RSVP IconMessenger: icon-based brain-interfaced alternative and augmentative communication

Article · October 2014

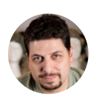
DOI: 10.1080/2326263X.2014.996066

8 authors, including:

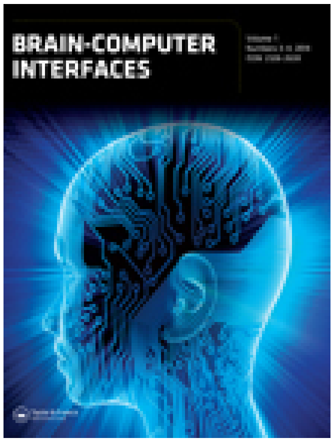
Murat Akcakaya
University of Pittsburgh

Hooman Nezamfar
Northeastern University

Rupal Patel
Northeastern University



Deniz Erdogmus
Northeastern University



Brain-Computer Interfaces
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tbci20

RSVP IconMessenger: icon-based brain-interfaced alternative and augmentative communication

Asieh Ahani[a], Karl Wiegand[b], Umut Orhan[a], Murat Akcakaya[a], Mohammad Moghadamfalahi[a], Hooman Nezamfar[a], Rupal Patel[c] & Deniz Erdogmus[a]
[a] Cognitive Systems Laboratory, Northeastern University, Boston, MA, USA
[b] Department of Computer Science, Northeastern University, Boston, MA, USA
[c] Department of Speech Language Pathology and Audiology, Northeastern University, Boston, MA, USA
Published online: 22 Dec 2014.

Brain-Computer Interfaces, 2014
Vol. 1, Nos. 3–4, 192–203, http://dx.doi.org/10.1080/2326263X.2014.996066

RSVP IconMessenger: icon-based brain-interfaced alternative and augmentative communication

Asieh Ahani[a]*, Karl Wiegand[b], Umut Orhan[a], Murat Akcakaya[a], Mohammad Moghadamfalahi[a], Hooman Nezamfar[a], Rupal Patel[c] and Deniz Erdogmus[a]

[a]*Cognitive Systems Laboratory, Northeastern University, Boston, MA, USA;* [b]*Department of Computer Science, Northeastern University, Boston, MA, USA;* [c]*Department of Speech Language Pathology and Audiology, Northeastern University, Boston, MA, USA*

(*Received 14 August 2014; accepted 4 December 2014*)

One of the principal application areas for brain-computer interface (BCI) technology is augmentative and alternative communication (AAC), typically used by people with severe speech and physical disabilities (SSPI). Existing word- and phrase-based AAC solutions that employ BCIs that utilize electroencephalography (EEG) are sometimes supplemented by icons. Icon-based BCI systems that use binary signaling methods, such as P300 detection, combine hierarchical layouts with some form of scanning. The rapid serial visual presentation (RSVP) IconMessenger combines P300 signal detection with the icon-based semantic message construction system of iconCHAT. Language models are incorporated in the inference engine and some modifications that facilitate the use of RSVP were performed such as icon semantic role order selection and the tight fusion of language evidence and EEG evidence. The results of a study conducted with 10 healthy participants suggest that the system has potential as an AAC system in real-time typing applications. Ability to construct messages with reduced physical movement demands due to RSVP and increased message construction speed and accuracy due to the incorporation of an icon-based language model in the inference process are the significant findings of this study.

Keywords: rapid serial visual presentation; augmentative and alternative communication; brain computer interface; P300 detection

1. Introduction

Augmentative and alternative communication (AAC) systems are primarily non-verbal, and are typically used by people who have severe speech and physical disabilities (SSPI). This population may include people with cerebral palsy (CP), amyotrophic lateral sclerosis (ALS), locked-in syndrome (LIS), stroke, and other physical or neurological impairments. According to some estimates, approximately 53% of people with CP [1] and 75% of people with ALS [2] use some form of AAC.

One form of AAC is speech-generating devices (SGD), or voice output communication aids (VOCA), that consist of electronic systems with buttons representing letters, words, or phrases. Users activate the buttons, such as via a touchscreen or joystick, to create messages that can be spoken aloud via a text-to-speech (TTS) engine. There is a time and generativity trade-off between letter-based systems and word- or phrase-based systems. Although letter-based systems allow for full expression in alphabetic languages, they require literacy and are often slower than word- or phrase-based systems.[3] Word- and phrase-based systems are often preferred for face-to-face communication and are usually supplemented by icons to aid in visual search and accommodate visually impaired or pre-literate users.[4]

There are several different types of BCI for potential use in AAC, including those that detect slow cortical potentials (SCP), steady-state visually evoked potentials (SSVEP), event-related synchronization and desynchronization (ERS/ERD), and event-related potentials (ERP). Of the event-related potentials, the P300 is one of the most studied waveforms. The brain generates a P300 response, a positive deflection in the scalp voltage mostly occurring in the centro-parietal region with a typical latency just over 300 ms, as a response to infrequent novel/target stimuli.[5] The focus of this work is on using the P300 signal for word- and phrase-based communication in conjunction with a tightly integrated language model for icon collections representing sentences, collectively referred to as icon-based communication.

Although many BCI communication systems are letter-based, such as the P300 Matrix Speller,[12] there has been prior and related work on icon-based BCI systems.[6,7] These icon-based BCI communication systems, however, often use layouts very similar to those of their non-BCI counterparts, supplemented by some form of scanning, such as linear or row-column.[8] Because

*Corresponding author. Email: ahani@ece.neu.edu

the entire vocabulary cannot usually be displayed all at once in these layouts, it is typically organized as arrays of icons in hierarchically nested pages, categorized according to lexical, semantic, or thematic similarity. Navigating these interfaces, even using a scanning BCI, requires that users visually locate their target icon from among the many options on the screen. The difficulty of this task is compounded as the screen size becomes larger, icons become smaller, or the navigation hierarchy becomes more complex to accommodate larger vocabularies. For users with extremely limited physical mobility or control, it can be fatiguing or prohibitively difficult to repeatedly perform the necessary head, neck, or eye movements when attempting to locate target items.

The RSVP IconMessenger interface is derived from a presentation layout called RSVP-iconCHAT, an icon-based AAC interface that was designed to minimize the amount of head, neck, and eye movements required for efficient interaction.[14] Instead of displaying all the available vocabulary, not all of which may be applicable to the current situation, RSVP-iconCHAT uses the available screen space to focus on the message under construction. Semantic frames [15] are used to subdivide the message into its major semantic roles (e.g. actor, action, object, and modifiers) and provide a set of visual fixation areas on the screen. Within each of these areas, applicable vocabulary options are displayed using rapid serial visual presentation (RSVP), a technique commonly used in experimental psychology to obscure non-relevant information and avoid distractions. RSVP has been used in letter-based AAC systems [1] and has been combined with P300 detection for letter-based communication.[9–11]

However, low typing speed in letter-based typing systems is usually an issue, especially once children and people with severe disability and fatigue use the system. In the RSVP-iconCHAT interface, the amount of required screen space and the amount of physical movement are uncoupled from vocabulary size and instead tied to the length of the desired message. Additionally, RSVP-iconCHAT does not constrain users to left-to-right message construction; instead, users are encouraged to first select a verb, or action, and may then populate any of the available semantic roles in any order.

In this study, we use the world *field* to address the verb (the action) and the corresponding semantic roles. In the current work, RSVP IconMessenger, we have not introduced field selection as this requires careful consideration regarding cognitive overhead that task switching will induce in the user. We have instead opted for a linear message-construction approach for a closed-vocabulary context, in which the fields are greedily sorted to minimize the uncertainty of the rest of the sentence given any number of initial fields. Furthermore, to account for the uncertainty in message construction order, and to improve P300-based intent inference accuracy in this closed-vocabulary context, the current work uses a language model called semantic grams, which supports utterance-based, order-agnostic word prediction.[32] The tight fusion language evidence and EEG evidence for P300 detection yields a novel icon-based BCI communication proposition that requires less screen space and less physical movement (none required, but some users with SSPI might choose to exploit EMG artifacts in EEG to their advantage) than existing systems. The results of a study with 10 participants suggest that the RSVP IconMessenger system is potentially well suited as an AAC approach for users with SSPI; however, field testing on actual target users is out of the scope of this first report on the system operation and will be considered in future work to make the ultimate assessment.

2. System description

2.1. System components

RSVP IconMessenger consists of four main components: visual presentation, EEG feature extraction, language model, and classifier. A visual presentation is employed to aid the user to construct a sentence by showing different icons. The data acquisition runs in real time during icon selection and both EEG classifiers and language model jointly make the decision. The visual representation of the system paradigm is shown in Figure 1.

2.1.1. Visual presentation

For users with SSPI, however, even this amount of gaze control can be fatiguing or prohibitively difficult. RSVP IconMessenger interface design significantly reduces the amount of head, neck, and eye movement required to visually locate the user's target item. Specifically, it reduces the number of required visual fixation areas to one large central area where RSVP of icon sequences is performed, but the user may also fixate, if able, to other locations showing previously typed icons corresponding to various semantic roles in the current utterance.

RSVP is a presentation technique in which icons are displayed as a temporal sequence at a fixed location on the screen. In RSVP IconMessenger, all icons are displayed in a single fixation area located in the center of the screen as shown in Figure 2. Available semantic roles for the current message are displayed in a horizontal bar at the top of the screen. To create an utterance using this new scheme, first the semantic role is selected automatically by the machine and then users maintain visual focus on the center of the screen while message construction proceeds for semantic role population selection. When one of the roles is selected, that field becomes activated and the system starts the

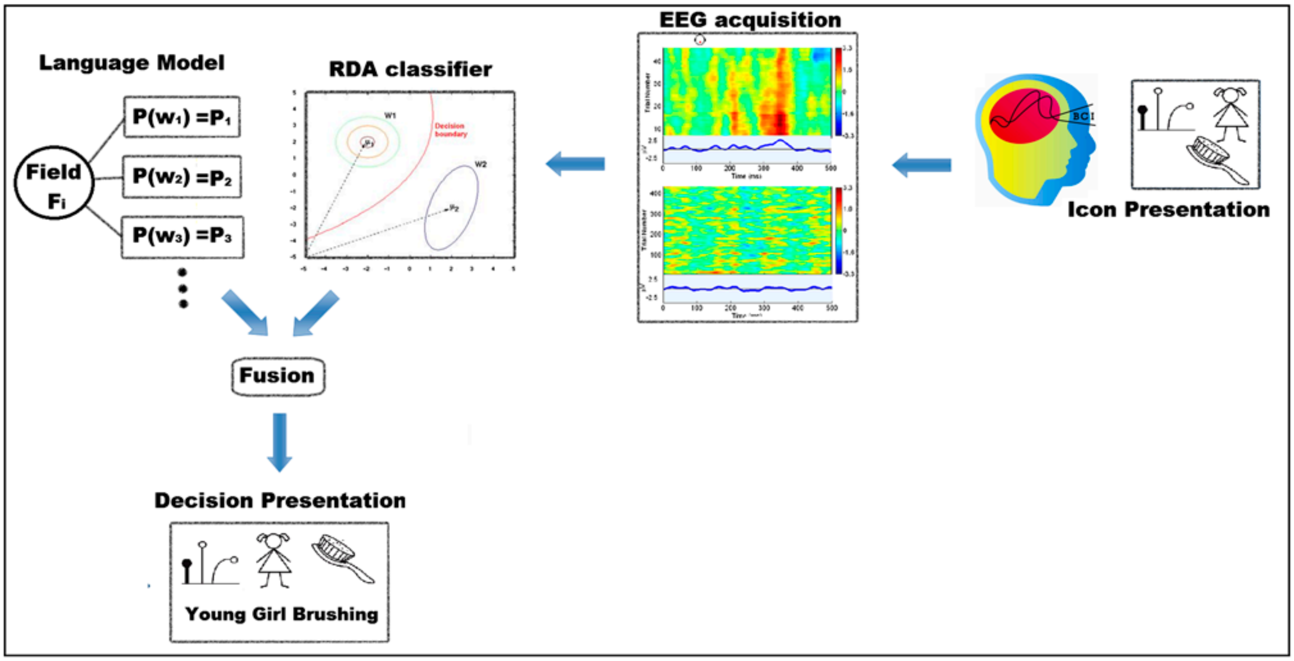


Figure 1. (Color online) The visual representation of the system paradigm. A visual presentation is employed to aid the user to construct a sentence by showing different icons. The data acquisition runs in real time during icon selection and the decision is made by both EEG classifiers and language model assistance.

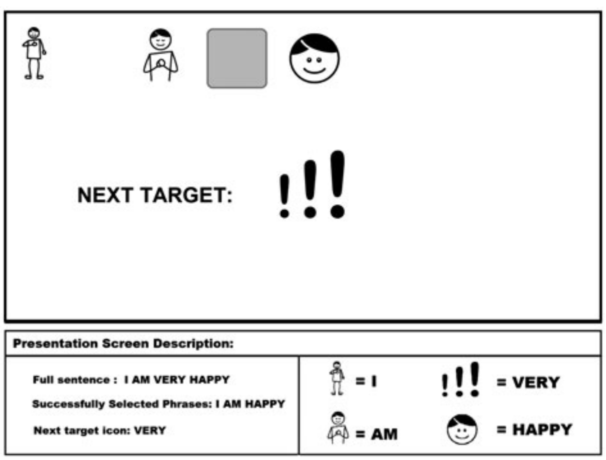


Figure 2. RSVP IconMessenger system presentation screen. On top of the figure a sample presentation screen is shown and the bottom describes the icons on the screen. The target sentence in this sample screen (as described at the bottom of figure) is 'I AM VERY HAPPY'. The corresponding participant has successfully selected 'I AM HAPPY' and the next target icon to select is 'VERY'.

role population mode selection. During role population, the system uses RSVP to display icons, representing vocabulary words that are applicable to the currently selected role. Users may select one of these available vocabulary words to populate the role or the special 'blank' value to indicate no selection. Figure 2 shows an example of the presentation screen. The target sentence is 'I AM VERY HAPPY' and the corresponding user has successfully selected 'I AM HAPPY' and the next target icon to select is 'VERY'.

For automatic field selection, using the language model properties (see section 2.1.2) of the words corresponding to different semantic role fields, we performed field ordering that minimizes the conditional entropy of remaining fields given the initial set of fields. Let F_K: $\Omega \rightarrow \mathcal{F}$ be the random variable from the vocabulary domainΩ to randomly sorted fields $\mathcal{F}$ and F_K be a field with $K = \{1, \ldots, K\}$, and K be the number of fields to be presented. Now let's assume $\mathcal{F}^*$ be the ordered set of fields based on language model probabilities and F_K^* is the kth field added to $\mathcal{F}^*$. We use algorithm 1 to find the ordering of the fields.

In algorithm (1), $\boldsymbol{H}(\bar{\mathcal{F}} \mid \bar{\mathcal{F}}^*)$ is the conditional entropy of remaining fields assuming that $\bar{\mathcal{F}}^*$ are selected. To estimate the entropy for each $\bar{\mathcal{F}}$ we define $P_{\bar{\mathcal{F}}}(\bar{\omega})$ with $\bar{\omega}$ being the combination of words (icons) corresponding to different fields within $\bar{\mathcal{F}}$. At each step, algorithm (1) calculates the entropy as:

Algorithm 1. Greedy forward selection method for field ordering.

Given $\mathcal{F} = \{F_K\}, k = \{1, ..., K\}$, and $\mathcal{F}^* = \{F_K^*\}$
for i=1:K **do**
$F_i^* = argmin_{F \in \mathcal{F}} H(\bar{\mathcal{F}}|\bar{\mathcal{F}}^*)$ with $\bar{\mathcal{F}}^* = \mathcal{F}^* - \{F\}$ and $\bar{\mathcal{F}}^* = \mathcal{F}^* \cup \{F\}$
$\mathcal{F} \leftarrow \mathcal{F} - \{F_i^*\}$
$\mathcal{F}^* \leftarrow \mathcal{F}^* \cup \{F_i^*\}$
end for

$$\begin{aligned}
H(\bar{\mathcal{F}}|\bar{\mathcal{F}}^*) &= \sum_{\bar{\omega}^*} P_{\bar{\mathcal{F}}^*}(\bar{\omega}^*)\left[-\sum_{\omega} P_{\bar{\mathcal{F}}|\bar{\mathcal{F}}^*}(\bar{\omega}|\bar{\omega}^*) log P_{\bar{\mathcal{F}}|\bar{\mathcal{F}}^*}(\bar{\omega}|\bar{\omega}^*)\right] \\
&= -\sum_{\bar{\omega}^*}\sum_{\bar{\omega}^*} P_{\bar{\mathcal{F}}^*}(\bar{\omega}^*) \frac{P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*)}{P_{\bar{\mathcal{F}}^*}(\bar{\omega}^*)} log \frac{P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*)}{P_{\bar{\mathcal{F}}^*}(\bar{\omega}^*)} \\
&= -\sum_{\bar{\omega}^*}\sum_{\bar{\omega}^*} P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*) log \frac{P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*)}{P_{\bar{\mathcal{F}}^*}(\bar{\omega}^*)} \\
&= -\sum_{\bar{\omega}^*}\sum_{\bar{\omega}^*} P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*) log P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*) \\
&\quad + \sum_{\bar{\omega}^*}\sum_{\bar{\omega}^*} P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*) log P_{\bar{\mathcal{F}}^*}(\bar{\omega}^*)
\end{aligned} \tag{1}$$

The first term in the result of (1) is constant while $\bar{\mathcal{F}}$ and $\bar{\mathcal{F}}^*$ change; therefore, the problem of minimizing (1) reduces to:

$$\min \boldsymbol{H}(\bar{\mathcal{F}}|\bar{\mathcal{F}}^*) = \min \sum_{\bar{\omega}^*}\sum_{\bar{\omega}^*} P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*) log P_{\bar{\mathcal{F}}^*}(\bar{\omega}^*) \tag{2}$$

where for each $\bar{\omega}^*$:

$$P_{\bar{\mathcal{F}}^*}(\bar{\omega}^*) = \frac{N_{\bar{\omega}^*}}{N} \tag{3}$$

$$P_{\bar{\mathcal{F}}\bar{\mathcal{F}}^*}(\bar{\omega},\bar{\omega}^*) = \sum_{\bar{\omega}} \frac{N_{\bar{\omega}\bar{\omega}^*}}{N} \tag{4}$$

Here N is the number of all possible combinations of words within vocabulary, $N_{\bar{\omega}^*}$ is the number of repetitions of word $\bar{\omega}^*$ and $N_{\bar{\omega}\bar{\omega}^*}$ is the number of repetitions of $\bar{\omega}$ and $\bar{\omega}^*$ together in a sentence.

In the current presentation scheme, we consider each presented stimulus (icon) as a *trial*. We then define a *sequence* as a group of consecutive trials. The length of a sequence can differ in general and the maximum size of a sequence is the number of all the icons in the corresponding field. An *epoch* is a group of consecutive sequences that leads to a decision. A decision can be made after a single sequence or after a number of sequences until a confidence is reached based on classifier results (see section 2.1.4).

2.1.2. Language model

There is a well-documented dearth of large corpora of authentic utterances by AAC users,[17] and so approximations or simulations are often used.[18,19] The Crowdsourced Corpus of AAC-Like Communications was chosen as a training corpus because it is freely available and contains relatively recent language constructions.[20] A counting language model using semantic grams of all lengths 1 through 5, with plus-one smoothing, was constructed based on this corpus. Semantic grams are similar to a 'bag of words' language model that functions at the level of utterances.

We chose the 150 most frequently appearing words from the Crowdsourced AAC corpus.[20] This list ignored synonyms, tenses, negations, and conjugation and it was further limited by using PorterStemmer, a modified implementation of the original Porter stemming algorithm [21] to a stem that was available as an icon in the Widgit icon set.[22]

2.1.3. Feature extraction

RSVP relies on temporal separation of the stimuli rather than spatial, and hence feature extraction starts by extracting stimulus-time-locked EEG responses corresponding to the visual stimuli. Physiologically the most relevant ERP signal components are expected to occur within the first 500 ms following the stimuli. Therefore, we apply a bandpass filter with linear phase to EEG data, downsample the data to 128 Hz and extract a [0,500) ms portion of the filtered data following each stimulus. After time-locked temporal information extraction, we apply principal component analysis (PCA) for dimensionality reduction to get rid of zero or negligible power components (frequency bands, since PCA on time-delay vectors acts as a bank of FIR bandpass filters). We concatenate the data from all channels to form the feature vector for each stimulus.

We use regularized discriminant analysis (RDA) to further decrease the dimension of the feature vectors for use in fusion with the language model. RDA is a modification of quadratic discriminant analysis (QDA). QDA yields the optimal minimum-expected-risk Bayes classifier under the assumption of multivariate Gaussian class distributions. This classifier depends on the inverses of covariance matrices for each class, which are estimated from training data. In BCI, to keep the calibration phase short, few training samples are acquired – especially for the positive intent class. Therefore, the sample covariance estimates may become singular or ill-conditioned for high-dimensional feature vectors. RDA applies shrinkage and regularization on class covariance estimates.[23]

Upon shrinkage and regularization, based on the initial assumption that class conditional distributions follow Gaussian distributions, RDA scores are calculated by log-likelihood ratio of the class conditional distributions to be used as EEG evidence.

$$\delta(x) = log\left(\frac{f(\mathrm{x}|\hat{\mu}_1, \bar{\Sigma}_1(\lambda,\gamma))\hat{\pi}_1}{f(\mathrm{x}|\mu_0, \bar{\Sigma}_0(\lambda,\gamma))\hat{\pi}_0}\right) \tag{5}$$

where $\delta(x)$ is the RDA score for data point, x $\hat{\pi}_c$ is the estimated prior probability for class $c = \{0, 1\}$ (corresponding to the non-target and target classes, respectively), λ is the shrinkage parameter, γ is the regularization parameter, $\bar{\Sigma}_c(\lambda,\gamma)$ is the maximum-likelihood estimate covariance matrix of class after shrinkage and regularization and $f(\mathrm{x}|\hat{\mu}_c, \bar{\Sigma}_c(\lambda,\gamma))$ is the multivariate class conditional Gaussian distribution function with estimated mean $\hat{\mu}_c$ and covariance $\bar{\Sigma}_c(\lambda,\gamma)$.

In order to use in fusion with the language model probabilities, we calculate the class conditional probability density functions of RDA scores using kernel density estimation on training data. That is for each RDA score in class $c = \{0, 1\}$, we calculate

$$f(\delta(x)|c) = \frac{1}{n_c - 1}\sum_{\mathrm{x}' \in C_c, \mathrm{x}' \neq \mathrm{x}} K_{h_k}(\delta(\mathrm{x}) - \delta(\mathrm{x}')) \tag{6}$$

where $f(\delta(x)|c)$ is the conditional probability density function of the RDA scores given the class c, C_c is the set of data points in class c and K_{h_k} is the kernel function with bandwidth K_h. We use Silverman's rule of thumb to calculate the bandwidth.[24]

2.1.4. *Classifier*

The RSVP IconMessenger system fuses EEG evidence and language model information to make a joint decision. Let $c = \{0, 1\}$ be the random variable representing the class labels corresponding to target and non-target classes, respectively, and $\mathbf{x} : \Omega \to \mathbb{R}^d$ is the EEG evidence, such that $\mathbf{x}_{\mathrm{t,i,r}}$ corresponds to epoch $t \in \mathbb{N}$ icon $i \in I$ and repetition $r = \{1, \ldots, R_i\}$. Now, the posterior probability for the class being c for the ith icon in epoch t is:

$$P(c_{t,i} = \varsigma|\delta(\mathbf{x}_{\mathrm{t,i}}), \boldsymbol{w}_t = \boldsymbol{\omega}) = \frac{f(\delta(\mathrm{x}_{\mathrm{t,i}}), \boldsymbol{w}_t = \boldsymbol{\omega}|c_{t,i} = \varsigma)P(c_{t,i} = \varsigma)}{f(\delta(\mathbf{x}_{\mathrm{t,i}}), \boldsymbol{w}_t = \boldsymbol{\omega})} \tag{7}$$

where $\delta(\mathbf{x}_{\mathrm{t,i}}) = \{\delta(\mathbf{x}_{\mathrm{t,i,1}}), \ldots, \delta(\mathbf{x}_{\mathrm{t,i,R_i}})\}$, i_t is the number of icons selected before epoch t, $\boldsymbol{w}_t = \{w_{i_t}, w_{i_{t-1}}, \ldots, w_1\}$ is the random variable of all previously selected icons before the tth epoch and $\boldsymbol{\omega} = \{\omega_1, \omega_2, \ldots, \omega_{i_t}\}$ corresponds to the set of icons.

Here we make our first assumption: *the EEG evidence for the current icon and the context information coming from the language model due to the previously selected icons are independent given the class label.* Using this assumption, we write (7) as:

$$P(c_{t,i} = \varsigma|\delta(\mathbf{x}_{\mathrm{t,i}}), \boldsymbol{w}_t = \boldsymbol{\omega}) = \frac{f(\delta(\mathrm{x}_{\mathrm{t,i}})|c_{t,i} = \varsigma)P(\boldsymbol{w}_t = \omega|c_{t,i} = \varsigma)p(c_{t,i} = \varsigma)}{f(\delta(\mathbf{x}_{\mathrm{t,i}}), w_t = \omega)} \tag{8}$$

Furthermore, we make our second assumption: *EEG features for an icon in different sequences are conditionally independent given the class label*. Therefore, by the Bayes rule we have:

$$P(c_{t,i} = \varsigma|\delta(\mathbf{x}_{\mathrm{t,i}}), w_t = \omega) = P(c_{t,i} = \varsigma)\frac{\left(\prod_{r=1}^{R_i} f(\delta(\mathrm{x}_{\mathrm{t,i}})|(c_{t,i} = \varsigma)\right)P(w_t = \omega|c_{t,i} = \varsigma)}{f(\delta(\mathbf{x}_{\mathrm{t,i}}), w_t = \omega)} \tag{9}$$

Now, we make our third assumption: *at each sequence, there is only one target response.* This assumption is reasonable since a user is expected to respond only for the target symbol. With this assumption and (9) we have:

$$P(i{=}1|\delta(\mathbf{x}_{\mathrm{t,i}}),\boldsymbol{w}_t{=}\boldsymbol{\omega}) = \frac{\frac{P(c_{t,i}=1|\delta(\mathrm{x}_{\mathrm{t,i}}),\boldsymbol{w}_t=\boldsymbol{\omega})}{P(c_{t,i}=0|\delta(\mathrm{x}_{\mathrm{t,i}}),\boldsymbol{w}_t=\boldsymbol{\omega})}}{\sum_{i' \in \mathrm{I}}\frac{P(c_{t,i}=1|\delta(\mathrm{x}_{\mathrm{t,i'}}),\boldsymbol{w}_t=\boldsymbol{\omega})}{P(c_{t,i}=0|\delta(\mathbf{x}_{\mathrm{t,i'}}),\boldsymbol{w}_t=\boldsymbol{\omega})}} \tag{10}$$

Finally, the decision for an icon to be the target is made by:

$$\hat{i} = argmax_i P(i = 1|\delta(\mathbf{x}_{\mathrm{t,i}}), w_t = \boldsymbol{\omega}) \tag{11}$$

where i represents the selected icon.

To avoid frequent correction requirements, we use (11) as the confidence measure for the selected icon. We use this metric as a dynamic stopping criterion for each epoch, which allows the system to have a variable number of sequences for each epoch, until the confidence threshold is achieved. In this experiment, we have set this threshold as 0.9, which means that the posterior probability of the most likely icon must exceed 0.9 for the epoch to stop. Relaxing this threshold by lowering it could result in faster selections at the cost of more errors, so net typing speed needs to be assessed with simulations (but we don't discuss that here as it is out of the scope of this paper).

2.2. ***Operation modes***

The RSVP IconMessenger system has three major modes of operation: calibration, copy icon, and free closed-vocabulary expression.

2.2.1. Calibration

In calibration mode, the statistics for the EEG data for a user are learned. In this mode, the user is asked to show positive intent to predefined targets in a series of sequences. Using the EEG data as a response to the predefined target and the non-target symbols, the RDA scores and the corresponding class conditional distributions are calculated to be fused with the language model in one of the testing modes.

2.2.2. Copy icon

Copy icon is one of the testing modes in which the user is presented with two sentences constructed by a sequence of icons. The first sentence is a complete phrase and the second sentence is a copy of the first sentence, missing one or more icons. The user is asked to complete the second sentence by selecting the missing icons from a series of sequences. This mode has two major tasks:

- Mastery task: in this task, the users are asked to complete the given sentence by selecting only one missing icon. The sample sentences presented to the users are divided into four groups based on the difficulty of icon selection according to language model probabilities. For example, in level 1 the language model probability for the target icon is relatively high compared to the non-target icon with the highest probability. As the level value increases, the language model probability of the target icon decreases while the non-target probabilities increase. At the higher levels, the EEG evidence should be dominant for a correct selection. There is a delete icon with fixed probability that users can select to undo their previous selection.
- Full Phrase Task: the users are asked to select all the icons in a given sentence. The fields are automatically selected based on the method described in section 2.1.1. Users can select the delete icon to go back to the previous field and make a new icon selection for that field.

2.2.3. Free closed-vocabulary expression

The third operating mode is free selection mode in which users can make free icon selections as they desire. The system automatically chooses the field based on the field selection method described in section 2.1.1. Users can choose the delete icon to go back to the previous field and choose the blank icon if the corresponding field is not a part of their desired sentence.

2.2.4. Simulation

In the simulation mode, we use the target and non-target EEG responses from the precollected calibration data and perform a kernel density estimation on these responses. Then during simulation we draw samples from these densities to obtain EEG evidence for target and non-target symbols. We fuse the EEG evidence with the language model to compute the posterior probabilities of the symbols that are used for decision-making (see also section 2.1.4). We apply our simulation to the copy icon task and report the estimated performance as number of sequences to write a symbol, time to complete the task, and probability of phrase completion. We utilize the simulation mode to optimize system parameters such as backspace probability, maximum number of sequences in an epoch, etc., for different users.

3. Methods

3.1. Experimental setup

For this experiment, the system was configured with a total of 124 icons that can construct different sentences of one to five words in length. Each icon corresponds to a word that belongs to a different field in the sentence. In the current system, fields are verb, subject, object, subject participant, and object modifier. In addition to icons for these five fields, there is a blank symbol for no selection and a delete icon for deleting the previous selection.

3.1.1. Participants

Ten healthy adults between 24 and 31 years of age were recruited from the university student population (none involved in this project directly). These participants consisted of seven women and three men, with a mean age of 26 years (standard deviation 2.4 and range 7 years).

3.1.2. EEG recording and protocol

We used g.USBamp biosignal amplifiers with active g.Butterfly electrodes attached to g.GAMMAcaps from g.Tec (Graz, Austria). The EEG was sampled at a 256 Hz rate. According to the International 10/20 electrode placement system, we used sites Fp1, Fp2, F3,F4, Fz, Fc1, Fc2, Cz, P1, P2, C1, C2, Cp3, Cp4, P5, and P6. The amplifiers' built-in nonlinear-phase (Butterworth) 0.5–60 Hz bandpass filter and 60 Hz notch filters were enabled.

3.1.3. Session instructions

All users participated in six sessions of data collection. Each session lasted approximately 2 hours. The first four sessions consisted only of multiple calibration

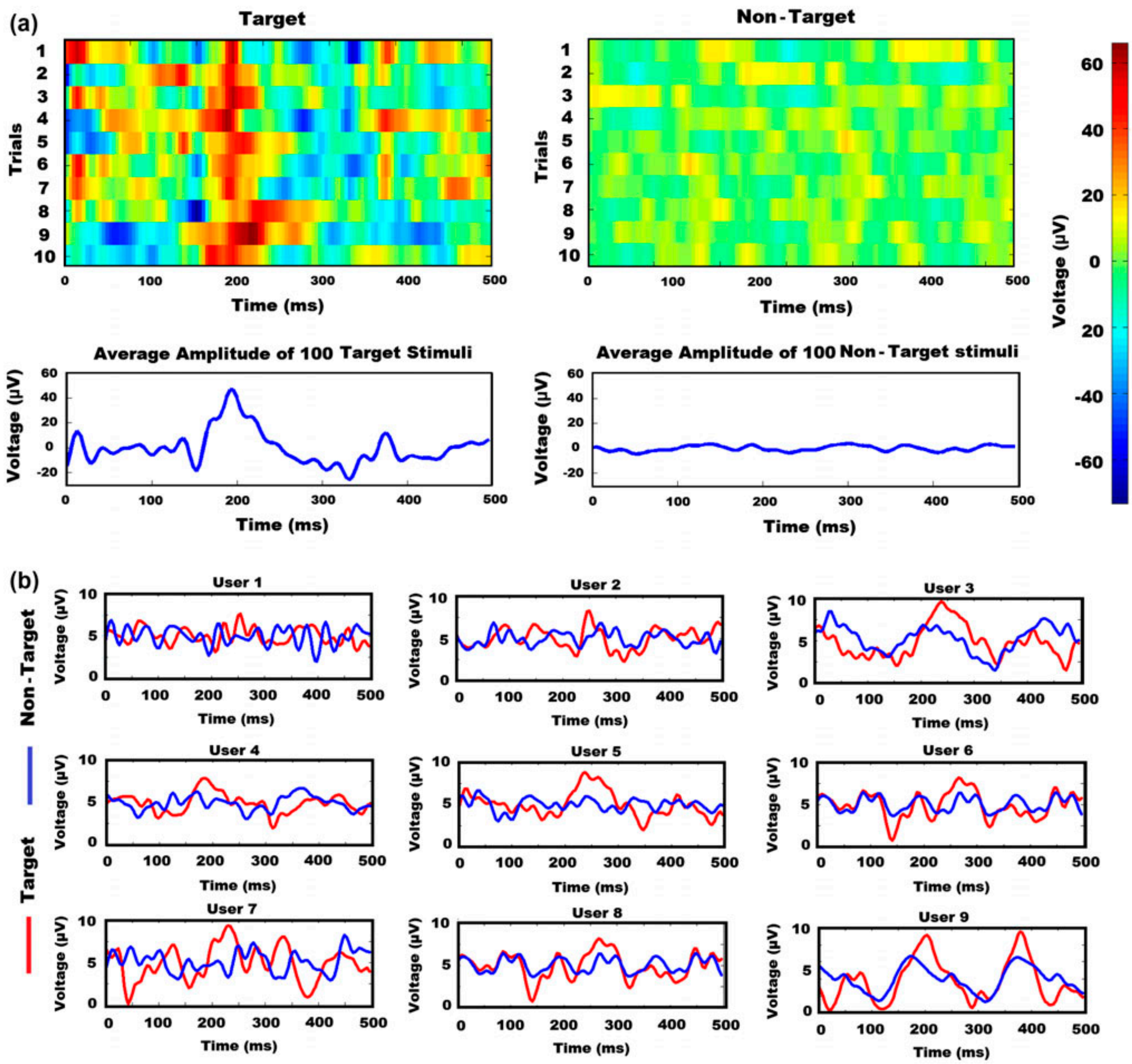


Figure 3. (Color online) (a) The ERP images for subject 10 at channel Cz. The images show electrical activity following for target and non-target classes. The top figures show the voltage value for 100 trials and the bottom traces are the EEG signals averaged over trials. The zero point corresponds to the stimulus onset. (b) ERP response corresponding to target (red line) and non-target (blue line) averaged over 100 trails for users 1–9 at channel Cz.

phases. In each individual calibration session, users were presented with 100 sequences of icons as stimuli with inter-stimulus intervals of 200 ms. Users were asked to rest between calibration phases and continue once they felt ready.

Using simulation, as explained in section 2.2.4, on the calibration data collected from each subject, the maximum number of sequences in an epoch and the delete icon probability parameters were optimized for each subject.

In the fifth session, first, the users had a single calibration phase to train the classifier, and the classifier parameters were calculated accordingly. The maximum number of sequences in an epoch and the delete icon probability parameters were optimized using the simulation mode on the calibration data collected from the first four sessions for each subject. Then, the users were asked to complete all the levels of the Mastery Task. Each level had eight sentences, with a total of 32 sentences, and a level is successfully completed when at

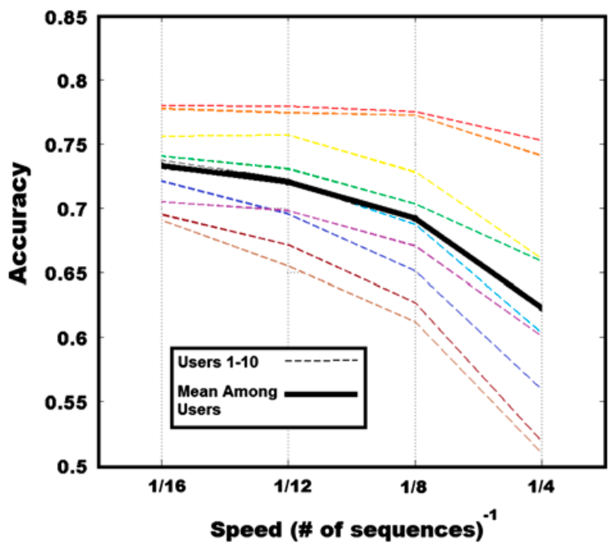


Figure 4. (Color online) Accuracy of typing versus inverse of maximum allowed number of sequences in an epoch.

least 75% of the sentences at each level (at least six out of eight) have been successfully spelled. If the users could not satisfy the level completion requirement, the sentences that were unsuccessful were repeated until the requirement was satisfied. For each sentence, if the selected icon was wrong, the user could delete it by selecting the delete icon and trying again. There were two stopping criteria for each sentence: (1) three consecutive wrong icon selections and (2) a time limit of 60 seconds for each sentence. If either of these two criteria was reached, the system skipped to the next sentence. The task is considered to be successful if all the levels are completed in the session. Each sequence in an epoch included 20 icons with the highest language model probability corresponding to the specific field under consideration. We made sure that the sentences selected always included the target icon in the top 20 according to the language model. Extensions for larger vocabulary contexts are possible, but are not considered in this work.

In the sixth session, first, the users had a single calibration phase to train the classifier. Then they were asked to perform the Full Phrase Task. The system skips to the next sentence if a user reaches the time limit of 5 minutes for each sentence or if the number of selected icons is equal to the number of icons in the original sentence. Similarly to the fifth session, the inter-stimulus interval was 200 ms and each sequence had 20 icons corresponding to each field.

4. Results and discussion

Our offline analysis on field ordering selection (as explained in section 2.1.1) showed that the optimum ordering of the fields for the vocabulary used is given as: verb, subject, object, subject participant, and object modifier. Retrospectively, this makes sense and validates the outcome offered by the greedy forward search algorithm used and the uncertainty objective employed.

Figure 3(a) shows the ERP images for subject 10 at channel Cz. The images show electrical activity following for target and non-target classes. The top figures show the voltage value for 100 trials and the bottom traces are the EEG signals averaged over trials. The zero point corresponds to the stimulus onset. Figure 3(b) shows the ERP response corresponding to target (red line) and non-target (blue line) averaged over 100 trials at channel Cz for users 1–9.

For each user, Monte Carlo simulations (as explained in section 2.2.4) are applied on the calibration data collected in the first four sessions. The results of simulations are calculated for four different values for the maximum number of sequences in an epoch: 4, 8, 12 and 16. These results are reported in terms of accuracy vs speed in Figure 4 for each user. The speed is defined as the inverse of the maximum number of sequences in an epoch, and accuracy is calculated as the ratio of the correctly completed sentences to the total number of sentences. Figure 4 also shows the accuracy results averaged over all users. We observe that as the speed decreases, on average the system can achieve around 75% accuracy.

The results of the fifth session demonstrated that seven out of 10 users could successfully complete all four levels of the Mastery Task. The remaining three users were only successful in passing the first three levels. We report the typing performance for each sentence as the number of sequences required to successfully select a desired icon – including all errors and deletions until the correct desired icon is achieved. These typing performance values for all users during the Mastery Task are summarized in Figure 5. The green bars show the number of sequences for a successful icon selection. If the level completion requirement (six correct completions out of eight sentences) is satisfied after presenting fewer than eight sentences, some sentences are not presented to the user. These sentences are marked with blue bars in the figure (fixed at the value of 10 for demonstration purposes). Red bars show the sentences that were unsuccessful after one or multiple repetitions of that sentence until the stopping criteria are reached.

We report the typing performance of the sixth session for each icon in each sentence as the number of sequences required to successfully select that icon – including all errors and deletions until the correct desired icon is achieved. These typing performance values for all users during the Full Phrase Task are summarized in Table 1. ‘N’ means ‘No Icon’ and marks the empty fields. This means that the corresponding field was not a

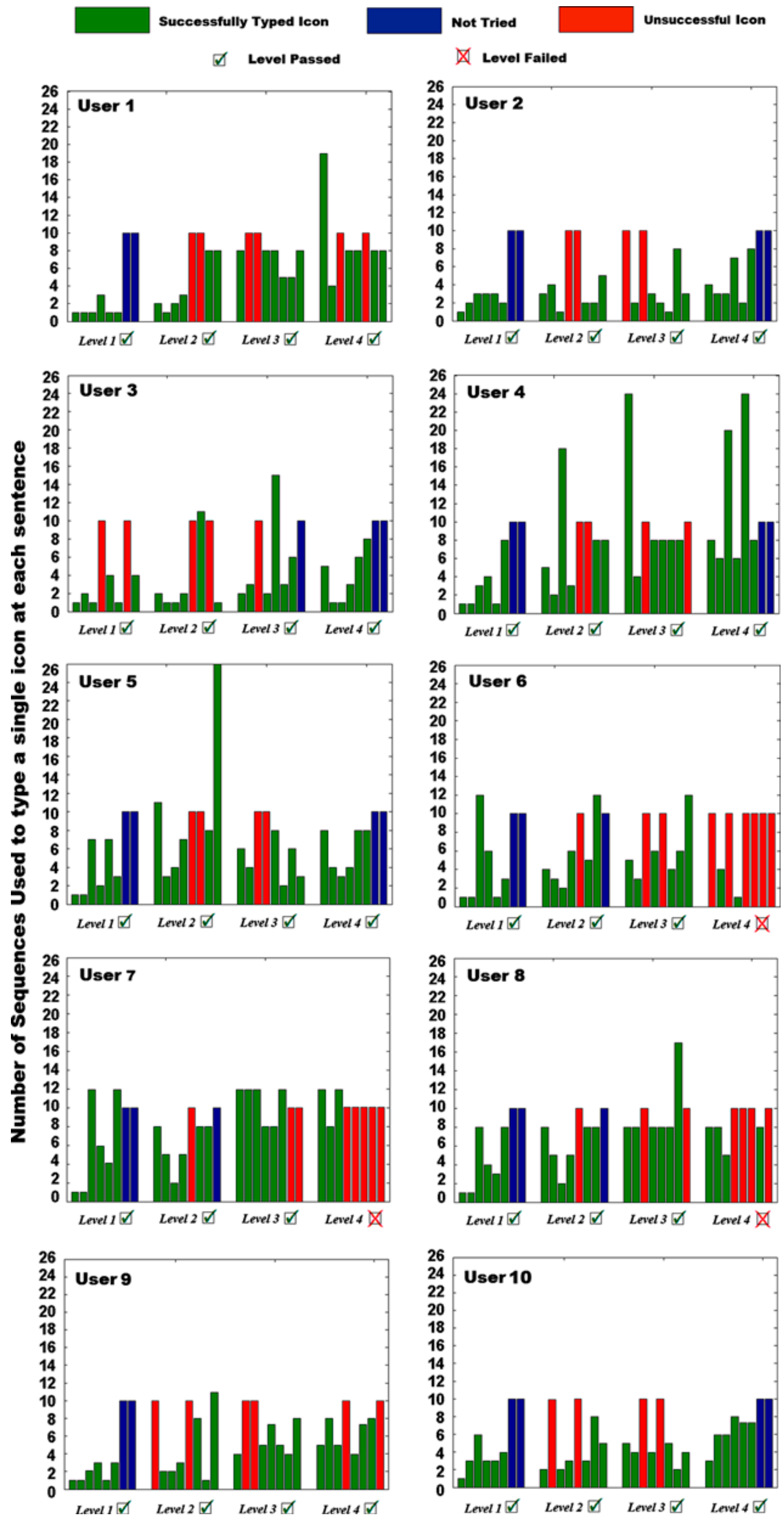


Figure 5. (Color online) Typing performance of all users after session five, Mastery Task. Green bars show the number of sequences for a successful icon selection. Blue bars mark the sentences that were not tried by the user due to the achievement of the level completion criterion. Red bars show the sentences that were unsuccessful.

Table 1. Typing performance of all users after session six, Full Phrase Task. Numbers show the number of sequences for a successful icon selection at each sentence. 'N' means 'No Icon' and marks the empty fields. This means that field was not a part of the corresponding sentence. 'W' means 'Wrong Selection' and marks the icons that were unsuccessful.

	User 1					User 2				
	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5
Sentence 1	1	N	8	2	8	2	N	8	2	8
Sentence 2	N	N	W	W	W	N	N	4	1	2
Sentence 3	W	N	W	N	W	8	N	3	N	2
Sentence 4	1	N	8	N	W	1	N	8	N	3
Sentence 5	1	N	8	N	8	1	N	7	N	7
Sentence 6	W	N	W	N	W	1	N	7	N	2
Sentence 7	1	N	8	W	W	1	N	4	1	8
Sentence 8	W	N	W	N	W	1	N	1	N	7
	User 3					User 4				
	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5
Sentence 1	W	N	W	W	W	1	N	8	1	20
Sentence 2	N	N	8	1	2	N	N	8	1	3
Sentence 3	1	N	8	N	2	3	N	8	N	1
Sentence 4	8	N	8	N	1	3	N	8	N	8
Sentence 5	1	N	8	N	W	1	N	8	N	6
Sentence 6	1	N	8	N	3	W	N	W	N	W
Sentence 7	1	N	8	W	W	8	N	8	2	21
Sentence 8	15	N	8	N	5	8	N	8	N	W
	User 5					User 6				
	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5
Sentence 1	5	N	6	3	2	5	N	3	2	4
Sentence 2	N	N	8	1	1	N	N		1	3
Sentence 3	1	N	2	N	1	4	N	61	N	1
Sentence 4	1	N	2	N	1	1	N	5	N	W
Sentence 5	1	N	4	N	1	3	N	8	N	1
Sentence 6	1	N	5	N	3	3	N	7	N	6
Sentence 7	1	N	3	6	6	2	N	6	4	6
Sentence 8	2	N	7	N	4	8	N	8	N	6
	User 7					User 8				
	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5
Sentence 1	W	N	W	W	W	W	N	W	W	W
Sentence 2	N	N	7	7	7	N	N	W	W	W
Sentence 3	W	N	W	N	W	8	N	17	N	8
Sentence 4	W	N	W	N	W	W	N	W	N	W
Sentence 5	W	N	W	N	W	8	N	8	N	8
Sentence 6	1	N	6	N	8	8	N	8	N	W
Sentence 7	8	N	8	2	21	8	N	8	5	28
Sentence 8	24	N	8	N	35	8	N	8	N	8
	User 9					User 10				
	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5	Icon 1	Icon 2	Icon 3	Icon 4	Icon 5
Sentence 1	W	N	W	W	W	1	N	8	8	6
Sentence 2	N	N	7	1	5	N	N	8	1	4
Sentence 3	W	N	W	N	W	W	N	W	N	W
Sentence 4	7	N	7	N	6	1	N	5	N	4
Sentence 5	1	N	7	N	8	1	N	5	N	8
Sentence 6	1	N	6	N	5	1	N	2	N	7
Sentence 7	W	N	W	W	W	1	N	4	W	W
Sentence 8	8	N	8	N	6	2	N	3	N	6

part of the sentence. 'W' mean 'Wrong Selection' and marks the icons that were unsuccessful.

Farwell and Donchin [27] showed that ERPs that contain P300 can be exploited for designing EEG-based BCI typing systems. Following their approach our results show good classifier performance and high accuracies in such typing interfaces for user intent detection systems by employing P300 detection (Figure 4).

The approach that most typing systems use is mainly to distribute symbols across a matrix (matrix speller).[6,7,12] There are some disadvantages in matrix speller systems. One is poor P300 signal quality as a result of flashing rows and columns for the presentation of a symbol. Most importantly, lack of precise gaze control is very common in many potential users and results in reduced performance of the system.[25,26] Different BCI typing systems use different presentation schemes to overcome this problem and the results are usually compared with the matrix presentation paradigm in terms of speed and accuracy.[28–31] We proposed using RSVP, in which icons appear on the screen sequentially, at a predefined fixed location on the screen and in a pseudo-random order. Real-time typing tasks in our study during sessions 5 and 6 show successful implementation of RSVP paradigm in typing systems, as shown in Figure 5 and Table 1.

The language model can enhance typing speeds and improve BCI typing systems' performance. We employed a probabilistic language model during the intent detection process to define a priori the potential target characters during the classification task.

5. Conclusion

We described RSVP IconMessenger, which is a brain-interfaced icon by icon (word by word) typing system. The complete BCI system consisted of three major modules: data acquisition, presentation, and inference. The system currently has three major modes of operation: calibration, copy icon, and free closed-vocabulary expression. Our study examined EEG data from 10 healthy users and the results indicate that the overall system concept is successfully validated for different modes of operation. Of course, healthy users are no substitute for actual users with SSPI, therefore future work with field tests on target populations remains as the next step.

Calibration results showed up to 94% accuracy of typing in some of the sessions among some users. This system uses a visual stimulus presentation paradigm that does not require precise gaze control. Increase in communication speed due to icon-based expression compared to letter-by-letter typing systems, such as our RSVP KeyboardTM, could improve the efficacy of communication through this interface in closed-vocabulary settings. A system that combines letter and icon typing could enable users to switch between fast closed-vocabulary communication and relatively slower open-vocabulary communication.

The current language model trained on a corpus that is not ideal as, once again, pretend AAC-system users are no substitute for real AAC-system users. Personalized and adaptive language models based on individual users' language styles and preferences and context-dependent closed-vocabulary options and models would provide significant improvement for the system and remain as future work.

Acknowledgements

This work was supported by the National Science Foundation [IIS-0914808, CNS-1136027, IIS-1149570, SMA-0835976], the National Institutes of Health [R01DC009834], and the National Institute on Disability and Rehabilitation [H133E140026].

References

[1] Jinks A, Sinteff B. Consumer response to AAC devices: acquisition, training, use, and satisfaction. Augment Altern Commun. 1994;10:184–190.
[2] Ball LJ, Beukelman DR, Pattee GL. Acceptance of augmentative and alternative communication technology by persons with amyotrophic lateral sclerosis. Augment Altern Commun. 2004;20:113–122.
[3] Todman J, Alm N, Elder L. Computer-aided conversation: a prototype system for nonspeaking people with physical disabilities. Appl Psycholinguist. 1994;15:45–73.
[4] Beukelman D, Mirenda P. Augmentative and alternative communication: supporting children and adults with complex communication needs. Baltimore (MD): Brookes; 2006.
[5] Sutton S, Braren M, Zubin J, John E. Evoked potential correlates of stimulus uncertainty. Science. 1965;150:1187–1188.
[6] Bayliss JD, Inverso SA, Tentler A. Changing the P300 brain computer interface. Cyberpsychol Behav. 2004;7:694–704.
[7] Hoffmann U, Vesin J, Ebrahimi T, Diserens K. An efficient P300-based brain–computer interface for disabled subjects. J Neurosci Methods. 2008;167:115–125.
[8] Nijboer F, Sellers EW, Mellinger J, Jordan MA, Matuz T, Furdea A, Halder S. A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. Clin Neurophysiol. 2008;119:1909–1916.
[9] Orhan U, Erdogmus D, Roark B, Oken B, Purwar S, Hild K, Fowler A, Fried-Oken M. Improved accuracy using recursive bayesian estimation based language model fusion in ERP-based BCI typing systems. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, 2012;2497–2500.
[10] Orhan U, Erdogmus D, Roark B, Oken B, Fried-Oken M. Offline analysis of context contribution to ERP-based typing BCI performance. J Neural Eng. 2013;10(6):066003.
[11] Orhan U, Hild KE, Erdogmus D, Roark B, Oken B, Fried-Oken M. RSVP keyboard: an EEG based typing interface. In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing, 2012;645–648.

[12] Acqualagna L, Treder MS, Schreuder M, Blankertz B. A novel brain-computer interface based on the rapid serial visual presentation paradigm. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, 2010; 2686–2689.

[13] Krusienski DJ, Sellers EW, McFarland DJ, Vaughan TM, Wolpaw JR. Toward enhanced P300 speller performance. J Neurosci Methods. 2008;167:15–21.

[14] Wiegand K, Patel R, Erdogmus D. Leveraging semantic frames and serial icon presentation for message construction. In Proceedings of the ISAAC, 2010.

[15] Fillmore CJ. Frame semantics and the nature of language. Ann N Y Acad Sci. 1976;280:20–32.

[16] Mackenzie IS, Felzer T. SAK: scanning ambiguous keyboard for efficient one-key text entry. ACM Trans Comput Hum Interact. 2010;17:11:1–11:39.

[17] Lesher G, Sanelli C. A web-based system for autonomous text corpus generation. In Proceedings of ISAAC. 2000, Washington DC, USA.

[18] Wandmacher T, Antoine J. Training Language Models without Appropriate Language Resources: Experiments with an AAC system for disabled people. In Proceedings of LREC. 2006.

[19] Trnka K, McCoy K. Corpus studies in word prediction. In Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility. 2007;195–202.

[20] Vertanen K, Kristensson PO. The imagination of crowds: conversational AAC language modeling using crowdsourcing and large data sources. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2011;700–711.

[21] Porter MF. An algorithm for suffix stripping. Program. 1980;14:130–137.

[22] Detheridge T, Whittle H, Detheridge C. Widgit Rebus Symbol Collection. Cambridge (UK): Widgit Software; 2002.

[23] Friedman J. Regularized discriminant analysis. J Am Stat Assoc. 1989;84(405):165–175.

[24] Silverman B. Density estimation for statistics and data analysis. Boca Raton: Chapman and Hall/CRC; 1998.

[25] Aloise F, Schettini F, Arico P, Salinari S, Babiloni F, Cincotti F. A comparison of classification techniques for a gaze-independent p300-based brain–computer interface. J Neural Eng. 2012;9(4):045012.

[26] Kaufmann T, Schulz S, Grunzinger C, Kubler A. Flashing characters with famous faces improves erp-based brain–computer interface performance. J Neural Eng. 2011;8(5):056016.

[27] Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol. 1988;70:510–523.

[28] Treder M, Schmidt N, Blankertz B. Gaze-independent brain–computer interfaces based on covert attention and feature attention. J Neural Eng. 2011;8(6):066003.

[29] Liu Y, Zhou Z, Hu D. Gaze independent brain–computer speller with covert visual search tasks. Clin Neurophysiol. 2011;122:1127–1136.

[30] S. Chennu A, Alsufyani M, Filetti A, Owen M, Bowman H. The cost of space independence in p300-bci spellers. J Neuroeng Rehabil. 2013;10(82):1–13.

[31] Zhou Z, Yin E, Liu Y, Jiang J, Hu D. A novel task-oriented optimal design for P300-based brain–computer interfaces. J Neural Eng. 2014;11(5):056003.

[32] Wiegand K, Patel R. Non-syntactic word prediction for AAC. Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies, Association for Computational Linguistics. 2012;28–36.