Improved Accuracy Using Recursive Bayesian Estimation Based Language Model Fusion in ERP-Based BCI Typing Systems

U. Orhan[1], D. Erdogmus[1], B. Roark[2], B. Oken[2], S. Purwar[2], K. E. Hild II, A. Fowler[2], M. Fried-Oken[2]

***Abstract*— RSVP Keyboard$^{\text{TM}}$ is an electroencephalography (EEG) based brain computer interface (BCI) typing system, designed as an assistive technology for the communication needs of people with locked-in syndrome (LIS). It relies on rapid serial visual presentation (RSVP) and does not require precise eye gaze control. Existing BCI typing systems which uses event related potentials (ERP) in EEG suffer from low accuracy due to low signal-to-noise ratio. Henceforth, RSVP Keyboard$^{\text{TM}}$ utilizes a context based decision making via incorporating a language model, to improve the accuracy of letter decisions. To further improve the contributions of the language model, we propose recursive Bayesian estimation, which relies on non-committing string decisions, and conduct an offline analysis, which compares it with the existing naïve Bayesian fusion approach. The results indicate the superiority of the recursive Bayesian fusion and in the next generation of RSVP Keyboard$^{\text{TM}}$ we plan to incorporate this new approach.**

I. INTRODUCTION

Brain computer interfaces (BCI) is an emerging technology that is being designed for people with severe speech and physical impairments for communication and computer control. When the BCI is coupled with a communication system, the device may offer people who are completed paralyzed a means to generate expressive language [1], [2], [3], [4]. Unfortunately the use of noninvasive BCI techniques on letter-by-letter spelling systems suffers from low accuracies for symbol selection due to low signal to noise ratio and variability of background brain activity. Researchers have turned to various hierarchical symbol trees to improve spelling efficiency [3], [5], [6]. Slow throughput greatly diminishes the practical usability of such systems. Incorporation of a language model into the decision-making process to predict the next letter using the previous letters can greatly affect the performance of these systems by improving both accuracy and speed [7]. If the symbol decisions are based on EEG evidence only, they will not be accurate enough, thus reducing the value of any text prediction. We propose a new BCI, the RSVP Keyboard$^{\text{TM}}$, based on rapid serial visual presentation (RSVP), where there is a fusion of EEG data and a sophisticated language model to affect decision making for symbol selection [8].

In this paper, we propose a new fusion technique based on recursive Bayesian estimation for RSVP Keyboard$^{\text{TM}}$, which is also applicable for any BCI typing system. This technique aims to make the fusion more robust to uncorrected errors to accommodate the needs of the user population who prefer continuing with their mistakes instead of correcting them. To quantitatively measure this effect, we investigate context based decision making, without any erasure option, in an offline/simulated manner. EEG classification of the event related potentials (ERP) corresponding to stimuli for RSVP is done using regularized discriminant analysis (RDA). Offline analysis is done by randomly sampling words from a language lexicon with appropriate frequencies. For each sampled word, a typing scenario, which the EEG data is ordered to simulate the letter-by-letter typing, is created. Context based letter probabilities and EEG classification scores are merged using a naïve Bayesian estimation approach and a recursive Bayesian estimation approach, comparatively. We present a performance analysis that compares different scenarios with varying numbers of visual presentation sequences used in EEG classification and initial results are very encouraging.

II. RSVP BASED BCI AND ERP CLASSIFICATION

RSVP is an experimental psychophysics technique in which visual stimulus sequences are displayed on a screen over time on a fixed focal area and in rapid succession. The Matrix-P300-Speller [1] used by Wadsworth and Graz groups (especially G.tec) opts for a spatially distributed presentation of possible symbols, highlighting them in different orders and combinations to elicit P300 responses. Berlin BCI's recent variation utilizes a 2-layer tree structure [3] where the subject chooses among six units (symbols or sets of these) where the options are laid out on the screen while the subject focuses on a central focal area that uses an RSVP-like paradigm to elicit P300 responses. In contrast, our approach is to distribute the stimuli temporally and present one symbol at a time using RSVP and seek a binary answer to find the desired letter in a right-branching tree. The latter method has the advantage of not requiring the user to look at different areas of the screen.

In the current study, which is an offline analysis, our RSVP paradigm utilizes stimulus sequences consisting of letters in the English alphabet presented sequentially with random ordering where the user is expected to show positive intent for only one predesignated letter for each epoch (see details below). When the user sees the predesignated infrequent (1 in 26) target, the brain generates an event related potential (ERP) in the EEG; the most prominent component of this ERP is the P300 wave, which is a positive deflection in the scalp voltage primarily over centro-parietal cortex that generally occurs with a latency of about 300 ms. This natural response of the brain to the event of visual stimulus matching the rare sought target allows us to make binary decisions about user's intent.

The intent detection problem becomes a signal classification problem when the EEG signals are windowed in a stimulus-time-locked manner over a duration with sufficient length. After some preprocessing (details are explained in [9]) the signals acquired from each EEG channel will be incorporated and classified to determine the class label: ERP or non-ERP.

Regularized Discriminant Analysis (RDA) [10] is used to determine a classification discriminant score for each stimulus indicating whether it is a response to a target letter or not; this score is used in conjunction with a language model to make the final Bayesian decision on the class label of each letter. RDA is a modified quadratic discriminant analysis (QDA) model, where the feature vectors from each class assumed to belong to a multivariate normal distribution. The singularities in the covariance matrices estimated

*This work is supported by NIH under grant 1R01DC009834-01. The opinions presented here are solely those of the authors and do not necessarily reflect the opinions of the funding agency.

[1]Cognitive Systems Laboratory, Northeastern University, Boston, MA
[2]Oregon Health and Science University, Portland, OR, USA

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

$$P(W_i = s|\boldsymbol{\delta}_{\text{RDA},i}, \mathbf{W}'_{i-1}) = P\left(c_s = 1|\boldsymbol{\delta}_{\text{RDA},i}, \mathbf{W}'_{i-1}, \sum_{t\in\mathcal{S}} c_t = 1\right) \tag{2}$$

$$= \frac{P(c_s = 1|\boldsymbol{\delta}_{\text{RDA},i}(\mathbf{x}_s), \mathbf{W}'_{i-1})/P(c_s = 0|\boldsymbol{\delta}_{\text{RDA},i}(\mathbf{x}_s), \mathbf{W}'_{i-1})}{\sum_{t\in\mathcal{S}} P(c_t = 1|\boldsymbol{\delta}_{\text{RDA},i}(\mathbf{x}_t), \mathbf{W}'_{i-1})/P(c_t = 0|\boldsymbol{\delta}_{\text{RDA},i}(\mathbf{x}_t), \mathbf{W}'_{i-1})} \tag{3}$$

from the calibration data are removed using the following shrinkage and regularization operations. The shrinkage procedure makes the class covariances closer to the overall data covariance, and therefore to each other, thus making the quadratic boundary closer to a linear one. Shrinkage is applied as

$$\hat{\mathbf{\Sigma}}_c(\lambda) = \frac{(1-\lambda)\hat{\mathbf{S}}_c + \lambda\hat{\mathbf{S}}}{(1-\lambda)N_c + \lambda N}, \quad \hat{\mathbf{S}}_c = N_c\hat{\mathbf{\Sigma}}_c, \quad \hat{\mathbf{S}} = \hat{\mathbf{S}}_0 + \hat{\mathbf{S}}_1 \tag{4}$$

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

$$\hat{\mathbf{\Sigma}}_c(\lambda, \gamma) = (1-\gamma)\hat{\mathbf{\Sigma}}_c(\lambda) + \frac{\gamma}{d}\text{tr}[\hat{\mathbf{\Sigma}}_c(\lambda)]\mathbf{I}, \tag{5}$$

where γ is the regularization parameter, tr$[\cdot]$ is the trace function and d is the dimension of the data vector.

After carrying out the regularization and shrinkage on the estimation covariance matrices, the Bayesian classification rule [11] is defined as the comparison of the log-of-the-posterior-ratio using the posterior probability distributions with a threshold, which can incorporate the relative risks or costs of making an error for each class. The corresponding log-of-the-posterior-ratio is given by

$$\delta_{\text{RDA}}(\mathbf{x}) = \log\frac{f_{\mathcal{N}}(\mathbf{x}; \hat{\boldsymbol{\mu}}_1, \hat{\mathbf{\Sigma}}_1(\lambda, \gamma))\hat{\pi}_1}{f_{\mathcal{N}}(\mathbf{x}; \hat{\boldsymbol{\mu}}_0, \hat{\mathbf{\Sigma}}_0(\lambda, \gamma))\hat{\pi}_0}, \tag{6}$$

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 normal distribution.

The letter candidates, which contain all possible selectors, can be shown multiple times to achieve a higher classification accuracy in EEG-scores by making use of independent visual stimulus trial responses, as is commonly the case in EEG-based spellers[1]. We define a sequence to be a randomly ordered set of all letters shown as stimuli. Since the randomness of the target stimulus position in any given sequence is key to eliciting an ERP, a random permutation of the letters is used for each sequence in our experiments.

Thereafter all or some of the sequences can be used to classify if a letter is target or non-target, depending on the operational mode of the ERP classifier, that is whether it is using a single-trial, 2-trial, or 3-trial approach. The log-of-the-posterior-ratios can then be used in a fusion with a language model explained in the following section.

III. LANGUAGE MODELING

Language modeling is very important for many text processing applications, such as speech recognition, machine translation, as well as for the kind of typing application being investigated here [12]. Typically, the prefix string (what has already been typed) is used to predict the next symbol(s) to be typed. The next letters to be typed become highly predictable in certain contexts, particularly word-internally. In applications where text generation/typing speed is very slow, the impact of language modeling can become much more significant. BCI-spellers, including the RSVP Keyboard™ paradigm presented here, can be extremely low-speed letter-by-letter writing systems, and thus can greatly benefit from the incorporation of probabilistic letter predictions from an accurate language model during the writing process.

The language model used in this paper is based on the n-gram sequence modeling paradigm, very widely used in all of the application areas mentioned above. It estimates the conditional probability of any letter in a sequence given $n-1$ previous letters using a Markov model of order $n-1$. Let W be a sequence of letters where W_i is the i^{th} letter and let $\mathcal{S}$ be the set of candidate symbols. For an n-gram model, the conditional probability of W_i given previously written symbols is obtained using a regularized relative frequency estimation from a large text corpus.

For the current study, all n-gram language models were estimated from a one million sentence (210M character) sample of the NY Times portion of the English Gigaword corpus. Corpus normalization and smoothing methods were as described in [12]. Most importantly for this work, the corpus was case normalized, and we used Witten-Bell [13] smoothing for regularization.

IV. LANGUAGE MODEL AND EEG FUSION

A. Naïve Bayesian Approach

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 becomes (1), where c_s is the candidate class label of letter s, n_{LM} is the order of the language model, $\mathbf{W}'_{i-1} = [W'_{i-1}, W'_{i-2}, \cdots, W'_{i-n_{LM}+1}]$ represents the already selected symbols and $\boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s) = [\delta_{\text{RDA}}(\mathbf{x}_{s,1}), \delta_{\text{RDA}}(\mathbf{x}_{s,2}), \cdots, \delta_{\text{RDA}}(\mathbf{x}_{s,N_S})]$. This is obtained assuming 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. 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 [14].

Finally, while making our decisions we assume that exactly one of the candidate symbols is target, which is reasonable since the user is expected to look for only one target symbol,

[1]The typical number of repetitions of visual stimuli is usually 8 or 16, although G.tec claims one subject is able to achieve reliable operation with 2-trials (verbal communication). We had similar subjects who can type accurately with 1 or 2 trials per symbol.

$$P(\mathbf{W}_i = \mathbf{s}|\boldsymbol{\delta}_{\text{RDA},1},\cdots,\boldsymbol{\delta}_{\text{RDA},i}) \;\propto\; p(\boldsymbol{\delta}_{\text{RDA},i}|W_i = s_{last})\cdot \sum_{\mathbf{s}'\in\mathcal{S}^{n_{LM}}} P(\mathbf{W}_i = \mathbf{s}|\mathbf{W}_{i-1} = \mathbf{s}')P(\mathbf{W}_{i-1} = \mathbf{s}'|\boldsymbol{\delta}_{\text{RDA},1},\cdots,\boldsymbol{\delta}_{\text{RDA},i-1}) \tag{7}$$

$$P(\mathbf{W}_i = \mathbf{s}|\mathbf{W}_{i-1} = \mathbf{s}') = \begin{cases} P(W_i = s_{last}|\mathbf{W}_{i-1} = s_{1:last-1}) & ,\ s_{1:last-1} = s'_{1:last} \\ 0 & ,\ \text{otherwise} \end{cases} \tag{8}$$

$$P(\mathbf{W}_i = \mathbf{s}|\boldsymbol{\delta}_{\text{RDA},1:i}) \propto p(\boldsymbol{\delta}_{\text{RDA},i}|W_i = s_{last})P(W_i = s_{last}|\mathbf{W}_{i-1} = s_{1:last-1})P(\mathbf{W}_{i-1} = s_{1:last-1}|\boldsymbol{\delta}_{\text{RDA},1:i-1}) \tag{9}$$

$$P(\mathbf{W}_i = \mathbf{s}|\mathbf{W}_{i-1} = \mathbf{s}') = \begin{cases} P(W_i = s_{last}|\mathbf{W}_{i-1} = s_{1:last-1}) & ,\ s_{1:last-1} = s'_{2:last} \\ 0 & ,\ \text{otherwise} \end{cases} \tag{10}$$

$$P(\mathbf{W}_i = \mathbf{s}|\boldsymbol{\delta}_{\text{RDA},1:i}) \;\propto\; p(\boldsymbol{\delta}_{\text{RDA},i}|W_i = s_{last})P(W_i = s_{last}|W_{i-1} = s_{last-1},\cdots,W_{i-n_{LM}+1} = s_1)\cdot \sum_{s'_0\in\mathcal{S}} P(W_{i-1} = s_{last-1},\cdots,W_{i-n_{LM}+1} = s_1, W_{i-n_{LM}} = s'_0|\boldsymbol{\delta}_{\text{RDA},1:i-1}) \tag{11}$$

and class labels for different symbols are independent given all the evidence. The posterior probability of the symbol is given as (2) where $\boldsymbol{\delta}_{\text{RDA},i} = \{\boldsymbol{\delta}_{\text{RDA}}(\mathbf{x}_s) : \forall s \in \mathcal{S}\}$, for i^{th} epoch. 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 (2). Correspondingly, the most likely symbol is

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

B. Recursive Bayesian Approach

The naïve approach has the disadvantage of committing to one selection of a string of symbols, which might potentially cause the language model to be more harmful then it is beneficial after an erroneous selection. Therefore as a remedy to this problem, we propose to construct the hidden Markov model (HMM) where the substring of last n_{LM} letters of the current epoch constitute the latent variable and EEG scores of all symbols corresponding to an epoch constitute the observed variable. An illustration of the corresponding HMM is given in Fig. 1, where $\boldsymbol{\delta}_{\text{RDA},i}$ represents $\boldsymbol{\delta}_{\text{RDA}}$ for epoch i.

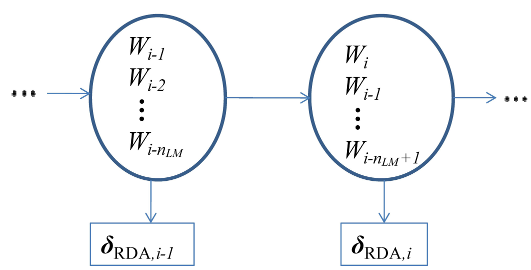

Fig. 1: Hidden Markov Model used in recursive Bayesian approach for the fusion of language model with EEG scores

Using recursive Bayesian estimation [15], marginalizing over older state variable the current state probability mass given all observed EEG scores obtained in (7) where $\mathbf{s} = \{s_1, s_2, \cdots, s_{last}\}$ is a specific state. There are common variables between consecutive latent states, therefore $P(\mathbf{W}_i = \mathbf{s}|\mathbf{W}_{i-1} = \mathbf{s}')$ becomes zero for many $\mathbf{s}'$ which decreases the number of summations required. If $i \leq n_{LM}$ it becomes (8), consequently the posterior probability of state variable becomes (9). If $i > n_{LM}$, similarly we obtain (10) and (11).

Assuming all EEG scores for an epoch are conditionally independent given the class labels, we can calculate $p(\boldsymbol{\delta}_{\text{RDA},i}|W_i = s)$ as (12), where $p(\delta_{\text{RDA},i}(\mathbf{x}_{s,n_s}|c_s = c)$ is estimated using KDE as in IV-A, $P(W_i = s_{last}|W_{i-1} = s_{last-1},\cdots,W_{i-n_{LM}+1} = s_1)$ or $P(W_i = s_{last}|\mathbf{W}_{i-1} = s_{1:last-1})$ is estimated using n-gram language model and $P(\mathbf{W}_{i-1} = s_{1:last-1}|\boldsymbol{\delta}_{\text{RDA},1:i-1})$ is calculated recursively.

Finally, at any epoch, we can estimate the latest substring using maximum likelihood,

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

Similarly, at i^{th} epoch, we can estimate W_j for $1 \leq j \leq i - n_{LM}$ using maximum likelihood as (13).

V. EXPERIMENTAL RESULTS

For this study, two healthy subjects were recruited for two sessions each. In each session 200 letters are selected (with replacement, out of 26) according to their frequencies in the English language and randomly ordered to be used as target letters in each epoch. In each epoch, the designated target letter and a fixation sign are shown for 1s each and followed by 3 sequences of randomly ordered 26 letters of the English alphabet with 150 ms inter-stimuli interval. Subjects are asked to look for the target letter shown at the beginning of the epoch.

The signals are recorded using a g.USBamp biosignal amplifier using active g.Butterfly electrodes from G.tec (Graz, Austria) at 256Hz. The EEG channels positioned according to the International 10/20 System were O1, O2, F3, F4, FZ, FC1, FC2, CZ, P1, P2, C1 C2, CP3, CP4. 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.

Let us denote by e_j the jth epoch in a given session and let $\mathbb{E}$ be the ordered set containing all epochs in the session. $\mathbb{E}$ is partitioned into 10 equal-sized nonintersecting blocks, $\mathbb{E}_k$; for every e_j there is exactly one k_j such that $e_j \in \mathbb{E}_{k_j}$. For every e_j acting as a test sample, the ERP classifier is trained on the set $\mathbb{E}\backslash\mathbb{E}_{k_j}$. During training, the classifier parameters λ and γ are determined using 10-fold validation and grid search within the set $\mathbb{E}\backslash\mathbb{E}_{k_j}$. The kernel density estimates of the conditional probabilities of classification scores for EEG classifiers are obtained using scores obtained from $\mathbb{E}\backslash\mathbb{E}_{k_j}$. The trained classifiers are applied to their respective test epochs to get the 10-fold cross-validation test results presented in the tables.

The language model was trained as described in III. 36000 words are randomly drawn from a word lexicon weighted according to word occurrences and letter count. Each sampled word is considered to be a typing scenario and for each letter in the word we simulate the fusion of EEG responses and the language model in the following way: (i) each letter is assumed to be the target letter of a typing process using BCI; (ii) an EEG epoch with a matching

$$p(\boldsymbol{\delta}_{\text{RDA},i}|W_i = s) = \left(\prod_{n_s=1}^{N_S} p(\delta_{\text{RDA},i}(\mathbf{x}_{s,n_s})|c_s = 1)\right) \prod_{s' \in S \setminus \{s\}} \prod_{n_{s'}=1}^{N_S} p(\delta_{\text{RDA},i}(\mathbf{x}_{s',n_{s'}})|c_{s'} = 0) \tag{12}$$

$$\begin{aligned} \hat{W}_j &= \arg\max_{s_0 \in \mathcal{S}} P(W_j = s_0|\boldsymbol{\delta}_{\text{RDA},1:i}) = \arg\max_{s_0 \in \mathcal{S}} P(W_j = s_0|\boldsymbol{\delta}_{\text{RDA},1:j+n_{LM}}) \\ &= \arg\max_{s_0 \in \mathcal{S}} \sum_{s'_1,\cdots,s'_{n_{LM}-1} \in \mathcal{S}} P(W_{j+n_{LM}-1} = s'_{n_{LM}-1}, \cdots, W_{j+1} = s'_1, W_j = s_0|\boldsymbol{\delta}_{\text{RDA},1:j+n_{LM}}) \end{aligned} \tag{13}$$

target symbol is randomly selected from the epochs collected[2]; (iii) under the assumption of independence of EEG responses with the previous letters selected, for each epoch, the EEG responses for every letter is converted to EEG classifier scores; (iv) the matching model probabilities for each letter are obtained from the language model using letters selected priorly with the same process; (v) and the fusion of ERP classifier scores and language models was achieved as described above. There was no erasure operation, and consequently during the simulation it is assumed that the subjects continued to type the next planned symbol even if the previous one selected incorrectly. Fusion results were obtained for n-gram model

	1 sequence	2 sequences	3 sequences
Naïve Bayesian fusion	(0.26, 0.36)	(0.46, 0.64)	(0.63, 0.82)
Recursive Bayesian fusion	(0.35, 0.49)	(0.64, 0.73)	(0.76, 0.87)

TABLE I: The minimum and the maximum values of the correct symbol selection ratios over different sessions for varying the number of sequences used in EEG classification for naïve Bayesian fusion and recursive Bayesian fusion approaches.

order of 4. The EEG scores were assumed to have been evaluated for N_S = 1, 2, and 3 sequences (to evaluate the contribution of multi-trial information) to decide if a letter under evaluation was a desired target letter or not. In the results, only EEG data from the first N_S sequences of the epoch was used for classification for each selected sequence count. Ratio of total number of letters selected correctly and the total number of letters is presented in Table I for all sessions.

VI. Conclusion

In previous work [9], we demonstrated that incorporation of context based information using naïve fusion to support ERP classification can improve the symbol classification accuracy upto 3-fold compared. Moreover, we demonstrated a real time system in action using this naïve fusion approach [16]. In this paper we demonstrated with a simulation study that using a recursive Bayesian estimation to estimate the last substring might considerably improve the typing accuracy, consequently speed by using less number of sequences to type each symbol, over our previously implemented naïve fusion.

In all of the sessions, the approach with recursive Bayesian estimation outperformed considerably the naïve fusion rule for all sequences per epoch, where the performances increase monotonically with the number of sequences as expected. The results are very promising, however it is important to note that there is no erasure symbol in these simulations. Since the naïve approach is less resistant to errors, existence of an erasure symbol might improve its performance more considerably than the other. As future work, we are planning to collect data for a similar analysis including the erasure symbol and using different model orders, which would result with a more accurate comparison.

The incorporation of recursive Bayesian approach to the real time system would also be useful to assess its performance. However this approach has some disadvantages which might become important during the real time implementation. Firstly, it requires a large amount of memory since it keeps all the states from the previous selection. Secondly, the interface needs to be more confusing for the subject since previously selected symbols are ambiguous. As we go forward with improvements to our existing RSVP Keyboard™ system, all of these considerations will be important.

VII. Acknowledgment

We would like to acknowledge Dr. Murat Akcakaya for valuable comments.

References

[1] D. Krusienski, E. Sellers, D. McFarland, T. Vaughan, and J. 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. 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, p. 28, 2010.

[4] S. Fager, D. R. Beukelman, M. Fried-Oken, T. Jakobs, and J. Baker, "Access interface strategies," *Assistive Technology*, vol. 24, no. 1, pp. 25–33, 2012.

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

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

[7] B. Roark, A. Fowler, and M. Fried-Oken, "Language models in keybooard emulation," *Computers and Human Interaction Conference*, 2012, to appear.

[8] U. Orhan, K. Hild, D. Erdogmus, B. Roark, B. Oken, and M. Fried-Oken, "RSVP keyboard: An EEG based typing interface," *ICASSP*, 2012, to appear.

[9] U. Orhan, D. Erdogmus, B. Roark, S. Purwar, K. 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] J. Friedman, "Regularized discriminant analysis," *Journal of the American statistical association*, vol. 84, no. 405, pp. 165–175, 1989.

[11] R. Duda, P. Hart, and D. Stork, *Pattern classification*. Citeseer, 2001.

[12] 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*, 2010, pp. 28–36.

[13] I. Witten and T. 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.

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

[15] A. Jazwinski, *Stochastic processes and filtering theory*. Academic Pr, 1970, vol. 63.

[16] K. 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*. Association for Computational Linguistics, 2011, pp. 38–43.

[2]Since subjects only focus to a single target letter without knowing the predecessor letters of the typing process in this experiment, it is assumed that the EEG responses created during an epoch are independent from the predecessors.