Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-86

SPELLING IN PARALLEL: TOWARDS A RAPID, SPATIALLY INDEPENDENT BCI

J.M. Stivers[1], V.R. de Sa[1]
[1] Cognitive Science Department, University of California, San Diego

E-mail: jstivers@ucsd.edu

ABSTRACT: BCI Spellers for end-users utilize numerous different techniques, but many require that stimuli in different areas of the screen be foveated for best performance. Spatial independence, however, is of considerable value for patients suffering from locked-in syndrome, which significantly attenuates their capacity for voluntary movements. To this end, we have designed a 10-segment library of letter subsets, which combinatorially create the letters of the alphabet. Segments can thus be centrally presented, allowing letters to be cued in parallel while maintaining the spatial independence of RSVP-style spellers. A 68% segment-classification accuracy yields a reasonably rapid speller, with several avenues for maximizing accuracy and information transfer rate.

INTRODUCTION

Neurodegenerative diseases have an increasingly significant impact on public health as life expectancies and treatment strategies improve. Locked-in syndrome (LIS) in particular – whether caused by injury or illness – poses significant challenges for patient and healthcare professionals. While the inability to communicate needs or discomforts can have a deleterious effect on one's health, the lack of social interaction can also pose a significant issue. P300 BCI spellers are a popular technique for ameliorating these challenges. In the original P300 system developed by Farwell and Donchin [1] the user observes a screen with a grid of symbols; individual rows and columns are flashed pseudo-randomly and the user is told to count the times their target symbol flashes. Due to the large size of the letter grid and the small size of the flashed letters, eye movements must be made to the vicinity of the desired letter.

This can be a problem for late-stage ALS patients who – even if they have some residual voluntary eye movement capacity – are not always able to make voluntary gaze shifts to direct overt attention [2,3]. To address this issue, rapid serial visual presentation (RSVP) spellers have been developed, which serially present whole letters flashed in the center of the screen [4,5]. While alleviating the problem of eye movements, the lack of simultaneously flashed items results in less combinatorial efficiency and a lower information transfer rate for these systems [5]. In this paper we describe preliminary experiments to develop a hybrid system with the benefits of combinatorial efficiency as well as centrally presented stimuli. As segments occur in many letters, we have the combinatorial advantage of one flash probing for many letters.

MATERIALS AND METHODS

Segment Library: Our stimuli or segment library was similar to the work of Minett et al. [6], wherein Chinese stroke-based text systems are used as the basis of a BCI. Since English letters are not composed of a series of ordinal strokes, a new system needed to be designed. Moreover, since our segments would by definition be arbitrary, it was important to design segments that are both simple to visualize, yet sufficiently distinct to allow easy comparison. To this end, we projected all 26 letters of the English alphabet onto a 7x5 grid of circular nodes, using a derivative of the scoreboard font. This allowed us to reduce the spatial complexity of characters into more general elements.

Our speller used a 10-segment library (Fig. 1) of letter subsets as query stimuli. Each segment consisted of 2-5 contiguous nodes on the 7x5 grid. The segments were assigned a specific color, which – along with their positions on the grid – was invariant. The color-segment mapping allowed participants to identify segments either through their colors or their relative spatial positions, minimizing the difficulty in making a correct discrimination. Not all nodes were contained within a segment; said independent nodes were colored white (see Fig 1).

Relative to a given letter, segments could be classified as "targets" or "non-targets", based on whether they are subsections of that letter (Figure 1). As most letters are made of a unique combination of segments, the responses to individual stimulations can be used to predict the target letter probabilistically via Bayesian inference. With our current library, O and D cannot be discriminated purely through segments, which could be a challenge in a standalone system. Language modeling and other techniques (see "Output-Letter Checks", Discussion) can easily resolve this shortcoming.

Experimental Paradigm: At the start of each block, the participant was assigned a random letter of the alphabet. For this initial test of the system, we excluded I, V, X, and Y from the list of potential targets due to their

 DOI: 10.3217/978-3-85125-533-1-86

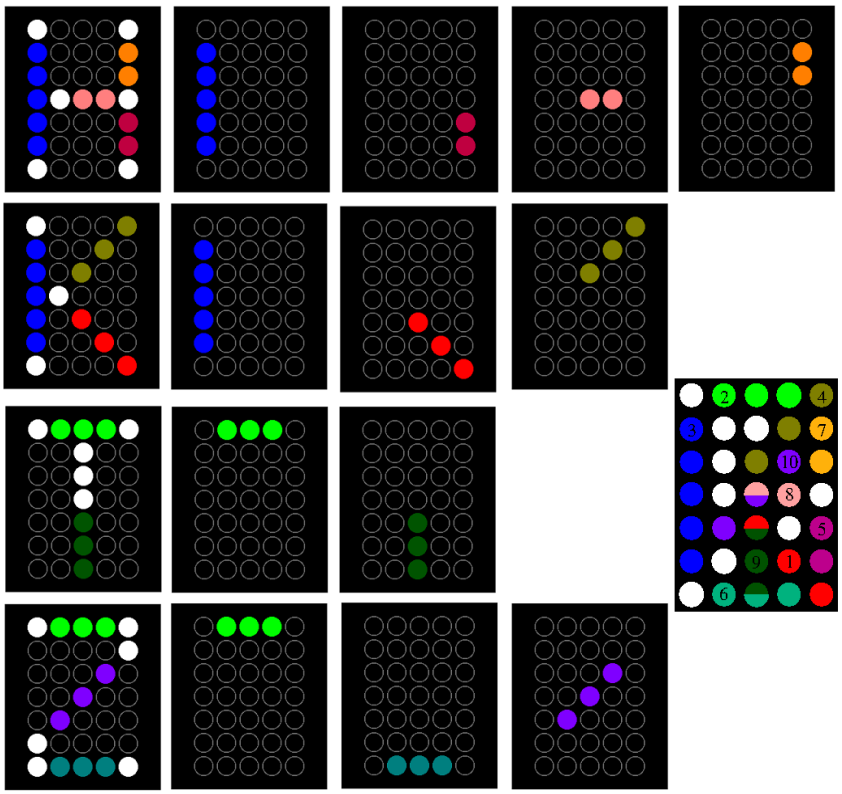

Figure 1: The segment library consists of 10 unique, invariant letter subsets. Each letter can be spelled with its own combination of segments. Depending on the letter to be spelled, component segments are deemed 'targets', whereas all others are 'non-targets'.

unique morphologies. The target letter was displayed for 2.5 seconds, with the component segments colored appropriately. This served to inform the subject of their target letter and target segments (shape and color). Afterwards, individual segments were presented serially, with a stimulus duration of 390 ms, and an inter-stimulus interval of 180 ms (total stimulus onset asynchrony 570 seconds).

The experiment consisted of an offline "training" phase, wherein data were collected to train a classifier, and an online "testing" phase, wherein the users' responses were analyzed and fed back to the system in real time. During the training phase, target segments were presented 30% of the time, and a total of 30 segments were presented before a block ended. During the analysis of the testing phase, stimuli were flashed until a) a letter was identified by the segment model or b) a total of 30 segments had been presented. During the experiment, incorrect letter selections were not a block-stopping criterion, in order to acquire more data for pseudo-online analysis.

Data Collection and Analysis: Data were collected from 6 undergraduates (4 Female, 1 Left-handed, mean age 19.5.) Stimulus presentation and timing were coordinated via the Simulation and Neuroscience Application Platform (SNAP, SCCN). The subjects' EEG data were collected using a BrainAmp (BrainVision) 64-channel active electrode system. Data were collected at 5 kHz. The marker and data streams were synchronized via LabRecorder, a Lab Streaming Layer derivative. Offline data analysis was performed using EEGLAB v 13.6.5b[7]. Data were downsampled to 500 Hz. For the topographic plots, the Artifact Subspace reconstruction designed by Christian Kothe [8] (clean_rawdata, EEGLAB) was used to clean artifactual data, and data were bandpass filtered from .1 to 5 Hz using EEGLABs hamming window sinc FIR filter (implemented in pop_eegfiltnew, EEGLAB). For visualizing plotted traces (Figure 2), data were re-referenced to the mastoids, and bandpass filtered from .1 to 10 Hz using the same filter.

Classifier: To train the classifier, class means were specified using 5 windows (100ms length) from 300ms to 800ms post-stimulus. For training of the classifier, data were downsampled to 100 Hz and bandpass filtered from .1 to 5 Hz using BCILABs [9] built-in FFT filter. The FFT filter has a much shorter length; beneficial for online filtering of 100 Hz downsampled data. LDA with automated shrinkage determination [10] as implemented in BCILAB was used. Due to the 30% target rate, a random subset of the nontarget trials were used to train to the classifier, in order to avoid erroneous solutions derived from imbalanced training classes. While the exact number of target trials (and thus nontarget trials) used for training varied slightly between subjects, about 320 trials for each class were used as training data.

Segment Model: During the online phase, a probability vector keeps track of the probabilities given to each possible letter of the alphabet. At the beginning of a block, the system assumes a uniform probability over all letters. In an end-user's speller application, this can be replaced with the probability mass function for initial letters in the language of the user; future letters can be initialized by a conditional probability function conditioned on the previous character.

Given the responses and results of the trained classifier on the serially presented segments, the model updates the letter probabilities based on the classifier response. In the case of a "target" decision by the classifier:

$$P(l\,|\text{"}target\text{"}, seg) = \frac{P(\text{"}target\text{"}|seg,l) \times P(l)}{P(\text{"}target\text{"}|\,seg)}$$

where P("*target*"|*seg*,*l*) = target segment hit rate for letters (*l*) with segment *seg* in them and
P("*target*"|*seg*,*l*) = target segment false alarm rate

 DOI: 10.3217/978-3-85125-533-1-86

for letters (*l*) without segment *seg* in them, P(*l*) is the prior probability for letter *l* before receiving the response to the flashed segment *seg*.
P("*target*"|*seg*) is the normalizing factor that keeps the total probabilities over all letters summed to 1.

Likewise in the case of a "non-target" decision by the classifier:

$$P(l|\text{"}nontarget\text{"},seg) = \frac{P(\text{"}nontarget\text{"}|seg,l) \times P(l)}{P(\text{"}nontarget\text{"}|seg)}$$

where P("*nontarget*"|*seg*,*l*) = target miss rate for letters (*l*) with segment *seg* in them and P("*nontarget*"|*seg*,*l*) = *target correct rejection rate* for letters (*l*) without segment *seg* in them.

P(*l*|*seg*) = P(*l*|"*target*",*seg*)when a target is detected and P(*l*|"*nontarget*",*seg*) when the classifier declares a non-target. These can be updated in parallel for all letters. Letter selection can be based on P(*l*|*seg*) exceeding a given threshold (e.g. 0.5) or P(l_1|*seg*) being more than a threshold above P(l_2|*seg*) where $l_1$is deemed the most probable letter, and $l_2$the second most probable. For these experiments, we used this latter rule with threshold of .2.

Note that for the purposes of these analyses (and the segment selection discussed below), we assume that the hit rates are the same for all segments that are present in the letters and the false alarm rate is also the same for all segments that are not present in the letters. For this work, we assumed a hit rate of 65% and a false alarm rate of 35%. This is very close to what was observed in the training data.

Given the letter probabilities and the mapping of segments to letters as well as estimates of the false-alarm and miss rates, the expected information gain acquired by receiving the response to each flashed segment can be computed. Segments are chosen to maximize:

$$E_{seg}(KL(P(\text{"}targ\text{"}|seg)P(l|\text{"}targ\text{"},seg) + P(\text{"}nontarg\text{"}|seg)P(l|\text{"}nontarg,seg),P(l))$$

That is, we maximize the expected Kullback-Leibler divergence between the expected letter probabilities after the response and the current letter probabilties.

Table 1: Per-subject global segment accuracy (acc), as well as class confusion performance (T/F – True, False; P/N – Positive, Negative) from the online testing phase.

	acc	TP	TN	FP	FN
S1	0.628	0.625	0.630	0.370	0.375
S2	0.664	0.639	0.675	0.325	0.361
S3	0.692	0.647	0.718	0.282	0.353
S4	0.712	0.740	0.697	0.303	0.260
S5	0.723	0.689	0.740	0.260	0.311
S6	0.597	0.559	0.615	0.385	0.441

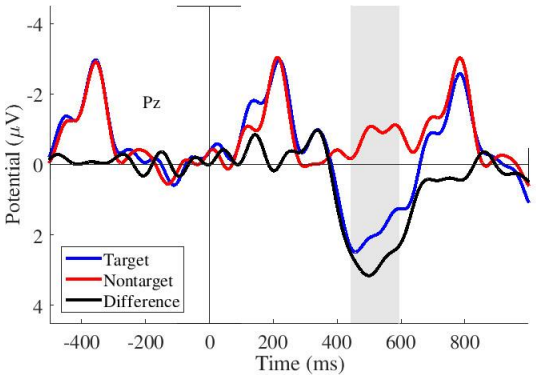

Figure 2: Grand average traces for the target (blue) and nontarget (red) classes, with the difference in black. Shaded regions reflect regions of significant difference between classes (p < .01, without correction).

Note that in the case of ties, the segment with the lowest index was selected. This does not hold true, however, if said segment was presented within the last two trials. Instead, the segment with the next lowest index was chosen.

Pseudo-Online Letter Selection: To probe the 21x25 class letter selection accuracy (Table 2), we decided to merge each subject's offline and online data, roughly doubling the number of blocks evaluated. As the experimental design between the offline and online phase is – to the participant – identical, this should be a reasonable approach. For training the pseudo-online classifier, we balanced the classes by splitting nontargets trials into two disjoint sets, and trained these sets separately, using all target-class trials. Using the same classifier, we split the data into 10 folds, using 2 folds of the data for testing.

Classifier outputs were yielded for each trial, reflecting P("*target*"|*seg*) vs P("*nontarget*"|*seg*). As each nontarget trial was in the test set twice, and each target trial was in the test set four times, the classifier outputs for each given trial are averaged. Then, seperately for each block, classifier outputs are fed into the segment model to infer the expected target. Table 2 reflects selections made a) when P(l_1|*seg*) more than .2 P(l_2|*seg*) or b) at end of a block, where P(l_1|*seg*) is selected as the model's output.

RESULTS

Offline analyses: Comparing responses from the target class (class 1) and the nontarget class (class 2), we see the expected significant difference in subject responses at Pz, averaged across all subjects (Figure 2). While the onset time of this stimulus is significantly longer than the 300ms that lends its name, it is not an unreasonable onset latency for a visual task [11]. A distinct N2 can be seen for the time-locked stimulus, as well as for the prior and subsequent letter stimulations. The N2 appearing at a typical latency improves our confidence in our observed high-latency P3. Fig, 2 also hints at a potential issue for the current system. The offset latency of the positive-

 DOI: 10.3217/978-3-85125-533-1-86

trending difference wave is – relative to stimulus onset – later than the onset of the following stimulus.

The per-subject topographic plots show an interesting trend. Excluding subject 4, all show a positive-amplitude posterior signal, reflecting a higher class 1 response amplitude consistent with a P300. The low amplitude of signal shown in Subject 6's (S6) plots may explain the poor classification results (Table 1). S4 shows a significantly different response pattern, relative to all other subjects. While some very posterior, positive-trending activity can be noticed in windows 2-4, its spatial pattern is distinct, with no immediately apparent dipole.

Online analyses: As seen from Table 1 the true negative rate was equal to or greater than true positive rate for most subjects. The one exception – S4 – also possessed a unique spatial pattern in their [target - nontarget] class responses (Fig. 3). S6 also has a somewhat unique topography; the differences between the class means appear attenuated in this subject. This could explain the uniquely poor classification results for S6.

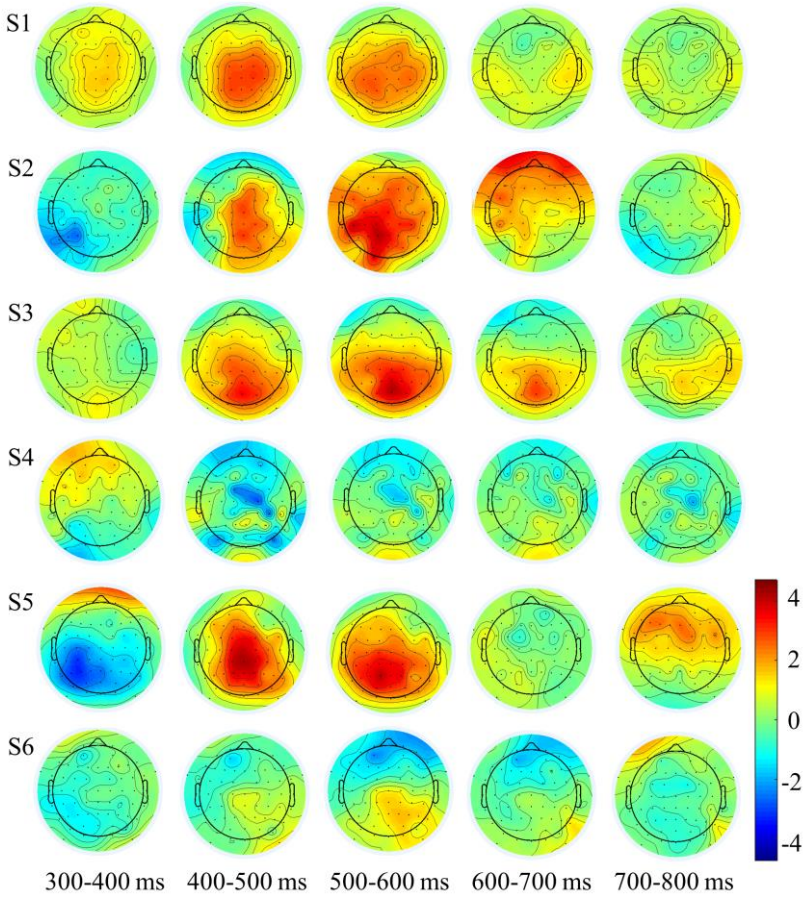

Figure 3: Per-subject topographic plots for the offline phase. Each row of plots is a scrolling average (100 ms window, 100 ms step between plots) extending from 300 ms to 800 ms. Due to the planar depiction of the 3-D electrode locations, electrodes further down the head extend beyond the head model.

Of the actual online blocks using the threshold method, 61.4% ended with the segment model outputting a potential letter. Of those output letters, 31.7% matched the blocks target. Letter selection accuracies could be improved by increasing the threshold for selection of an "output letter". Implemented along with more segment presentations per target, correct target accuracy could increase dramatically.

Pseudo-Online Results: As can be seen from Table 2, the incorrectly selected output-letters tend to share common characteristics. Sums along the columns – especially relative to diagonals – reflect high false selection rates for a given letter. Of the 261 letters selected output by the model, 88 (33.7%) matched the target letter for their given block.

We can see from this array (Table 2) that dense-segment letters – particularly B, E, G, and R – suffer from poor selection accuracies. Erroneous outputs tend to share many shape characteristics with the true target, however, and each of said letters shares at least 2 segments. Longer trials or a more conservative threshold could lead to increases in ITR, even at the cost of increasing time.

DISCUSSION

Spatially independent spellers pay a non-trivial cost as they restrict themselves to specific regions of the visual field [3,5]. In many cases this cost must be assessed, as directing overt attention towards a target is not feasible for all end-use scenarios [2]. Moreover, BCIs that require accurate eye movements must compete with eye trackers that overcome many of an EEG BCI's shortcomings. Our system was designed with patients suffering from LIS in mind. A high information transfer rate despite the spatial independence is nevertheless important, especially when designing a channel of communication.

The classifier's discrimination of subject responses benefits from the distinct class-specific posterior potential, which we believe to be a P300. In this case of S5, however, no significant posterior response is elicited. Despite this, the online accuracy of the subject is slightly above the rest of the participant cohort's average. It is possible that the anterior negativity present in the second and third windows nevertheless allows the classes to be discriminated. Alternatively, the ocular activity in the prefrontal channels, or other artefactual sources may be responsible. The 5 Hz lowpass attenuates most of the muscle activity, but it is also possible that unconscious reflexes elicited in some class-specific manner could be driving classification. These peripheral signals should be more salient in the topographic plots, however, so we

 DOI: 10.3217/978-3-85125-533-1-86

find this unlikely. Increasing single-segment classification is an obvious goal moving forward. A Markov Chain, coding previous trial class as state could be fruitful, as the duration of the subjects' responses to targets are longer than the SOA. This would be especially useful if stimulus rates grow faster than 2 Hz; with a 300 ms SOA, the onsets of the preceding P3 (here, latency 500 seconds) and the subsequent N200 would overlap.

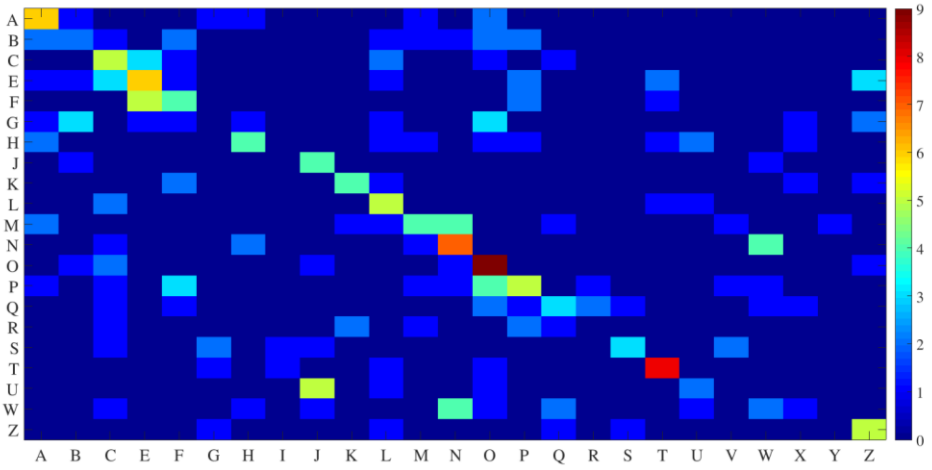

Figure 4: Visualization of the confusion matrix shown in table 2. Rows represent the target letter for a given block, whereas columns represent the letter output by the segment model. Thus, (1,1)/ (A,A) is the sum of all A's output by the segment model while A was the target.

Output-Letter Checks: Letter selection accuracies could be improved by increasing the threshold for selection of an "output letter". We have previously tested output letter feedback via full letter presentations. These presentations can either be timed identically to the segment stimuli, or the output-letter checks can be flashed with longer preceding and anteceding inter-stimulus intervals. Preliminary tests on small numbers of trials suggest that both yield responses distinct from the segments. This is not surprising, as while the segments are subsets of letters, the complete letter arrays are – by design – relatively complex images [11].

Consequently, a second independently-trained classifier will be the most appropriate implementation. It is possible that responses to these flashed "test letters" would be more like an error-related response than a P300

Table 2: This confusion matrix reflects all output-letters selected by the segment model. The rows correspond to target stimulus, and the columns correspond to classifier output. 'Blocks' reflects the number of time a given letter was a target. Shaded version in Fig. 4.

	A	*B*	*C*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*	*Blocks*
A	6	1	0	0	0	1	1	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	12
B	2	2	1	0	2	0	0	0	0	0	1	1	1	2	2	0	0	0	0	0	0	0	0	0	0	14
C	0	0	5	3	1	0	0	0	0	0	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	13
E	1	1	3	6	1	0	0	0	0	0	1	0	0	0	2	0	0	0	2	0	0	0	0	0	3	20
F	0	0	0	5	4	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	12
G	1	3	0	1	1	0	1	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	1	0	2	14
H	2	0	0	0	0	0	4	0	0	0	1	1	0	1	1	0	0	0	1	2	0	0	1	0	0	14
J	0	1	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	6
K	0	0	0	0	2	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	9
L	0	0	2	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	1	1	0	0	0	0	0	9
M	2	0	0	0	0	0	0	0	0	1	1	4	4	0	0	1	0	0	0	0	1	0	0	1	0	15
N	0	0	1	0	0	0	2	0	0	0	0	1	7	0	0	0	0	0	0	0	0	4	0	0	0	15
O	0	1	2	0	0	0	0	0	1	0	0	0	1	9	0	0	0	0	0	0	0	0	0	0	1	15
P	1	0	1	0	3	0	0	0	0	0	0	1	1	4	5	0	1	0	0	0	1	1	0	0	0	19
Q	0	0	1	0	1	0	0	0	0	0	0	0	0	2	1	3	2	1	0	0	0	1	1	0	0	13
R	0	0	1	0	0	0	0	0	0	2	0	1	0	0	2	1	0	0	0	0	0	0	0	0	0	7
S	0	0	1	0	0	2	0	1	1	0	0	0	0	0	0	0	0	3	0	0	2	0	0	0	0	10
T	0	0	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	8	0	0	0	0	0	0	12
U	0	0	0	0	0	0	0	0	5	0	1	0	0	1	0	0	0	0	0	2	0	0	0	0	0	9
W	0	0	1	0	0	0	1	0	1	0	0	0	4	1	0	2	0	0	0	1	0	2	1	0	0	14
Z	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	5	9

 DOI: 10.3217/978-3-85125-533-1-86

response.

Furthermore, given the letter selection frequency in the 30-trial online blocks, a larger maximum trial cap is an obvious change going forward. As the discrimination difficulty and input fatigue make the task somewhat strenuous, brief breaks may be necessary as blocks lengthen. Current, trials last no longer than 20 seconds, and responses may grow increasingly non-stationary due to fatigue if that duration significantly increases.

One advantage of the segment speller is that errors tend to be to visually similar letters (letters with similar subsets of segments) as opposed to neighboring letters in the grid for standard P300 grid and hexagonal spellers. This means that perfect selection of letters may not be necessary for typed words to be readable, as replacing letters by visually similar ones can still be quite legible.

An important question to consider is whether the combinatorial advantage of using segments justifies the increase in task difficulty. The complexity of the oddball task increases significantly when moving from letters to segments, and the chances of misidentification also increase. Moreover, task difficulty has been shown to attenuate P3 amplitude [12], especially that of the more frontal P3a [13]. It is possible that the high latency, posterior distribution of our responses is a consequence of this difficult categorization task. Consequently, future experiments could compare responses to an easier oddball task, to help contextualize the data.

CONCLUSION

The online classification of target vs nontarget segments proved possible for all subjects, with an average segment accuracy of 68%. The segment model – designed to probabilistically infer the target letter based on segment classification – outputs a target 61.4% of the time. Of these output targets, the correct letter was selected 31.7% of the time. As only 30 segments could be queried per target letter, we expect a longer selection block paired with a higher threshold should significantly improve final letter accuracy and rate.

ACKNOWLEDGEMENTS

This study would not have been possible without the support of our funding agencies. This work was supported by the NSF grants SMA 1041775 and IIS 1528214.

REFERENCES

[1] Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, *70*(6), 510-523.

[2] Cohen, B., & Caroscio, J. (1982). Eye movements in amyotrophic lateral sclerosis. *Journal of neural transmission. Supplementum*, *19*, 305-315.

[3] Brunner, P., Joshi, S., Briskin, S., Wolpaw, J. R., Bischof, H., & Schalk, G. (2010). Does the 'P300'speller depend on eye gaze?. *Journal of neural engineering*, *7*(5), 056013.

[4] Orhan, U., Hild, K. E., Erdogmus, D., Roark, B., Oken, B., & Fried-Oken, M. (2012, March). RSVP keyboard: An EEG based typing interface. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (pp. 645-648). IEEE.

[5] Chennu, S., Alsufyani, A., Filetti, M., Owen, A. M., & Bowman, H. (2013). The cost of space independence in P300-BCI spellers. *Journal of neuroengineering and rehabilitation*, *10*(1), 82.

[6] Minett, J. W., Zheng, H. Y., Fong, M. C., Zhou, L., Peng, G., & Wang, W. S. (2012). A Chinese text input brain–computer interface based on the P300 speller. *International Journal of Human-Computer Interaction*, *28*(7), 472-483.

[7] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

[8] Kothe, C. A. E., & Jung, T. P. (2014). *U.S. Patent Application No. 14/895,440*.

[9] Kothe, Christian Andreas, and Scott Makeig. "BCILAB: a platform for brain–computer interface development." *Journal of neural engineering* 10.5 (2013): 056014.

[10] Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, *88*(2), 365-411.

[11] Polich, J., Ellerson, P. C., & Cohen, J. (1996). P300, stimulus intensity, modality, and probability. *International Journal of Psychophysiology*, *23*(1), 55-62.

[12] Johnson, R., & Donchin, E. (1980). P300 and stimulus categorization: two plus one is not so different from one plus one. *Psychophysiology*, *17*(2), 167-178.

[13] Baur, N., Metzner, M. F., & Śmigasiewicz, K. (2014). The hard oddball: Effects of difficult response selection on stimulus-related P3 and on response-related negative potentials. *Psychophysiology*, *51*(11), 1089-1100.

[14] Kim, K. H., Kim, J. H., Yoon, J., & Jung, K. Y. (2008). Influence of task difficulty on the features of event-related potential during visual oddball task. *Neuroscience letters*, *445*(2), 179-183.