letter in different sequences are conditionally independent given the class label, using Bayes' Theorem multiple times, the posterior probability becomes (3). The conditional probability density functions of RDA scores given the class labels, $P(\delta_{\text{RDA}}(\mathbf{x}_{s,n_s}) | c_s = c)$, are estimated using kernel density estimation on the scores of training data, using a Gaussian kernel whose bandwidth is selected using Silverman's rule of thumb that assumes the underlying density has the same average curvature with a matching-variance normal distribution [16].

Finally, while making our decisions we assume that there is exactly one target symbol in an epoch, which is reasonable since the user is expected to look for only one target symbol, and class labels for different symbols are independent given all the evidence. The posterior probability of the symbol is given as (4) where $\boldsymbol{\delta}_{\text{RDA}} = \{\boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s) : \forall s \in \mathcal{S}\}$. If $P(c_s = 1 | \boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s), \mathbf{W}'_{i-1}) \neq 1$ $\forall s \in \mathcal{S}$, after using our assumptions and Bayes' Theorem we obtain (4). Correspondingly, the most likely symbol is

$$\hat{W}_i = \arg\max_{s \in \mathcal{S}} P(W_i = s | \boldsymbol{\delta}_{\text{RDA}}, \mathbf{W}'_{i-1}).$$

[2]The symbol selection accuracy of our current prototype is around $85\% - 95\%$ for various subjects, therefore a deletion probability of 10% is reasonable.

In real-time typing, at the end of each sequence (a presentation of candidate symbols via RSVP) $P(W_i = \hat{W}_i | \boldsymbol{\delta}_{\text{RDA}}, \mathbf{W}'_{i-1})$ is used as the confidence of selecting the target symbol correctly, which gives us a stopping criterion for the epoch and allows us to have variable number of sequences for each typed symbol. In this study, the confidence threshold is set as 0.9, i.e when the accumulated conditional probability of the most likely symbol exceeds 0.9, the epoch is stopped and that symbol is typed. Using the maximum value in a probability mass function as the measure of certainty corresponds to utilizing Renyi's order-infinity entropy definition as the measure of uncertainty in the decision we make. One could use other definitions of entropy to measure uncertainty/certainty in the current decision.

III. EXPERIMENTS AND RESULTS

The EEG signals are recorded using a g.USBamp biosignal amplifier using active g.Butterfly and g.Ladybird electrodes with cap application from G.Tec (Graz, Austria) at 256 Hz. The EEG channels used were O1, O2, F3, F4, FZ, FC1, FC2, CZ, P1, P2, C1, C2, CP3 and CP4 according to International 10/20 system. Signals were filtered by nonlinear-phase 0.5-60 Hz bandpass filter and 60 Hz notch filter (G.tec's built-in design), afterwards signals filtered further by 1.5-42 Hz linear-phase bandpass filter (our design). The filtered signals were downsampled to 128Hz. For each channel, stimulus-onset-locked time windows of [0,500)ms following each image onset was taken as the stimulus response.

We conducted several typing experiments in which two healthy subjects and one locked-in subject typed some sentences using the proposed system acting in real-time followed by a short training session with predeclared targets consisting of 50 epochs with 3 sequences each, i.e 150 sequences of symbols. During the typing tests, subjects were given freedom to type what they want. However they were asked not to change the text they initially planned and to correct all the mistakes using the backspace symbol. For the classification, the effect of language modeling was also decreased by taking the square root of corresponding conditional probabilities to make the symbols more balanced. The inter-stimuli interval had been selected to be 150 ms, however it was increased to 400 ms due to the uncomfortable feeling of some of the subjects. During the typing session, each epoch was upper bounded to 6 sequences, which was decreased to 2 as the preference of the locked-in subject.

Typing performance is quantified by the number of sequences required to successfully type a desired symbol - including all errors and deletions until the correct desired symbol is achieved. All subjects wrote two short phrases in two sessions in two separate days. For the first healthy subject (HS1), whom inter-stimulus interval was 150 ms, the training session data from the first day was used to train the EEG classifier of the second day as well.

The typing performances of that subject are given in Fig. 2 and the average number of sequences needed per symbol were 2.93 and 1.58, respectively.[3] The typing performances of the second healthy subject (HS2), whom inter-stimulus interval was 400 ms, are given in Fig. 3 and the average number of sequences needed per symbol were 1.4 and 3.06, respectively. The locked-in subject successfully wrote HELLO_THERE_PEOP and THIS_IS_FAMILY_, and average number of sequences needed per symbol were 6.12 and 3.69, respectively [4].

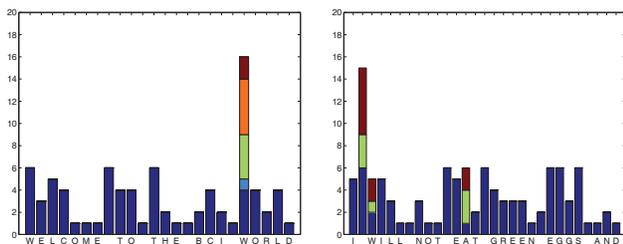Fig. 2.   HS1 number of sequences used to type each symbol.

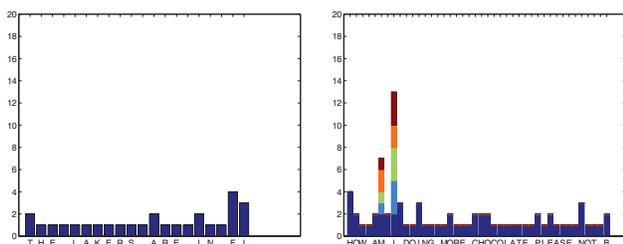

Fig. 3.   HS2 number of sequences used to type each symbol.

IV. DISCUSSION

In this paper, we presented the description of the first RSVP Keyboard prototype and some initial results from its real-time operations. Our experience had been that symbol selection accuracy can go up to around 95% for healthy subjects and around 85% for locked-in subjects (based on a few examples each). Training of subjects, especially those who are locked-in, to use the system effectively has been determined to be an unexpected major challenge. Although not discussed in this paper, we have identified that visual feedback about EEG-classifier activity during BCI calibration helps subjects greatly in adjusting their mental activity to help achieve better accuracy collaboratively with the system. The RSVP Keyboard uses a visual stimulus presentation paradigm that does not require precise gaze control - a capability that is not present in many patients with locked-in syndrome.

Both off-line and on-line results indicate that symbols that occur later in a word can be selected with fewer stimulus presentations (sequences). This demonstrates that the language model evidence is particularly useful for these symbols - once the first few letters are typed, the remaining letters in the word become highly predictable using a rudimentary language model and this helps EEG classification tremendously.

---

[3]Here, 1.00 would be the minimum possible, corresponding to single-trial typing of the desired text.

[4]The exact number of sequences per symbol was not recorded, however the total number of sequences used was extracted from the triggers recorded.

In future revisions of the RSVP Keyboard, we are contemplating the incorporation of probabilistic word completion in the RSVP paradigm (no drop down menus as in typical commercial systems, but instead a tight coupling with the RSVP Keyboard concept). We are also contemplating the optimization of presented sequences and saving time by selecting a subset of likely symbols in each sequence. All decisions will have to be made probabilistically in order to minimize expected time to type a symbol.

REFERENCES

[1] D.J. Krusienski, E.W. Sellers, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw, "Toward enhanced P300 speller performance," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 15–21, 2008.

[2] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, and M. Pregenzer, "Current trends in Graz brain-computer interface (BCI) research," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 216–219, 2000.

[3] M.S. Treder and B. Blankertz, "(C) overt attention and visual speller design in an ERP-based brain-computer interface," *Behavioral and Brain Functions*, vol. 6, no. 1, pp. 28, 2010.

[4] S. Mathan, D. Erdogmus, Y. Huang, M. Pavel, P. Ververs, J. Carciofini, M. Dorneich, and S. Whitlow, "Rapid image analysis using neural signals," in *CHI'08 extended abstracts on Human factors in computing systems*. ACM, 2008, pp. 3309–3314.

[5] S. Mathan, P. Ververs, M. Dorneich, S. Whitlow, J. Carciofini, D. Erdogmus, M. Pavel, C. Huang, T. Lan, and A. Adami, "Neurotechnology for Image Analysis: Searching for Needles in Haystacks Efficiently," *Augmented Cognition: Past, Present, and Future,*, 2006.

[6] H. Serby, E. Yom-Tov, and G.F. Inbar, "An improved P300-based brain-computer interface," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 13, no. 1, pp. 89–98, 2005.

[7] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[8] B. Roark, J. de Villiers, C. Gibbons, and M. Fried-Oken, "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*. Association for Computational Linguistics, 2010, pp. 28–36.

[9] U. Orhan, D. Erdogmus, B. Roark, S. Purwar, KE Hild, B. Oken, H. Nezamfar, and M. Fried-Oken, "Fusion with language models improves spelling accuracy for ERP-based brain computer interface spellers," in *IEEE Engineering in Medicine and Biology Society Conference Proceedings*. IEEE, 2011, pp. 5774–5777.

[10] K.E. Hild, U. Orhan, D. Erdogmus, B. Roark, B. Oken, S. Purwar, H. Nezamfar, and M. Fried-Oken, "An ERP-based brain-computer interface for text entry using rapid serial visual presentation and language modeling," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations*. Association for Computational Linguistics, 2011, pp. 38–43.

[11] P. Brunner, S. Joshi, S. Briskin, JR Wolpaw, H. Bischof, and G. Schalk, "Does the'p300'speller depend on eye gaze?," *Journal of Neural Engineering*, vol. 7, pp. 056013, 2010.

[12] L.A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[13] J.H. Friedman, "Regularized discriminant analysis," *Journal of the American statistical association*, vol. 84, no. 405, pp. 165–175, 1989.

[14] J. Nocedal and S.J. Wright, *Numerical optimization*, Springer verlag, 1999.

[15] I.H. Witten and T.C. Bell, "The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression," *Information Theory, IEEE Transactions on*, vol. 37, no. 4, pp. 1085–1094, 1991.

[16] B.W. Silverman, *Density estimation for statistics and data analysis*, Chapman & Hall/CRC, 1998.