IMPROVING INFORMATION TRANSFER RATE IN ACTIVE BCIS

V. R. de Sa[1]

[1]Cognitive Science, UC San Diego, La Jolla, CA, United States

E-mail: desa@ucsd.edu

ABSTRACT: We consider the case of a noisy binary EEG-based brain-computer interface where a human attempts to generate two discriminable control signals but the received signals are noisy and the optimal classification boundary (or decoder) is not known or changing. In such situations, it is common for the computer to accumulate evidence over time before performing an action. Intermediate feedback can be given to inform the user of the current decoding along the way. Under these conditions we have shown via Markov chain analysis that the information transfer rate is higher when the user and computer attach the responsive (to the intermediate feedback) meanings of "continue/good" and "change-direction/bad" to the two classes of noisy signals they generate instead of direct commands "left" and "right". In this paper, we analyze the first step of these systems and show that when there is not yet a computer-interpreted response to respond to, the "right/left" commands are most informative and that a system where a first "right/left" step is combined with future "good/bad" interactive commands gives the highest information transfer rate. Finally we show that this hybrid approach can be seen as a natural game-like interface.

INTRODUCTION

We consider the case of a human-controlled interface to a computer where the received signals are noisy and not perfectly classifiable. Such signals arise in EEG-based brain-computer interfaces (BCIs) but could also arise in other noisy interfaces such as gesture recognition or speech recognition in noisy environments or in some clinical populations. For this paper, we restrict analyis to the binary control case, where humans are trying to generate one of two signals as is commonly used in motor-imagery EEG-based brain computer interfaces.

For concreteness, we consider the scenario of a user using an EEG-based motor-imagery brain-computer interface with the goal of selecting one of two targets on the right and left of the screen. Subjects require a certain level of selection accuracy (e.g. 70% [5]) for performance to be considered acceptable, and there are applications where it is costly to output the wrong answer [11]. For these reasons, it is common to accumulate evidence, and feedback can be given to the subject over time. A common method is to have the cursor move a step every processing window (processing windows are usually 500ms to 1 second long). The number of steps between the two target endpoints can be varied to trade off accuracy and selection speed or can be set to optimize information transfer rate (ITR) for a given discriminability between the distributions of signals for the two classes [2]. In text and figures below, we refer to the number of possible cursor positions (NCP) which is equal to twice the number of steps from the center to a target plus one.

While EEG signals are high dimensional, after standard signal processing [9, 1, 7] the signals are reduced to a lower dimensional space, and it is common to approximate them as Normal/Gaussian distributions and use simple linear classifiers in this projected space [10, 9, 8]. Once a classifier is defined, the critical variable for classification, is the side of the classification boundary (or more generally the signed perpendicular distance from the classification boundary). This is represented as the (single-step) one-dimensional probability distributions diagrammed on the right side of Figures 1, 2, 3, and 8. and given by the blue dashed curves in Figures 4, 6, 7.

In any noisy machine learning problem with finite data, a classifier will not be able to find the exact optimal classification boundary between the distributions. In problems with non-stationary data such as EEG [3, 10], this issue is especially true. In these cases, we have shown that by changing the meaning of the single-step signals generated by the humans, the information transfer rate can be increased by changing the meanings of the two signals that the user generates to “continue” and “change direction” instead of “left” or “right” [2]. What this means is that the same signals that the user generates are given different semantics in the communication strategy. So for example, right hand imagery could be used to mean “change direction (I am disatisfied with the current movement)” and left hand imagery used to signal “continue (I am satisfied with the current movement direction)”. This changes the communication protocol from direct commands to interactive commands that respond to the current interpretation as reflected by the feedback of the cursor movement. We have shown that when a control signal is used “that depends intimately on what has already been transmitted, interpreted, and received”, a much more robust communication system results [2]. In particular the information transfer rate is demonstrably higher with the interactive communication system than with the standard direct communication system when the classification boundary is not in the optimal position. As in [2], we will refer to the direct method of controlling with “move right” and “move left” signals as an R/L control system and the interactive method of con-

trolling with “change-direction/bad/**D**issatisfaction” and “continue/good/**S**atisfaction” as a D/S control system.

We examine the mathematics for single step systems (where a target is reached in one step or NCP=3), and based on this analysis propose a new hybrid approach, where the starting cursor direction is fixed and known but D/S commands are used. We show that in this case the first step is equivalent to a “Right/Left” commanded step, and the system makes less processing demand of the user and results in a higher information transfer rate than the previously proposed D/S method in [2] using random start direction.

METHODS

The standard R/L method of control where the user generates one signal (e.g. right hand motor imagery) to mean “Move the cursor to the Right” and another discriminable signal (e.g. left hand motor imagery) to mean “Move the cursor to the Left” can be shown to be modeled by a Markov chain as shown in Figure 1 [2] whereas the interactive D/S method of control that uses the same underlying signals of right and left hand motor imagery can be modeled by a Markov chain shown in Figure 2. In this case, the state of the system contains the position and cursor direction information.

If the system does not use multiple steps to reach the goal (in other words, if NCP=3. See Figures 1,2,3), then the interactive nature does not come into play. We can still consider how a D/S control system might work in a one-step system. In a D/S system the cursor would appear moving in one direction (or simply appear as an arrow instead of a circular cursor, and the user would generate a "continue" (if they like that direction) or "change direction" (if they don’t) signal which would influence the one and only step. It might seem that the most natural method would be to have the initial cursor direction be drawn randomly from right/left as in [2] (and shown in Figure 2 by the 0.5 probabilities for each possible starting state). In this case, the accuracies are actually the same as a function of the classification boundary for the D/S and R/L system. However the error-rates when considered separately for the left and right classes can be quite different as the boundary is moved from the crossing points of the distributions. In the R/L system, if the boundary is offset so "Right" is output more than it should be, the error rate for the Left class will be higher than the error rate for the right class. This is shown in Figure 4.

However for the D/S system, if the classification boundary is not at the crossing point of the distributions, it will be more likely to output "continue" or "change-direction". If, the start direction (to which continue or change-direction are responded to) is chosen randomly and the two classes have equal prior probabilities, then the error-rate for the left and right classes will be equal, as whatever happens to the left class when the cursor starts in the left direction will be matched by what happens to the right class when the cursor starts in the right direction and what happens to the left class when the cursor starts in the right direction will be matched by what happens to the right class when the cursor starts in the left direction. This can be seen in Figure 4.

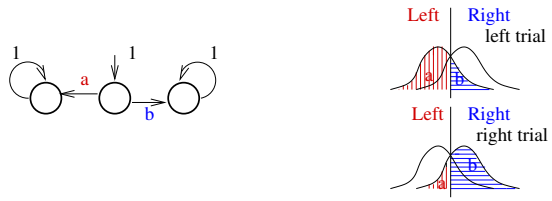

Figure 1: This figure shows the Markov chain model for the R/L control method for one step (NCP=3).

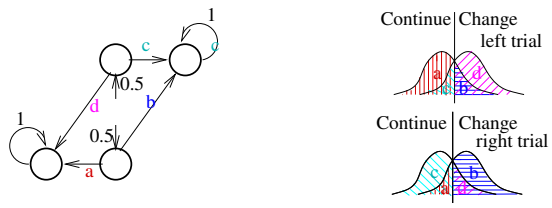

Figure 2: This figure shows the Markov chain model for the original D/S control method for one step (NCP=3) with random start direction (RS). Notice that this model is different from the Markov chain for the one step R/L control method.

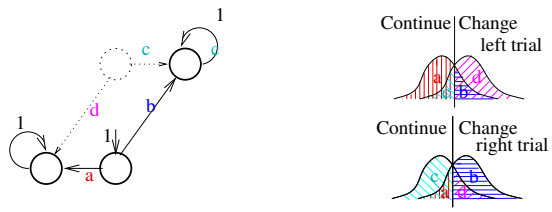

Figure 3: This figure shows the Markov chain model for the D/S control method with one step (NCP=3) and constant left direction start (CS). The node and connections with dashed lines represents states and transitions that are not possible but are possible in the random start model. Note how this Markov chain is equivalent to that of the one step R/L control method with transition probabilities given only by a and b.

Note, however, that if the cursor always starts in the left direction, the one-step D/S method of control is equivalent to the R/L method of control (with continue equivalent to left and change-direction equivalent to right). (This is shown in Figure 3). (Similarly if the cursor always starts in the right direction, the one-step D/S method of control is equivalent to the R/L method with continue equivalent to right and change-direction equivalent to left.)

The number of steps (one in this case) are the same for either starting strategy with the D/S method, but because the accuracies are different for the two classes for the R/L control and the D/S (random start), the information transfer rate (ITR) ends up being different for the two systems. As our classification rates can differ for the right and left classes, we use the general equation for computing the ITR.

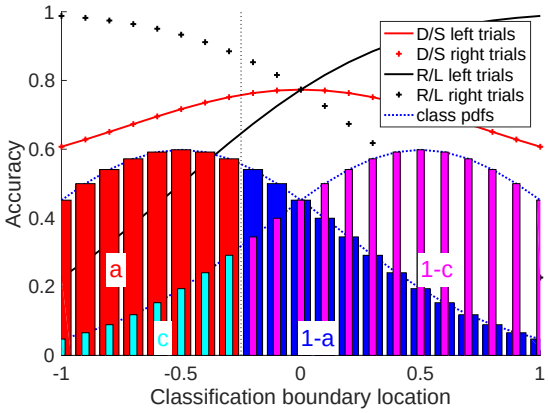

Figure 4: This figure shows the probability distributions for two classes (bottom) and the error rate for the left and right classes as a function of the classification boundary (top). Note that with the R/L control method the error rate is different for right and left trials whereas for the D/S control system with random start direction, the error rates are the same for the two classes as discussed in the text. At the possible classification boundary given by the dotted vertical line, the R/L control accuracy for the left class is 61.8% and the accuracy for the right class is 88.5%. The accuracy for both left and right classes for the D/S control (random-start) is the average of these two (75.14%) Also shown in the bottom of the figure are the variables used in the mathematical analysis of the one-step system. Note that b and c will change as a function of the classification boundary.

$$ITR = \left(\sum_{j=1}^{C} -p(y_j)\log_2(p(y_j)) + \sum_{i=1}^{C}\sum_{j=1}^{C} p(x_i)p(y_j|x_i)log_2(p(y_j|x_i)) \right) /T$$

where $p(y_j) = \sum_{i=1}^{C} p(x_i)p(y_j|x_i)$ and x_i represents intended class i and y_j represents decoded class j. (For our example $C = 2$). T is the total time (including overhead/set up time) for the full trial to select either target. Figure 6 shows by analysis of the Markov chain model [2] that for the one-step system, the ITR is better for the R/L (and equivalent D/S same-start) control than the D/S random start method. We now prove this mathematically. For the purposes of the proof we will consider the probabilities under the pdfs given by the distributions in Figure 4. For R/L control the class on the left corresponds to "Left trials" and the class on the right "Right trials" and thus for R/L control $P(y_1|x_1) = a, P(y_1|x_2) = c, P(y_2|x_1) = 1-a, P(y_2|x_2) = 1-c$. For the one-step D/S system with random start (RS) the distributions actually represent the "Continue" and "Change Direction" distributions. In order to determine $P(\text{"Left"}|left)$ we must compute the expected value over both right and left start directions. If the cursor starts right, then the accuracy is given by the area under the change direction distribution on the correct side of the boundary (i.e. $1-c$). If the cursor starts moving left, then the accuracy is given by the area under the continue distribution on the correct side of the boundary (i.e. b). So for equal probability of each starting direction $P(\text{"Left"}|left) = .5a + .5(1-c)$. Likewise $P(\text{"Right"}|right) = .5a + .5(1-c)$. That is the error rates are equal for the right and left classes and equal to $\frac{c+(1-a)}{2}$.

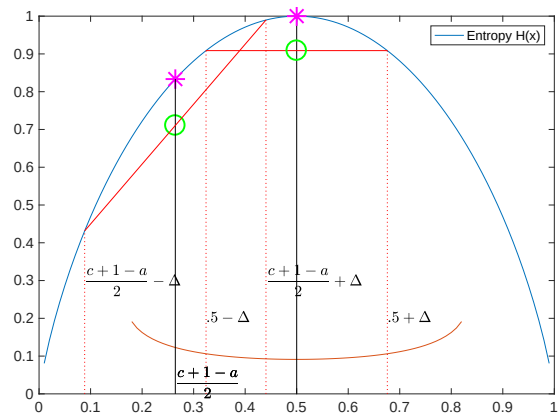

Figure 5: Graphical view of the entities in IT_{RS} and IT_{CS}. The magenta asterisks show $H(x)$ and the Green circles show $G(x,\Delta)$ where $\Delta = \frac{1-a-c}{2}$. IT_{RS} Subtracts the lower magenta asterisk from the higher one (at x=.5). IT_{CS} subtracts the lower green circle from the higher one (at x=.5). As $H(x) - G(x,\Delta)$ (the difference between the magenta asterisks and the green circles (a subset is shown by the red curve near the bottom of the Figure)) has a minimum at 0.5 on $(0,1)$, $IT_{CS} \geq IT_{RS}$. When $\Delta = 0$ $(c = 1-a)$, $IT_{CS} = IT_{RS}$.

$$\begin{aligned} IT_{RS} = & \sum_{j=1}^{2} -p(y_j)log_2(p(y_j)) \\ & + \sum_{i=1}^{2}\sum_{j=1}^{2} p(x_i)p(y_j|x_i)log_2(p(y_j|x_i)) \\ = & H(.5) - H\left(\frac{c+1-a}{2}\right) \end{aligned}$$

where $H(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ is the discrete entropy function for a Bernoulli random variable with probability of one class given by p (and probability of the other class given by $(1-p)$).

For the one-step D/S system with consistent left start (CS) similarly assuming equal number of class 1 and class 2 patterns, the error rates are different for the right and left classes (and are identical to an R/L system) and are respectively given by $1-a$ and c we have:

$$
\begin{aligned}
IT_{CS} = & \sum_{j=1}^{2} -p(y_j)log_2(p(y_j)) \\
& + \sum_{i=1}^{2}\sum_{j=1}^{2} p(x_i)p(y_j|x_i)log_2(p(y_j|x_i)) \\
= & -\left(\frac{a+c}{2}\right) * \log_2\left(\frac{a+c}{2}\right) \\
& -\left(1-\frac{a+c}{2}\right) * \log_2\left(1-\frac{a+c}{2}\right) \\
& + .5(1-a)log_2(1-a) + .5(a)log_2(a) \\
& + .5(c)log_2(c) + .5(1-c) * log_2(1-c) \\
= & H\left(\frac{a+c}{2}\right) - .5H(a) - .5H(c) \\
= & .5H(.5 - \frac{1-a-c}{2}) + .5H(.5 + \frac{1-a-c}{2}) \\
& - .5H(\frac{1-a+c}{2} + \frac{1-a-c}{2}) \\
& - .5H(\frac{1-a+c}{2} - \frac{1-a-c}{2}) \\
= & G\left(.5, \frac{1-a-c}{2}\right) - G\left(\frac{c+1-a}{2}, \frac{1-a-c}{2}\right)
\end{aligned}
$$

where we define

$$G(p,\Delta) = \frac{H(p+\Delta)+H(p-\Delta)}{2}$$

for $\Delta = \frac{1-a-c}{2}$ and we use $H(x) = H(1-x)$ for $x \in (0,1)$. If we compare the IT_{CS} with consistent start to the IT_{RS} with random start,

$$
\begin{aligned}
IT_{RS} - IT_{CS} = H(.5) - G\left(.5, \frac{1-a-c}{2}\right) - \\
H\left(\frac{c+1-a}{2}\right) + G\left(\frac{c+1-a}{2}, \frac{1-a-c}{2}\right)
\end{aligned}
$$

To see whether IT_{CS} is larger or smaller than IT_{RS} we check how $H(x) - G(x,\Delta)$ varies as a function of x. Looking at the first derivative of the function. $H(x) - G(x,\Delta)$, we have

$$
\begin{aligned}
\frac{d(H(p)-G(p,\Delta))}{dp} = & -\log_2(p) - 1 + \log_2(1-p) + 1 \\
& - .5(-\log_2(p+\Delta) + \log_2(1-(p+\Delta))) \\
& - .5(-\log_2(p-\Delta) + \log_2(1-(p-\Delta)))
\end{aligned}
$$

which equals 0 at $p = .5$.
The second derivative:

$$
\begin{aligned}
\frac{d^2(H(p)-G(p,\Delta))}{dp^2} = & -\frac{1}{p} - \frac{1}{1-p} + \frac{.5}{(p+\Delta)} \\
& + \frac{.5}{(1-p-\Delta)} + \frac{.5}{(p-\Delta)} + \frac{.5}{(1-p+\Delta)} > 0
\end{aligned}
$$

is positive for $p \in (0,1)$ and $\Delta = \frac{1-a-c}{2}$ by the concavity (Jensen's inequality [4]) of $\frac{1}{x}$. Therefore $H(x) - G(x,\Delta)$ where $\Delta = \frac{1-a-c}{2}$ has its minimum for $x \in (0,1)$ at $x = 0.5$ and therefore $IT_{RS} \leq IT_{CS}$ (and equality only when $c = 1-a$ at the optimal crossing point). That is more information is transferred in the single step system using the standard R/L control method and consistent start D/S method than the random start D/S method. Note that by symmetry, the direction of the start does not matter; it just matters that it always starts in the same direction.

Following the results of [2] showing that in the multi-step case, an interactive D/S control method is preferable to the standard R/L control method for improved ITR in the presence of noise, the one-step result showing that the consistent start D/S control (equivalent to R/L control on the first step) is better than the random start D/S control might be somewhat surprising, but the reason for improved performance of interactive commands in the multi-step D/S systems is because the user is responding to the computer's classification error. However in the one-step systems, the user response is to a randomly generated direction, not the result of the computer's interpretation of the user's command, and so does not provide the benefit of revealing the computer's bias.

As the first step in a multi-step system (NCP>3) is equivalent to a one-step system we conclude that for multi-step systems it is also better to start with a consistent direction. For reasons discussed below, we will consider the start direction to be left. This new suggested control strategy is shown in Figure 8. The ITR curves computed from the Markov chain analysis [2] for chains with NCP=7 are shown in Figure 7. The plots show that there is an ITR benefit to having a consistent start direction, though the effect is less with the models with more NCPs as should be expected with a change that only changes the first step. The difference is larger on the side when "continue" is more likely to be output (on the right side of the figure) as that is when fewer steps are taken, and an effect from the first step will have a greater effect.

DISCUSSION

While for systems with larger NCPs the ITR difference is not large between the random-start and same-start D/S systems, there are also other human factors considerations. When the original D/S control method was introduced, there was a concern that the control system would require more time for the user to process each step as they would need time to determine whether to give a "change" or "continue" command [2]. Thinking about each step compared to constantly generating a "right" or "left" command requires more processing and is less automatic. However, starting the cursor movement in the same direction allows the user responses to be more automatic. For instance if the user knows that the cursor will always start moving left and the change direction command is given by right hand motor imagery, then for right targets, the user will start with right motor imagery and will continue right im-

agery until the cursor changes direction at which point they will switch to left hand motor imagery (see Figure 9). In this way, the user feels like they are pushing the cursor (moving the hand that the cursor is moving away from). If the subject desires the left target, he will start with left motor imagery and again only change imagery when the cursor changes direction. In this mode (left target), the user feels like they are trying to bat the target back and forth (moving the hand that the cursor is moving towards). The user does not need to think about which hand is continue or change direction, they simply have to start with the motor imagery of the hand in the desired direction (e.g. left hand imagery for left target) and change whenever the cursor changes direction.

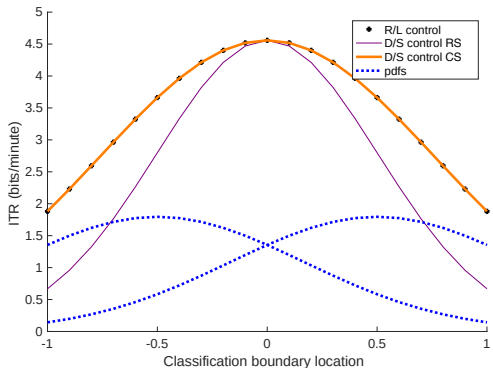

Figure 6: This figure shows the probability distributions for two classes and the information transfer rate for the R/L and D/S (random start (RS) and left/consistent start (CS)) control methods with NCP=3 (the one-step systems). Note the R/L and D/S left start curves are on top of each other.

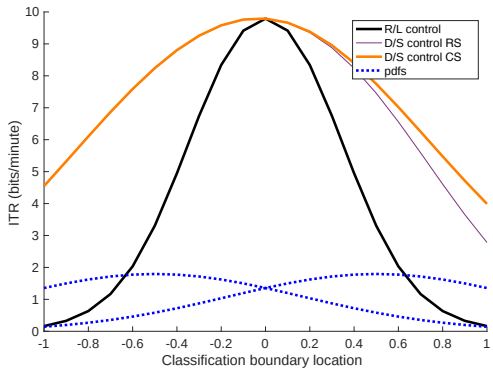

Figure 7: This figure shows the probability distributions for two classes and the information transfer rate for the R/L and D/S (random start (RS) and left/consistent start (CS)) control methods with NCP=7. The orange curve shows the results with D/S control with consistent left-moving start. The thin purple line, gives the result for D/S control with random start direction, and the thick black line shows the result using the standard R/L paradigm.

The idea behind the interactive communication method is to allow a user to compensate for inaccuracies or biases in the classification boundary that are inevitable with noisy and non-stationary systems. This is similar to the idea used in [11] where Wolpaw and colleagues required users to activate two commands on opposite sides of a possibly biased classification boundary. Notice in the multi-step D/S control method, the forces pushing the cursor in each direction are caused by transitions from each side of the classification boundary (e.g. a and d cause transitions in the same direction as do b and c in Figures 2, 3, and 8.)

It is also similar to the way that humans naturally change to a more interactive style when giving directions to a non-native speaker. If the person does not understand, we don't repeat the same sentences but attempt to give the same instructions with different words. We also monitor understanding and change our directions to react to their understanding. Some HCI systems have a feature like this; when performing a risky computer operation that may have been incorrectly activated (e.g. delete a file), the interface does not ask you to press the same button that was originally (possibly mistakenly pressed), but to respond to "Are you sure? (Yes/No)."

There are other practical advantages to the D/S control system for many modalities of signal generation. In EEG, for example, emotional, error, and frustration, responses are combined with any signal the user is actively trying to generate. Frustration with loss of control has been shown to induce non-stationarity in EEG [3]. It has been shown that some of these signals are in the same frequency bands as the signals to be actively detected and are difficult to separate [6]. By using the D/S control method, these unconscious signals actually add to the discriminability of the signals rather than subtracting from them. This may also be true in gesture-based systems where a person who is happy with the progress may make more animated, or otherwise somewhat different, moves than when unhappy. Similarly in speech-based systems, affect is generally reflected in people's speech signals.

We see this work as important for considering the human and computer as cooperative agents. While the human may be limited by the discriminability of the signals they can generate and the computer is limited in its ability to learn the best classification boundary (given finite data), the two together can have greater information transfer (from human to computer) by changing the semantics of the signals the human generates.

CONCLUSIONS

To conclude, we have analyzed the first step and revealed that the direct R/L control is better for this first step (where the feedback is random and can't provide the computer's interpretation of the user's signal). Through incorporation of this knowledge to restrict the intial cursor direction in a D/S system, we have further improved the information transfer rate of the D/S interactive control method over the standard direct R/L control method. At the same time, this change reduces the real-time processing requirements of the user and reduces the task to a more reactive task requiring less conscious effort. The change in how the first step is handled maintains the other advantage of the D/S in-

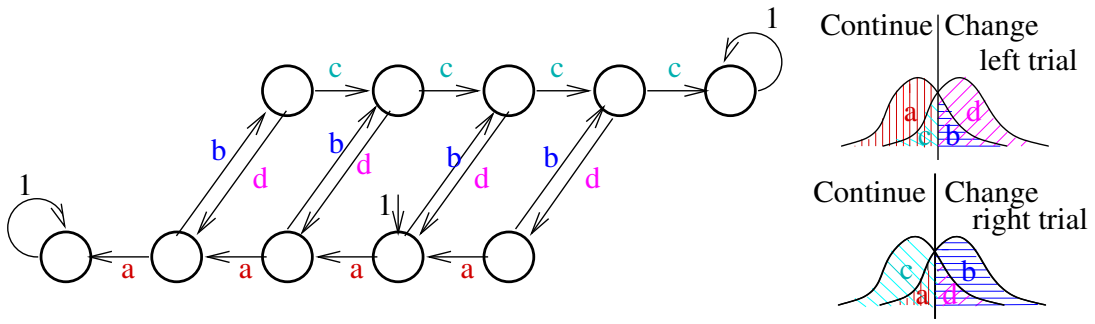

Figure 8: This Figure shows the Markov chain model for the new proposed D/S control method with constant left start direction.

teractive control strategy: changes in emotion/frustration help the classification instead of distracting from it.

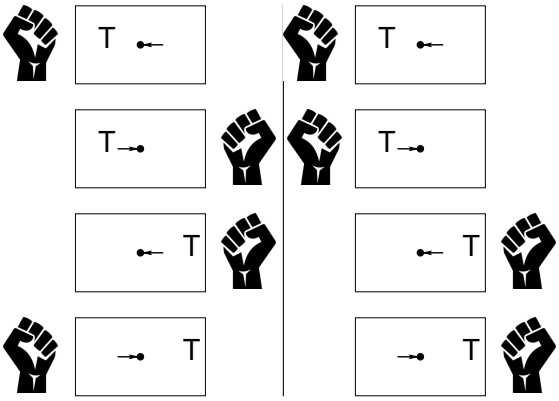

Figure 9: This figure shows the commands a user would use after specific cursor movements with the D/S control method (LEFT) and R/L control method (RIGHT). The hand icon on the right/left side of the drawn screen represents right/left-hand motor imagery. The T represents the desired target location/direction and the arrow represents the direction of the last cursor movement. Note that with the D/S control the user changes command when the cursor changes direction and that right hand targets can be viewed/felt as the user performing a pushing behavior giving the imagery of the hand that the cursor is moving away from (as if to push it away). When the user desires the left hand target, the user performs a batting back and forth behavior where they perform imagery of the hand that the cursor is moving towards (as if to bat it back).

ACKNOWLEDGMENTS

This work is supported by NSF grants IIS 1219200, SMA 1041755, IIS 1528214, and UCSD FISP G2171.

REFERENCES

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller. Optimizing spatial filters for robust EEG single-trial analysis. *Signal Processing Magazine, IEEE*, 25(1):41–56, 2008.

[2] V. de Sa. An interactive control strategy is more robust to non-optimal classification boundaries. In *Proceedings of ICMI'12*, pages 579–586, 2012.

[3] S. Jatzev, T. Zander, M. DeFlilippis, C. Kothe, S. Welke, and M. Roetting. Examining causes for non-stationarities: The loss of controllability is a factor which induces non-stationarities. In *In Proc. of the 4th Int. BCI Workshop & Training Course*. Graz University of Technology Publishing House, 2008.

[4] J. Jensen. Sur les fonctions convexes et les inegalites entre les valeurs moyennes. *Acta Mathematica*, 30:175–193, 1906.

[5] A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. P. Birbaumer. Brain-computer communication: self-regulation of slow cortical potentials for verbal communication. *Archives of physical medicine and rehabilitation*, 82(11):1533–1539, 2001.

[6] M. Mousavi, A. S. Koerner, Q. Zhang, E. Noh, and V. R. de Sa. Improving motor imagery BCI with user response to feedback. *Brain-Computer Interfaces*, 2017.

[7] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *J Clin Neurophysiol*, 110(5):787–98, May 1999.

[8] E. Noh and V. de Sa. Discriminative dimensionality reduction for analyzing EEG data. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pages 1090–1095, 2014.

[9] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. da Silva. Mu rhythm (de)synchonization and EEG single-trial classification of different motor imagery tasks. *Neuroimage*, 31:153–159, 2006.

[10] P. Shenoy, M. Krauledat, B. Blankertz, R. Rao, and K.-R. Müller. Towards adaptive classification for BCI. *J Neural Eng*, 3:R13–R23, 2006.

[11] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller. EEG-based communication: improved accuracy by response verification. *IEEE transactions on Rehabilitation Engineering*, 6(3):326–333, 1998.