

Language-Model Assisted Brain Computer Interface for Typing: A Comparison of Matrix and Rapid Serial Visual Presentation

Mohammad Moghadamfalahi, *Student Member, IEEE*, Umut Orhan, *Member, IEEE*,
Murat Akcakaya, *Member, IEEE*, Hooman Nezamfar, *Student Member, IEEE*,
Melanie Fried-Oken, and Deniz Erdogmus, *Senior Member, IEEE*

Abstract—Noninvasive electroencephalography (EEG)-based brain-computer interfaces (BCIs) popularly utilize event-related potential (ERP) for intent detection. Specifically, for EEG-based BCI typing systems, different symbol presentation paradigms have been utilized to induce ERPs. In this manuscript, through an experimental study, we assess the speed, recorded signal quality, and system accuracy of a language-model-assisted BCI typing system using three different presentation paradigms: a 4×7 matrix paradigm of a 28-character alphabet with row-column presentation (RCP) and single-character presentation (SCP), and rapid serial visual presentation (RSVP) of the same. Our analyses show that signal quality and classification accuracy are comparable between the two visual stimulus presentation paradigms. In addition, we observe that while the matrix-based paradigm can be generally employed with lower inter-trial-interval (ITI) values, the best presentation paradigm and ITI value configuration is user dependent. This potentially warrants offering both presentation paradigms and variable ITI options to users of BCI typing systems.

Index Terms—Brain-computer interface, event-related potential, matrix speller, P300, RSVP keyboard.

I. INTRODUCTION

NONINVASIVE brain-computer interfaces (BCIs), specifically those based on electroencephalography (EEG), have become popular to safely enable people with severe motor and speech impairments to communicate with their social networks and interact with their environments [1]–[3]. Typing is one of the most widely explored applications for EEG-based BCI systems [1]. Event-related potentials (ERPs), specifically

the P300 component of these EEG responses, are commonly exploited by such typing interfaces for user intent detection [4]–[7].

The pioneering work of Farwell and Donchin showed that ERPs containing the P300 response can be used to design EEG-based BCI typing systems [4]. They distributed 36 symbols consisting of the 26 letters in the English alphabet and 10 numerical digits across a 6×6 matrix. The rows and columns of the matrix are flashed in a random fashion to generate an oddball paradigm such that when the row or column that includes the symbol that the user intends to select is flashed, an ERP containing the P300 component is elicited. This ERP is then used for target symbol detection. P300 is a positive deflection in the scalp voltage with a typical latency around 300 ms after the onset of an infrequent target stimuli [8].

Despite the practice being the benchmark in matrix spellers, flashing rows and columns for the presentation of a symbol may result in poor P300 signal quality, and a single character flashing paradigm enhances the P300 response [9]. Studies also demonstrated that the performance of a BCI typing system that employs a matrix presentation paradigm depends on the gaze of the user [10], [11]. Many potential users from the target population, unfortunately, lack precise gaze control, and for these users, it is anticipated that matrix paradigms will suffer from reduced performance. To overcome this dependency in BCI typing systems, different presentation schemes have been explored and shown to have comparable performances with the matrix presentation paradigm in terms of speed and accuracy [11]–[13]. Rapid serial visual presentation (RSVP) is one of these paradigms, in which symbols are presented sequentially in time, at a predefined fixed location on the screen and in a pseudorandom order [5], [14]–[18].

BCI typing systems can benefit greatly from a language model in order to enhance typing speed. A probabilistic language model can be employed to incorporate predictive word completion during the intent detection process [19]–[21], or to define a prior on potential target characters during the classification task [22]–[24]. Our system, the RSVP keyboard, originally developed based on the RSVP paradigm and now also featuring the matrix presentation paradigm, probabilistically fuses context evidence with physiological evidence to infer user intent. A symbol n-gram language model trained on a large corpus provides probabilities for each character in the

Manuscript received June 09, 2014; revised November 13, 2014; accepted February 04, 2015. This work was supported by NIH grant R01DC009834 and NSF grants CNS-1136027, IIS-1149570, SMA-0835976. The package, including the code and data associate with this paper, can be find at <https://repository.lib.neu.edu/collections/neu:rx913r029>.

M. Moghadamfalahi, U. Orhan, H. Nezamfar, and D. Erdogmus are with Northeastern University, Boston, MA, USA (e-mail: moghamdam@ece.neu.edu; orhan@ece.neu.edu; nezamfar@ece.neu.edu; erdogmus@ece.neu.edu).

M. Akcakaya is with University of Pittsburgh, Pittsburgh, PA, USA.

M. Fried-Oken is with Oregon Health and Science University, Portland, OR, USA. **[AU: Please provide all missing postal codes--ed.]**

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNSRE.2015.2411574

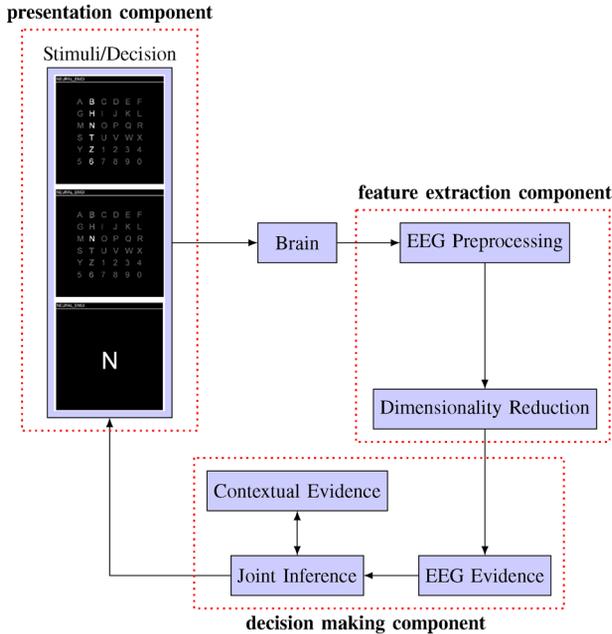


Fig. 1. The in-house BCI block diagram.

alphabet, which are fused tightly in a Bayesian fashion with EEG evidence [5], [15], [16].

In this paper, we utilize two different matrix schemes (row column flash and single symbol flash) and one RSVP scheme in a BCI typing interface and compare the differences in measured signal quality, typing speed, and accuracy. In a similar study, Chennu *et al.*, through an offline study, have shown that the classification accuracy is comparable between RSVP and matrix based paradigms, but without a language model the typing speed is relatively low while utilizing the RSVP paradigm [13]. In this study, we also compare the typing performance during online typing of both RSVP and matrix paradigms, using the aforementioned language-model-assisted BCI.

The contributions of this paper areas follows:

- 1) building a unified framework for different presentation paradigms that utilize EEG and language model evidence for joint decision making;
- 2) conducting real-time and offline comparisons among different presentation schemes;
- 3) analyzing the effect of different presentation paradigms on the EEG signal quality.

II. GENERAL SYSTEM SPECIFICATIONS

The complete operational flowchart of the language-model-assisted BCI typing system is illustrated in Fig. 1. The system has the following main components: (A) a **presentation component** that controls the presentation scheme, (B) a **feature extraction component** that converts raw EEG evidence into a likelihood for Bayesian fusion and (C) a **decision making component** that fuses EEG (physiology) and language evidence to infer user intent. In the following, we describe these components in some more detail.

A. Presentation Component

1) *Definitions:* Let $\mathcal{A} = \{a_1, a_2, a_3, \dots, a_N\}$ be the set of all possible symbols, typically including the letters in the (English) alphabet, numerical symbols, space and backspace symbols (represented here by $_$ and \langle , respectively). Let $\mathcal{F} = \{f_1, f_2, \dots, f_{2^{|\mathcal{A}|}}\}$ be the set of all subsets of \mathcal{A} ; $f_i \subset \mathcal{A}$. $|\mathcal{A}|$ represents the cardinality of \mathcal{A} .

A “trial” in the matrix based presentation scheme flashes a subset f_i that can contain multiple characters, i.e., $|f_i| \geq 1$, and in RSVP, it presents a single symbol; i.e., $|f_i| = 1$. A “flash” is the presentation of a trial. A “sequence” is a series of consecutive flashes of trials with no gap in between. After presenting each sequence, the system updates the posterior probabilities of every symbol in the alphabet \mathcal{A} using the new EEG evidence and tries to make an inference about user intent. However, a decision is not made until a predefined confidence level is reached.¹ Therefore, the system may need to present multiple sequences before a decision can be made. We define the collection of sequences, at the end of which one symbol is selected, as an “epoch”.

2) *Matrix Presentation:* Typically, in noninvasive EEG-based typing BCIs with the matrix presentation paradigm, symbols are arranged in an $\mathcal{R} \times \mathcal{C}$ matrix with \mathcal{R} number of rows and \mathcal{C} number of columns [1]. Subsets of these symbols are intensified usually in pseudorandom order to produce an odd ball paradigm to induce ERP responses.

Trials $f_{(1)}, f_{(2)}, \dots, f_{(n)}$ in a sequence typically cover all the symbols in the matrix, that is $\bigcup_{i=1}^n f_{(i)} = \mathcal{A}$. When each trial $f_{(i)}$ contains exactly all the symbols in a row or a column of the matrix layout with $n = (\mathcal{R} + \mathcal{C})$ [4], this setup is known as the row-and-column presentation (RCP) paradigm. RCP requires that all the symbols in \mathcal{A} would be flashed twice and $|f_i \cap f_j| \leq 1, i \neq j$. In this study, we utilize a matrix of size 4×7 , which leads to the best coverage of the widescreen monitors used in our experiments. It has been claimed that the probability of target character in each sequence’s flash set should be lower than 25% to induce the P300 response [4]. In this grid setup for RCP, each sequence contains 11 flashes, two of which include the target symbol. Therefore, the probability of each target trial in each sequence is $(2/11) \simeq 0.18$, which satisfies the threshold suggested above.

A single-character presentation (SCP) paradigm is also a widely used scheme. SCP was shown to increase the P300 signal quality compared with RCP [9]. In this paradigm, each trial contains single symbols, i.e., $|f_i| = 1$, and assuming there is no repetition in a sequence, $f_i \cap f_j = \emptyset; i \neq j$. With enough number of flashes ($n \geq 5$) in a sequence, we can satisfy the suggested condition for target probability.

3) *Rapid Serial Visual Presentation (RSVP):* RSVP is a presentation technique in which trials are presented one at a time at a fixed predefined location on the screen at a rapid rate and in a pseudorandom order [1], [5]. If a BCI user’s desired symbol exists in a sequence of trials presented in RSVP fashion, a P300

¹In the current implementation, confidence is measured by the maximum posterior probability over \mathcal{A} ; this corresponds to using Renyi entropy of order ∞ as the measure of uncertainty. Other entropy definitions such as Shannon’s could also be used.

response is elicited by the target in the EEG signal. RSVP is similar to SCP in that each presentation subset includes only a single symbol; however, RSVP decreases the dependency on gaze control. Presenting 28 symbols in an RSVP paradigm is time consuming; therefore, a typical RSVP-based BCI system can only achieve a speed of five symbols/minute if each sequence contains the entire alphabet [5], [16]–[18]. However, recent efforts to speed up typing with this presentation paradigm showed that using context information (such as a language model) and careful selection of subsets of a in each sequence may significantly improve typing speed and accuracy [5], [15], [16], [19], [21].

B. Feature Extraction Component

The EEG signals are acquired using a g.USBamp biosignal amplifier with active g.Butterfly electrodes at a sampling rate of 256 Hz, from 16 EEG sites (according to the International 10/20 configuration): Fp1, Fp2, F3, F4, Fz, Fc1, Fc2, Cz, P1, P2, C1, C2, Cp3, Cp4, P5, and P6. To improve the signal-to-noise ratio (SNR) and to eliminate drifts, signals were filtered by an FIR linear-phase bandpass filter passing [1.5, 42] Hz with zero dc gain and a notch filter at 60 Hz.

In order to capture the P300 while omitting the possible motor EEG [8], EEG from a time window of [0,500] ms after each flash’s onset is processed as the corresponding raw data for each trial. As we explain later in Section III, we test our system with healthy users; therefore the window length is chosen short to avoid any discriminative contributions of motor-activity-related EEG response, if any. EEG data processing continues with i) downsampling by 2, ii) projection to a lower dimensional space using principle component analysis (PCA) to remove directions with negligible variance, and iii) concatenation of data from all channels corresponding to the same trial to form a feature vector for each trial.

C. Decision Making Component

Evidence from EEG is supported with evidence from language structure. These two information sources are fused using a Naïve Bayes’ assumption to make a joint decision using MAP inference. Optimal classifier parameters for target detection are learned using the calibration data.

1) *EEG Feature Extraction and Classification*: To improve intent detection performance, the EEG feature vectors computed as described above are projected in to a one-dimensional space, which attempts to maximize the separation between target and nontarget classes according to a measure. Specifically, assuming that, in each class, feature vectors follow a multivariate Gaussian distribution,² quadratic discriminant analysis (QDA) is used to project the data to minimize the expected risk. QDA requires the inverse of the empirical covariance for each class. Estimating an invertible covariance is not feasible in the practical usage of the typing system due to the high dimensionality of the EEG feature vectors and low number of calibration samples in each class. This issue has been addressed by employing regularized discriminant

analysis (RDA), which provides full-rank covariance estimates for each class [25].

RDA uses shrinkage and regularization. Shrinkage is a linear combination of each class covariance matrix and the overall class-mean-subtracted covariance. Considering $\mathbf{x}_i \in \mathbf{R}^p$ as a p -dimensional feature vector and l_i as its label, which can take values of 0 and 1 for nontarget and target classes, respectively, the maximum-likelihood estimator for mean and covariance of each class are

$$\begin{aligned}\boldsymbol{\mu}_k &= \frac{1}{N_k} \sum_{i=1}^N \mathbf{x}_i \delta_{l_i, k} \\ \boldsymbol{\Sigma}_k &= \frac{1}{N_k} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \delta_{l_i, k}\end{aligned}\quad (1)$$

where $k \in \{0, 1\}$, N_k is the number of training feature vectors in class k , and thus N , the total number of feature vectors, will be $N_0 + N_1$ and $\delta_{\langle \cdot, \cdot \rangle}$ is the Kronecker- δ . **[AU: Previous sentence edited correctly for meaning?]** The shrinkage procedure manipulates the covariance matrices by

$$\hat{\boldsymbol{\Sigma}}_k(\lambda) = \frac{(1 - \lambda)N_k \boldsymbol{\Sigma}_k + (\lambda) \sum_{k=0}^1 N_k \boldsymbol{\Sigma}_k}{(1 - \lambda)N_k + (\lambda) \sum_{k=0}^1 N_k}.\quad (2)$$

Here, $\lambda \in [0, 1]$ is the shrinkage parameter that defines the similarity of two classes’ covariance. $\lambda = 1$ leads to equal covariance matrices for both classes, which turns RDA to linear discriminant analysis (LDA). The regularization procedure is as follows:

$$\hat{\boldsymbol{\Sigma}}_k(\lambda, \gamma) = (1 - \gamma)\hat{\boldsymbol{\Sigma}}_k(\lambda) + (\gamma)\frac{1}{p} \text{tr}[\hat{\boldsymbol{\Sigma}}_k(\lambda)] \mathbf{I}_p.\quad (3)$$

$\text{tr}[\cdot]$ is the trace operator, \mathbf{I}_p is a $p \times p$ identity matrix, and $\gamma \in [0, 1]$ is the regularization parameter, which determines the circularity of the covariance matrix.

Correspondingly, the discriminant score function defined as

$$d_{\text{RDA}}(\mathbf{x}) = \log \frac{f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}_1, \hat{\boldsymbol{\Sigma}}_1(\lambda, \gamma)) \hat{\pi}_1}{f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}_0, \hat{\boldsymbol{\Sigma}}_0(\lambda, \gamma)) \hat{\pi}_0}\quad (4)$$

where $f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the Gaussian probability density function when $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\hat{\pi}_k$ is the prior probability of class k . In our system, we use $\hat{\pi}_1 = \hat{\pi}_0$. To find the class conditional probability distributions of RDA scores, we use kernel density estimation (KDE) [16]. Each class conditional KDE is calculated over the RDA scores of EEG evidence recorded for the representative trials of that class in the calibration data set. Finally, the conditional probability density function for each class is defined as

$$\begin{aligned}f(\mathbf{x} = \mathbf{y} | l = k) &= \\ f_{\text{KDE}}(d_{\text{RDA}}(\mathbf{x}) = d_{\text{RDA}}(\mathbf{y}) | l = k) &= \\ \frac{1}{N_k} \sum_{i=1}^N \mathcal{K}_{h_k}(d_{\text{RDA}}(\mathbf{x}_i), d_{\text{RDA}}(\mathbf{y})) \delta_{l_i, k}.\end{aligned}\quad (5)$$

Here, $\mathcal{K}_{h_k}(\cdot, \cdot)$ is a suitable kernel function with bandwidth h_k . A Gaussian kernel is used in our system, and accordingly the

²The Gaussian distribution assumption here is a direct consequence of the assumption that filtered EEG is a Gaussian random process.

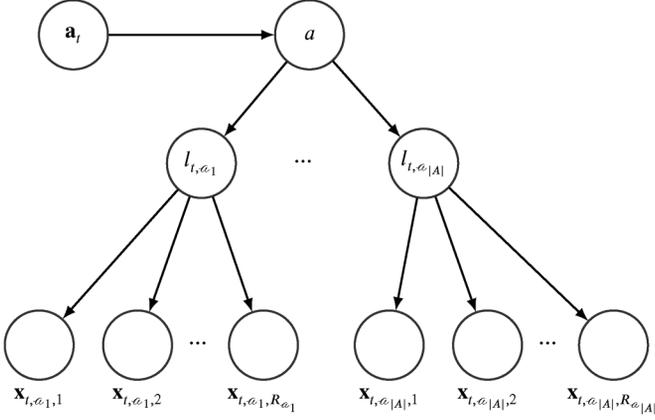


Fig. 2. Probabilistic graphical model of the fusion rule.

kernel bandwidth h_k for each class is calculated using the Silverman rule of thumb [26] over the RDA scores for the corresponding class.

2) *Language Model*: The system utilizes a letter n-gram model in an iterative Bayesian framework to increase the typing speed by prioritizing the symbols to be presented in each sequence and by providing a prior context for intent detection. A letter n-gram model estimates the conditional probability of every letter in the alphabet based on $n - 1$ previously typed letters in a Markov model framework [27].

Therefore, in a letter n-gram model, the conditional probability of each character, according to the Bayes rule, is given by

$$p(a_t = a | \mathbf{a}_t = \mathbf{a}) = \frac{p(a_t = a, \mathbf{a}_t = \mathbf{a})}{p(\mathbf{a}_t = \mathbf{a})} \quad (6)$$

where a_t is the symbol (yet) to be typed at epoch t and \mathbf{a}_t is the string of previously written $n - 1$ symbols. In our system, we use a 6-gram letter model, which is trained on the *New York Times* portion of the English Gigaword corpus [27].

3) *Fusion*: Assume \mathbf{x}_{t,r,a_i} represents the EEG feature vector of a trial, which contains $a_i \in \mathcal{A}$, at repetition $r \in \{1, 2, \dots, R_{a_i}\}$ in epoch t where R_{a_i} represents the total number of repetitions of trials containing the character a_i in the same epoch. Moreover, define l_{t,a_i} as the class label for $a_i \in \mathcal{A}$ in epoch t . The probabilistic graphical model that we use for fusion is shown in Fig. 2.

Let $\mathcal{X}_{t,a_i} = [\mathbf{x}_{t,1,a_i}, \mathbf{x}_{t,2,a_i}, \dots, \mathbf{x}_{t,R_{a_i},a_i}]$ represent a $(p \times R_{a_i})$ matrix of observed EEG feature vectors in epoch t . Here, p is the length of each feature vector. Accordingly, assume $\mathcal{X}_t = [\mathcal{X}_{t,a_1}, \mathcal{X}_{t,a_2}, \dots, \mathcal{X}_{t,a_{|A|}}]$ is a $(p \times N)$ matrix, where N is the number of total flashes in epoch t . Define \mathcal{X} as a possible outcome for matrix \mathcal{X}_t . Using Bayes' rule, we can define the posterior probability conditioned on the prior typed text and the observed EEG feature vectors as

$$Q = p(a_t = a | \mathcal{X}_t = \mathcal{X}, \mathbf{a}_t = \mathbf{a}) \propto p(\mathcal{X}_t = \mathcal{X}, \mathbf{a}_t = \mathbf{a} | a_t = a) P(a_t = a). \quad (7)$$

³Lower levels consist of copying phrases that have letters that are assigned high probabilities by the language model. As the level increases, the language model probabilities become increasingly adversarial. Level 3 is neutral on average.

Using the proposed graphical model, given the intended symbol a , the EEG evidence and previously typed text are conditionally independent. Moreover, given a , the EEG evidences for each trial $\mathbf{x}_{t,a_i,1}, \mathbf{x}_{t,a_i,2}, \dots, \mathbf{x}_{t,a_i,R_{a_i}}$ are independent, as follows:

$$Q \propto \left(\prod_{a_i \in \mathcal{A}} \prod_{r=1}^{R_{a_i}} f(\mathbf{x}_{t,a_i,r} = \mathbf{x}_{a_i,r} | a_t = a) \right) P(a_t = a | \mathbf{a}_t = \mathbf{a}). \quad (8)$$

$\mathbf{x}_{a_i,r}$ is the possible EEG evidence for r th repetition of character a_i . Also for given a , l_{t,a_i} s are deterministically defined. With this assumption, (8) can be simplified as

$$Q \propto \left(\prod_{r=1}^{R_a} \frac{f(\mathbf{x}_{t,a,r} = \mathbf{x}_{a,r} | l_{t,a} = 1)}{f(\mathbf{x}_{t,a,r} = \mathbf{x}_{a,r} | l_{t,a} = 0)} \right) P(a_t = a | \mathbf{a}_t = \mathbf{a}). \quad (9)$$

At the end of each sequence, $p(a_t = a | \mathcal{X}_t = \mathcal{X}, \mathbf{a}_t = \mathbf{a})$ is calculated for all the symbols; if the maximum of these posterior probabilities is higher than a predefined confidence threshold, a decision to type the corresponding symbol is made. Otherwise, sequences are repeated until the required confidence level is reached. If the confidence level is not reached in a predefined maximum number of repetitions bound for sequences, the symbol with the maximum *a posteriori* probability is chosen as the desired symbol.

D. System Operation Modes

The developed typing interface can currently be utilized in four different modes.

- i) *Calibration mode*: During calibration, the users are asked to attend to predefined target symbols within randomly ordered sequences to record labeled EEG data. The data acquired in this mode are then used in the estimation of classifier parameters to be used in other system operation modes. The shrinkage and regularization parameters are optimized during calibration using k-fold cross-validation to maximize area under the ROC curve.
- ii) *Copy phrase task mode*: In this task, the users are given a set of predefined phrases. Each phrase includes a missing word and the users are asked to complete these words. This task is designed to assess the system and/or user performance in terms of speed and accuracy in the presence of a language model.
- iii) *Mastery task mode*: Users are trained to use the system in this mode. It is similar to the copy phrase task mode in that the users are asked to type a set of predefined phrases. In contrast, the phrases used in this task have been carefully selected and divided into five difficulty levels based on their predictability by the language model. As the user completes the phrases in a level, the task continues with the next level with more difficult sentences.³
- iv) *Free spelling mode*. This mode allows the users to type their desired text.
- v) *Simulation mode*: In this mode, the copy phrase task is completed using samples drawn from the KDE of class conditional EEG feature distributions as computed in (5). These samples simulate EEG evidence and are fused with the language model probabilities for decision making as in regular operation [16]. Probability of completing the

task and expected task completion durations are reported as estimated performance measures using Monte Carlo simulations.

In this paper, we use all modes of the system for the following experiments, except free spelling.

III. EXPERIMENTAL RESULTS

A. Experiment

In this study, we assess the system performance in three presentation scenarios:

- 1) 4×7 matrix row and column presentation (RCP) paradigm;
- 2) 4×7 matrix single-character presentation (SCP) paradigm;
- 3) rapid serial visual presentation (RSVP) paradigm.

The comparison is based on three dependent variables: signal quality, system accuracy, and typing speed. Following a group-based analysis, we utilize paired t-tests to determine if the system performance varies significantly due to changes in the presentation paradigm or inter-trial interval (ITI) values. In addition, we perform paired t-tests within each user to assess the variations in P300 responses due to different ITIs.

Twelve healthy volunteers, nine males and three females, between the ages of 24 and 38 years, consented to participate in this study, which is conducted following an IRB-approved protocol. Each user participated in three sessions, each session on a different day with the various presentation paradigms. It is possible for a participant to gradually obtain skills to handle the system more efficiently, thereby introducing learning effects from session to session. To control for this effect, we relied on quasi-randomization; we distributed the presentation paradigms over the experimental sessions such that the number of users who attended a session with a specific presentation paradigm on a specific session order is kept the same (balanced). Every session that a user attended included calibration tasks with four different ITI values of {200, 150, 100, 85} ms. These values are chosen to be compatible with a 60-Hz monitor refresh rate and cover the range of possible optimum inter-trial durations. To account for the effect of user fatigue on typing performance, we randomized the order of ITI values for each presentation scenario and among all users. We used a duty cycle of 75% for each flash.

After calibration, each session proceeded with the mastery task [28] followed by the copy phrase task with eight sentences. We use a level 1 mastery task to familiarize the users with the copy phrase task. To prevent long sessions, the system marks a phrase as unsuccessful if more than four wrong letter selections occur in a row, and the next phrase is presented to the user.

B. Results

1) *Signal Quality*: In their work, Sellers *et al.* show that ITI effectively modifies the shape of the P300 response [29]. To investigate the effect of ITI on the P300 response, we analyzed the signal quality for every presentation scheme and ITI combination using the calibration data collected for different ITI values. For such combinations, we computed the area under the

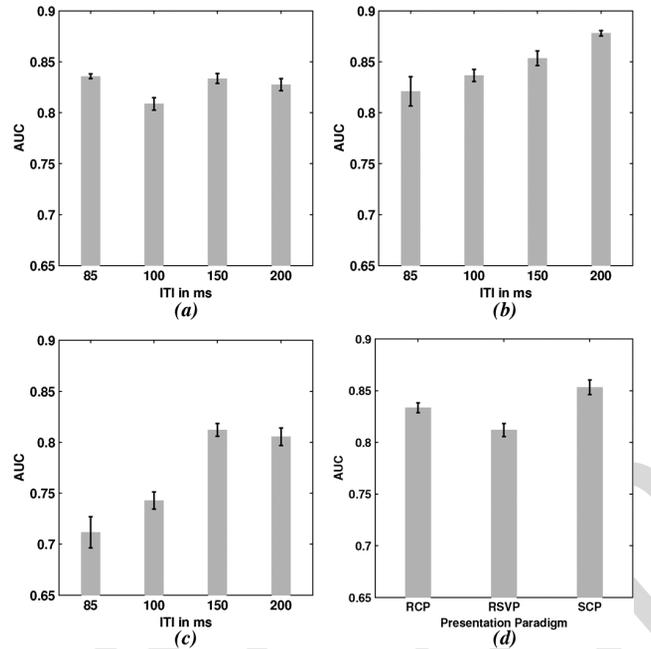


Fig. 3. Bar charts of average AUC with error bars. (a), (b), and (c) demonstrate the accuracy statistics for each ITI, respectively, for RCP, SCP, and RSVP paradigms. (d) reports the AUC statistics for different presentation paradigms at ITI = 150 ms. (a) ITI comparison in RCP; (b) ITI comparison in SCP; (c) ITI comparison in RSVP; and (d) paradigm comparison at ITI = 150 ms.

TABLE I
HYPOTHESIS TESTING RESULTS BETWEEN DIFFERENT ITIS WITHIN EACH PARADIGM. THE NULL HYPOTHESIS IS THAT THE EXPECTED AUC DIFFERENCE OF THE TWO CONSIDERED ITIS IS ZERO. HERE, WE USED $\alpha = 0.05$

Paired t-test results between different ITIs in each paradigm			
ITI_1 v.s. ITI_2	<i>P</i> -values in RCP	<i>P</i> -values in SCP	<i>P</i> -values in RSVP
85 v.s. 100 ms	0.068	0.593	0.157
85 v.s. 150 ms	0.803	0.927	0.001
85 v.s. 200 ms	0.550	0.240	0.009
100 v.s. 150 ms	0.075	0.673	0.0008
100 v.s. 200 ms	0.053	0.053	0.027
150 v.s. 200 ms	0.570	0.236	0.693

curve (AUC) for the ROC as the classification accuracy measure. Within each presentation paradigm, we applied a paired t-test over these accuracy values. The results are reported in Fig. 3(a)–(c) and Table I for each paradigm. The three subfigures correspond to different presentation paradigms, and in each subfigure, the average accuracies for different ITIs are presented using bar-graphs with error bars. Table I summarizes the paired t-test results between every ITI pair for each presentation paradigm.

From Table I, we observe that the group-based hypothesis testing does not show significant variations among classification accuracies due to changes in ITI values for the RCP paradigm. The results also suggest that the ITI value of 85 ms is the best candidate for matrix RCP paradigm. This ITI offers shorter sequence times and a consistent higher average AUC (averaged across users) as shown in Fig. 3(a). Our observations suggest that ERP responses in the RCP paradigm are more robust to the changes in ITI values.

TABLE II
MULTIVARIATE PAIRED T-TEST RESULTS OF P300 PEAK VALUE FOR EACH SUBJECT AMONG DIFFERENT ITIS WITHIN EACH PRESENTATION PARADIGM. THE NULL HYPOTHESIS IS THAT THE EXPECTED AUC DIFFERENCE OF THE TWO CONSIDERED ITIS IS ZERO WITHIN EACH PARADIGM. HERE, WE USED $\alpha = 0.05$, I.E. $H = 1$ IF $P < 0.05$ AND $H = 0$ OTHERWISE

Users	U1		U2		U3		U4		U5		U6		U7		U8		U9		U10		U11		U12	
RCP	H	P																						
85 v.s. 100 ms	0	0.18	0	0.09	0	0.19	0	0.27	1	0.01	0	0.85	0	0.16	1	0	1	0.02	0	0.16	1	0.02	1	0.04
85 v.s. 150 ms	0	0.61	1	0	1	0.03	0	0.89	1	0	0	0.28	1	0.01	1	0	1	0	0	0.32	0	0.18	0	0.73
85 v.s. 200 ms	1	0.02	0	0.27	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0.03	1	0.02	0	0.79
100 v.s. 150 ms	0	0.13	1	0	1	0.01	0	0.25	1	0	0	0.31	1	0.01	1	0	1	0	1	0.03	0	0.16	0	0.11
100 v.s. 200 ms	0	0.07	1	0	1	0	1	0	1	0	1	0	1	0	0	0.06	1	0	1	0	1	0	0	0.11
150 v.s. 200 ms	1	0	0	0.15	1	0	1	0	1	0	1	0	1	0	1	0	0	0.25	0	0.14	1	0	0	0.11
SCP	H	P																						
85 v.s. 100 ms	0	0.35	0	0.13	1	0	0	0.31	0	0.88	0	0.1	1	0	1	0	0	0.75	1	0.04	1	0	0	0.08
85 v.s. 150 ms	0	0.47	1	0	0	0.95	0	0.15	0	0.72	0	0.11	1	0	1	0	0	0.33	1	0.03	1	0	0	0.06
85 v.s. 200 ms	0	0.18	1	0	1	0	0	0.33	1	0	1	0	1	0	1	0	1	0.03	0	0.16	1	0	0	0.08
100 v.s. 150 ms	0	0.05	1	0	1	0	1	0.05	0	0.85	1	0.04	1	0.01	1	0	0	0.38	0	0.29	0	0.93	0	0.28
100 v.s. 200 ms	1	0.02	1	0	1	0	0	0.55	0	0.1	1	0	1	0	1	0	0	0.47	0	0.39	0	0.09	1	0.03
150 v.s. 200 ms	1	0	1	0	0	0.36	1	0.04	1	0.01	1	0	1	0	1	0	1	0.05	0	0.18	0	0.48	0	0.2
RSVP	H	P																						
85 v.s. 100 ms	0	0.06	0	0.69	0	0.1	0	0.38	1	0.02	1	0	0	0.42	0	0.14	0	0.57	0	0.64	0	0.24	0	0.07
85 v.s. 150 ms	0	0.63	0	0.09	1	0.02	0	0.07	0	0.36	1	0	1	0	1	0.04	1	0.05	0	0.19	0	0.37	0	0.05
85 v.s. 200 ms	1	0	0	0.15	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0.08	0	0.09
100 v.s. 150 ms	0	0.16	1	0.01	0	0.08	1	0.01	0	0.21	1	0	0	0.12	1	0.01	0	0.1	0	0.13	0	0.24	0	0.82
100 v.s. 200 ms	1	0	0	0.28	1	0	1	0	0	0.06	1	0	1	0	1	0	1	0.01	1	0.02	0	0.93	0	0.37
150 v.s. 200 ms	1	0	0	0.2	1	0	1	0	1	0.03	1	0	1	0	1	0	1	0	0	0.45	0	0.1	0	0.62

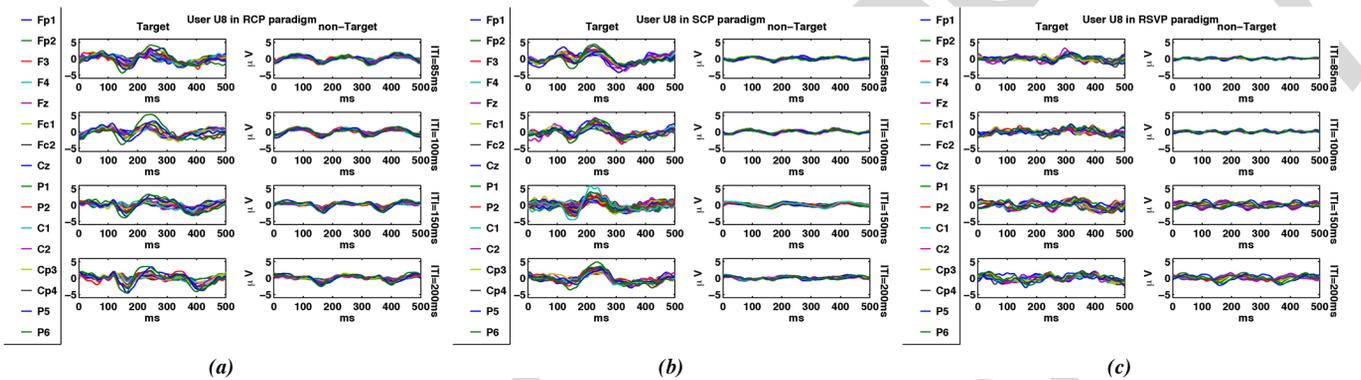


Fig. 4. Average ERP response to target and nontarget stimuli, for each presentation paradigm and ITI pairs for user “U8”. From top to bottom, the ITI is increasing monotonically. (a) RCP; (b) SCP; and (c) RSVP.

The SCP paradigm with an ITI of 200 ms demonstrates the highest average AUC with the lowest variance [Fig. 3(b)]. Although average AUCs across users show an increasing trend from ITI of 85 to 200 ms, pairwise comparisons between different ITIs do not show statistically significant variations in population AUCs (see Table I).

Generally, in matrix-based presentation paradigms, variations in ITI values seem to have a negligible effect on system AUC. The usage of smaller ITIs might be preferable due to a possible decrease in the sequence length, which might improve the speed of the typing interface. Moreover, it might be viable to optimize the matrix subset flashes based on context information to have shorter sequence lengths and higher classification confidence by increasing the number of flashes of probable characters, which can lead to faster target detections.

On the other hand, accuracies with the RSVP paradigm tend to be more sensitive to changes in ITI values (as shown in Table I). The most significant increase in AUC happens from ITI = 100 ms to ITI = 150 ms. The accuracy deviations between ITI = 85 ms and ITI = 100 ms and also between ITI = 150 ms and ITI = 200 ms are not significant as reported in Table I. Consequently, among the ITI values tested with the RSVP paradigm, ITI = 150 ms is the best choice for system design, since the accuracies between ITIs of 150 and 200 ms do not significantly change while ITI = 150 ms provides

better speed. This is consistent with our previous work using RSVP for image search [14]. In contrast with matrix-based presentation paradigms, in the RSVP paradigm, users need to recognize the target symbols, which induces the weaker P300 signals, especially at lower ITIs, as shown in Fig. 3(c).

To investigate signal quality variations due to ITI changes in each presentation scheme, we extract the P300 peak values for every target stimulus at all channels per user, for different combinations of ITI values and presentation paradigms. In this process, we filter the EEG signal using a Gaussian low pass filter with ($\sigma = 5$ samples) to increase the signal-to-noise ratio (SNR). For each target trial, we define a (16×1)-dimensional feature vector with the i th element containing the peak value of the EEG at channel i in the time window [250, 350] ms after stimulus onset. For every user and presentation paradigm, we use these feature vectors in a multivariate paired t-test to investigate the P300 amplitude deviations across different ITI values. We report the results in Table II. Comparing the results for different paradigms, we do not observe a consistent change in P300 amplitude among different ITI values. For instance, variations in ITI can significantly change the P300 peak values of user U8 at every presentation paradigm, as illustrated in Fig. 4, while this is not true for user U12 (see Fig. 5). Consequently, to acquire the best performance, we recommend that optimum ITI be defined uniquely for each user.

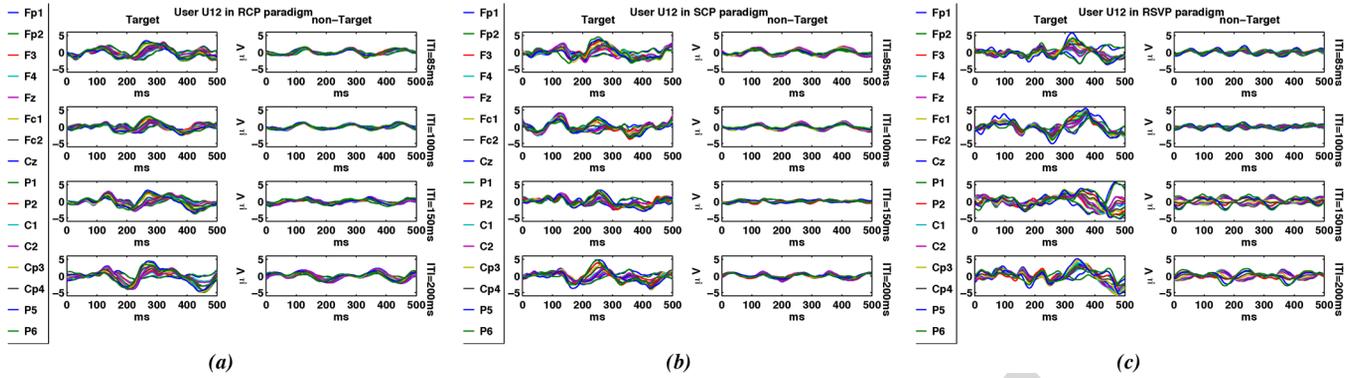


Fig. 5. Average ERP response to target and nontarget stimuli, for each presentation paradigm and ITI pairs for user “U12”. From top to bottom, the ITI is increasing monotonically. (a) RCP; (b) SCP; and (c) RSV.

TABLE III
HYPOTHESIS TESTING RESULTS BETWEEN DIFFERENT PRESENTATION PARADIGMS AT ITI = 150 ms. THE NULL HYPOTHESIS IS THAT THE EXPECTED AUC DIFFERENCE OF THE TWO CONSIDERED PARADIGMS IS ZERO. HERE, WE USED $\alpha = 0.05$

Paired t-test results between different presentation paradigms		
compaired paradigms	H_0 rejected	P -value
RCP v.s. RSV	No	0.362
RCP v.s. SCP	No	0.453
RSVP v.s. SCP	No	0.188

2) *System Accuracy Based on Presentation Paradigm*: We analyzed the changes in system classification accuracy across different presentation paradigms. Similar to the signal quality analysis, we employed AUC values as the measure of accuracy. We set ITI = 150 ms, which provides good performance for all paradigms and is close to the ITI value typically used for matrix based presentation paradigms (125 ms) [4]. We analyzed the changes in AUC values across different presentation paradigms using a paired t-test between different presentation paradigms. We report the results in Table III. These results do not illustrate significant changes due to different presentation paradigms. Moreover, in Fig. 3(d), we plot the AUC values averaged over all the users for different presentation schemes. This figure shows that the average AUC values in matrix-based paradigms are higher than in the RSV paradigm. However, the paired t-test outcomes do not confirm statistically significant separations among these average AUC values. Based on these results, we propose that the system accuracy might be more dependent on the user than the presentation paradigm at ITI = 150 ms.

In Fig. 6, we plot the channel by channel significance levels for the paired t-tests between different presentation paradigms. From this figure, we first observe that there is no statistically significant difference between different presentation paradigms. This result is consistent with the results that we report in Table III.

We also observe that there is no consistent electrode subset that shows significant difference among different presentation paradigms. We also plot the AUC values calculated using each channel separately, for each user, for different presentation paradigms, in Fig. 7. For a specific presentation paradigm, this

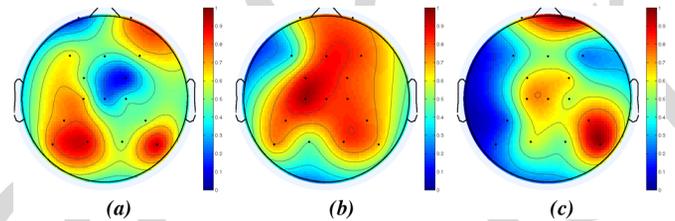


Fig. 6. Topography map of $(1 - p)$ resulting from paired t-tests for each channel’s AUC between each paradigm pair and across users for ITI = 150 ms. Here, red denotes $1 - p = 1$, and blue represents $1 - p = 0$. (a) RCP versus RSV; (b) RCP versus SCP; and (c) RSV versus SCP.

figure does not show a consistent region on the scalp, across different users, for high accuracy.⁴ Based on these results, we suggest that an optimum presentation paradigm is user dependent.

3) *Typing Speed*: We analyze the differences in the typing speed across different presentation paradigms employing the average number of sequences per target trial as the measure of speed-inverse (time spent per letter). Conventionally, in the RCP paradigm, during each sequence, all rows and columns are flashed once (which results in 11 flashes during a sequence in this study). On the other hand, the RSV paradigm has previously demonstrated almost optimized performance with eight trials in a sequence [15]. Accordingly, to keep the analysis equitable, we use the average number of sequences per target trial as the measure of time spent. We set ITI = 150 ms. In addition to the experimental results, we also perform 20 Monte Carlo simulations of the copy phrase task for every user under different presentation paradigms, using the corresponding calibration EEG data to generate simulated EEG evidence.

We report both the simulation and experimental results in Fig. 9. For different presentation paradigms, this figure shows the average number of sequences per target trial and task completion probabilities versus AUC values of different users. We observe that both minimum and maximum values of user AUCs are smaller in the RSV paradigm than the matrix-based presentation schemes. In the RCP paradigm, each symbol is represented twice in a sequence. Subsequently, the number of data

⁴One may need to optimize best electrode locations for each paradigm–user combination to maximize performance.

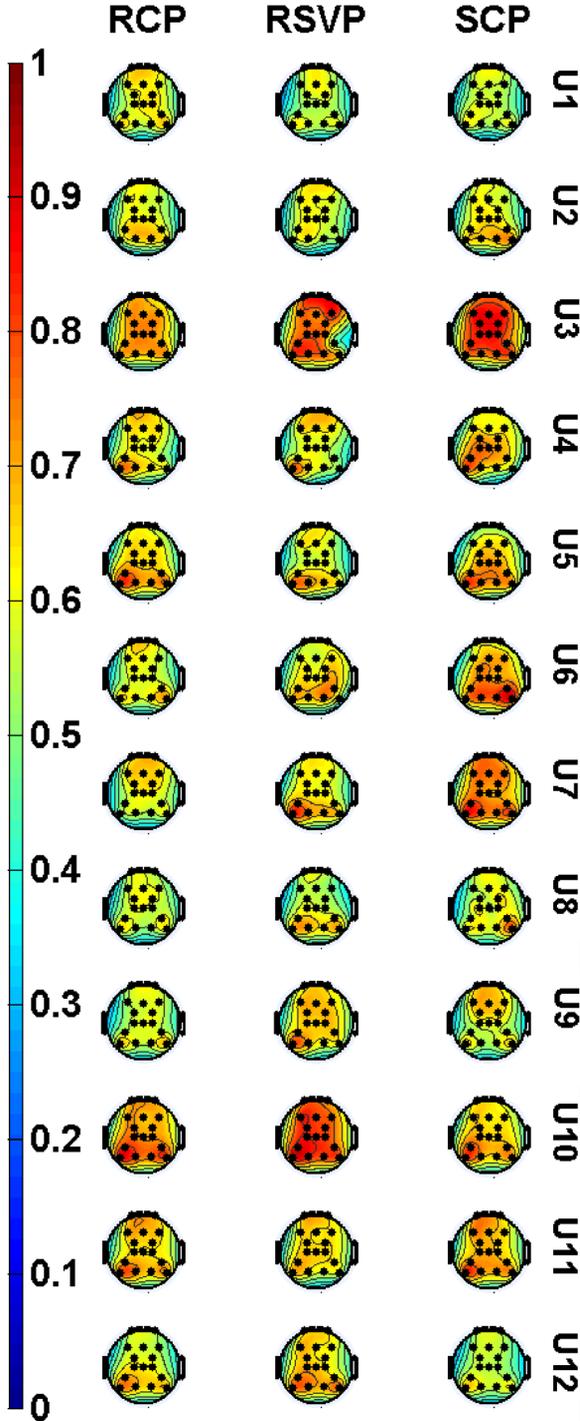


Fig. 7. Topography of channel based AUCs for each user at ITI = 150 ms.

points from the target class for recorded EEG during the calibration task is twice the other paradigms. This can lead to a more accurate estimation of classifier parameters, which then leads to smaller average numbers of sequences per target trial and higher task completion probability.

In general, actual typing performance in the SCP paradigm shows a behavior consistent with simulation results. In all paradigms, simulation results are reasonably predictive of the actual typing task statistics for larger AUCs. Mismatch between simulation results and actual user typing speeds is

more frequent in the RSVP paradigm. This maybe because the user AUCs are generally lower for the RSVP paradigm, since the requirement to recognize the target symbol might impose more cognitive load and require more attention from the user.⁵ However, some participants still show faster typing performance with RSVP than the matrix-based presentation schemes (see Table IV).

From this table, user U7 shows better typing performance when using the RSVP paradigm [see Fig. 9(a)–(c)], while users U3 and U9 spelled target phrases with a lower average number of sequences when using SCP and RCP paradigms [see Fig. 9(d)–(f) and (g)–(i), respectively]. Accordingly, the choice for the best presentation scheme should be user dependent.

4) *Effect of Language Model on Typing Duration:* We employ the simulation mode of the system to assess the effect of the language model on the (estimated) performance of each presentation paradigm. We preform $[AU: "perform"?]$ ten Monte Carlo simulations (of the copy phrase task) with and without the language model to estimate the typing speed under both conditions using calibration EEG data from each user. We represent the typing speed as the average number of sequences for correctly typing a character, N_{avg} . The results of N_{avg} shown in Figs. 10(a)–(c) indicate that the language model significantly improves the performance for all three presentation paradigms. This is seen in the form of reduced average sequence counts required to type a target symbol correctly, as well as reduced variance. That is, without a language model, the mean values of N_{avg} are larger for all the users, and the standard deviations of N_{avg} are larger for most of the participants. For RCP, users with lower AUC (larger sequence counts for the without-LM axis) seem to increasingly benefit from the assistance of the language model in this task [see Fig. 10(a)]. In the case of SCP [Fig. 10(b)] and RSVP [Fig. 10(c)], while the same trend is observed for high to moderately good AUCs, for users with the lowest several AUCs (appearing on the right-most side of their respective plots), the consistency of language model assistance is not as good as that in the case of RCP. This inconsistent behavior seems to occur due to low AUCs (for $AUC < .74$). This suggests that, for some users with low classification performance, we may need to collect more training samples in the calibration session for them to be able to benefit from the language model assistance.

IV. CONCLUSION

In this paper, we compared three different presentation paradigms: i) a 4×7 matrix row and column; ii) a 4×7 matrix single character; and iii) a rapid serial visual presentation, utilizing a language-model-assisted EEG-based letter-by-letter typing BCI. The underlying intent inference engine used tight fusion of language and EEG evidence, as described in earlier papers on the RSVP keyboard [5], [15], [16]. Twelve participants were recruited to use the system in four different ITIs of {85, 100, 150, 200} ms for each presentation scheme. The order of paradigm presentations for each session, and each user were

⁵This claim is mainly based on the users' feedback after each session. They described it as more challenging to spot the desired character in the RSVP paradigm, especially for the sessions with smaller ITIs.

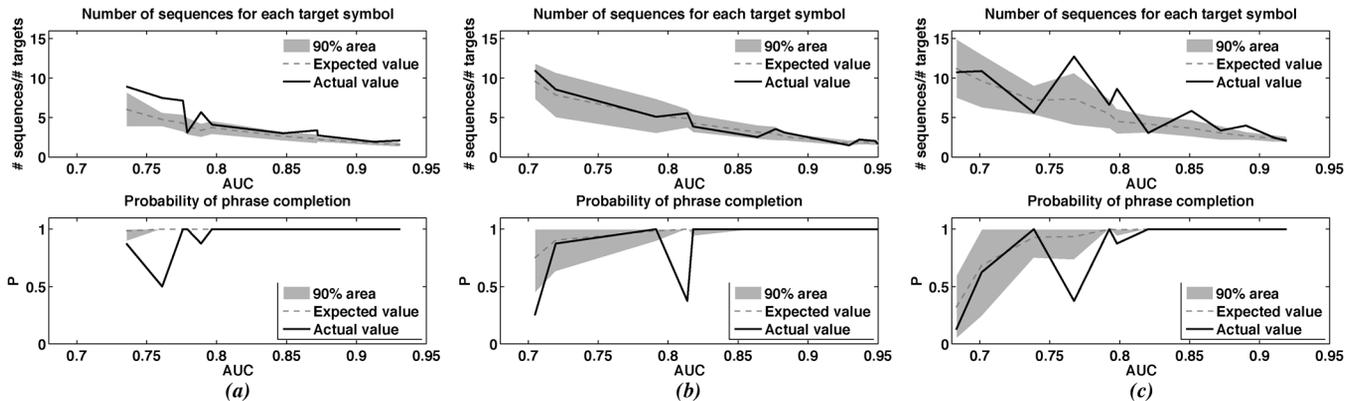


Fig. 8. Typing speed analysis results. Average number of sequences per (typed) target character (lower means faster typing) and probability of phrase completion (higher means more accuracy) are shown. Simulation results are used to define the shaded 90% confidence area shown. The dashed line shows the expected value from simulation for each variable, and the solid line shows actual typing outcomes in a single experimental run that follows. (a) RCP; (b) SCP; and (c) RSVP.

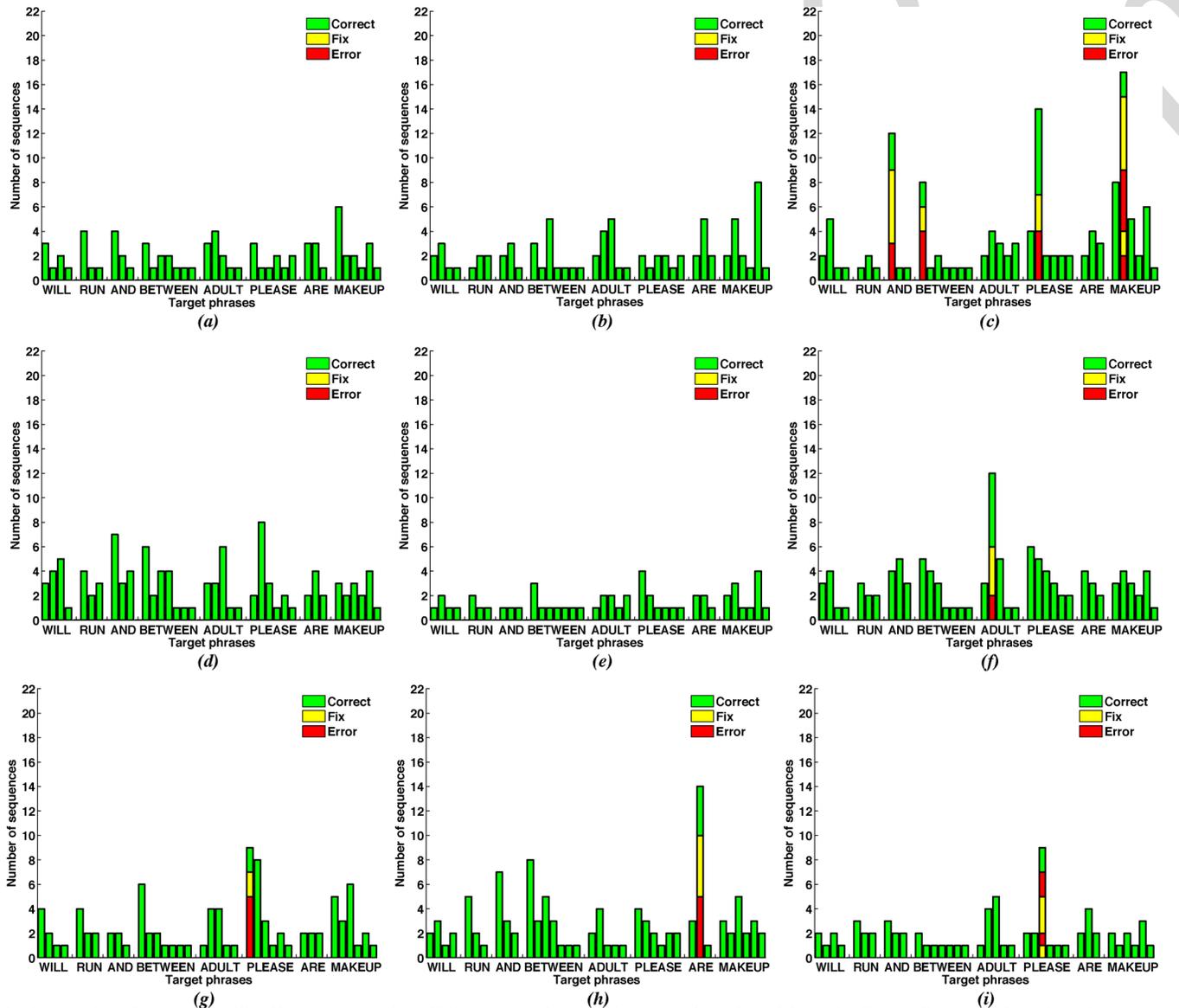


Fig. 9. Number of sequences utilized by users U7, U3, and U9 to type each target character using RSVP, SCP, and RCP paradigms. Red bars show the sequence counts for epochs that typed a wrong character, and yellow bars show the number of sequences used to fix the error before typing the correct target. Green bars show the number of sequences in epochs that resulted in correct selection of target symbols (lower means faster typing). (a) U7, RSVP; (b) U7, SCP; (c) U7, RCP; (d) U3, RSVP; (e) U3, SCP; (f) U3, RCP; (g) U9, RSVP; (h) U9, SCP; and (i) U9, RCP.

quasi-randomized. The same classifier, language model, and fusion rule were used for all paradigms and ITI combinations.

Through this study, we illustrated that the best presentation paradigm and ITI combination among the ones presented in this

TABLE IV
 TYPING SPEED RESULTS FOR EACH USER AND PARADIGM COMBINATION. HERE, "AVERAGE \pm STANDARD DEVIATION"
 OF SEQUENCE COUNT PER TARGET (CORRECTLY TYPED) SYMBOL IS REPORTED

	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12
RSVP	3.98 \pm 2.7	12.74 \pm 5.74	3.05 \pm 0.69	10.74 \pm 3.35	5.6 \pm 1.46	3.35 \pm 1.48	2.04 \pm 0.35	8.64 \pm 5.15	2.44 \pm 0.76	10.89 \pm 4.3	6.59 \pm 3.76	5.84 \pm 3.87
SCP	1.29 \pm 0.26	10.96 \pm 4.9	1.48 \pm 0.32	3.55 \pm 1.78	5.54 \pm 1.97	3.85 \pm 2.04	2.21 \pm 0.62	5.08 \pm 2.44	3.1 \pm 1.36	2.52 \pm 0.95	2.03 \pm 0.42	8.55 \pm 4.26
RCP	2.73 \pm 0.85	5.67 \pm 4.53	3.1 \pm 0.83	8.92 \pm 3.83	7.48 \pm 3.55	4.1 \pm 1.32	3.38 \pm 1.68	2.09 \pm 0.57	2.09 \pm 0.57	3.01 \pm 1.11	1.93 \pm 0.5	7.14 \pm 4.09

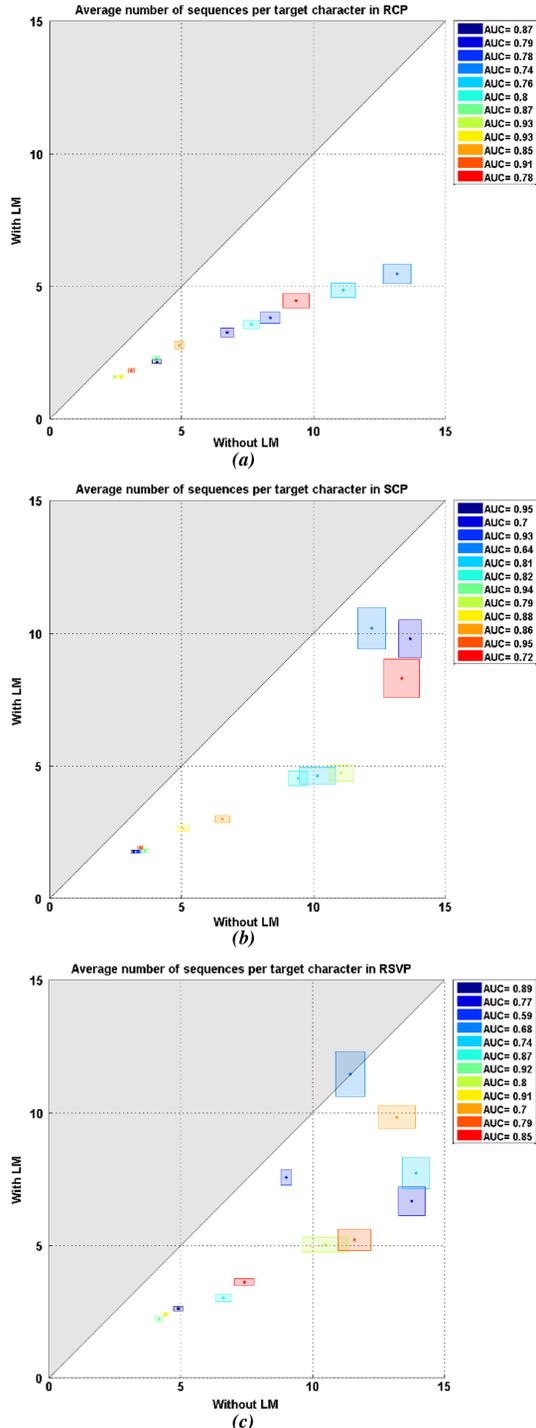


Fig. 10. Scatter plot of the average number of sequences for correctly typing a target character. The x -axis demonstrates the mean number of sequences per target character when no language model is used, y -axis represents the mean number of sequences required per target character while a 6-gram language model is utilized. Each point on the figure shows the average of the mean number of sequences per target from ten Monte Carlo simulations. Horizontal skewness of each box around a point is the standard deviation of the number of sequences per target character for typing while no language model was used, and the vertical skewness is the standard deviation in presence of the language model. (a) RCP; (b) SCP; and (c) RSVP.

study should be identified for each user individually to achieve the best performance. Also, we showed that the performance of the RSVP paradigm is comparable to matrix-based presentation paradigms with healthy users. Based on our results, we propose that BCI typing systems capable of employing multiple presentation schemes including both RSVP and matrix presentation paradigms are inevitable. This system, after individual clinical assessments, should be able to determine the best presentation option and the best ITI value for each user, according to user preferences, capabilities, EEG signal statistics, and simulations. Moreover, the length of the calibration session might need to be increased based on the classification performance for a user at each presentation paradigm.

A side product of this work is that we now have a unified BCI typing interface that has both RSVP and matrix presentation options along with a MAP intent inference engine that tightly fuses n -gram symbol and EEG evidence. It is an open vocabulary typing interface with the potential to be individualized by personal language models and the incorporation of supplementary physiological and behavioral evidence about intent, for instance via EMG or switches. Other open problems include improved signal models for more accurate performance simulations and run-time intent inference, optimized dynamic selection of stimulus subsets to be presented in each trial for the upcoming sequence, and rigorous field testing to compare RSVP and matrix presentation paradigms on potential user populations.

REFERENCES

- [1] M. Akcakaya, B. Peters, M. Moghadamfalahi, A. Mooney, U. Orhan, B. Oken, D. Erdogmus, and M. Fried-Oken, "Noninvasive brain computer interfaces for augmentative and alternative communication," *IEEE Rev. Biomed. Eng.*, vol. 7, no. 1, pp. 31–49, 2014.
- [2] C.-T. Lin, L.-W. Ko, M.-H. Chang, J.-R. Duann, J.-Y. Chen, T.-P. Su, and T.-P. Jung, "Review of wireless and wearable electroencephalogram systems and brain-computer interfaces—A mini-review," *Gerontology*, vol. 56, no. 1, pp. 112–119, 2010.
- [3] S. Moghimi, A. Kushki, A. M. Guerguerian, and T. Chau, "A review of EEG-based brain-computer interfaces as access pathways for individuals with severe disabilities," *Assistive Technol.: Official J. RESNA*, vol. 25, no. 2, pp. 99–110, 2012.
- [4] L. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, pp. 510–523, 1988.
- [5] U. Orhan, K. E. Hild, D. Erdogmus, B. Roark, B. Oken, and M. Fried-Oken, "RSVP keyboard: An EEG based typing interface," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 645–648.
- [6] E. Sellers, G. Schalk, and E. Donchin, "The P300 as a typing tool: Tests of brain computer interface with an ALS patient," *Psychophysiology*, vol. 40, p. 77, 2003.
- [7] E. W. Sellers and E. Donchin, "A P300-based brain-computer interface: Initial tests by ALS patients," *Clin. Neurophysiol.*, vol. 117, no. 3, pp. 538–548, 2006.
- [8] S. Sutton, M. Braren, J. Zubin, and E. John, "Evoked-potential correlates of stimulus uncertainty," *Science*, vol. 150, no. 3700, pp. 1187–1188, 1965.

- [9] C. Guan, M. Thulasidas, and J. Wu, "High performance P300 speller for brain-computer interface," in *Proc. IEEE Int. Workshop Biomed. Circuits Syst.*, 2004, pp. S3–5.
- [10] **[AU: Provide more information on type of source and where to find--ed.]** M. S. Treder and B. Blankertz, "Research (c) Overt Attention and Visual Speller Design in an ERP-Based Brain-Computer Interface," 2010.
- [11] Y. Liu, Z. Zhou, and D. Hu, "Gaze independent brain-computer speller with covert visual search tasks," *Clin. Neurophysiol.*, vol. 122, no. 6, pp. 1127–1136, 2011.
- [12] M. S. Treder, N. M. Schmidt, and B. Blankertz, "Gaze-independent brain-computer interfaces based on covert attention and feature attention," *J. Neural Eng.*, vol. 8, no. 6, p. 066003, 2011.
- [13] S. Chennu, A. Alsufyani, M. Filetti, A. M. Owen, and H. Bowman, "The cost of space independence in P300-BCI spellers," *J. Neuroeng. Rehab.*, vol. 10, no. 82, pp. 1–13, 2013.
- [14] Y. Huang, "Event-related potentials in electroencephalography: Characteristics and single-trial detection for rapid object search," Ph.D. dissertation, Oregon Health & Sci. Univ., Portland, OR, USA, 2010.
- [15] U. Orhan, D. Erdogmus, B. Roark, B. Oken, S. Purwar, K. E. Hild, A. Fowler, and M. Fried-Oken, "Improved accuracy using recursive Bayesian estimation based language model fusion in ERP-based BCI typing systems," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2012, pp. 2497–2500.
- [16] U. Orhan, D. Erdogmus, B. Roark, B. Oken, and M. Fried-Oken, "Offline analysis of context contribution to ERP-based typing BCI performance," *J. Neural Eng.*, vol. 10, no. 6, p. 066003, 2013.
- [17] L. Acqualagna, M. S. Treder, M. Schreuder, and B. Blankertz, "A novel brain-computer interface based on the rapid serial visual presentation paradigm," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2010, pp. 2686–2689.
- [18] L. Acqualagna and B. Blankertz, "Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP)," *Clin. Neurophysiol.*, vol. 124, no. 5, pp. 901–908, 2013.
- [19] D. B. Ryan, G. Frye, G. Townsend, D. Berry, S. Mesa-G, N. A. Gates, and E. W. Sellers, "Predictive spelling with a P300-based brain-computer interface: Increasing the rate of communication," *Int. J. Human-Comput. Interact.*, vol. 27, no. 1, pp. 69–84, 2010.
- [20] T. Kaufmann, S. Völker, L. Gunesch, and A. Kübler, "Spelling is just a click away—a user-centered brain-computer interface including auto-calibration and predictive text entry," *Frontiers in Neurosci.*, vol. 6, 2012.
- [21] S. Lee and H.-S. Lim, "Predicting text entry for brain-computer interface," in *Future Information Technology*. New York, NY, USA: Springer, 2011, pp. 309–312.
- [22] E. Samizo, T. Yoshikawa, and T. Furuhashi, "A study on application of RB-ARQ considering probability of occurrence and transition probability for P300 speller," in *Foundations of Augmented Cognition*. New York, NY, USA: Springer, 2013, pp. 727–733.
- [23] W. Speier, C. Arnold, J. Lu, R. K. Taira, and N. Pouratian, "Natural language processing with dynamic classification improves P300 speller accuracy and bit rate," *J. Neural Eng.*, vol. 9, no. 1, p. 016004, 2012.
- [24] C. Ulas and M. Cetin, "Incorporation of a language model into a brain computer interface based speller through HMMs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013, pp. 1138–1142.
- [25] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Statist. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.
- [26] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL, USA: CRC Press, 1986, vol. 26.
- [27] B. Roark, J. D. Villiers, C. Gibbons, and M. Fried-Oken, "Scanning methods and language modeling for binary switch typing," in *Proc. NAACL HLT Workshop Speech Lang. Process. Assist. Technol.*, 2010, pp. 28–36, Association for Computational Linguistics.
- [28] B. S. Oken, U. Orhan, B. Roark, D. Erdogmus, A. Fowler, A. Mooney, B. Peters, M. Miller, and M. B. Fried-Oken, "Brain-computer interface with language model-electroencephalography fusion for locked-in syndrome," *Neurorehabilitat. Neural Repair*, p. 1545968313516867, 2013.
- [29] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A P300 event-related potential brain-computer interface (bci): The effects of matrix size and inter stimulus interval on performance," *Biologic. Psychol.*, vol. 73, no. 3, pp. 242–252, 2006.



Mohammad Moghadamfalahi (S'14) received the B.Sc. degree in electrical engineering from Amir-kabir University (Tehran Polytechnics), Tehran, Iran, in 2008. Since January 2012, he has been working towards the Ph.D. student in Electrical and Computer Engineering Department of Northeastern University, Boston, MA, USA.

From 2008 to 2011, he worked at the Mobile Communication Company of Iran (MCCI). Currently, he is a Research Assistant at Cognitive System Laboratory (CSL), Northeastern University. His fields of interests are brain-computer interfaces, machine learning, and statistical signal processing.



Umut Orhan (M'09) received the B.S. degree in electrical and electronics engineering from Bilkent University, Turkey, in 2009 and the Ph.D. degree in electrical and computer engineering from Northeastern University, Boston, MA, USA, in 2014.

He currently contracts to Honeywell Laboratories as a Research and Development Engineer. His current research interests are signal processing and machine learning for biomedical and neural applications.

Dr. Orhan is a member of Eta Kappa Nu.



Murat Akcakaya (M'11) received the B.Sc. degree from the Electrical and Electronics Engineering Department of Middle East Technical University, Ankara, Turkey, in 2005 and the M.Sc. and the Ph.D. degrees in electrical engineering from Washington University in St. Louis, in May and December 2010, respectively.

He is currently an Assistant Professor at the University of Pittsburg, Pittsburg, PA, USA. His research interests include statistical signal processing and machine learning with applications to noninvasive electroencephalography (EEG) based brain-computer interface (BCI) systems, array signal processing, and physiological signal analysis for health informatics.

Dr. Akcakaya was the winner of the student paper contest awards at the 2010 IEEE Radar Conference; the 2010 IEEE Waveform Diversity and Design Conference; and the 2010 Asilomar Conference on Signals, Systems and Computers.



Hooman Nezamfar (S'14) received the B.S. degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 2004 and the M.S. degree in communication systems from the Iran University of Science and Technology, Tehran, Iran, in 2008. He has been working towards the Ph.D. degree in the Electrical and Computer Engineering Department at Northeastern University, Boston, MA, USA, since 2010.

He is also a member of Cognitive Systems Laboratory (CSL) at Northeastern University. His research has been on the brain-computer interfaces, probabilistic modeling, and embedded systems with a special focus on the steady-state visually evoked potentials (SSVEPs).



Melanie Fried-Oken received the B.A. degree from the University of Rochester, Rochester, NY, USA; the M.A. degree in speech-language pathology from Northwestern University, Boston, MA, USA; and the Ph.D. degree in psycholinguistics from Boston University, Boston, MA, USA, in 1984. **[AU: Provide years in which all degrees were received--ed.]**

She is an international clinical researcher in the Augmentative and Alternative Communication (AAC) **[AU: At OHSU?]**, where she provides expertise about assistive technology for persons who cannot use speech or writing for expres-

sion. She is a Professor of Neurology, Pediatrics, Biomedical Engineering and Otolaryngology at the Oregon Health & Science University (OHSU), Portland, OR, USA; Director of OHSU Assistive Technology Program, and clinical speech-language pathologist. She has written over 30 peer-reviewed articles and six chapters and has given over 200 national and international lectures on AAC for persons with various medical conditions. She has been PI on 16 federal grants and co-investigator on six federal grants to research communication technology and behavior for persons with developmental disabilities, neurodegenerative diseases, and the normally aging population.



Deniz Erdogmus received the B.S. degree in electric engineering and mathematics and the M.S. degree in electric engineering from the Middle East Technical University, Ankara, Turkey, in 1997 and 1999, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2002.

He was a postdoctoral researcher at the University of Florida until 2004. He was subsequently an Assistant Professor of BME at the Oregon Health and Science University (2004–2008). Since then, he has been with Northeastern University, Boston, MA, USA, where he is currently an Associate Professor of electrical and computer engineering. His research focuses on statistical signal processing and machine learning with applications to biomedical signal/image processing and cyberhuman systems.

Dr. Erdogmus has served as Associate Editor and technical committee member for various journals and conferences.

IEEE Pre-proof
Web Version

Language-Model Assisted Brain Computer Interface for Typing: A Comparison of Matrix and Rapid Serial Visual Presentation

Mohammad Moghadamfalahi, *Student Member, IEEE*, Umut Orhan, *Member, IEEE*,
Murat Akcakaya, *Member, IEEE*, Hooman Nezamfar, *Student Member, IEEE*,
Melanie Fried-Oken, and Deniz Erdogmus, *Senior Member, IEEE*

Abstract—Noninvasive electroencephalography (EEG)-based brain-computer interfaces (BCIs) popularly utilize event-related potential (ERP) for intent detection. Specifically, for EEG-based BCI typing systems, different symbol presentation paradigms have been utilized to induce ERPs. In this manuscript, through an experimental study, we assess the speed, recorded signal quality, and system accuracy of a language-model-assisted BCI typing system using three different presentation paradigms: a 4×7 matrix paradigm of a 28-character alphabet with row-column presentation (RCP) and single-character presentation (SCP), and rapid serial visual presentation (RSVP) of the same. Our analyses show that signal quality and classification accuracy are comparable between the two visual stimulus presentation paradigms. In addition, we observe that while the matrix-based paradigm can be generally employed with lower inter-trial-interval (ITI) values, the best presentation paradigm and ITI value configuration is user dependent. This potentially warrants offering both presentation paradigms and variable ITI options to users of BCI typing systems.

Index Terms—Brain-computer interface, event-related potential, matrix speller, P300, RSVP keyboard.

I. INTRODUCTION

NONINVASIVE brain-computer interfaces (BCIs), specifically those based on electroencephalography (EEG), have become popular to safely enable people with severe motor and speech impairments to communicate with their social networks and interact with their environments [1]–[3]. Typing is one of the most widely explored applications for EEG-based BCI systems [1]. Event-related potentials (ERPs), specifically

the P300 component of these EEG responses, are commonly exploited by such typing interfaces for user intent detection [4]–[7].

The pioneering work of Farwell and Donchin showed that ERPs containing the P300 response can be used to design EEG-based BCI typing systems [4]. They distributed 36 symbols consisting of the 26 letters in the English alphabet and 10 numerical digits across a 6×6 matrix. The rows and columns of the matrix are flashed in a random fashion to generate an oddball paradigm such that when the row or column that includes the symbol that the user intends to select is flashed, an ERP containing the P300 component is elicited. This ERP is then used for target symbol detection. P300 is a positive deflection in the scalp voltage with a typical latency around 300 ms after the onset of an infrequent target stimuli [8].

Despite the practice being the benchmark in matrix spellers, flashing rows and columns for the presentation of a symbol may result in poor P300 signal quality, and a single character flashing paradigm enhances the P300 response [9]. Studies also demonstrated that the performance of a BCI typing system that employs a matrix presentation paradigm depends on the gaze of the user [10], [11]. Many potential users from the target population, unfortunately, lack precise gaze control, and for these users, it is anticipated that matrix paradigms will suffer from reduced performance. To overcome this dependency in BCI typing systems, different presentation schemes have been explored and shown to have comparable performances with the matrix presentation paradigm in terms of speed and accuracy [11]–[13]. Rapid serial visual presentation (RSVP) is one of these paradigms, in which symbols are presented sequentially in time, at a predefined fixed location on the screen and in a pseudorandom order [5], [14]–[18].

BCI typing systems can benefit greatly from a language model in order to enhance typing speed. A probabilistic language model can be employed to incorporate predictive word completion during the intent detection process [19]–[21], or to define a prior on potential target characters during the classification task [22]–[24]. Our system, the RSVP keyboard, originally developed based on the RSVP paradigm and now also featuring the matrix presentation paradigm, probabilistically fuses context evidence with physiological evidence to infer user intent. A symbol n-gram language model trained on a large corpus provides probabilities for each character in the

Manuscript received June 09, 2014; revised November 13, 2014; accepted February 04, 2015. This work was supported by NIH grant R01DC009834 and NSF grants CNS-1136027, IIS-1149570, SMA-0835976. The package, including the code and data associate with this paper, can be found at <https://repository.lib.neu.edu/collections/neu:rx913r029>.

M. Moghadamfalahi, U. Orhan, H. Nezamfar, and D. Erdogmus are with Northeastern University, Boston, MA, USA (e-mail: moghadam@ece.neu.edu; orhan@ece.neu.edu; nezamfar@ece.neu.edu; erdogmus@ece.neu.edu).

M. Akcakaya is with University of Pittsburgh, Pittsburgh, PA, USA.

M. Fried-Oken is with Oregon Health and Science University, Portland, OR, USA. **[AU: Please provide all missing postal codes--ed.]**

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNSRE.2015.2411574

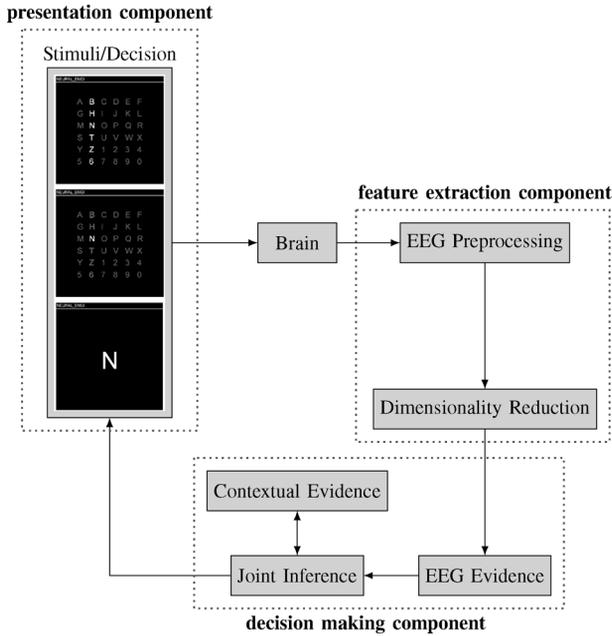


Fig. 1. The in-house BCI block diagram.

alphabet, which are fused tightly in a Bayesian fashion with EEG evidence [5], [15], [16].

In this paper, we utilize two different matrix schemes (row column flash and single symbol flash) and one RSVP scheme in a BCI typing interface and compare the differences in measured signal quality, typing speed, and accuracy. In a similar study, Chennu *et al.*, through an offline study, have shown that the classification accuracy is comparable between RSVP and matrix based paradigms, but without a language model the typing speed is relatively low while utilizing the RSVP paradigm [13]. In this study, we also compare the typing performance during online typing of both RSVP and matrix paradigms, using the aforementioned language-model-assisted BCI.

The contributions of this paper areas follows:

- 1) building a unified framework for different presentation paradigms that utilize EEG and language model evidence for joint decision making;
- 2) conducting real-time and offline comparisons among different presentation schemes;
- 3) analyzing the effect of different presentation paradigms on the EEG signal quality.

II. GENERAL SYSTEM SPECIFICATIONS

The complete operational flowchart of the language-model-assisted BCI typing system is illustrated in Fig. 1. The system has the following main components: (A) a **presentation component** that controls the presentation scheme, (B) a **feature extraction component** that converts raw EEG evidence into a likelihood for Bayesian fusion and (C) a **decision making component** that fuses EEG (physiology) and language evidence to infer user intent. In the following, we describe these components in some more detail.

A. Presentation Component

1) *Definitions*: Let $\mathcal{A} = \{a_1, a_2, a_3, \dots, a_N\}$ be the set of all possible symbols, typically including the letters in the (English) alphabet, numerical symbols, space and backspace symbols (represented here by $_$ and $\<$, respectively). Let $\mathcal{F} = \{f_1, f_2, \dots, f_{2^{|\mathcal{A}|}}\}$ be the set of all subsets of \mathcal{A} ; $f_i \subset \mathcal{A}$. $|\mathcal{A}|$ represents the cardinality of \mathcal{A} .

A “trial” in the matrix based presentation scheme flashes a subset f_i that can contain multiple characters, i.e., $|f_i| \geq 1$, and in RSVP, it presents a single symbol; i.e., $|f_i| = 1$. A “flash” is the presentation of a trial. A “sequence” is a series of consecutive flashes of trials with no gap in between. After presenting each sequence, the system updates the posterior probabilities of every symbol in the alphabet \mathcal{A} using the new EEG evidence and tries to make an inference about user intent. However, a decision is not made until a predefined confidence level is reached.¹ Therefore, the system may need to present multiple sequences before a decision can be made. We define the collection of sequences, at the end of which one symbol is selected, as an “epoch”.

2) *Matrix Presentation*: Typically, in noninvasive EEG-based typing BCIs with the matrix presentation paradigm, symbols are arranged in an $\mathcal{R} \times \mathcal{C}$ matrix with \mathcal{R} number of rows and \mathcal{C} number of columns [1]. Subsets of these symbols are intensified usually in pseudorandom order to produce an odd ball paradigm to induce ERP responses.

Trials $f_{(1)}, f_{(2)}, \dots, f_{(n)}$ in a sequence typically cover all the symbols in the matrix, that is $\bigcup_{i=1}^n f_{(i)} = \mathcal{A}$. When each trial $f_{(i)}$ contains exactly all the symbols in a row or a column of the matrix layout with $n = (\mathcal{R} + \mathcal{C})$ [4], this setup is known as the row-and-column presentation (RCP) paradigm. RCP requires that all the symbols in \mathcal{A} would be flashed twice and $|f_i \cap f_j| \leq 1, i \neq j$. In this study, we utilize a matrix of size 4×7 , which leads to the best coverage of the widescreen monitors used in our experiments. It has been claimed that the probability of target character in each sequence’s flash set should be lower than 25% to induce the P300 response [4]. In this grid setup for RCP, each sequence contains 11 flashes, two of which include the target symbol. Therefore, the probability of each target trial in each sequence is $(2/11) \simeq 0.18$, which satisfies the threshold suggested above.

A single-character presentation (SCP) paradigm is also a widely used scheme. SCP was shown to increase the P300 signal quality compared with RCP [9]. In this paradigm, each trial contains single symbols, i.e., $|f_i| = 1$, and assuming there is no repetition in a sequence, $f_i \cap f_j = \emptyset; i \neq j$. With enough number of flashes ($n \geq 5$) in a sequence, we can satisfy the suggested condition for target probability.

3) *Rapid Serial Visual Presentation (RSVP)*: RSVP is a presentation technique in which trials are presented one at a time at a fixed predefined location on the screen at a rapid rate and in a pseudorandom order [1], [5]. If a BCI user’s desired symbol exists in a sequence of trials presented in RSVP fashion, a P300

¹In the current implementation, confidence is measured by the maximum posterior probability over \mathcal{A} ; this corresponds to using Renyi entropy of order ∞ as the measure of uncertainty. Other entropy definitions such as Shannon’s could also be used.

response is elicited by the target in the EEG signal. RSVP is similar to SCP in that each presentation subset includes only a single symbol; however, RSVP decreases the dependency on gaze control. Presenting 28 symbols in an RSVP paradigm is time consuming; therefore, a typical RSVP-based BCI system can only achieve a speed of five symbols/minute if each sequence contains the entire alphabet [5], [16]–[18]. However, recent efforts to speed up typing with this presentation paradigm showed that using context information (such as a language model) and careful selection of subsets of a in each sequence may significantly improve typing speed and accuracy [5], [15], [16], [19], [21].

B. Feature Extraction Component

The EEG signals are acquired using a g.USBamp biosignal amplifier with active g.Butterfly electrodes at a sampling rate of 256 Hz, from 16 EEG sites (according to the International 10/20 configuration): Fp1, Fp2, F3, F4, Fz, Fc1, Fc2, Cz, P1, P2, C1, C2, Cp3, Cp4, P5, and P6. To improve the signal-to-noise ratio (SNR) and to eliminate drifts, signals were filtered by an FIR linear-phase bandpass filter passing [1.5, 42] Hz with zero dc gain and a notch filter at 60 Hz.

In order to capture the P300 while omitting the possible motor EEG [8], EEG from a time window of [0,500] ms after each flash's onset is processed as the corresponding raw data for each trial. As we explain later in Section III, we test our system with healthy users; therefore the window length is chosen short to avoid any discriminative contributions of motor-activity-related EEG response, if any. EEG data processing continues with i) downsampling by 2, ii) projection to a lower dimensional space using principle component analysis (PCA) to remove directions with negligible variance, and iii) concatenation of data from all channels corresponding to the same trial to form a feature vector for each trial.

C. Decision Making Component

Evidence from EEG is supported with evidence from language structure. These two information sources are fused using a Naïve Bayes' assumption to make a joint decision using MAP inference. Optimal classifier parameters for target detection are learned using the calibration data.

1) *EEG Feature Extraction and Classification*: To improve intent detection performance, the EEG feature vectors computed as described above are projected in to a one-dimensional space, which attempts to maximize the separation between target and nontarget classes according to a measure. Specifically, assuming that, in each class, feature vectors follow a multivariate Gaussian distribution,² quadratic discriminant analysis (QDA) is used to project the data to minimize the expected risk. QDA requires the inverse of the empirical covariance for each class. Estimating an invertible covariance is not feasible in the practical usage of the typing system due to the high dimensionality of the EEG feature vectors and low number of calibration samples in each class. This issue has been addressed by employing regularized discriminant

analysis (RDA), which provides full-rank covariance estimates for each class [25].

RDA uses shrinkage and regularization. Shrinkage is a linear combination of each class covariance matrix and the overall class-mean-subtracted covariance. Considering $\mathbf{x}_i \in \mathbf{R}^p$ as a p -dimensional feature vector and l_i as its label, which can take values of 0 and 1 for nontarget and target classes, respectively, the maximum-likelihood estimator for mean and covariance of each class are

$$\begin{aligned}\boldsymbol{\mu}_k &= \frac{1}{N_k} \sum_{i=1}^N \mathbf{x}_i \delta_{l_i, k} \\ \boldsymbol{\Sigma}_k &= \frac{1}{N_k} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \delta_{l_i, k}\end{aligned}\quad (1)$$

where $k \in \{0, 1\}$, N_k is the number of training feature vectors in class k , and thus N , the total number of feature vectors, will be $N_0 + N_1$ and $\delta_{\langle \cdot, \cdot \rangle}$ is the Kronecker- δ . **[AU: Previous sentence edited correctly for meaning?]** The shrinkage procedure manipulates the covariance matrices by

$$\hat{\boldsymbol{\Sigma}}_k(\lambda) = \frac{(1 - \lambda)N_k \boldsymbol{\Sigma}_k + (\lambda) \sum_{k=0}^1 N_k \boldsymbol{\Sigma}_k}{(1 - \lambda)N_k + (\lambda) \sum_{k=0}^1 N_k}.\quad (2)$$

Here, $\lambda \in [0, 1]$ is the shrinkage parameter that defines the similarity of two classes' covariance. $\lambda = 1$ leads to equal covariance matrices for both classes, which turns RDA to linear discriminant analysis (LDA). The regularization procedure is as follows:

$$\hat{\boldsymbol{\Sigma}}_k(\lambda, \gamma) = (1 - \gamma)\hat{\boldsymbol{\Sigma}}_k(\lambda) + (\gamma)\frac{1}{p} \text{tr}[\hat{\boldsymbol{\Sigma}}_k(\lambda)] \mathbf{I}_p.\quad (3)$$

$\text{tr}[\cdot]$ is the trace operator, \mathbf{I}_p is a $p \times p$ identity matrix, and $\gamma \in [0, 1]$ is the regularization parameter, which determines the circularity of the covariance matrix.

Correspondingly, the discriminant score function defined as

$$d_{\text{RDA}}(\mathbf{x}) = \log \frac{f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}_1, \hat{\boldsymbol{\Sigma}}_1(\lambda, \gamma)) \hat{\pi}_1}{f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}_0, \hat{\boldsymbol{\Sigma}}_0(\lambda, \gamma)) \hat{\pi}_0}\quad (4)$$

where $f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the Gaussian probability density function when $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\hat{\pi}_k$ is the prior probability of class k . In our system, we use $\hat{\pi}_1 = \hat{\pi}_0$. To find the class conditional probability distributions of RDA scores, we use kernel density estimation (KDE) [16]. Each class conditional KDE is calculated over the RDA scores of EEG evidence recorded for the representative trials of that class in the calibration data set. Finally, the conditional probability density function for each class is defined as

$$\begin{aligned}f(\mathbf{x} = \mathbf{y} | l = k) &= \\ f_{\text{KDE}}(d_{\text{RDA}}(\mathbf{x}) = d_{\text{RDA}}(\mathbf{y}) | l = k) &= \\ \frac{1}{N_k} \sum_{i=1}^N \mathcal{K}_{h_k}(d_{\text{RDA}}(\mathbf{x}_i), d_{\text{RDA}}(\mathbf{y})) \delta_{l_i, k}.\end{aligned}\quad (5)$$

Here, $\mathcal{K}_{h_k}(\cdot, \cdot)$ is a suitable kernel function with bandwidth h_k . A Gaussian kernel is used in our system, and accordingly the

²The Gaussian distribution assumption here is a direct consequence of the assumption that filtered EEG is a Gaussian random process.

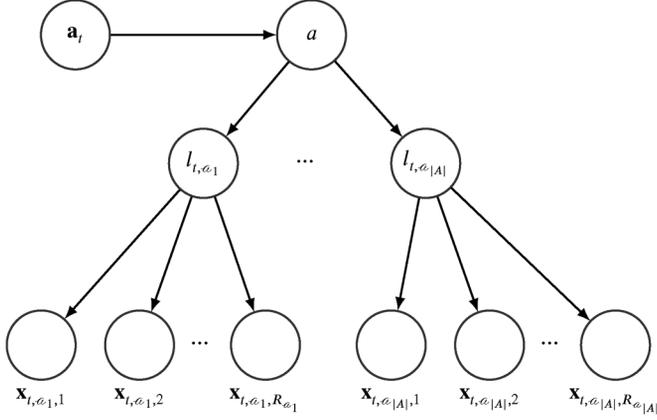


Fig. 2. Probabilistic graphical model of the fusion rule.

kernel bandwidth h_k for each class is calculated using the Silverman rule of thumb [26] over the RDA scores for the corresponding class.

2) *Language Model*: The system utilizes a letter n-gram model in an iterative Bayesian framework to increase the typing speed by prioritizing the symbols to be presented in each sequence and by providing a prior context for intent detection. A letter n-gram model estimates the conditional probability of every letter in the alphabet based on $n - 1$ previously typed letters in a Markov model framework [27].

Therefore, in a letter n-gram model, the conditional probability of each character, according to the Bayes rule, is given by

$$p(a_t = a | \mathbf{a}_t = \mathbf{a}) = \frac{p(a_t = a, \mathbf{a}_t = \mathbf{a})}{p(\mathbf{a}_t = \mathbf{a})} \quad (6)$$

where a_t is the symbol (yet) to be typed at epoch t and \mathbf{a}_t is the string of previously written $n - 1$ symbols. In our system, we use a 6-gram letter model, which is trained on the *New York Times* portion of the English Gigaword corpus [27].

3) *Fusion*: Assume \mathbf{x}_{t,r,a_i} represents the EEG feature vector of a trial, which contains $a_i \in \mathcal{A}$, at repetition $r \in \{1, 2, \dots, R_{a_i}\}$ in epoch t where R_{a_i} represents the total number of repetitions of trials containing the character a_i in the same epoch. Moreover, define l_{t,a_i} as the class label for $a_i \in \mathcal{A}$ in epoch t . The probabilistic graphical model that we use for fusion is shown in Fig. 2.

Let $\mathcal{X}_{t,a_i} = [\mathbf{x}_{t,1,a_i}, \mathbf{x}_{t,2,a_i}, \dots, \mathbf{x}_{t,R_{a_i},a_i}]$ represent a $(p \times R_{a_i})$ matrix of observed EEG feature vectors in epoch t . Here, p is the length of each feature vector. Accordingly, assume $\mathcal{X}_t = [\mathcal{X}_{t,a_1}, \mathcal{X}_{t,a_2}, \dots, \mathcal{X}_{t,a_{|A|}}]$ is a $(p \times N)$ matrix, where N is the number of total flashes in epoch t . Define \mathcal{X} as a possible outcome for matrix \mathcal{X}_t . Using Bayes' rule, we can define the posterior probability conditioned on the prior typed text and the observed EEG feature vectors as

$$Q = p(a_t = a | \mathcal{X}_t = \mathcal{X}, \mathbf{a}_t = \mathbf{a}) \propto p(\mathcal{X}_t = \mathcal{X}, \mathbf{a}_t = \mathbf{a} | a_t = a) P(a_t = a). \quad (7)$$

³Lower levels consist of copying phrases that have letters that are assigned high probabilities by the language model. As the level increases, the language model probabilities become increasingly adversarial. Level 3 is neutral on average.

Using the proposed graphical model, given the intended symbol a , the EEG evidence and previously typed text are conditionally independent. Moreover, given a , the EEG evidences for each trial $\mathbf{x}_{t,a_i,1}, \mathbf{x}_{t,a_i,2}, \dots, \mathbf{x}_{t,a_i,R_{a_i}}$ are independent, as follows:

$$Q \propto \left(\prod_{a_i \in \mathcal{A}} \prod_{r=1}^{R_{a_i}} f(\mathbf{x}_{t,a_i,r} = \mathbf{x}_{a_i,r} | a_t = a) \right) P(a_t = a | \mathbf{a}_t = \mathbf{a}). \quad (8)$$

$\mathbf{x}_{a_i,r}$ is the possible EEG evidence for r th repetition of character a_i . Also for given a , l_{t,a_i} s are deterministically defined. With this assumption, (8) can be simplified as

$$Q \propto \left(\prod_{r=1}^{R_a} \frac{f(\mathbf{x}_{t,a,r} = \mathbf{x}_{a,r} | l_{t,a} = 1)}{f(\mathbf{x}_{t,a,r} = \mathbf{x}_{a,r} | l_{t,a} = 0)} \right) P(a_t = a | \mathbf{a}_t = \mathbf{a}). \quad (9)$$

At the end of each sequence, $p(a_t = a | \mathcal{X}_t = \mathcal{X}, \mathbf{a}_t = \mathbf{a})$ is calculated for all the symbols; if the maximum of these posterior probabilities is higher than a predefined confidence threshold, a decision to type the corresponding symbol is made. Otherwise, sequences are repeated until the required confidence level is reached. If the confidence level is not reached in a predefined maximum number of repetitions bound for sequences, the symbol with the maximum *a posteriori* probability is chosen as the desired symbol.

D. System Operation Modes

The developed typing interface can currently be utilized in four different modes.

- i) *Calibration mode*: During calibration, the users are asked to attend to predefined target symbols within randomly ordered sequences to record labeled EEG data. The data acquired in this mode are then used in the estimation of classifier parameters to be used in other system operation modes. The shrinkage and regularization parameters are optimized during calibration using k-fold cross-validation to maximize area under the ROC curve.
- ii) *Copy phrase task mode*: In this task, the users are given a set of predefined phrases. Each phrase includes a missing word and the users are asked to complete these words. This task is designed to assess the system and/or user performance in terms of speed and accuracy in the presence of a language model.
- iii) *Mastery task mode*: Users are trained to use the system in this mode. It is similar to the copy phrase task mode in that the users are asked to type a set of predefined phrases. In contrast, the phrases used in this task have been carefully selected and divided into five difficulty levels based on their predictability by the language model. As the user completes the phrases in a level, the task continues with the next level with more difficult sentences.³
- iv) *Free spelling mode*. This mode allows the users to type their desired text.
- v) *Simulation mode*: In this mode, the copy phrase task is completed using samples drawn from the KDE of class conditional EEG feature distributions as computed in (5). These samples simulate EEG evidence and are fused with the language model probabilities for decision making as in regular operation [16]. Probability of completing the

task and expected task completion durations are reported as estimated performance measures using Monte Carlo simulations.

In this paper, we use all modes of the system for the following experiments, except free spelling.

III. EXPERIMENTAL RESULTS

A. Experiment

In this study, we assess the system performance in three presentation scenarios:

- 1) 4×7 matrix row and column presentation (RCP) paradigm;
- 2) 4×7 matrix single-character presentation (SCP) paradigm;
- 3) rapid serial visual presentation (RSVP) paradigm.

The comparison is based on three dependent variables: signal quality, system accuracy, and typing speed. Following a group-based analysis, we utilize paired t-tests to determine if the system performance varies significantly due to changes in the presentation paradigm or inter-trial interval (ITI) values. In addition, we perform paired t-tests within each user to assess the variations in P300 responses due to different ITIs.

Twelve healthy volunteers, nine males and three females, between the ages of 24 and 38 years, consented to participate in this study, which is conducted following an IRB-approved protocol. Each user participated in three sessions, each session on a different day with the various presentation paradigms. It is possible for a participant to gradually obtain skills to handle the system more efficiently, thereby introducing learning effects from session to session. To control for this effect, we relied on quasi-randomization; we distributed the presentation paradigms over the experimental sessions such that the number of users who attended a session with a specific presentation paradigm on a specific session order is kept the same (balanced). Every session that a user attended included calibration tasks with four different ITI values of {200, 150, 100, 85} ms. These values are chosen to be compatible with a 60-Hz monitor refresh rate and cover the range of possible optimum inter-trial durations. To account for the effect of user fatigue on typing performance, we randomized the order of ITI values for each presentation scenario and among all users. We used a duty cycle of 75% for each flash.

After calibration, each session proceeded with the mastery task [28] followed by the copy phrase task with eight sentences. We use a level 1 mastery task to familiarize the users with the copy phrase task. To prevent long sessions, the system marks a phrase as unsuccessful if more than four wrong letter selections occur in a row, and the next phrase is presented to the user.

B. Results

1) *Signal Quality*: In their work, Sellers *et al.* show that ITI effectively modifies the shape of the P300 response [29]. To investigate the effect of ITI on the P300 response, we analyzed the signal quality for every presentation scheme and ITI combination using the calibration data collected for different ITI values. For such combinations, we computed the area under the

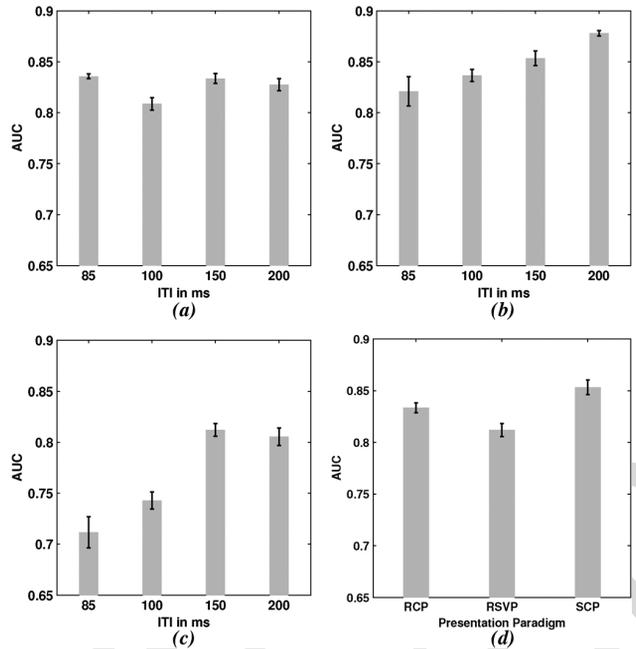


Fig. 3. Bar charts of average AUC with error bars. (a), (b), and (c) demonstrate the accuracy statistics for each ITI, respectively, for RCP, SCP, and RSVP paradigms. (d) reports the AUC statistics for different presentation paradigms at ITI = 150 ms. (a) ITI comparison in RCP; (b) ITI comparison in SCP; (c) ITI comparison in RSVP; and (d) paradigm comparison at ITI = 150 ms.

TABLE I
HYPOTHESIS TESTING RESULTS BETWEEN DIFFERENT ITIS WITHIN EACH PARADIGM. THE NULL HYPOTHESIS IS THAT THE EXPECTED AUC DIFFERENCE OF THE TWO CONSIDERED ITIS IS ZERO. HERE, WE USED $\alpha = 0.05$

Paired t-test results between different ITIs in each paradigm			
ITI_1 v.s. ITI_2	<i>P</i> -values in RCP	<i>P</i> -values in SCP	<i>P</i> -values in RSVP
85 v.s. 100 ms	0.068	0.593	0.157
85 v.s. 150 ms	0.803	0.927	0.001
85 v.s. 200 ms	0.550	0.240	0.009
100 v.s. 150 ms	0.075	0.673	0.0008
100 v.s. 200 ms	0.053	0.053	0.027
150 v.s. 200 ms	0.570	0.236	0.693

curve (AUC) for the ROC as the classification accuracy measure. Within each presentation paradigm, we applied a paired t-test over these accuracy values. The results are reported in Fig. 3(a)–(c) and Table I for each paradigm. The three subfigures correspond to different presentation paradigms, and in each subfigure, the average accuracies for different ITIs are presented using bar-graphs with error bars. Table I summarizes the paired t-test results between every ITI pair for each presentation paradigm.

From Table I, we observe that the group-based hypothesis testing does not show significant variations among classification accuracies due to changes in ITI values for the RCP paradigm. The results also suggest that the ITI value of 85 ms is the best candidate for matrix RCP paradigm. This ITI offers shorter sequence times and a consistent higher average AUC (averaged across users) as shown in Fig. 3(a). Our observations suggest that ERP responses in the RCP paradigm are more robust to the changes in ITI values.

TABLE II
MULTIVARIATE PAIRED T-TEST RESULTS OF P300 PEAK VALUE FOR EACH SUBJECT AMONG DIFFERENT ITIS WITHIN EACH PRESENTATION PARADIGM. THE NULL HYPOTHESIS IS THAT THE EXPECTED AUC DIFFERENCE OF THE TWO CONSIDERED ITIS IS ZERO WITHIN EACH PARADIGM. HERE, WE USED $\alpha = 0.05$, I.E. $H = 1$ IF $P < 0.05$ AND $H = 0$ OTHERWISE

Users	U1		U2		U3		U4		U5		U6		U7		U8		U9		U10		U11		U12	
RCP	H	P																						
85 v.s. 100 ms	0	0.18	0	0.09	0	0.19	0	0.27	1	0.01	0	0.85	0	0.16	1	0	1	0.02	0	0.16	1	0.02	1	0.04
85 v.s. 150 ms	0	0.61	1	0	1	0.03	0	0.89	1	0	0	0.28	1	0.01	1	0	1	0	0	0.32	0	0.18	0	0.73
85 v.s. 200 ms	1	0.02	0	0.27	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0.03	1	0.02	0	0.79
100 v.s. 150 ms	0	0.13	1	0	1	0.01	0	0.25	1	0	0	0.31	1	0.01	1	0	1	0	1	0.03	0	0.16	0	0.11
100 v.s. 200 ms	0	0.07	1	0	1	0	1	0	1	0	1	0	1	0	0	0.06	1	0	1	0	1	0	0	0.11
150 v.s. 200 ms	1	0	0	0.15	1	0	1	0	1	0	1	0	1	0	1	0	0	0.25	0	0.14	1	0	0	0.11
SCP	H	P																						
85 v.s. 100 ms	0	0.35	0	0.13	1	0	0	0.31	0	0.88	0	0.1	1	0	1	0	0	0.75	1	0.04	1	0	0	0.08
85 v.s. 150 ms	0	0.47	1	0	0	0.95	0	0.15	0	0.72	0	0.11	1	0	1	0	0	0.33	1	0.03	1	0	0	0.06
85 v.s. 200 ms	0	0.18	1	0	1	0	0	0.33	1	0	1	0	1	0	1	0	1	0.03	0	0.16	1	0	0	0.08
100 v.s. 150 ms	0	0.05	1	0	1	0	1	0.05	0	0.85	1	0.04	1	0.01	1	0	0	0.38	0	0.29	0	0.93	0	0.28
100 v.s. 200 ms	1	0.02	1	0	1	0	0	0.55	0	0.1	1	0	1	0	1	0	0	0.47	0	0.39	0	0.09	1	0.03
150 v.s. 200 ms	1	0	1	0	0	0.36	1	0.04	1	0.01	1	0	1	0	1	0	1	0.05	0	0.18	0	0.48	0	0.2
RSVP	H	P																						
85 v.s. 100 ms	0	0.06	0	0.69	0	0.1	0	0.38	1	0.02	1	0	0	0.42	0	0.14	0	0.57	0	0.64	0	0.24	0	0.07
85 v.s. 150 ms	0	0.63	0	0.09	1	0.02	0	0.07	0	0.36	1	0	1	0	1	0.04	1	0.05	0	0.19	0	0.37	0	0.05
85 v.s. 200 ms	1	0	0	0.15	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0.08	0	0.09
100 v.s. 150 ms	0	0.16	1	0.01	0	0.08	1	0.01	0	0.21	1	0	0	0.12	1	0.01	0	0.1	0	0.13	0	0.24	0	0.82
100 v.s. 200 ms	1	0	0	0.28	1	0	1	0	0	0.06	1	0	1	0	1	0	1	0.01	1	0.02	0	0.93	0	0.37
150 v.s. 200 ms	1	0	0	0.2	1	0	1	0	1	0.03	1	0	1	0	1	0	1	0	0	0.45	0	0.1	0	0.62

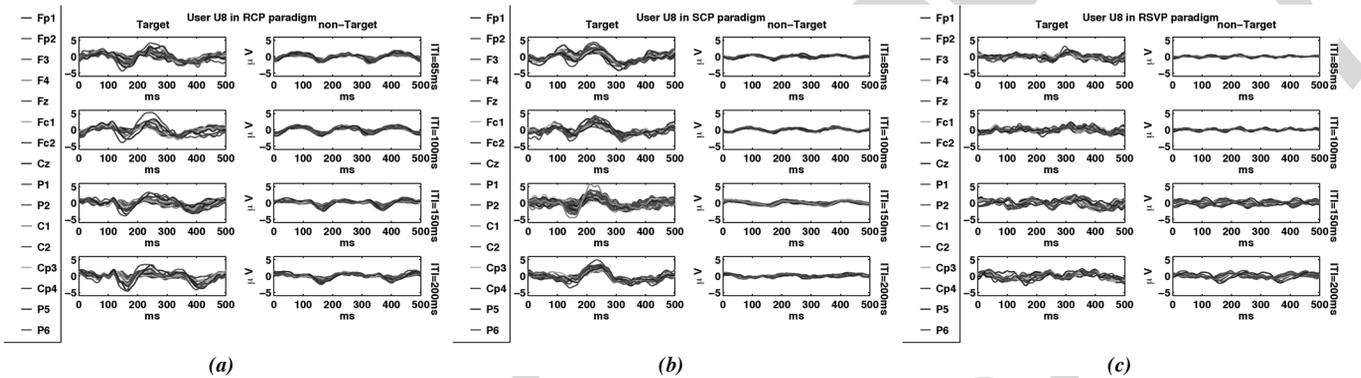


Fig. 4. Average ERP response to target and nontarget stimuli, for each presentation paradigm and ITI pairs for user “U8”. From top to bottom, the ITI is increasing monotonically. (a) RCP; (b) SCP; and (c) RSVP.

The SCP paradigm with an ITI of 200 ms demonstrates the highest average AUC with the lowest variance [Fig. 3(b)]. Although average AUCs across users show an increasing trend from ITI of 85 to 200 ms, pairwise comparisons between different ITIs do not show statistically significant variations in population AUCs (see Table I).

Generally, in matrix-based presentation paradigms, variations in ITI values seem to have a negligible effect on system AUC. The usage of smaller ITIs might be preferable due to a possible decrease in the sequence length, which might improve the speed of the typing interface. Moreover, it might be viable to optimize the matrix subset flashes based on context information to have shorter sequence lengths and higher classification confidence by increasing the number of flashes of probable characters, which can lead to faster target detections.

On the other hand, accuracies with the RSVP paradigm tend to be more sensitive to changes in ITI values (as shown in Table I). The most significant increase in AUC happens from ITI = 100 ms to ITI = 150 ms. The accuracy deviations between ITI = 85 ms and ITI = 100 ms and also between ITI = 150 ms and ITI = 200 ms are not significant as reported in Table I. Consequently, among the ITI values tested with the RSVP paradigm, ITI = 150 ms is the best choice for system design, since the accuracies between ITIs of 150 and 200 ms do not significantly change while ITI = 150 ms provides

better speed. This is consistent with our previous work using RSVP for image search [14]. In contrast with matrix-based presentation paradigms, in the RSVP paradigm, users need to recognize the target symbols, which induces the weaker P300 signals, especially at lower ITIs, as shown in Fig. 3(c).

To investigate signal quality variations due to ITI changes in each presentation scheme, we extract the P300 peak values for every target stimulus at all channels per user, for different combinations of ITI values and presentation paradigms. In this process, we filter the EEG signal using a Gaussian low pass filter with ($\sigma = 5$ samples) to increase the signal-to-noise ratio (SNR). For each target trial, we define a (16×1) -dimensional feature vector with the i th element containing the peak value of the EEG at channel i in the time window [250, 350] ms after stimulus onset. For every user and presentation paradigm, we use these feature vectors in a multivariate paired t-test to investigate the P300 amplitude deviations across different ITI values. We report the results in Table II. Comparing the results for different paradigms, we do not observe a consistent change in P300 amplitude among different ITI values. For instance, variations in ITI can significantly change the P300 peak values of user U8 at every presentation paradigm, as illustrated in Fig. 4, while this is not true for user U12 (see Fig. 5). Consequently, to acquire the best performance, we recommend that optimum ITI be defined uniquely for each user.

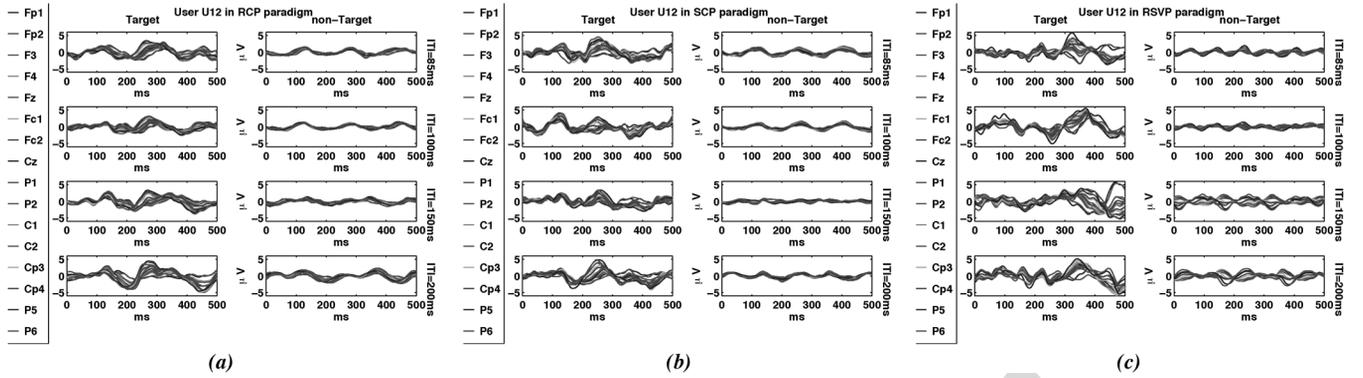


Fig. 5. Average ERP response to target and nontarget stimuli, for each presentation paradigm and ITI pairs for user “U12”. From top to bottom, the ITI is increasing monotonically. (a) RCP; (b) SCP; and (c) RSV.

TABLE III
HYPOTHESIS TESTING RESULTS BETWEEN DIFFERENT PRESENTATION PARADIGMS AT ITI = 150 ms. THE NULL HYPOTHESIS IS THAT THE EXPECTED AUC DIFFERENCE OF THE TWO CONSIDERED PARADIGMS IS ZERO. HERE, WE USED $\alpha = 0.05$

Paired t-test results between different presentation paradigms		
compared paradigms	H_0 rejected	P -value
RCP v.s. RSV	No	0.362
RCP v.s. SCP	No	0.453
RSVP v.s. SCP	No	0.188

2) *System Accuracy Based on Presentation Paradigm*: We analyzed the changes in system classification accuracy across different presentation paradigms. Similar to the signal quality analysis, we employed AUC values as the measure of accuracy. We set ITI = 150 ms, which provides good performance for all paradigms and is close to the ITI value typically used for matrix based presentation paradigms (125 ms) [4]. We analyzed the changes in AUC values across different presentation paradigms using a paired t-test between different presentation paradigms. We report the results in Table III. These results do not illustrate significant changes due to different presentation paradigms. Moreover, in Fig. 3(d), we plot the AUC values averaged over all the users for different presentation schemes. This figure shows that the average AUC values in matrix-based paradigms are higher than in the RSV paradigm. However, the paired t-test outcomes do not confirm statistically significant separations among these average AUC values. Based on these results, we propose that the system accuracy might be more dependent on the user than the presentation paradigm at ITI = 150 ms.

In Fig. 6, we plot the channel by channel significance levels for the paired t-tests between different presentation paradigms. From this figure, we first observe that there is no statistically significant difference between different presentation paradigms. This result is consistent with the results that we report in Table III.

We also observe that there is no consistent electrode subset that shows significant difference among different presentation paradigms. We also plot the AUC values calculated using each channel separately, for each user, for different presentation paradigms, in Fig. 7. For a specific presentation paradigm, this

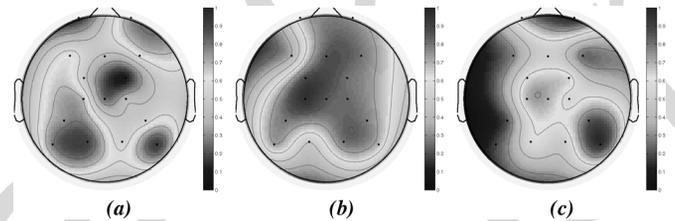


Fig. 6. Topography map of $(1 - p)$ resulting from paired t-tests for each channel’s AUC between each paradigm pair and across users for ITI = 150 ms. Here, red denotes $1 - p = 1$, and blue represents $1 - p = 0$. (a) RCP versus RSV; (b) RCP versus SCP; and (c) RSV versus SCP.

figure does not show a consistent region on the scalp, across different users, for high accuracy.⁴ Based on these results, we suggest that an optimum presentation paradigm is user dependent.

3) *Typing Speed*: We analyze the differences in the typing speed across different presentation paradigms employing the average number of sequences per target trial as the measure of speed-inverse (time spent per letter). Conventionally, in the RCP paradigm, during each sequence, all rows and columns are flashed once (which results in 11 flashes during a sequence in this study). On the other hand, the RSV paradigm has previously demonstrated almost optimized performance with eight trials in a sequence [15]. Accordingly, to keep the analysis equitable, we use the average number of sequences per target trial as the measure of time spent. We set ITI = 150 ms. In addition to the experimental results, we also perform 20 Monte Carlo simulations of the copy phrase task for every user under different presentation paradigms, using the corresponding calibration EEG data to generate simulated EEG evidence.

We report both the simulation and experimental results in Fig. 9. For different presentation paradigms, this figure shows the average number of sequences per target trial and task completion probabilities versus AUC values of different users. We observe that both minimum and maximum values of user AUCs are smaller in the RSV paradigm than the matrix-based presentation schemes. In the RCP paradigm, each symbol is represented twice in a sequence. Subsequently, the number of data

⁴One may need to optimize best electrode locations for each paradigm–user combination to maximize performance.

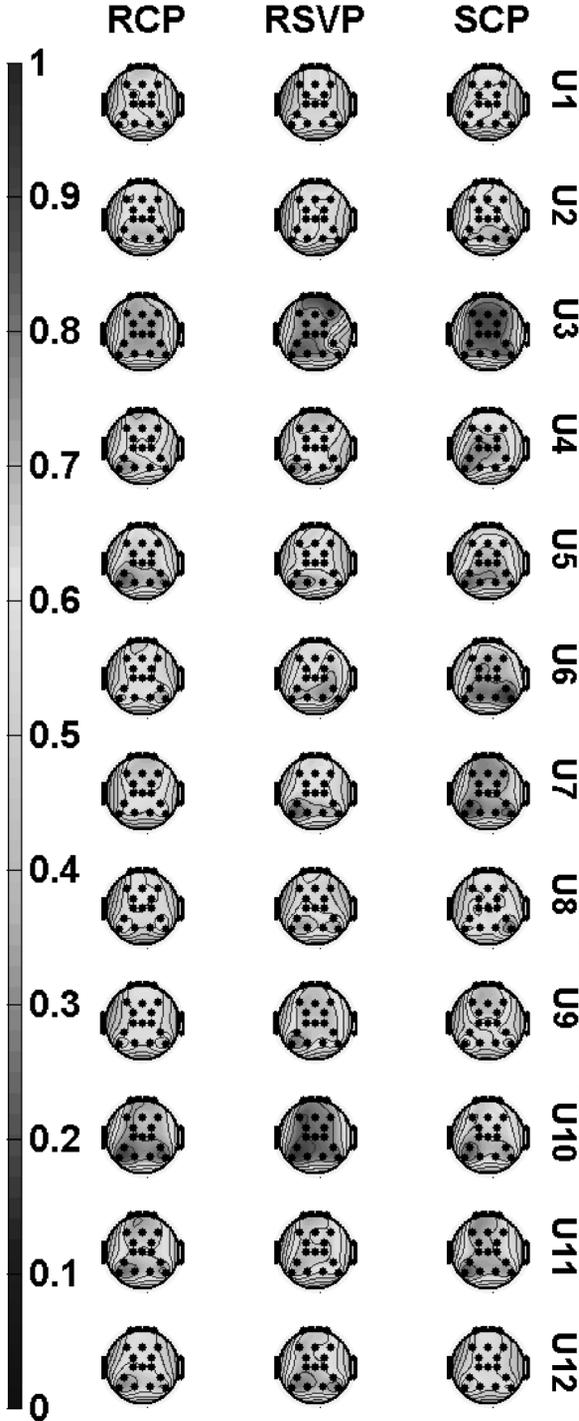


Fig. 7. Topography of channel based AUCs for each user at ITI = 150 ms.

points from the target class for recorded EEG during the calibration task is twice the other paradigms. This can lead to a more accurate estimation of classifier parameters, which then leads to smaller average numbers of sequences per target trial and higher task completion probability.

In general, actual typing performance in the SCP paradigm shows a behavior consistent with simulation results. In all paradigms, simulation results are reasonably predictive of the actual typing task statistics for larger AUCs. Mismatch between simulation results and actual user typing speeds is

more frequent in the RSVP paradigm. This maybe because the user AUCs are generally lower for the RSVP paradigm, since the requirement to recognize the target symbol might impose more cognitive load and require more attention from the user.⁵ However, some participants still show faster typing performance with RSVP than the matrix-based presentation schemes (see Table IV).

From this table, user U7 shows better typing performance when using the RSVP paradigm [see Fig. 9(a)–(c)], while users U3 and U9 spelled target phrases with a lower average number of sequences when using SCP and RCP paradigms [see Fig. 9(d)–(f) and (g)–(i), respectively]. Accordingly, the choice for the best presentation scheme should be user dependent.

4) *Effect of Language Model on Typing Duration:* We employ the simulation mode of the system to assess the effect of the language model on the (estimated) performance of each presentation paradigm. We preform **[AU: "perform"?)** ten Monte Carlo simulations (of the copy phrase task) with and without the language model to estimate the typing speed under both conditions using calibration EEG data from each user. We represent the typing speed as the average number of sequences for correctly typing a character, N_{avg} . The results of N_{avg} shown in Figs. 10(a)–(c) indicate that the language model significantly improves the performance for all three presentation paradigms. This is seen in the form of reduced average sequence counts required to type a target symbol correctly, as well as reduced variance. That is, without a language model, the mean values of N_{avg} are larger for all the users, and the standard deviations of N_{avg} are larger for most of the participants. For RCP, users with lower AUC (larger sequence counts for the without-LM axis) seem to increasingly benefit from the assistance of the language model in this task [see Fig. 10(a)]. In the case of SCP [Fig. 10(b)] and RSVP [Fig. 10(c)], while the same trend is observed for high to moderately good AUCs, for users with the lowest several AUCs (appearing on the right-most side of their respective plots), the consistency of language model assistance is not as good as that in the case of RCP. This inconsistent behavior seems to occur due to low AUCs (for $AUC < .74$). This suggests that, for some users with low classification performance, we may need to collect more training samples in the calibration session for them to be able to benefit from the language model assistance.

IV. CONCLUSION

In this paper, we compared three different presentation paradigms: i) a 4×7 matrix row and column; ii) a 4×7 matrix single character; and iii) a rapid serial visual presentation, utilizing a language-model-assisted EEG-based letter-by-letter typing BCI. The underlying intent inference engine used tight fusion of language and EEG evidence, as described in earlier papers on the RSVP keyboard [5], [15], [16]. Twelve participants were recruited to use the system in four different ITIs of {85, 100, 150, 200} ms for each presentation scheme. The order of paradigm presentations for each session, and each user were

⁵This claim is mainly based on the users' feedback after each session. They described it as more challenging to spot the desired character in the RSVP paradigm, especially for the sessions with smaller ITIs.

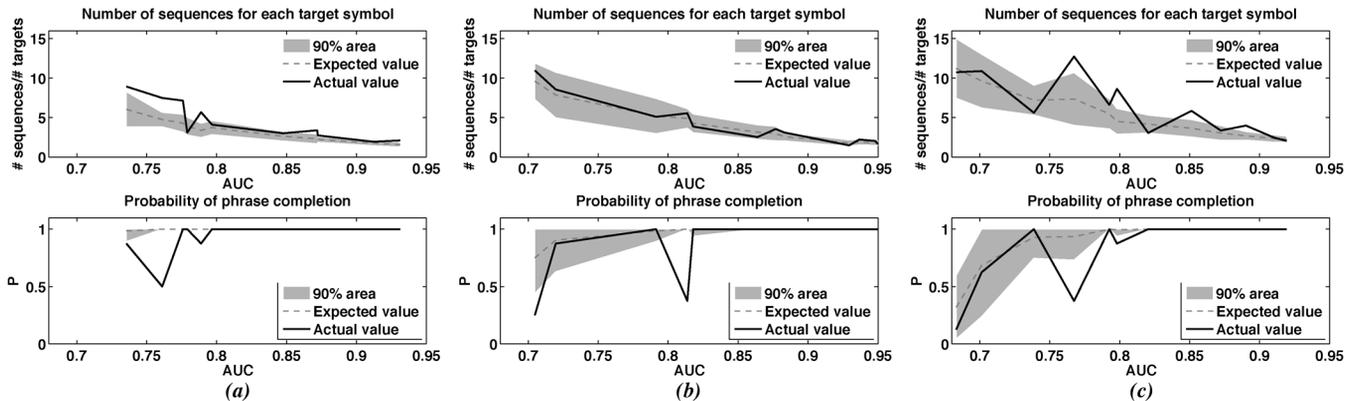


Fig. 8. Typing speed analysis results. Average number of sequences per (typed) target character (lower means faster typing) and probability of phrase completion (higher means more accuracy) are shown. Simulation results are used to define the shaded 90% confidence area shown. The dashed line shows the expected value from simulation for each variable, and the solid line shows actual typing outcomes in a single experimental run that follows. (a) RCP; (b) SCP; and (c) RSVP.

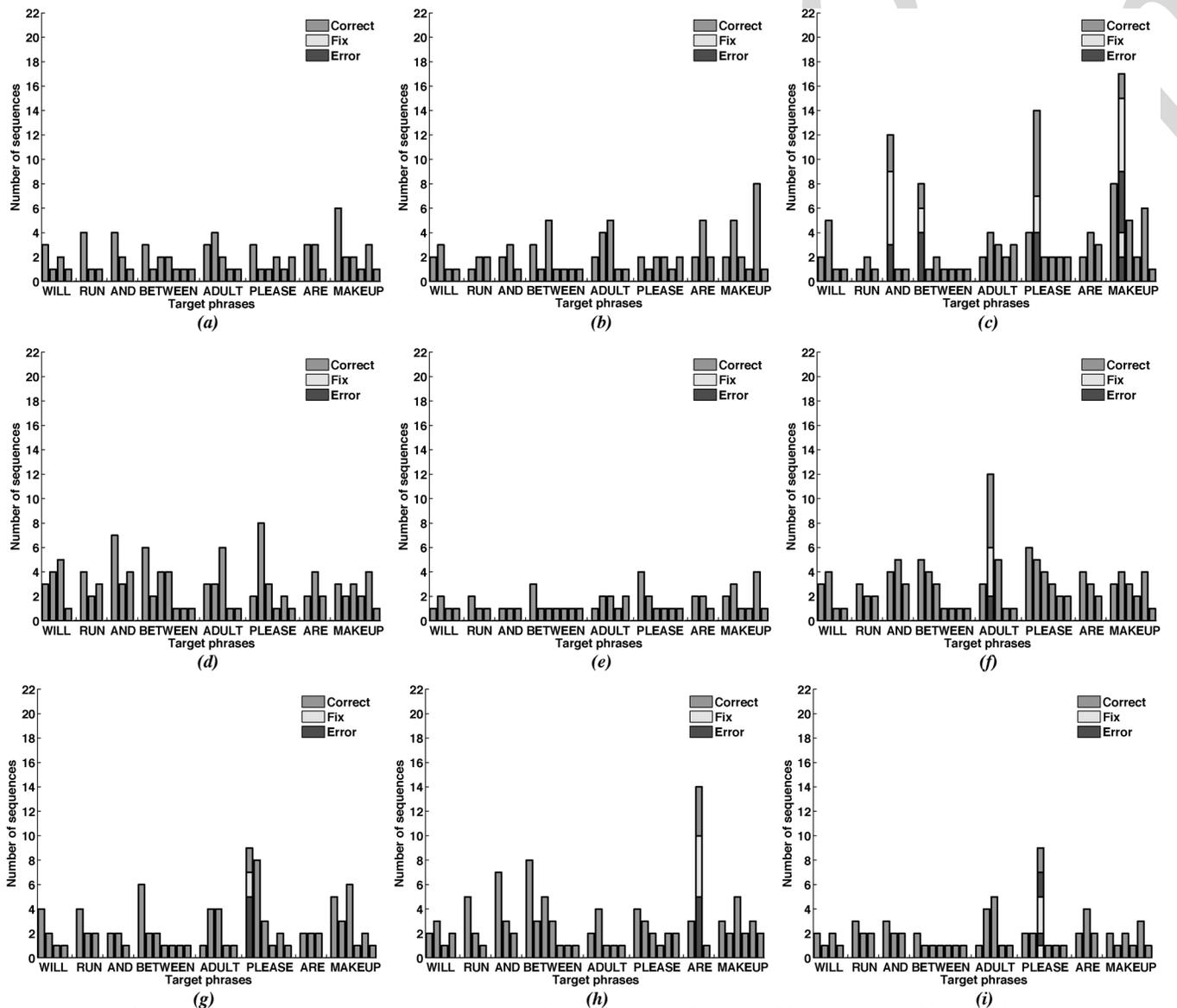


Fig. 9. Number of sequences utilized by users U7, U3, and U9 to type each target character using RSVP, SCP, and RCP paradigms. Red bars show the sequence counts for epochs that typed a wrong character, and yellow bars show the number of sequences used to fix the error before typing the correct target. Green bars show the number of sequences in epochs that resulted in correct selection of target symbols (lower means faster typing). (a) U7, RSVP; (b) U7, SCP; (c) U7, RCP; (d) U3, RSVP; (e) U3, SCP; (f) U3, RCP; (g) U9, RSVP; (h) U9, SCP; and (i) U9, RCP.

quasi-randomized. The same classifier, language model, and fusion rule were used for all paradigms and ITI combinations.

Through this study, we illustrated that the best presentation paradigm and ITI combination among the ones presented in this

TABLE IV
 TYPING SPEED RESULTS FOR EACH USER AND PARADIGM COMBINATION. HERE, "AVERAGE \pm STANDARD DEVIATION"
 OF SEQUENCE COUNT PER TARGET (CORRECTLY TYPED) SYMBOL IS REPORTED

	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12
RSVP	3.98 \pm 2.7	12.74 \pm 5.74	3.05 \pm 0.69	10.74 \pm 3.35	5.6 \pm 1.46	3.35 \pm 1.48	2.04 \pm 0.35	8.64 \pm 5.15	2.44 \pm 0.76	10.89 \pm 4.3	6.59 \pm 3.76	5.84 \pm 3.87
SCP	1.29 \pm 0.26	10.96 \pm 4.9	1.48 \pm 0.32	3.55 \pm 1.78	5.54 \pm 1.97	3.85 \pm 2.04	2.21 \pm 0.62	5.08 \pm 2.44	3.1 \pm 1.36	2.52 \pm 0.95	2.03 \pm 0.42	8.55 \pm 4.26
RCP	2.73 \pm 0.85	5.67 \pm 4.53	3.1 \pm 0.83	8.92 \pm 3.83	7.48 \pm 3.55	4.1 \pm 1.32	3.38 \pm 1.68	2.09 \pm 0.57	2.09 \pm 0.57	3.01 \pm 1.11	1.93 \pm 0.5	7.14 \pm 4.09

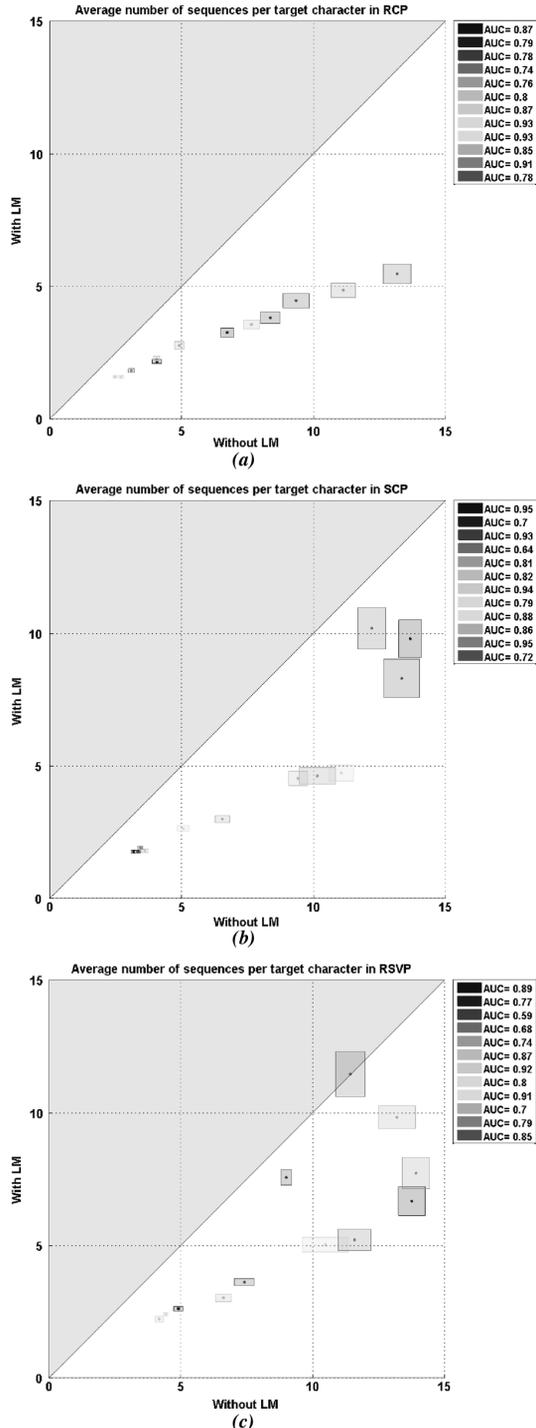


Fig. 10. Scatter plot of the average number of sequences for correctly typing a target character. The x -axis demonstrates the mean number of sequences per target character when no language model is used, y -axis represents the mean number of sequences required per target character while a 6-gram language model is utilized. Each point on the figure shows the average of the mean number of sequences per target from ten Monte Carlo simulations. Horizontal skewness of each box around a point is the standard deviation of the number of sequences per target character for typing while no language model was used, and the vertical skewness is the standard deviation in presence of the language model. (a) RCP; (b) SCP; and (c) RSVP.

study should be identified for each user individually to achieve the best performance. Also, we showed that the performance of the RSVP paradigm is comparable to matrix-based presentation paradigms with healthy users. Based on our results, we propose that BCI typing systems capable of employing multiple presentation schemes including both RSVP and matrix presentation paradigms are inevitable. This system, after individual clinical assessments, should be able to determine the best presentation option and the best ITI value for each user, according to user preferences, capabilities, EEG signal statistics, and simulations. Moreover, the length of the calibration session might need to be increased based on the classification performance for a user at each presentation paradigm.

A side product of this work is that we now have a unified BCI typing interface that has both RSVP and matrix presentation options along with a MAP intent inference engine that tightly fuses n -gram symbol and EEG evidence. It is an open vocabulary typing interface with the potential to be individualized by personal language models and the incorporation of supplementary physiological and behavioral evidence about intent, for instance via EMG or switches. Other open problems include improved signal models for more accurate performance simulations and run-time intent inference, optimized dynamic selection of stimulus subsets to be presented in each trial for the upcoming sequence, and rigorous field testing to compare RSVP and matrix presentation paradigms on potential user populations.

REFERENCES

- [1] M. Akcakaya, B. Peters, M. Moghadamfalahi, A. Mooney, U. Orhan, B. Oken, D. Erdogmus, and M. Fried-Oken, "Noninvasive brain computer interfaces for augmentative and alternative communication," *IEEE Rev. Biomed. Eng.*, vol. 7, no. 1, pp. 31–49, 2014.
- [2] C.-T. Lin, L.-W. Ko, M.-H. Chang, J.-R. Duann, J.-Y. Chen, T.-P. Su, and T.-P. Jung, "Review of wireless and wearable electroencephalogram systems and brain-computer interfaces—A mini-review," *Gerontology*, vol. 56, no. 1, pp. 112–119, 2010.
- [3] S. Moghimi, A. Kushki, A. M. Guerguerian, and T. Chau, "A review of EEG-based brain-computer interfaces as access pathways for individuals with severe disabilities," *Assistive Technol.: Official J. RESNA*, vol. 25, no. 2, pp. 99–110, 2012.
- [4] L. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, pp. 510–523, 1988.
- [5] U. Orhan, K. E. Hild, D. Erdogmus, B. Roark, B. Oken, and M. Fried-Oken, "RSVP keyboard: An EEG based typing interface," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 645–648.
- [6] E. Sellers, G. Schalk, and E. Donchin, "The P300 as a typing tool: Tests of brain computer interface with an ALS patient," *Psychophysiology*, vol. 40, p. 77, 2003.
- [7] E. W. Sellers and E. Donchin, "A P300-based brain-computer interface: Initial tests by ALS patients," *Clin. Neurophysiol.*, vol. 117, no. 3, pp. 538–548, 2006.
- [8] S. Sutton, M. Braren, J. Zubin, and E. John, "Evoked-potential correlates of stimulus uncertainty," *Science*, vol. 150, no. 3700, pp. 1187–1188, 1965.

- [9] C. Guan, M. Thulasidas, and J. Wu, "High performance P300 speller for brain-computer interface," in *Proc. IEEE Int. Workshop Biomed. Circuits Syst.*, 2004, pp. S3–5.
- [10] **[AU: Provide more information on type of source and where to find--ed.]** M. S. Treder and B. Blankertz, "Research (c) Overt Attention and Visual Speller Design in an ERP-Based Brain-Computer Interface," 2010.
- [11] Y. Liu, Z. Zhou, and D. Hu, "Gaze independent brain-computer speller with covert visual search tasks," *Clin. Neurophysiol.*, vol. 122, no. 6, pp. 1127–1136, 2011.
- [12] M. S. Treder, N. M. Schmidt, and B. Blankertz, "Gaze-independent brain-computer interfaces based on covert attention and feature attention," *J. Neural Eng.*, vol. 8, no. 6, p. 066003, 2011.
- [13] S. Chennu, A. Alsufyani, M. Filetti, A. M. Owen, and H. Bowman, "The cost of space independence in P300-BCI spellers," *J. Neuroeng. Rehab.*, vol. 10, no. 82, pp. 1–13, 2013.
- [14] Y. Huang, "Event-related potentials in electroencephalography: Characteristics and single-trial detection for rapid object search," Ph.D. dissertation, Oregon Health & Sci. Univ., Portland, OR, USA, 2010.
- [15] U. Orhan, D. Erdogmus, B. Roark, B. Oken, S. Purwar, K. E. Hild, A. Fowler, and M. Fried-Oken, "Improved accuracy using recursive Bayesian estimation based language model fusion in ERP-based BCI typing systems," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2012, pp. 2497–2500.
- [16] U. Orhan, D. Erdogmus, B. Roark, B. Oken, and M. Fried-Oken, "Offline analysis of context contribution to ERP-based typing BCI performance," *J. Neural Eng.*, vol. 10, no. 6, p. 066003, 2013.
- [17] L. Acqualagna, M. S. Treder, M. Schreuder, and B. Blankertz, "A novel brain-computer interface based on the rapid serial visual presentation paradigm," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2010, pp. 2686–2689.
- [18] L. Acqualagna and B. Blankertz, "Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP)," *Clin. Neurophysiol.*, vol. 124, no. 5, pp. 901–908, 2013.
- [19] D. B. Ryan, G. Frye, G. Townsend, D. Berry, S. Mesa-G, N. A. Gates, and E. W. Sellers, "Predictive spelling with a P300-based brain-computer interface: Increasing the rate of communication," *Int. J. Human-Comput. Interact.*, vol. 27, no. 1, pp. 69–84, 2010.
- [20] T. Kaufmann, S. Völker, L. Gunesch, and A. Kübler, "Spelling is just a click away—a user-centered brain-computer interface including auto-calibration and predictive text entry," *Frontiers in Neurosci.*, vol. 6, 2012.
- [21] S. Lee and H.-S. Lim, "Predicting text entry for brain-computer interface," in *Future Information Technology*. New York, NY, USA: Springer, 2011, pp. 309–312.
- [22] E. Samizo, T. Yoshikawa, and T. Furuhashi, "A study on application of RB-ARQ considering probability of occurrence and transition probability for P300 speller," in *Foundations of Augmented Cognition*. New York, NY, USA: Springer, 2013, pp. 727–733.
- [23] W. Speier, C. Arnold, J. Lu, R. K. Taira, and N. Pouratian, "Natural language processing with dynamic classification improves P300 speller accuracy and bit rate," *J. Neural Eng.*, vol. 9, no. 1, p. 016004, 2012.
- [24] C. Ulas and M. Cetin, "Incorporation of a language model into a brain computer interface based speller through HMMs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013, pp. 1138–1142.
- [25] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Statist. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.
- [26] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL, USA: CRC Press, 1986, vol. 26.
- [27] B. Roark, J. D. Villiers, C. Gibbons, and M. Fried-Oken, "Scanning methods and language modeling for binary switch typing," in *Proc. NAACL HLT Workshop Speech Lang. Process. Assist. Technol.*, 2010, pp. 28–36, Association for Computational Linguistics.
- [28] B. S. Oken, U. Orhan, B. Roark, D. Erdogmus, A. Fowler, A. Mooney, B. Peters, M. Miller, and M. B. Fried-Oken, "Brain-computer interface with language model-electroencephalography fusion for locked-in syndrome," *Neurorehabilitat. Neural Repair*, p. 1545968313516867, 2013.
- [29] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A P300 event-related potential brain-computer interface (bci): The effects of matrix size and inter stimulus interval on performance," *Biologic. Psychol.*, vol. 73, no. 3, pp. 242–252, 2006.



Mohammad Moghadamfalahi (S'14) received the B.Sc. degree in electrical engineering from Amir-kabir University (Tehran Polytechnics), Tehran, Iran, in 2008. Since January 2012, he has been working towards the Ph.D. student in Electrical and Computer Engineering Department of Northeastern University, Boston, MA, USA.

From 2008 to 2011, he worked at the Mobile Communication Company of Iran (MCCI). Currently, he is a Research Assistant at Cognitive System Laboratory (CSL), Northeastern University. His fields of interests are brain-computer interfaces, machine learning, and statistical signal processing.



Umut Orhan (M'09) received the B.S. degree in electrical and electronics engineering from Bilkent University, Turkey, in 2009 and the Ph.D. degree in electrical and computer engineering from Northeastern University, Boston, MA, USA, in 2014.

He currently contracts to Honeywell Laboratories as a Research and Development Engineer. His current research interests are signal processing and machine learning for biomedical and neural applications.

Dr. Orhan is a member of Eta Kappa Nu.



Murat Akcakaya (M'11) received the B.Sc. degree from the Electrical and Electronics Engineering Department of Middle East Technical University, Ankara, Turkey, in 2005 and the M.Sc. and the Ph.D. degrees in electrical engineering from Washington University in St. Louis, in May and December 2010, respectively.

He is currently an Assistant Professor at the University of Pittsburg, Pittsburg, PA, USA. His research interests include statistical signal processing and machine learning with applications to noninvasive electroencephalography (EEG) based brain-computer interface (BCI) systems, array signal processing, and physiological signal analysis for health informatics.

Dr. Akcakaya was the winner of the student paper contest awards at the 2010 IEEE Radar Conference; the 2010 IEEE Waveform Diversity and Design Conference; and the 2010 Asilomar Conference on Signals, Systems and Computers.



Hooman Nezamfar (S'14) received the B.S. degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 2004 and the M.S. degree in communication systems from the Iran University of Science and Technology, Tehran, Iran, in 2008. He has been working towards the Ph.D. degree in the Electrical and Computer Engineering Department at Northeastern University, Boston, MA, USA, since 2010.

He is also a member of Cognitive Systems Laboratory (CSL) at Northeastern University. His research

has been on the brain-computer interfaces, probabilistic modeling, and embedded systems with a special focus on the steady-state visually evoked potentials (SSVEPs).



Melanie Fried-Oken received the B.A. degree from the University of Rochester, Rochester, NY, USA; the M.A. degree in speech-language pathology from Northwestern University, Boston, MA, USA; and the Ph.D. degree in psycholinguistics from Boston University, Boston, MA, USA, in 1984. **[AU: Provide years in which all degrees were received--ed.]**

She is an international clinical researcher in the Augmentative and Alternative Communication (AAC) **[AU: At OHSU?]**, where she provides expertise about assistive technology for persons who cannot use speech or writing for expres-

sion. She is a Professor of Neurology, Pediatrics, Biomedical Engineering and Otolaryngology at the Oregon Health & Science University (OHSU), Portland, OR, USA; Director of OHSU Assistive Technology Program, and clinical speech-language pathologist. She has written over 30 peer-reviewed articles and six chapters and has given over 200 national and international lectures on AAC for persons with various medical conditions. She has been PI on 16 federal grants and co-investigator on six federal grants to research communication technology and behavior for persons with developmental disabilities, neurodegenerative diseases, and the normally aging population.



Deniz Erdogmus received the B.S. degree in electric engineering and mathematics and the M.S. degree in electric engineering from the Middle East Technical University, Ankara, Turkey, in 1997 and 1999, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2002.

He was a postdoctoral researcher at the University of Florida until 2004. He was subsequently an Assistant Professor of BME at the Oregon Health and Science University (2004–2008). Since then, he has been with Northeastern University, Boston, MA, USA, where he is currently an Associate Professor of electrical and computer engineering. His research focuses on statistical signal processing and machine learning with applications to biomedical signal/image processing and cyberhuman systems.

Dr. Erdogmus has served as Associate Editor and technical committee member for various journals and conferences.

IEEE Pre-proof
Print Version