Information-transfer rate modeling of EEG-based synchronized brain-computer interfaces

Julien Kronegg, Sviatoslav Voloshynovskiy, and Thierry Pun

***Abstract*— The information-transfer rate (ITR) is commonly used to assess the performance of brain-computer interfaces (BCI). Various studies have shown that the optimal number of mental tasks to be used is fairly low, around 3 or 4. We propose a formal approach and an experimental validation to demonstrate and confirm that this optimum is user and BCI design dependent. Even if increasing the number of mental tasks to the optimum indeed leads to an increase of the ITR, the gain remains small. This might not justify the added complexity in terms of protocol design.**

***Index Terms*—Brain-computer interface (BCI), electro-encephalogram (EEG) classification, information transfer rate, optimal number of mental tasks.**

I. Introduction

Brain-Computer Interfaces (BCIs) are input-devices that allow a user to communicate with a computer by way of thinking. Thoughts must be inferred from the neuronal electrical activity. Brain structure however is too complex to allow users to think to whatever they want and still having a system able to infer what was thought of. The "thoughts" vocabulary must therefore be limited to a few mental tasks with well characterized and localized neuronal activity, such as imagination of finger movement or of a rotating object. If properly recognized from the analysis of the EEG signals, each mental task can then be used as a command. Typical applications include wheelchair or virtual keyboard control in the context of rehabilitation for disabled people, and more recently entertainment.

Synchronized BCI systems are defined by Mason *et al.* [20] as BCIs that recognize unintentional control, and that are intermittently available for control. As opposed to self-paced BCIs, synchronized BCIs require the system to prompt the user for a response and therefore ignore unexpected user input. The performance of such systems is usually measured using the information-transfer rate (also called mutual information or bit-rate), as proposed by Wolpaw *et al.* [33]:

$$B = V\left[\log_2 N + P\log_2 P + (1-P)\log_2\left(\frac{1-P}{N-1}\right)\right] \tag{1}$$

where the information-transfer rate B corresponds to the amount of information reliably received by the system, V is the application speed in trials/second (i.e. how many thoughts are recognized per second), P is the classifier accuracy (i.e. how well thoughts are recognized) and N is the number of mental tasks (or symbols) used in the "thoughts" vocabulary. This definition is well suited to compare keyboard-based BCI applications, but most likely not for pointing device BCIs where a Fitts' law [1] could be more appropriate. Moreover, it is not suited to assess either the so-called "non-comparable" BCI devices [26] or the self-paced BCIs [21], and for evaluating experiments where the number of tasks N is greater than four [17]. Nevertheless, it has the advantage of being very simple; using this measure, commonly reported bit-rates range from 5 to 25 bits/min [32].

Whereas most BCI studies use Wolpaw's ITR as a performance measure, this metric itself has not received a lot of attention. Dornhege *et al* studied the error bounds effects on Eq. 1 and suggested that increasing N produces only tiny ITR increases, and that the ITR also depends on the error rate [8]. Kronegg *et al* compared the existing ITR definitions [17] and studied the influence of the protocol speed V on Eq. 1 for average-trials protocols [16]. Dornhege *et al* [8], McFarland *et al.* [23], and Obermaier *et al.* [27] have shown that the optimal number of symbols is around 3 to 4. Schlögl *et al* compared Eq. 1 with the mutual information [29], [30].

In this paper, we demonstrate that the optimal number of mental tasks N_{opt} depends on the user skills and on the BCI design. We also show that the ITR improvement from N=2 to N=N_{opt} is low. These two questions are addressed using a theoretical BCI model and an experimental validation.

II. Theoretical Modeling

A. ITR Improvement

We model the BCI as a channel with encoder and decoder, corrupted by Additive White Gaussian Noise (AWGN) [6] as in our previous work [17], [16]. A mental task W (e.g. "mental calculation") among N possible tasks is generated by the brain, encoded into a feature vector X by a feature extraction process, and then transmitted to a system that decodes the task W. The feature vector X is contaminated by an independent AWGN

Manuscript received December 20, 2005. This work is supported in part by the Swiss National Center of Competence in Research (IM)2, Interactive Multimodal Information Management (http://www.im2.ch/), as well as by the European Network of Excellence SIMILAR (http://www.similar.cc/).

J. Kronegg, S. Voloshynovskiy, T. Pun are with the Computer Vision and Multimedia Laboratory, University of Geneva, 24 Rue Général-Dufour, 1204 Geneva, Switzerland (corresponding author: Julien.Kronegg@cui.unige.ch, phone: +41(22)3797628; fax:41(22)3797780).

$Z \sim \mathcal{N}(0, s_z^2)$ induced by the background activity of the brain, yielding the noisy feature $Y=X+Z$ (see Fig. 1). We denote by $p(y|x_i)$ the probability that a feature y is correctly recognized when x_i is emitted; since Z is Gaussian, we have :

$$p(y \mid x_i) = 1 \big/ \left(\sqrt{2p}\, s_Z\right) e^{-(y-x_i)^2 / 2 s_Z^2} .$$

Whereas not all features are Gaussian distributed in practice, the Gaussian assumption is commonly accepted in the BCI community [5], [8], [9], [12], [17], [16], [28], [29], [30], [31].

We model the feature vector X as a Pulse Amplitude Modulated (PAM) signal of N equiprobable states with constrained energy $E[X^2] \leq s_x$. The equiprobable assumption is commonly accepted by the BCI community (e.g. [8], [9], [17], [16], [28], [30], [29], [33]). The PAM assumption is reasonable, considering that several features used in the BCI community can be reduced to such PAM signal (e.g. mu/beta rhythm modulation [23], power spectral density [8], [27]). If a Bayes classifier is used, the probability that a mental task w_i is recognized as the mental task w_j becomes

$$p\left(\hat{w}_j | w_i\right) = \int_{R_j} p\left(y | x_i\right) dy \ ,$$

R_j being the Bayes decision region for w_j [10]. The terms $p(w_j|w_i)$ constitute the transition matrix. The system accuracy P is then computed as the mean of the transition matrix diagonal, which allows to compute Wolpaw's ITR as a function of the SNR and N, given the signal-to-noise ratio $\text{SNR}=10 \cdot \log_{10}(s_x^2/s_z^2)$, see Fig. 2.

The assumptions underlying Wolpaw's ITR (equiprobable classes, same accuracy for all classes, error distributed equally on the remaining classes) are not valid in all practical cases. However, our previous work [17] has shown that Wolpaw's ITR is very close to Shannon's ITR when the number of tasks does not exceed about five. The use of Wolpaw's ITR is thus justified in this paper, as the maximum number of tasks currently used in BCIs is 6.

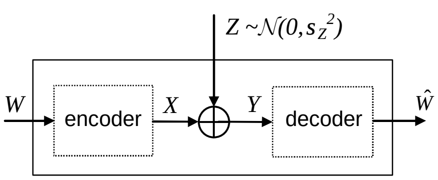

Fig. 1. BCI channel model.

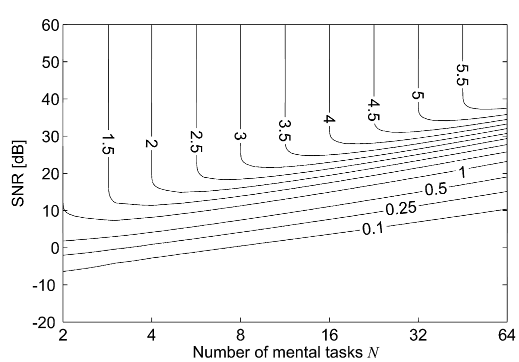

Fig. 2. Contour lines plot of constants Wolpaw's ITR in bits/trial; the ITR here varies between 0.1 and 5.5 bits/trial.

Using our channel model, we can expect that an increase in the number of mental tasks N will lead to an increase of the ITR, if and only if the SNR is sufficiently high. For example, if the SNR is around 40 dB for every N, the ITR will greatly increase as N increases (see Fig. 2). Conversely, if the SNR is around 0 dB for every N, the ITR will even decrease. Consequently, only BCIs with good accuracy (high SNR) will significantly benefit from an increase of the number of mental tasks.

B. Optimal Number of Tasks

We propose to describe the classification accuracy P as a function of the number of tasks N using a linear model:

$$P(N) = a\,N + b, \text{ with } a < 0 \text{ and } 1/N \leq P(N) \leq 1. \tag{2}$$

where the slope a depends on the user skill as well as on the BCI design, and the intercept b is such that $0.5 \leq P(2) \leq 1$ since in the case of N=2 classes one wants to have an accuracy higher than or equal to the random guess; thus, b varies between $b_{min}=0.5-2 \cdot a$ and $b_{max}=1-2 \cdot a$, see Fig. 3.

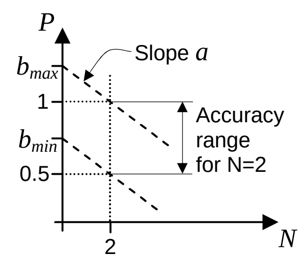

Fig. 3. Minimum and maximum values of intercept b.

The choice of a linear model is justified because the number of possible values of N is small, typically limited to 7±2 for protocols implying the memorization of mental tasks [25]. Consequently, using more complex models does not seem necessary. However, the use of linear, logarithmic, and second order polynomial models are discussed in Section IV.

For a given BCI design, users with differing skills will not perform in the same way. The user skills are measured here using the classifier accuracy P. The parameters a and b and thus $P(N)$ will therefore differ from one user to another. Similarly, for a given user, all BCI designs will not perform identically; different BCIs will exhibit differing a, b and thus $P(N)$. It is however difficult to differentiate the impact of the user skills from the impact of the BCI design: users are not usually shared by BCI researchers and BCI designs are tuned for a given set of test users. In this work, we will use the average participant performance on our own users to determine to which extent the BCI design influences the parameters a and b.

Using this accuracy model, we can compute the ITR from Eq. 1 and the corresponding optimal number of tasks $N_{opt} = \text{argmax}_N\, B(N, P(N), V\text{=constant})$ for each pair of $(a; b)$ values, see Fig. 4.

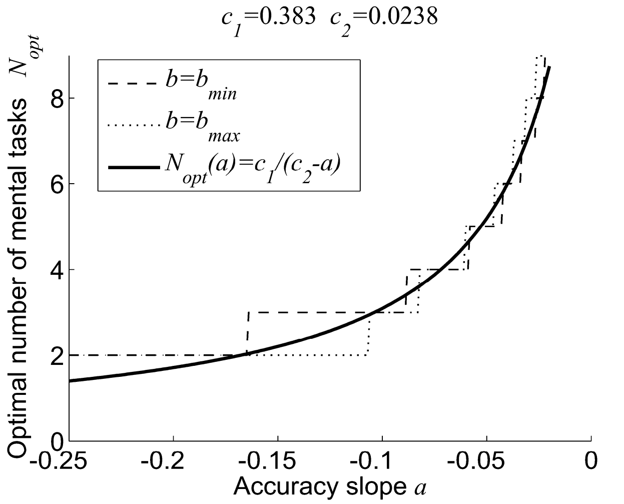

Fig. 4. Optimal number of tasks for several values of a and upper/lower values of b. The continuous line is the N_{opt} model.

As shown on Fig. 4, the optimal number of tasks N_{opt} mostly depends on a and to a lesser extent on b: the optimal number of tasks N_{opt} differs from at most one mental task for minimum and maximum values b.

The model described by Eq. 2 allows determining the feasibility of a specific BCI design. For example, if one wants to design a BCI with N_{opt} = 8 mental tasks, the linear model of Fig. 4 leads to a value of a=-0.04. Considering the best case $b=b_{max}$=1.08, the classifier accuracy must be $P(8)=-0.04\cdot 8+1.08=0.76=76\%$. As such accuracies cannot be reached for 8 classes using current BCI technologies, it does not seem worthwhile attempting to build an 8 tasks BCI.

Further, based on Fig. 4, the following model of the relationship between N_{opt} and a is proposed:

$$N_{opt}(a) = \frac{c_1}{c_2 - a}. \quad (3)$$

This function yielded the minimum least-squares error compared with other models (linear, second and third order polynomial, logarithmic, exponential). The parameter values c_1 and c_2 have been estimated, using the least-squares error method, to c_1=0.444 and c_2=0.034 for $b=b_{min}$, and to c_1=0.320 and c_2=0.011 for $b=b_{max}$. Given the low N_{opt} dependency on b, we chose to use the mean between these two extremes, hence c_1=0.383 and c_2=0.0238. This allows to predict the optimal number of mental tasks as well as to explain why some users can have $N_{opt}>2$.

III. Experimental Validation

A. Participants

Four healthy right-handed male humans A, B, C, and D, 24 to 29 years old, participated in the study. They had no previous experience using BCI systems, except participant B (4th experiment). The participants were selected among the laboratory members for their availability and interest in the study. They were not remunerated. All participants filled a consent form before the experiment.

B. Data acquisition

EEG data was recorded at 256 Hz using a 64 electrodes Biosemi Active Two [2] system with the ABC setup. The ABC setup is a derivation of the international 10/20 electrode placement system, and allows placing of up to 256 electrodes on the scalp. The mapping between the 64 electrodes and the 256-holders electrode cap is described in [14]. Each channel was filtered using a 4-45 Hz equiripple filter, which allows removal of the power line noise (50 Hz) and of the DC and low frequency components.

C. Experimental protocol

The participant was seated on a chair in front of a LCD computer screen, in a basement office offering reasonable immunity to electromagnetic noise. Following an off-line synchronized protocol without feedback, he was instructed to execute the mental task corresponding to a trigger image displayed on the screen and depicting the required task. The term "synchronized" is used here in the sense of [20]: an intermittently available protocol and a transducer without unintentional control support. Four mental tasks were chosen (see Fig. 5): exact calculation of repetitive additions (T1), imagination of left finger movement (T2), mental rotation of a cube (T3), and evocation of a non-verbal audio signal (T4, e.g. a cell phone ring tone). According to physiological studies [7], [11], [13], [22], these mental tasks generate activity in different brain regions, which ought to make classification easier.

Each mental task (trial) lasted for 4 seconds, followed by a 1-second pause. A session was composed of 36 randomly selected trials (9 for each task), followed by a 2-minutes pause (totalling 5 minutes), see Fig. 6. During pause periods, the user could think to whatever he wanted apart from the tasks T1 to T4. Usually participants were relaxing, changing position, moving, etc, thus these periods were often characterized by a lot of muscle artifacts. During long pauses, participants were also filling in the questionnaire and discussing with the experimenter. The full experiment comprised 12 sessions (1 hour), totalling 108 trials for each mental task. After each session, participants subjectively reported on the quality of their performance using questionnaires.

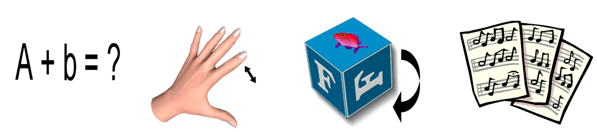

Fig. 5. Images used to instruct the user. From left to right: mental calculation, finger movement, cube rotation, auditory evocation.

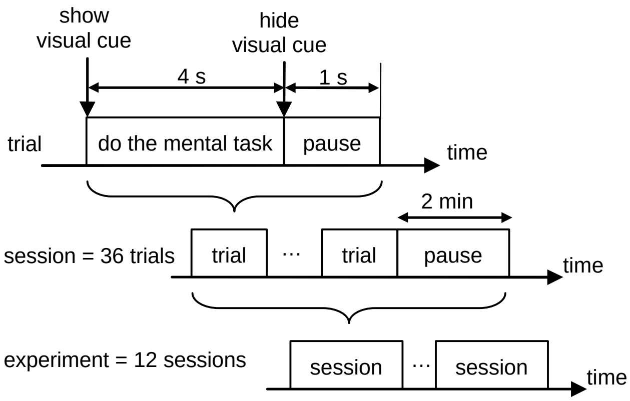

Fig. 6. Temporal structure of the protocol.

D. Features extraction

We used two feature extraction methods: power of the Short-Term Fourier Transform (STFT) and Power Spectral Density (PSD). Both feature sets were selected because of the time-frequency spatial behaviour of the brain when executing the selected mental tasks. The STFT is used instead of the Discrete Wavelet Transform because of its higher redundancy allowing easier analysis. We also preferred it over the Continuous Wavelet Transform for its simpler interpretation. We made the hypothesis that the EEG signal is stationary for periods of about 200 ms, thus each 4-second trial is decomposed in 20 timeframes by the STFT (Eq. 4). We used a 4.3 Hz frequency band width over the range 8.5-30 Hz. The number of features is consequently 7680 (64 electrodes x 20 timeframes x 6 frequency bands).

$$STFT_x^w(t',w) = \int_t \left[x(t)\cdot W(t-t')\right]\cdot e^{-jwt}dt \cdot \tag{4}$$

The PSD features were computed for 6 electrodes (C3, P3, Cz, Pz, C4, P4) referenced to the local neighbourhood mean signal, following a variant of [24]. We used the Welch periodogram method, which consists of averaging STFT periodograms along time, with a 256 samples Hamming window with 50% overlap (1 Hz resolution). The baseline PSD, computed from the EEG when the participant was not thinking to any of the protocol's mental tasks, was subtracted from the obtained PSD. The normalized PSD of the 23 frequency bands from 8 to 30 Hz were used as features, giving a total of 138 features per trial.

E. Classification

We used both a Classification And Regression Tree (CART) and a Support Vector Machine (SVM) classifier to model the features. The CART classifier is used with the Gini-index as a split criterion [4]. A five folds cross-validation was done to obtain the best classification tree by pruning. The SVM classifier is a linear SVM from the OSU-SVM Matlab toolbox [18]. All processing was done in Matlab. The SVM classifier has proven its efficiency in the BCI 2005 classification competition [3], [15]. CART was selected for its known efficiency on PSD features [24].

F. Validation

We used a sampled version of the leave-v-out stratified cross-validation (see Algorithm 1) with M=1000 sampling steps for the SVM classifier and only M=200 sampling steps for the CART classifier because CART is already cross-validated and takes more CPU power. At each pass, 80% of the instance set D is randomly taken for training (TRN_i) while the remaining 20% is left for testing (TST_i). This method allows computing the mean and variance of the classifier accuracy. We preferred it to the leave-one-out cross-validation which allows to compute the mean accuracy only, and to the *k*-fold cross-validation which allows to compute the mean accuracy and only an estimation of the variance. The drawback of the leave-v-out stratified cross-validation is its high computation time.

For *N*<4, all tasks combinations (e.g. [T1, T2], [T1, T3, T4]) have been tested. For each task combination, the best feature/classifier is identified using a 1-tailed *t*-test with 99.5% confidence level. As the *t*-test only compares only two feature/classifiers (hypothesis is “is feature/classifier A better than feature/classifier B?”), the best feature/classifier is determined by majority voting. Then, another 1-tailed *t*-test was conducted to determine the best task combination in term of accuracy.

```
for i=1 to M
  divide D in TRNi and TSTi by random
      sampling with same class probability
  hi= learn on TRNi
  accuracyi=hi(TSTi)
end for
m=mean(accuracyi)
v=var(accuracyi)
```

Algorithm 1. Sampled version of the leave-v-out stratified cross-validation. The model h_i is trained on TRN_i and tested on TST_i

IV. Results and Discussion

Participants' brain signals are classified using the two feature sets and the two classifiers described in the previous section. Participants reported a mislabelled trials rate lower than 1%. The mislabelled rate is the proportion of trials during which participant made the wrong mental task. Such low rate should not significantly affect the classification. Participants also reported that fatigue increases after 1.5 to 2 hours of experiment; only the first hour of data was thus used. The questionnaires seemed to indicate that well trained users are less subject to fatigue. However, fatigue was not our main concern in this paper, thus we did not investigate further its consequences. Tables I to IV show the mean and standard deviation of the accuracy for the best task combination for participants A to D respectively, selected by 1-tailed *t*-test. Bold font is used to highlight the best feature/classifier selected by *t*-test for each participant and for *N*=2 to 4. The *t*-test values are not shown in this paper.

TABLE I
MEAN ACCURACY AND STANDARD DEVIATION (BRACKETED) FOR USER A

N	Accuracy and std. dev. [%]			
	STFT		Welch	
	SVM	CART	SVM	CART
2 [T1,T4]	59.5 [6.9]	55.6 [7.1]	66.7 [6.6]	**77.3 [5.3]**
3 [T1,T2,T4]	39.0 [5.3]	37.3 [6.3]	46.4 [5.3]	**51.7 [5.4]**
4	29.6 [4.5]	29.2 [4.3]	34.0 [4.4]	**38.7 [4.6]**

TABLE II
MEAN ACCURACY AND STANDARD DEVIATION (BRACKETED) FOR USER B

N	Accuracy and std. dev. [%]			
	STFT		Welch	
	SVM	CART	SVM	CART
2 [T1,T3]	**79.3 [5.5]**	61.1 [7.1]	66.8 [6.1]	56.9 [7.1]
3 [T1,T2,T3]	**65.6 [5.3]**	44.8 [5.4]	51.8 [5.4]	50.4 [5.1]
4	**54.2 [4.9]**	33.7 [4.7]	40.2 [4.4]	40.8 [4.7]

TABLE III
MEAN ACCURACY AND STANDARD DEVIATION (BRACKETED) FOR USER C

N	Accuracy and std. dev. [%]			
	STFT		Welch	
	SVM	CART	SVM	CART
2 [T1,T3]	**68.3 [5.5]**	49.3 [6.6]	54.8 [6.2]	56.2 [6.4]
3 [T1,T3,T4]	**53.3 [5.5]**	32.7 [5.8]	38.5 [5.2]	38.9 [5.5]
4	**42.4 [4.8]**	26.0 [4.4]	31.3 [4.3]	33.1 [4.1]

TABLE IV
MEAN ACCURACY AND STANDARD DEVIATION (BRACKETED) FOR USER D

N	Accuracy and std. dev. [%]			
	STFT		Welch	
	SVM	CART	SVM	CART
2 [T1,T2]	**77.2 [6.0]**	57.9 [7.1]	50.2 [2.3]	50.6 [6.1]
3 [T1,T2,T3]	**48.2 [5.7]**	41.2 [5.7]	33.9 [1.6]	43.3 [5.8]
4	**35.0 [4.6]**	29.3 [4.6]	25.8 [1.4]	33.9 [4.6]

Depending on the participant and on method used, the difference between the minimum and maximum accuracies can be up to 30% for N=2, which is represented by the high standard deviations from Tables 1 to 4. This can be explained by the fact that the two classifiers used are known to have low bias but high variance. The best feature/classifier pair is user dependant: for three participants, the STFT features/SVM classifier pair gives the highest accuracy, while in one participant, the Welch features/CART performs better. This confirms Wolpert's "no free lunch theorems" [34], stating that "for any algorithm, any elevated performance over one class of problems is exactly paid for in performance over another class". The *significant* dependence of the classification results on the measurement session and on the subject allows considering distinct measurements as separate classification problems.

The information-transfer rates are computed using Eq. 1 in which the accuracies are those produced by the best classifier for each participant. They are reported in Table V.

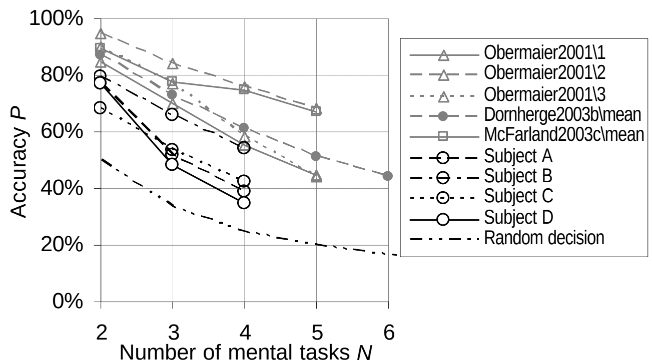

Fig. 7. Accuracies for the best task combination and the best feature/classifier for participants A to D and comparison with results from other studies.

TABLE V
ITRS FOR USERS A TO D. THE HIGHEST ITR IS HIGHLIGHTED IN BOLD FONT.

N	ITR [bits/trial]			
	A	B	C	D
2	**0.227**	0.264	0.099	**0.226**
3	0.103	**0.312**	**0.121**	0.068
4	0.066	0.279	0.104	0.036

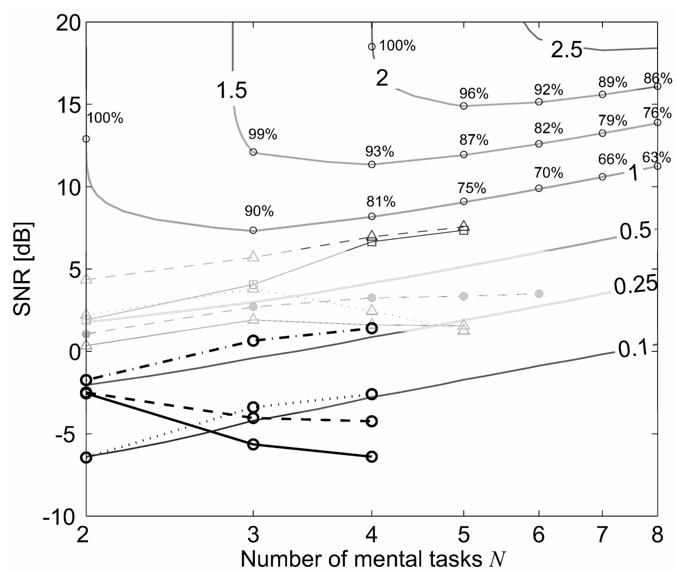

Fig. 8. ITRs for participants A to D and from other studies (same legend as in Fig. 7).

A. ITR Improvement

Not all participants exhibited an increase in ITR when the number of mental tasks increased. In our case, this optimal number of mental tasks is N=2 or 3 (0.12 to 0.31 bits/trial, see Table V).

Following the taxonomy from Mason *et al* [19], we selected several BCI systems ([8], [23], [27]) that are comparable, using the ITR, in the sense of [26]: all these BCIs use endogenous transducers and EEG signal, and have discrete classes without idle support, except [23] which uses discretized continuous classes. The average gain in increasing the number of tasks from N=2 to the optimal N_{opt} is in all cases relatively small (0.02 bits/trial in our case, 0.02 bits/trial for [8], 0.04 bits/trial for [27], and 0.25 bits/trial for [23]), because the ITR curves tend to “follow” the constant ITR curves from Fig. 8. Therefore, BCIs with "low" classification accuracy will not exhibit significant ITR improvements when increasing the number of mental tasks. This confirms Dornhege *et al.* theoretical framework [8] which also predicts that increasing the number of mental tasks to more than 3 will lead to small ITR improvements for typical BCI accuracies (i.e. more than 10% of errors for 2-class problems). This applies to all current state-of-the-art EEG-based BCIs, see Fig. 8. For BCIs relying on other types of electrodes, e.g. Electro-Cortico-Graphic or intra-cortical electrodes, it is likely that the signal-to-noise ratio will be higher than with scalp EEG signals, leading to a higher optimal number of mental tasks and to a higher ITR improvement.

B. *Optimal number of tasks*

The results obtained in the present as well as in other studies (see Fig. 7 and Tables I to IV) confirm that N and P are interdependent: P decreases when N increases. This confirms the existence of an optimal number of mental tasks N_{opt} for Eq. 1. The experimental accuracy (Fig. 7) is significantly correlated with the one given by the linear model proposed in Eq. 2 (mean R^2=0.981). The correlation with logarithmic and second order polynomial models are even higher (respectively R^2=0.990 and R^2=0.996), but as stated in the model definition, such complex models are not reasonable as the number of tasks is small, e.g. limited to 4 for our participants. However, if the number of mental tasks were to increase in the future with the progress in analysis techniques, the logarithmic or second order polynomial models could be more suited for our accuracy model; the linear model would represent the worst-case scenario. Moreover, while Eq. 1 underestimation of Shannon ITR increases with the number of tasks used [17], this underestimation remains small, e.g. about only 0.2 bits for N=10. This allows using the N_{opt} model (Eq. 3) to determine the optimal number of mental tasks, see Table VI.

TABLE VI
MEASURED AND MODELED OPTIMAL NUMBER OF SYMBOLS FOR OUR USERS AND OTHER STUDIES RESULTS.

Participant	Accuracy slope a	Measured N_{opt}	Modelled N_{opt}
A	-0.193	2	1.8
B	-0.126	3	2.6
C	-0.130	3	2.5
D	-0.211	2	1.6
Mean of [8]	-0.108	3	2.9
Mean of [23]	-0.070	4	4.1
Mean of [27]	-0.127	3	2.5

Fig. 9 compares the proposed accuracy model from Eq. 2 with real data, for Table VI maximum and minimum accuracy slopes a (-0.07 and -0.211 respectively). This figure shows that the optimal number of mental tasks is not necessarily 3 to 4 as stated in [8], [23], [27], but varies according to the accuracy slope a.

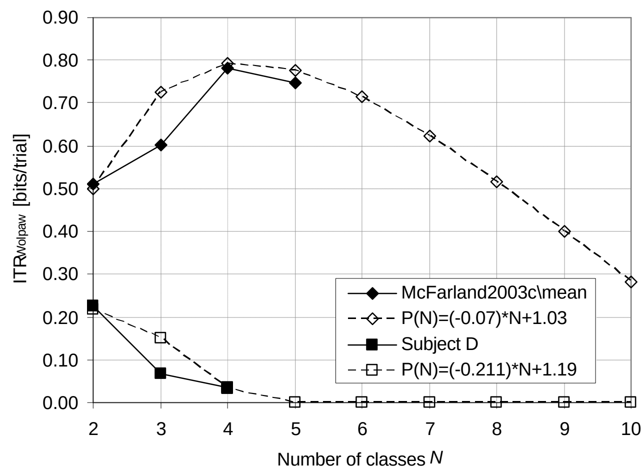

Fig. 9. Comparison between the proposed model from Eq. 2 (dashed lines) and real data (plain lines), for two different accuracy slopes a.

The modeled N_{opt} is strongly correlated with the measured N_{opt} (R^2=0.94, or R^2=1 when the modeled N_{opt} is rounded towards the next integer value). The proposed model can thus be considered as valid. This also confirms Dornhege *et al* suggestion that the optimal number of tasks is user dependant [8]. The average across participants over the accuracy slopes a computed from data reported in [8], [23],[27] differs from our own values; this confirms that the optimal number of mental tasks is also BCI dependant. However, as previously stated, this is only indicative since participants are not shared between researchers.

V. CONCLUSIONS

Commonly reported results state that the optimal number of mental tasks for BCI applications is 3 to 4. It is further experimentally demonstrated by some researchers that this optimal number could be user-dependant. In this paper, we proposed and validated a model which allows to compute the optimal number of mental tasks and which explains the dependency of this number on both user skills and BCI design.

Our model also showed that an increase from N=2 to N=N_{opt} mental tasks only produces a very limited improvement in terms of information-transfer rate. This is due to the fact that current BCI accuracies are too low to produce significant ITR increases. This is in accordance with Dornhege *et al.* theoretical model [8].

Even if increasing the number of tasks would lead to an increase in ITR, the small gain might not justify the added complexity in terms of protocol design. We can thus conclude that it is currently premature to aim at significantly increasing the information-transfer rate by increasing the number of tasks; improving classification accuracy should be the first target. This might prove to be difficult, given the accuracies that must be attained to obtain a given ITR, see Fig. 8. For instance, at N=4, the accuracy must be 81%, 93% and almost 100% to obtain an ITR of 1, 1.5 and 2 bits/trial, respectively.

ACKNOWLEDGMENT

The authors would like to acknowledge fruitful discussions with Dr. Teodor Iulian Alecu and Guillaume Chanel.

REFERENCES

[1] R. M. Baecker and W. A. S. Buxton, *Readings in Human-Computer Interaction: a multidisciplinary approach*. Los Altos, California: Morgan Kaufmann Publishers, Inc., 1987.

[2] Biosemi, "Biosemi web site," 2005, Available: http://www.biosemi.com/.

[3] B. Blankertz, "BCI competition III," 2005, Available: http://ida.first.fraunhofer.de/projects/bci/competition_iii/.

[4] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Boca Raton: Chapman and Hall, 1993.

[5] P. Celka, B. Boashash, and P. Colditz, "Preprocessing and Time-Frequency Analysis of Newborn EEG Seizures," *IEEE Engineering in Medicine and Biology*, vol. 20, pp. 30-39, 2001.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*: John Wiley & Sons Inc, 1991.

[7] S. Dehaene, E. Spelke, P. Pinel, R. Stanescu, and T. S., "Sources of mathematical thinking: Behavioral and brain-imaging evidence," *Science*, pp. 970-974, 1999.

[8] G. Dornhege, B. Blankertz, G. Curio, and M. K.-R., "Increase information transfer rates in BCI by CSP extension to multi-class," presented at Advances in Neural Inf. Proc. Sys. (NIPS03), 2003.

[9] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 993-1002, 2004.

[10] R. O. Duda, P. E. Hart, and S. D. G., *Pattern Classification*, Second Edition ed: John Wiley & Sons, Inc., 2000.

[11] A. R. Halpern and R. J. Zatorre, "When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies," *Cerebral Cortex*, vol. 9, pp. 697-704, 1999.

[12] J. Kohlmorgen and B. Blankertz, "Bayesian Classication of Single-Trial Event-Related Potentials in EEG," presented at Artifical Neural Networks (ICANN 2002), 2002.

[13] S. M. Kosslyn, G. J. Digirolamo, W. L. Thompson, and N. M. Alpert, "Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography," *Psychophysiology*, vol. 35, pp. 151-161, 1998.

[14] J. Kronegg, T. I. Alecu, and G. Chanel, "Electrode placement at the Computer Vision and Multimedia Lab," University of Geneva, Switzerland, *unpublished*, 2005, Available: http://vision.unige.ch/MMI/bci/papers/Kronegg2005__CVML_Electrodes_position_64_256_c.pdf.

[15] J. Kronegg and D. A. Rofes-Gonzalez, "BCI Competition III - dataset V - TR05-02," Computer Science Department Technical Report, University of Geneva, Switzerland, 2005.

[16] J. Kronegg, A. T., and T. Pun, "Information theoretic bit-rate optimization for average trial protocol Brain-Computer Interfaces," presented at HCI International, Crete, Greece, 2003.

[17] J. Kronegg, S. Voloshynovskiy, and T. Pun, "Analysis of bit-rate definitions for Brain-Computer Interfaces," presented at Int. Conf. on Human-computer Interaction (HCI'05), Las Vegas, Neveda, USA, 2005.

[18] J. Ma and Y. Zhao, "OSU-SVM Matlab toolbox v3.0," 2002, Available: http://www.ece.osu.edu/~maj/osu_svm/.

[19] S. G. Mason, A. Bashashati, M. Fatourechi, K. F. Navarro, and G. E. Birch, "A Comprehensive Survey of Brain Interface Technology Designs," *Annals of Biomedical Engineering (submitted for publication)*, 2005.

[20] S. G. Mason and G. E. Birch, "Temporal Control Paradigms for Direct Brain Interfaces – Rethinking the Definition of Asynchronous and Synchronous," presented at HCI International, Las Vegas, Nevada, USA, 2005.

[21] S. G. Mason, J. Kronegg, J. E. Huggins, A. Schlögl, M. Fatourechi, R. Kaidar, R. Scherer, and A. Buttfield, "Asynchronous BCI performance evaluation," *BCI-info.org research papers*, 2005, Available: http://bci-info.org/Research_Info/documents/articles/AsynchBCIDiscussionSummaryBCI2005.pdf.

[22] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topography*, vol. 12, pp. 177-186, 2000.

[23] D. J. McFarland, W. A. Samacki, and J. R. Wolpaw, "Brain-computer interface operation: Optimizing information transfer rates," *Biological Psychology*, vol. 63, pp. 237-251, 2003.

[24] J. d. R. Millan, M. Franzé, J. Mourino, F. Cincotti, and F. Babiloni, "Relevant EEG features for the classification of spontaneous motor-related tasks," *Biol. Cybern.*, vol. 86, pp. 89-95, 2002.

[25] G. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *The Psychological Review*, vol. 63, pp. 81-97, 1956, Available: http://www.well.com/user/smalin/miller.html.

[26] M. M. Moore-Jackson, S. G. Mason, and G. E. Birch, "Analyzing Trends in Brain Interface Technology: A Method to Compare Studies," *Annals of Biomedical Engineering (submitted for publication)*, 2005.

[27] B. Obermaier, C. Neuper, C. Guger, and G. Pfurtscheller, "Information Transfer Rate in a Five-Classes Brain-Computer Interface," *IEEE Trans. on Neural Sys. And Rehab. Eng.*, vol. 9, pp. 283-288, 2001.

[28] M. Sahin, "Noise tolerance as a measure of channel discrimination for multi-channel neural interfaces," presented at Annual Int. Conf. IEEE Eng. in Med. and Biol. Soc., 2001.

[29] A. Schlögl, C. Keinrath, R. Scherer, and G. Pfurtscheller, "Information transfer of an EEG-based Brain-computer interface," presented at Int. IEEE EMBS Conf. on Neural Engineering, 2003.

[30] A. Schlögl, C. Neuper, and G. Pfurtscheller, "Estimating the Mutual Information of an EEG-based Brain-Computer Interface," *Biomedizinische Technik*, vol. 47, pp. 3-8, 2002.

[31] R. C. Smith, "Electroencephalograph based Brain Computer Interfaces," Faculty of Electrical and Electronic Engineering, University College Dublin, Dublin 2004 2004.

[32] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. Hunter Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-Computer Interface Technology: A Review of the First International Meeting," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, pp. 164-173, 2000.

[33] J. R. Wolpaw, H. Ramoser, M. D. J., and G. Pfurtscheller, "EEG-Based Communication: Improved Accuracy by Response Verification," *IEEE Trans. Rehab. Eng.*, vol. 6, pp. 326-333, 1998.

[34] D. H. Wolpert and W. G. Macready, "No Free Lunch Theorems for Optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, pp. 67-82, 1997, Available: http://www.no-free-lunch.org/.