2005 Int. Conf. on Human-computer Interaction (HCI'05), Las Vegas, Nevada, USA, June 20-23, 2005.

http://vision.unige.ch/

Analysis of bit-rate definitions for Brain-Computer Interfaces

Julien Kronegg **Svyatoslav Voloshynovskiy** **Thierry Pun**

Computer Vision and Multimedia Laboratory, Computer Science Dept., University of Geneva, Switzerland

Email: {first name.surname}@cui.unige.ch, phone: +41 22 3797628, fax: +41 22 3797780

Abstract - A comparison of different bit-rate definitions used in the Brain-Computer Interface (BCI) community is proposed; assumptions underlying those definitions and their limitations are discussed. Capacity estimates using Wolpaw and Nykopp bit-rates are computed for various published BCIs. It appears that only Nykopp's bit-rate is coherent with channel coding theory. Wolpaw's definition might lead to underestimate the real bit-rate and to infer wrong conclusions about the optimal number of symbols; its use should be avoided. The usage of a proper bit-rate assessment is motivated and advocated. Finally, it is found that the typical signal-to-noise ratio of current BCIs lies around 0 dB.

Keywords: brain-computer interface, bit-rate, information transfer rate, number of classes, information theory.

1 Introduction

A Brain-Computer Interface (BCI) is an input device that allows a user to drive a specific application (e.g. virtual keyboard [10], cursor control [28], robot control [24]) using EEG data induced by thinking to a specific notion or mental state (e.g. mental calculation, imagination of movement, mental rotation of objects). This mental state is then recognized by the machine using a classifier. The first objective BCI performance measure is due to Wolpaw *et al* in 1998 [28], where the bit-rate, or information-transfer rate, was defined based on Shannon channel theory with some simplifying assumptions. Bit-rates commonly reported range from 5 to about 25 bits/minute [29].

In this article, we compare the bit-rate definitions used in the BCI domain and propose recommendations for optimizing the number of mental states in a BCI. The article is organized as follows: first, a review of the noisy channel theory is presented, as a support to the BCI model described. Existing bit-rates definitions used in the BCI domain are then presented and analyzed.

2 Noisy Channel Theory

A channel is a communication medium that allows the transmission of information from a sender A to a receiver B. Due to imperfections in that medium, the transmission process is subject to noise and B might receive information differing from the one emitted by A. The simplest noisy channel is the additive noise channel where the received signal Y is the sum of an emitted signal X and some independent noise Z here assumed Gaussian. Since we deal with real, physical input signals, the input signal energy is limited (which also implies that X has zero mean in order to minimize its energy, $E[X^2]\leq\sigma_X$). The information channel capacity is the quantity of reliable information carried by one symbol transmitted through the channel.

The channel capacity depends on the input signal distribution as well as on the signal-to-noise ratio (SNR) [5]. For continuous input signal and using $SNR=10\cdot\log_{10}(\sigma_X^2/\sigma_Z^2)$, the capacity (in bits/symbol) is:

$$C = 0.5\cdot\log_2\left(1+10^{SNR/10}\right) \quad (1)$$

For discrete Pulse Amplitude Modulated input with N symbols of a priori probability $p(X=x_i)=1/N$ (denoted $p(x_i)$), the capacity C_N is defined by Eq. 2.

$$C_N = \sum_{i=1}^{N}\int_{y=-\infty}^{+\infty} p(y\mid x_i)\,p(x_i)\log_2\frac{p(y\mid x_i)}{p(y)}dy \quad (2)$$

$$p(y\mid x_k) = \frac{1}{\sqrt{2\pi}\sigma_Z}e^{-(y-x_k)^2/2\sigma_Z^2}$$

$$p(y) = \sum_{j=1}^{N} p(x_j)\,p(y\mid x_j)$$

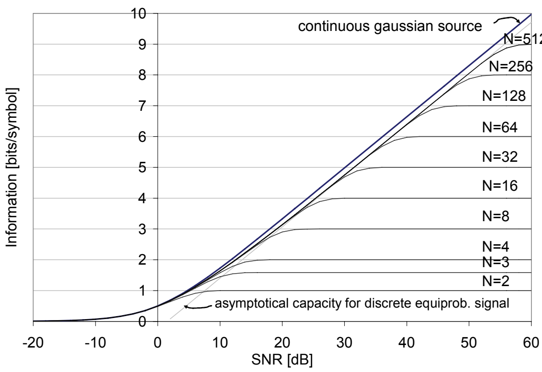

Figure 1: Comparison of the capacity for Gaussian input C and for discrete equiprobable input C_N as a function of the number of symbols N and of the SNR.

The probability $p(y|x_i)$ is the probability that the continuous symbol y is recognized when the symbol x_i is sent. The variance σ_Z^2 of the noise Z is given by $\sigma_Z^2 = \sigma_X^2 \cdot 10^{-SNR/10}$.

The capacity C_N has to be determined by numerical integration. Figure 1 compares the continuous capacity C (Eq. 1) and the discrete equiprobable capacity C_N (Eq. 2). There is an asymptotic difference of 1.53 dB between C and C_∞ (the discrete capacity for $N=\infty$), so called shaping loss. In all cases, the continuous capacity is greater than the discrete equiprobable capacity.

3 BCI Model

The BCI is modeled using an additive white Gaussian noise (AWGN) channel as follows [15], [25], [26], see Figure 2. A mental task w_i (e.g. mental calculation), selected amongst N possible mental tasks, is encoded by the brain, producing a discrete memoryless feature x_i (e.g. power spectral density, auto-regressive model coefficients). This feature is perturbed by an additive Gaussian noise $Z\sim\mathcal{N}(0,\sigma_z^2)$ induced by the background brain activity, considered as independent of X. The resulting feature $Y=X+Z$ is decoded as $\hat{W}$ by a classifier able to recognize M symbols ($M=N+1$ for classifiers with rejection capability and $M=N$ for classifiers without rejection capability). No prior knowledge is assumed about the occurrence probability of classes, therefore they can be in general considered as non-equiprobable ($p(w_i)\neq 1/N$). The capacity can thus be computed by numerical integration using Eq. 2.

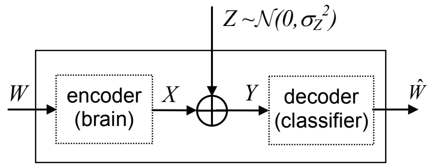

Figure 2: Model of the BCI using an AWGN channel.

The transition matrix $p(\hat{w}_j|w_i)$, also called confusion matrix, is computed during the classifier training phase. This matrix describe the probability that a mental task w_i is recognized as a mental task $\hat{w}_i$. The diagonal of the transition matrix is the classifier accuracy, thus a perfect classifier has an identity confusion matrix [20].

This BCI model however does not truly correspond to a real BCI application, but more to an ideal BCI. First, the source is not always memoryless. For example, in average-trial protocols, the user repeats the same symbol a given number of times: the probability that the next symbol is the same than the current one is high. Secondly, the type of noise (background brain activity) depends on the application, and the noise is probably not always Gaussian distributed. Finally yet importantly, the fact that symbol values are most of the time imposed by the used feature might impose additional constraints on x_i.

4 Review of Existing Bit-Rate Definitions

Farwell and Donchin have proposed the first bit-rate definition in 1988 [10] when they designed the first BCI. The second definition is due to Wolpaw *et al* (1998, [28]), the third one to Nykopp (2001, [20]). All theses definitions are based on Shannon channel capacity. All bit-rates are indicated in bits/symbol, and can be converted to bits/minute according to $B=V\cdot R$ with V being the classification speed (in symbols/minute) and R the information carried by one symbol (in bits/symbol).

The most popular definition is the one from Wolpaw, which is reasonably simple and has often been used ([27], [28], [29], [30] and papers in Table 1). The most generic definition is the one from Nykopp, which corresponds to Shannon channel capacity theory (see previous sections). Wolpaw's as well as Farwell and Donchin's definitions are in fact simplifications of Nykopp's definition.

4.1 Farwell & Donchin Definition

When designing the first BCI in 1988, Farwell & Donchin introduced a bit-rate definition that did not take the classifier accuracy into account. The assumptions made were of a classifier without rejection ($M=N$), and perfect (i.e. no classification error). This therefore leads to an identity transition matrix $p(\hat{w}_j|w_i)=I$ of size NxN. The mental states were assumed to be equiprobable, thus $p(w_i)=1/N$. From Eq. 6, these assumptions lead to the following bit-rate definition (in bits/symbol):

$$R_{Farwell\&Donchin} = \log_2 N .$$

4.2 Wolpaw Definition

In 1998, Wolpaw *et al* suggested that it could be interesting to consider the performance measurement not only from the accuracy standpoint but also from the information rate point of view [28]. Using the definition of the information rate proposed by Shannon (see Eq. 6) for noisy channels, they made some simplifying assumptions.

First, they supposed that N symbols are recognized if N symbols are emitted by the user. They did not consider additional symbols (such as a "not recognized" mental state) as would be the case for classifiers with rejection, or their equivalent erasure channels. The second assumption is that the symbols (or mental states) have all the same a priori occurrence probability $p(w_i)=1/N$. The third assumption is that the classifier accuracy P is the same for all target symbols ($p(\hat{w}_j|w_i)=P$ for $i=j$))[1]. The fourth assumption is that the classification error $1-P$ is equally distributed amongst all remaining symbols ($p(\hat{w}_j|w_i)=(1-P)/(N-1)$ for $i\neq j$). From Eq. 6, all theses assumptions lead to the following simplified bit-rate definition (in bits/symbol):

$$R_{Wolpaw} = \log_2 N + P\cdot\log_2 P + (1-P)\cdot\log_2\frac{1-P}{N-1}. \quad (5)$$

4.3 Nykopp Definition

Nykopp's capacity has been introduced in the framework of the Adaptive Brain Interface (ABI) project [20]. The ABI is a BCI with rejection capacity meaning that no decision is taken if the confidence level of the classification does not exceed a certain threshold. This is modeled by means of an erasure channel[2] where some symbols might be lost during transmission [5]. In summary, Nykopp's capacity is defined by (in bits/symbol):

$$C = \max_{p(x)} I(X;Y)$$

$$R_{Nykopp} = I(W;\hat{W}) = H(\hat{W}) - H(\hat{W}|W) \quad (6)$$

$$H(\hat{W}) = -\sum_{j=1}^{M} p(\hat{w}_j)\cdot\log_2 p(\hat{w}_j)$$

$$p(\hat{w}_j) = \sum_{i=1}^{N} p(w_i)\cdot p(\hat{w}_j|w_i)$$

$$H(\hat{W}|W) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(w_i)\cdot p(\hat{w}_j|w_i)\cdot\log_2 p(\hat{w}_j|w_i)$$

The a priori symbol probabilities $p(w_i)$ are computed by means of the Arimoto-Blahut optimization algorithm [14], in order to obtain the best bit-rate or capacity for the underlying channel specified by a given transition matrix.

5 Methods

In order to compare bit-rate definitions, we will compute the difference between the discrete equiprobable capacity and the bit-rates given by each definition. To this end, we model the discrete input signal as a Pulse Amplitude Modulation (PAM) signal with N symbols, the noise Z as independent and Gaussian distributed as in the proposed model. All symbols have the same probability $p(X=w_i)=1/N$, denoted by $p(w_i)$. This leads to the signal distribution shown in Figure 3.

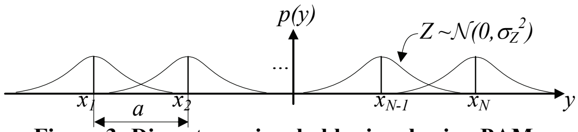

Figure 3: Discrete equiprobable signal using PAM.

Although PAM does not appropriately model all types of BCI features, it appears to be adequate for at least several classes of them (e.g. P300, SSVEP and Common Spatial Patterns; PAM modelling of other classes is under investigation). Furthermore, PAM is chosen since it models near to optimal features leading to an information transfer rate close to the capacity. More complex feature models like Quadrature Amplitude Modulation could be used to lower the shaping loss, but PAM is a good compromise between complexity and performance. Equiprobable symbols are used since it would be impractical to simulate every possible symbol distribution, and because Wolpaw's bit-rate definition does not support non-equiprobable symbols. In this case, the accuracy P is therefore computed as the maximum of the transition matrix diagonal; the use of the maximum instead of the mean models the worst case.

A Bayes hard-classifier has been used, since it is known to be the near-to-optimal hard-classifier if the underlying distribution of the data is known [9] (which is the case since Z is known, see Figure 3). Soft-classifier would produce even better results but a hard classifier is sufficient to demonstrate the differences between the bit-rate definitions. The classifier is considered to have no rejection capability ($M=N$) to ensure a fair comparison between bit-rate definitions.

6 Results and Discussion

As a hard-classifier is being used, we can expect that the bit-rate will be smaller than or equal as the discrete equiprobable capacity: the smaller the difference is, the better the bit-rate definition will be. The only difference between bit-rate definition parameters being the transition matrix, we can expect that the divergence between results will be due to the transition matrix only.

6.1 Discussion of Farwell&Donchin Bit-rate Definition

In Farwell & Donchin bit-rate definition, the classifier is considered perfect. This does of course not correspond to a real BCI. This definition leads to a strong bit-rate over-estimation in low SNRs (see Figure 4), especially because the typical BCI SNR is around 0 dB (see Table 1). This definition should therefore not be

[1] The mean or the maximum of the transition matrix diagonal is sometimes used when the classifier accuracy is not the same for all classes [21]. The use of the mean is only correct is the symbols are equiprobable; in the other case, the diagonal must be weighted by the a-priori probabilities.

[2] In this definition, it is considered that the transition matrix and the matrix that describes the channel are the same.

used in practice. Its main interest is to show the maximum achievable bit-rate for high SNRs.

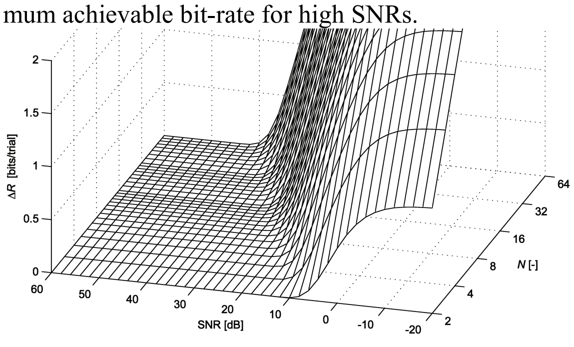

Figure 4: Farwell & Donchin bit-rate minus the discrete equiprobable capacity.

6.2 Discussion of Wolpaw's Bit-rate Definition

Figure 5 compares Wolpaw's bit-rate (Eq. 5) with the discrete signal capacity for equiprobable symbols. The difference should ideally be zero, but as seen in the previous section this is not the case due to the hard classifier and of PAM features. Furthermore, for Wolpaw's definition, this difference increases with the number of classes for specific values of signal-to-noise ratio. Wolpaw's bit-rate definition is therefore not appropriate for more than 4 or 5 symbols.

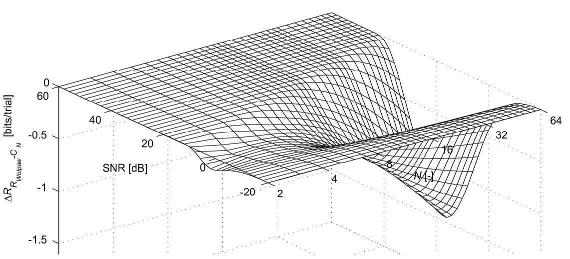

Figure 5: Wolpaw's bit-rate definition minus the discrete equiprobable capacity. The difference increases with N.

This increasing difference with the number of classes could be explained by the fact that Wolpaw's assumptions (Eq. 5) do not hold in a number of practical situations. First, the number of recognized symbols is not always equal to the number of input symbols, like in the ABI of Millán *et al* where the classifier has a rejection capability [16], [20], [24]. Secondly, the a priori occurrence probability $p(w_i)$ is not always the same for all symbols, as has been shown in a number of studies [12], [13], [20], [21], [23]. This is especially true when using the oddball paradigm [1], [2], [7], [10], if the application is a virtual keyboard (due to unequal letter appearance frequencies [7], [10]), or if average-trial protocols are used [1], [7], [10], [15], [19] where the probability of the next symbol depends on the current symbol. Thirdly, the classifier accuracy $p(\hat{w}_i | w_i)$ has also been shown to differ between symbols [2], [3] [11], [16], [17], [20], [21], [22], [24] and thus cannot be reduced to the scalar accuracy *P*.

Finally, the error is not equally distributed over the remaining symbols [2], [3], [11], [16], [17], [20], [21], [24]. In fact, given the theoretical framework presented here (discrete memoryless source, equiprobable symbols, M=N), the only difference between the results from our model and those using Wolpaw's assumptions is the classification error distribution on those remaining mental tasks. Thus, the difference shown on Figure 5 only depends on this error distribution.

6.3 Discussion of Nykopp's bit-rate definition

In Nykopp's bit-rate definition, the a priori symbol probabilities $p(w_i)$ are computed by means of the Arimoto-Blahut optimization algorithm [14], in order to obtain the best bit-rate or capacity for the underlying channel specified by a given transition matrix. This does not truly correspond to a real BCI because in most cases the application imposes the a priori symbols probability, e.g. [12], [20], [21]. Furthermore, an equiprobable a priori symbol probability is needed for comparison. The Arimoto-Blahut optimization algorithm was therefore not been used for comparison.

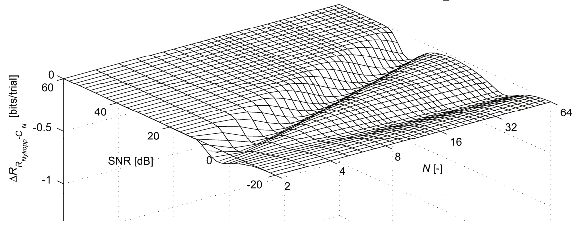

Figure 6: Nykopp's bit-rate definition minus the discrete equiprobable capacity.

Figure 6 compares Nykopp's bit-rate definition with the discrete signal capacity for equiprobable symbols. Ideally, the difference should be zero, but due to the use of a hard classifier and of PAM features, this difference is bounded by 0.3 bit/trial.

6.4 Comparison between Nykopp's and Wolpaw's bit-rate definitions

Using the previous hypotheses, Figure 7 presents the difference between Wolpaw's bit-rate and Nykopp's bit-rate $\Delta R = R_{Wolpaw} - R_{Nykopp}$. It shows that depending on the number of symbols and on the SNR, Wolpaw's bit-rate can be higher, lower or equal to Nykopp's bit-rate. For N=2 (error distributed on one symbol only) or when the SNR is very high or very low, both definitions lead to the same bit-rate. For typical BCI SNR values (about 0 dB, see Table 1), Wolpaw's definition is at most 0.1 bit/symbol greater than Nykopp's definition if $N \le 5$, and lower than Nykopp's definition for $N>5$ (the difference increases with N). Two BCIs with bit-rates computed using different definitions should therefore not be directly compared, even if the difference between Wolpaw and Nykopp definitions are small for typical number of classes and SNRs.

Based on the analysis of a number of experimental published protocols and results, it has been possible to compute Wolpaw's and Nykopp's bit-rates in bits/trial and in bits/minute (Table 1). This table only intends to compare

Table 1. Computation of the bit-rates for various published BCIs according to Wolpaw's and Nykopp's definitions (R_{Wolpaw} and R_{Nykopp}, in bits/trial). The mean accuracy $\overline{P}$ (in %) and speed $\overline{V}$ (in trials/minute) are calculated for a given experiment taking all users into account. The SNR (in dB) is determined using a simulated transition matrix. The bit-rate B is computed in bits/minute using R_{Wolpaw} and $\overline{V}$ for the sake of comparison with others studies. A plus (+) denotes articles where the transition matrix was simulated to allow for Nykopp's bit-rate calculation. "NA" stands for "Not Available".

BCI group	N	$\overline{P}$	R_{Wolpaw}	R_{Nykopp}	SNR	$\overline{V}$	B
[4]+	10	90.0	2.54	2.51	17.5	10.8	27.4
[19]+	36	80.0	3.42	3.78	25.0	11.1	38.0
BerlinBCI [8]+	2	87.5	0.46	0.46	1.2	13.3	6.1
BerlinBCI [8]+	3	73.1	0.47	0.46	0.8	13.3	6.3
BerlinBCI [8]+	4	61.2	0.42	0.42	0.1	13.3	5.6
BerlinBCI [8]+	5	51.9	0.36	0.37	-0.5	13.3	4.8
BerlinBCI [8]+	6	44.8	0.31	0.34	-0.9	13.3	4.1
Graz-BCI [21]+	2	91.0	0.56	0.56	2.5	9.5	5.4
Graz-BCI [21]+	3	78.0	0.60	0.58	2.3	9.5	5.7
Graz-BCI [21]+	4	63.0	0.46	0.46	0.6	9.5	4.4
Graz-BCI [21]	5	52.7	0.38	0.43	-0.2	9.5	3.7
Wadsworth [18]+	2	89.3	0.51	0.51	1.9	10.9	5.6
Wadsworth [18]+	3	77.9	0.60	0.57	2.2	10.9	6.6
Wadsworth [18]+	4	74.7	0.78	0.77	4.0	10.9	8.5
Wadsworth [18]+	5	67.0	0.75	0.77	3.8	10.9	8.2
ABI [20]	3	69.2	0.39	0.43	-0.6	NA	NA
ABI [16]	3	53.1	0.12	0.15	-5.8	11.0	1.3
ABI [24]	3	88.3	0.95	0.96	5.8	NA	NA

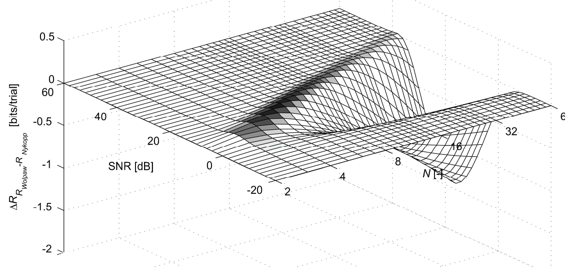

Figure 7. Difference ΔR between Wolpaw's bit-rates and Nykopp's bit-rates, in bits/trial. ΔR values are positive when Wolpaw's bit-rates are larger than Nykopp's bit-rates and negative when lower. Since positive values are very small, they are marked in dark.

Nykopp's and Wolpaw's bit-rate definitions; it does not compare different BCI paradigms. Most of the studies shown here use Wolpaw's bit-rate definition with N>2. The SNR has been determined using a simulated transition matrix (see the "methods" section). For published results which do not specify the transition matrix needed to compute Nykopp's bit-rate, this matrix was simulated using PAM features and the Bayes classifier theory, as previously.

These computations show that the typical BCI signal-to-noise ratio can be estimated to lie between –6 and 6 dB (see Table 1). An exception is for the first two BCIs of the Table where evoked potentials are used: these BCIs cannot really be compared to the others. The mean bit-rate $\overline{B}$ of all bit-rates B of Table 1 is 11 bits/minute, which is in the range of what is stated in [29] (about 5-25 bits/minute), but very low compared to a typical keyboard bit-rate. These SNR estimates are based on PAM features that appear to be valid for various types of BCIs (see previous section); in the other cases, the estimates might be indicative. Another method to compute the SNR is presented in [26].

7 Conclusions

In this paper, we made a comparison between existing bit-rate definitions used in the Brain-Computer Interface domain. A theoretical study has shown that bit-rates computed from Wolpaw's definition decreases when N increases for specific SNR, especially in the range of current state-of-the-art BCIs. The study of several BCIs showed that their average bit-rate is of the order of 11 bits/minute (using either Wolpaw's or Nykopp's definition) and that the typical SNR is mainly between –6 to 6 dB.

Theoretical as well as practical comparisons between Wolpaw's bit-rate definition and the discrete capacity for equiprobable symbols show that Wolpaw's bit-rates underestimates the real bit-rate for more than 4 symbols[1]. This could lead to wrong conclusions about the optimal number of symbols to use, so Wolpaw's definition should be avoided. Instead, Nykopp's bit-rate definition should be used, but without the Arimoto-Blahut optimization of the a priori symbols probability. As various BCIs currently employ Wolpaw's bit-rate, shifting to Nykopp's bit-rate would make BCI performance comparisons difficult. Therefore, to allow comparison with previous studies and if feasible, bit-rates computed according to both definitions should be indicated.

The bit-rate allows making objective comparisons between BCIs that are using different protocols or that

[1] This contradicts the classical channel theory stating that the bit-rate increases when the number of symbols increases.

are designed for different applications. It can also help to solve the features selection problem (the features with a high information-rate are more involved in the decision process) [6] or can be used as user feedback [26].

Although the bit-rate is an objective measure, its use is controversial. It may not be the right measure for all BCI applications: it is well adapted for keyboard applications, but the commonly used Fitt's law could be better suited for comparing pointing devices. For applications where a high classifier accuracy is needed (e.g. control of a wheelchair or of a robot in hostile environments), the bit-rate may not be the ideal performance measure [27] (e.g. in average-trial protocols, the speed decreases in order to increase the accuracy, leading to a low bit-rate [15]). In such critical applications, bit-rate definitions including cost-functions could lead to a better measure, but with the disadvantage of losing some "objectivity" since the cost of a wrong decision would be application dependent. For asynchronous BCIs where most of the time the user does not think to anything in particular ("idle state" [2]), the bit-rate will be lower than when using other BCI paradigms because of a higher $p(w_i)$ for that idle state. In addition, for BCIs relying on evoked potentials, the bit-rate is generally higher than for "normal" BCIs because these potentials produce higher SNRs, allowing for the use of more symbols. Such BCIs should thus not be compared with BCIs using other paradigms.

We can thus conclude that further work should be done in order to define which performance measure would be the most adapted to each BCI application.

8 Acknowledgements

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). The authors would also like to acknowledge fruitful discussions with Teodor Iulian Alecu.

9 References

[1] J. D. Bayliss, *A Flexible Brain-Computer Interface (PhD Thesis)*, University of Rochester, Department of Computer Science, Rochester, New York, 2001.

[2] G.E. Birch, and S.G. Mason, "Brain-computer interface research at the Neil Squire Foundation", *IEEE Trans. on Rehab. Eng.*, Vol. 8, No. 2, pp. 193-195, June 2000.

[3] Z. Bozorgzadeh, G.E. Birch, and S.G. Mason, "The LF-ASD brain computer interface: on-line identification of imagined finger flexions in the spontaneous EEG of able-bodied subjects", *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (ICASSP2000), Vol. 4, pp. 2385-2388, 2000.

[4] M. Cheng, D. Xu, X. Gao, and S. Gao, "Brain-Computer Interface with High Transfer Rates", *8th Int. Conference on Neural Information Processing (ICONIP2001)*, Shanghai, China, Nov. 14-18 2001.

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

[6] A. Date, "An Information Theoretic Analysis of 256-Channel EEG Recordings: Mutual Information and Measurement Selection Problem", *Int. Conf. on Independent Component Analysis and Blind Signal Separation* (ICA2001), San Diego, California, pp. 185-188, Dec. 9-12 2001.

[7] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The Mental Prosthesis: Assessing the Speed of a P300-Based Brain–Computer Interface", *IEEE Trans. on Rehab. Eng.*, Vol. 8, No. 2, pp. 174-179, June 2000.

[8] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Increase information transfer rates in BCI by csp extension to multi-class", *Advances in Neural Inf. Proc. Systems (NIPS03)*, Vol. 16, http://books.nips.cc/, 2003.

[9] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification, 2nd ed.*, John Wiley & Sons, Inc., Oct. 2000.

[10] L.A. Farwell, and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials", *Electroencephalography and Clinical Neurophysiology*, Vol. 70, pp. 510-523, 1988.

[11] G. N. Garcia Molina, T. Ebrahimi, U. Hoffmann, and J.-M. Vesin, "Direct brain-computer communication through EEG signals", Swiss Federal Institute of Technology (EPFL), To appear in the *IEEE EMBS Book Series on Neural Engineering*, 2003.

[12] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-Time EEG Analysis with Subject-Specific Spatial Patterns for a Brain–Computer Interface", *IEEE Trans. on Rehab. Eng.,* Vol. 8, No. 4, pp. 447-456, Dec. 2000.

[13] E. Haselsteiner, and G. Pfurtscheller, "Using Time-Dependent Neural Networks for EEG Classification", *IEEE Trans. on Rehab. Eng.*, Vol. 8, No. 4, pp. 457-463, Dec. 2000.

[14] L. Ip, "The Blahut-Arimoto Algorithm for the Calculation of the Capacity of a Discrete Memoryless Channel", Technical Report, Berkeley Uni., Dec. 10 1999.

[15] J. Kronegg, T. Alecu, and T. Pun, "Information theoretic bit-rate optimization for average trial protocol Brain-Computer Interfaces", *Proc. HCI International 2003*, Crete, Greece, pp. 1437-1440, June 22-27 2003.

[16] J. Lehtonen, *EEG-based Brain Computer Interfaces (M.Sc diploma)*, Helsinki University of Technology, Dep. of Electrical and Comm. Eng., May 3 2002.

[17] S.G. Mason, and G.E. Birch, "A brain-controlled switch for asynchronous control applications", *IEEE Trans. Biomed. Eng.*, Vol.47, No.10, pp.1297-1307, Oct. 2000.

[18] D. J. McFarland, W. A. Samacki, and J. R. Wolpaw, "Brain-computer interface (BCI) operation: Optimizing information transfer rates," *Biological Psychology*, Vol. 63, No. 3, pp. 237-251, July 2003.

[19] P. Meinicke, M. Kaper, F. Hoppe, M. Heumann, and H. Ritter, "Improving Transfer Rates in Brain Computer Interfacing: a Case Study", *Advances in Neural Inf. Proc. Systems (NIPS02)*, Vol. 15, S. Becker, S. Thrun and K. Obermayer eds., MIT Press, Cambridge (MA), pp. 1107-1114, http://books.nips.cc/, 2003.

[20] T. Nykopp, *Statistical Modelling Issues for The Adaptive Brain Interface (M.Sc diploma)*, Helsinki University of Technology, Dep. of Electrical and Comm. Eng., June 9 2001.

[21] 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, No. 3, pp. 283-288, Sept. 2001.

[22] J. Perelmouter, and N. Birbaumer, "A binary spelling interface with random errors", *IEEE Trans. on Rehab. Eng.*, Vol. 8, No 2, pp. 227-232, June 2000.

[23] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement", *IEEE Trans. on Rehab. Eng.*, Vol. 8, No. 4, pp. 441-446, Dec. 2000.

[24] F. Renkens, and J. del R. Millán, "Brain-Actuated Control of a Mobile Platform", *Int. Conf. on Simulation of Adaptive Behavior, Workshop on Motor Control in Humans and Robots*, Edinburgh, Aug. 4-11 2002.

[25] M. Sahin, "Noise tolerance as a measure of channel discrimination for multi-channel neural interfaces", *23th Annual Int. Conf. of the IEEE Eng. in Med. and Biol. Soc.*, Istanbul, Turkey, Oct 25-28 2001.

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

[27] T. M. Vaughan, W. J. Heetderks,L. J. Trejo, W. Z. Rymer, M. Wdnricli, M. M. Moore, A. Kubler. B. H. Oohkin, N. Birbaumer. E. Donchin, E. W. Wolpaw, and J. R. Wolpaw, "Brain-Computer Interface Technology: A Review of the Second International Meeting", *IEEE Trans. on Neural Sys. and Rehab. Eng.*, Vol. 11, No. 2, pp. 94-109, June 2003.

[28] J. R. Wolpaw, H. Ramoser, D. J. McFarland and G. Pfurtscheller, "EEG-Based Communication: Improved Accuracy by Response Verification", *IEEE Trans. on Rehab. Eng.*, Vol. 6, No. 3, pp. 326-333, Sept. 1998.

[29] 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 Trans. on Rehab. Eng.*, Vol. 8, No 2, pp. 164-173, June 2000.

[30] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control", *Clinical Neurophysiology*, Vol. 113, pp. 767–791, 2002.