Two-dimensional auditory P300 Speller with predictive text system

Johannes Höhne, Martijn Schreuder, Benjamin Blankertz and Michael Tangermann

Abstract— **P300-based Brain Computer Interfaces offer communication pathways which are independent of muscle activity. Mostly visual stimuli, e.g. blinking of different letters are used as a paradigm of interaction. Neural degenerative diseases like amyotrophic lateral sclerosis (ALS) also cause a decrease in sight, but the ability of hearing is usually unaffected. Therefore, the use of the auditory modality might be preferable. This work presents a multiclass BCI paradigm using two-dimensional auditory stimuli: cues are varying in pitch (high/medium/low) and location (left/middle/right). The resulting nine different classes are embedded in a predictive text system, enabling to spell a letter with a 9-class decision. Moreover, an unbalanced subtrial presentation is investigated and compared to the well-established sequence-wise paradigm. Twelve healthy subjects participated in an online study to investigate these approaches.**

I. INTRODUCTION

Using Brain-computer interfaces (BCI) one can send control signals without the use of any muscle. Recording electroencephalography (EEG), brain signals are acquired, analyzed and classified, thereby a direct connection between brain and computer is set up. Recently, most research in this field is aimed towards developing tools for patients with completely locked-in syndrome, who have lost their volitional control over all muscles. BCI might be the only technology which could establish a communication pathway for these patients.

There is a variety of different approaches to set up a BCI speller. These mostly differ in the measuring technology, feature extraction, data analysis and modality of interaction. In common type of BCI experiment, subjects are asked to attend to a specific cue while masking others. Using this oddball paradigm one reliably observes a positive deflection in voltage with a latency of about 300 ms after target stimulus onset, called P300. This innate ERP component mainly appears over central and parietal brain areas. The P300 speller [1] quantifies P300 responses to choose letters. In a visual paradigm, letters are ordered in a grid where rows and columns are randomly flashing up. This paradigm has been successfully studied for more than two decades [2], [3] and was successfully tested as a communication device for individuals with advanced ALS [4].

The visual modality requires the subject to be able to control the eyes. Since eye-movements, blinks and the adjustment of focus is -at least partly- relying on volitional muscle activity, some patients, including those suffering from late stage amyotrophic lateral sclerosis (ALS) are not eligible for a BCI applications with visual stimuli. The auditory modality could circumvent this problem.

Most auditory BCI applications are based on one-dimensional stimuli [5], [6], [7] with up to 6 alternative choices per trial. There are recent approaches for multiclass BCI paradigms [8] that include a second spatial dimension to increase discriminability of auditory cues. In this study, cues varied in two dimensions (pitch and direction) but both dimensions transmitted the same information, i.e. a cue with a specific pitch was always presented from the same direction.

The present study presents a multiclass auditory P300 speller with auditory stimuli differing in two independent dimensions: nine auditory cues are varying in pitch (high, medium, low) and location (left, middle, right). A spelling system, which is very similar to the T9 system on mobile phones was implemented. Using this system, subjects are able to spell a character with a 1 out of 9 multiclass decision. Subjects were asked to spell two sentences in an online experiment. We demonstrate that this auditory P300 speller is more accurate and faster than most of those previously reported. Moreover a novel method for subtrial selection is investigated, where the number of presentations of each cue is dependent on previous classifier outputs results within the same trial.

II. METHODS

A. Participants

12 healthy volunteers (9 male, mean age: 25.1, range: 21 - 34) participated in the BCI experiment which lasted for three to four hours. Subjects were not paid for participation, two subjects (VPmg and VPja) had previous experience with BCI. All subjects reported to not suffer from neurological diseases and to have normal hearing. Two subjects (VPnx and VPmg) were excluded from online experiments due to a poor classification performance in the calibration data.

B. Experimental design

Subjects sat in a comfortable chair facing a static screen that showed the visual representation of a 3x3 pad with 9 numbers ordered row wise. Each block in the experiment consisted of an auditory oddball task. Subjects were asked to minimize eye movements and other muscle contractions during the experiment. Nine auditory stimuli, lasting 100 ms were presented using a low-latency USB soundcard and light

neckband headphones.

While preparing the EEG cap, subjects got used to the sound and speed of the cues, listening to those auditory stimuli which were used in the spelling paradigm lateron.

We performed three calibration runs, each run consisted of nine trials with each cue being target once. One practice-run without recording was performed ahead. In one trial we first presented the target cue three times while the corresponding number on a 3x3 grid was highlighted. After a short pause of 2 seconds we presented 13 or 14 random sequences of all nine auditory cues. The last 12 sequences of each trial were used to train the classifier. ISI was 225ms and we assured the pitch to change with each subtrial and that there were at least 3 different cues in between the same stimulus.

After training the binary classifier we performed two online spelling runs. Subjects were asked to spell a short sentence ('Klaus geht zur Uni') and a long sentence ('Franz jagt im Taxi quer durch Berlin') in separate runs. The short sentence was spelled using the standard sequence-wise subtrial selection. For the long sentence an unbalanced subtrial selection was used. Each trial consisted of 135 subtrials. Order of the sentences was randomized. In the spelling runs subjects were asked to attend to the tone which represents the key (1-9) of the character they want to spell.

C. Auditory stimuli

The selection of stimuli is a crucial element for any kind of P300 BCI system. Since auditory perception is strongly varying within subjects, the selection of cues for an auditory BCI application is even more important. Therefore three tones varying in pitch (high/medium/low) and sound quality were carefully chosen in a way that they are -on a subjective scale- as different as possible to each other. Each of these tones was presented in three different locations: only on the left channel, only on the right channel and on both channels. This two-dimensional 3x3 design obeys a close analogy to the number pad of a standard mobile phone, where e.g. number 4 is represented by the middle tone (4-6) presented on the left channel only (1, 4, 7).

D. Data acquisition

EEG was recorded monopolarly using 64 Ag/AgCl electrodes. Channels were referenced to the nose. Electrooculogram (EOG) was recorded under the right eye. Signals were amplified using a Brain Products 64-channel amplifier, sampled at 1 kHz and filtered by an analog bandpass filter between 0.1 and 250 Hz. Further analyzes were done in Matlab. The online feedback was implemented as Pythonic Feedback Framework [9].

After filtering, data was down sampled to 100 Hz and epoched between -150 ms and 800 ms relative to stimulus onset, using the first 150 ms as baseline.

E. Unbalanced subtrial presentation

Aiming to possibly reduce the number of subtrials per trial we investigated a method of unbalanced subtrial selection: the number of presentations of each cue was dependent on previous classifier outputs results within the same trial. Thus cues initiating less significant classifier outputs were presented less frequently. After 5 complete sequences the unbalanced procedure started and the selection of the next cue was a random experiment with

$$p_i = c_{norm}(\exp{-\frac{\mu_i - \bar{\mu}}{\text{std}(\mu)}} + d), \quad (1)$$

where p_i is the probability of choosing i as the next cue. μ_i is the mean classifier output initiated by cue i and c_{norm} is a factor for normalization. d is a constant shift to weight the influence of the online classification results that was constantly set to 0.1 for this study. Additionally, p_i was set to 0 if the preceding subtrial had the same pitch or cue i was presented within the last three presentations.

F. Classification

Binary classification was done using the Fisher Discriminant (FD) algorithm. Due to the dimensionality of the features (up to 252 dimensions), we applied a shrinkage method [10]. In the online experiment, the 1 out of 9 multiclass decision was based on a fixed number (135) of subtrials and their classifier outputs. A one sided t-test with unequal variances [11] was applied for each key and the most significant key (i.e. lowest p-value) was chosen to be the target.

G. Predictive text system

For this BCI speller, the commonly used T9 predictive text system from mobile phones was applied in a modified version. The system was set up with a german dictionary of the 10,000 mostly used words in the german language. Since the standard T9 system uses more than 9 keys, it was modified in a way that exactly 9 keys are needed for spelling. Two different modes were implemented: A *spelling mode* in which key '2' to '9' represent the alphabet and activation of key '1' leads to the *selection mode*. In the selection mode, the user can choose between fitting words, go back to the spelling mode or delete previously entered keys. That way, errors in the multiclass selection can be undone with two additional selections. The system is constrained to words in the dictionary, which can be arbitrarily extended.

III. RESULTS

A. Binary accuracy

Accuracy of the binary problem was computed on the calibration data for each subject based on 327 target and 2592 nontarget subtrials minus the ones excluded with the artifact correction which was done with a simple variance threshold method. After excluding subjects VPnx and VPmg, cross validation analyses reveal that on average 69.5% of targets and 81.4% of nontargets were correctly identified.

B. Multiclass accuracy

Within all trials of the online experiments, 89.37% of the multiclass decisions were correct. Decision were made after 135 subttrials in each trial. The multiclass decision in

TABLE I

S

ubject-specific data and spelling performance. 'correct hits' and 'correct miss' refer to the accuracy of the binary classification problem which is estimated with cross validation. 'cl. error' is an estimate of multiclass classification error which is calculated on calibration data as well. For the online spelling results, the short sentence is marked with [a], whereas [b] specifies the long sentence. The number of decisions is varying since a false decision may require 1 to 3 unmistaken decision to be corrected. The spelling run marked with [x] was not completed because subject *VPoc* had a drop of accuracy after the 45th trial and failed to enter the last two remaining keys.

subject	VPnv	VPnw	VPnx	VPny	VPnz	VPmg	VPoa	VPob	VPoc	VPod	VPja	VPoe	μ
correct hits	73.4	64.8	42.4	67.5	68.4	47.8	65.5	72.7	75	61.6	77.7	68.7	65.5
correct miss	80.6	80.8	63.2	82.4	81.3	73.3	78.8	84.7	83.3	78.6	85.9	77.1	79.2
cl. error	0.158	0.199	0.461	0.199	0.18	0.354	0.195	0.131	0.13	0.232	0.095	0.19	0.2
# decisions[a]	29	31		23	29		38	26		28	23		28.4
time (min)[a]	25.1	23.0		15.4	21.7		26.9	18.1		19.1	17.9		20.9
# decisions[b]		63		53	97		51	49	45[x]	61	49	48	57.3
time (min)[b]		47.1		36.5	76.7		36.8	39.5	30.9[x]	48.9	36.2	38.6	43.5

the short sentence with a balanced subtrial presentation was slightly more accurate (92.51%) than the decision in the long sentence (87.98%). We observed that 3 subjects (VPnt, VPoc and VPoe) had a sudden drop in decision accuracy. Without any obvious reason, their classification performance dropped to zero. For VPnt and VPoe, this observation occurred in between the two sentences, accuracy of VPoc dropped at the end of the first (long) sentence when there were just 2 correct trials required to finish the sentence. These subjects reported that they could not concentrate anymore. Even longer pauses did not have any beneficial effect. Experiments were then stopped. Since VPoc almost finished the sentence the run was considered to be completed.

C. Location and latency of N200 and P300

An early negative and a late positive component could be found for each subject except for the two excluded ones. Although individual differences in location and latency of these components were observed, grand average analyses reveal strong evidence for structurally common signals. For a target cue, we find an early negative deflection 200-300ms after stimulus onset which is centered in the frontal-temporal area. Moreover we find a positive deflection 350-600 ms after stimulus onset which is centered in the central-parietal area. Fig. 2 depicts the ERPs at electrode 'Fz' and scalpmaps for four time frames.

Due to the cross-lateral processing of auditory stimuli we expected the N200 to vary for each stimulus. Fig. 1 depicts grand averaged ROC scalpmaps of N200 for each of the nine cues, illustrating that the early negative deflection is located in cross-lateral areas. As expected, the P300 component did not vary for the nine cues (scalpmaps not shown) since it is not related to primary but cognitive processing. Fig. 3 pictures the importance of the spacial and temporal dimension seperately for classification.

D. Bit rate, characters per minute and early-stopping

In the online spelling runs early stopping methods were not applied, thus each trial had 135 subtrials. It took 15 min to 26 min (μ=20.9) to spell the short sentence and 31 min to 76 min
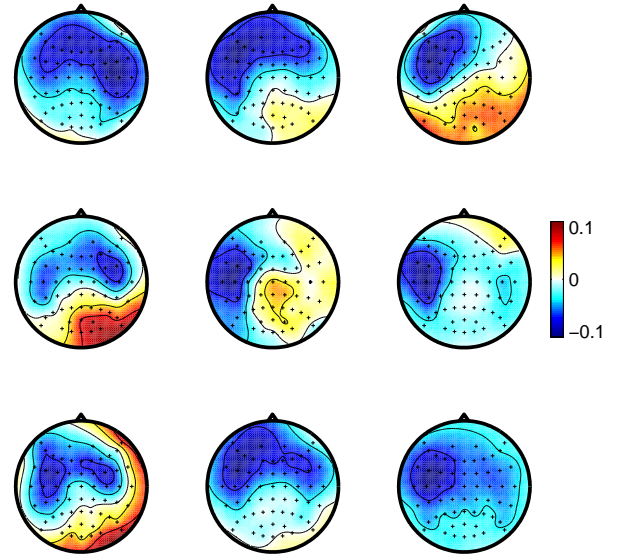

Fig. 1. ROC scalpmaps of the N200 component (interval: 200-300ms after stimulus onset) for each single cue. Design corresponds to the two-dimensional paradigm: The left plot in the second row maps the ROC generated by target cue 4 against all other nontargets.

(μ=43.5) for the long sentence. We find an average spelling speed of 0.845 characters/minute. Since the sentences were not spelled word by word but in one go, all kinds of pauses are taken into account: individual relaxation as well as fixed intertrial periods mainly influence the spelling speed. Furthermore the space character is considered as a valid character, resulting in 16 characters for the short sentence and 36 characters for the long sentence.

The rate of communication can also be assessed with the Information Transfer Rate (ITR) [12]. On average, a subject achieved a bit rate of 3.18 bits/minute in this unconstrained condition.

An offline early stopping method was simulated: A decision was made as soon as the minimum p-value fell below a given threshold. We find that introducing an early stopping method, we can intensely reduce the number of subtrial per decision:
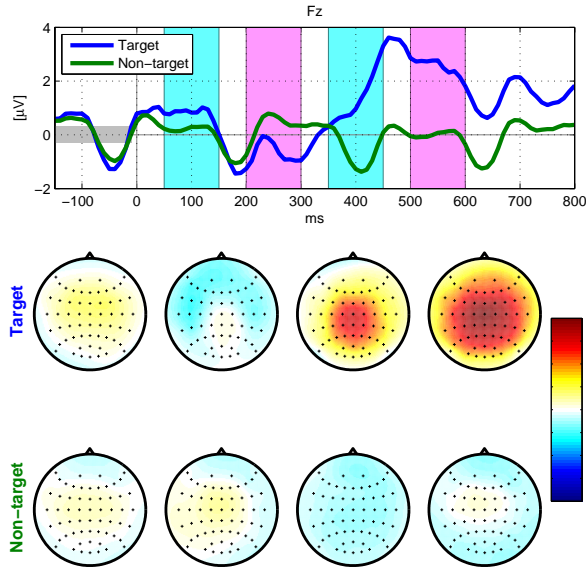
Fig. 2.   ERP at electrode 'Fz' and ERP scalpmaps for the 4 marked time frames
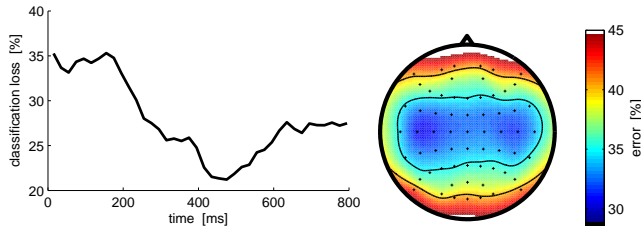

Fig. 3.   Spatial and temporal distribution of discriminative information. Loss obtained for a single temporal 40 ms averaged window (left plot). The loss obtained for each electrode separately is depicted as scalp topography (right plot).

the average bitrate got improved to 5.95 bits/minute (Fig. 4).

E. Balanced and unbalanced subtrial presentation

Next to presenting cues in a completely balanced random sequence we investigated an unbalanced trial presentation, using online results to present stimuli with conspicuous classifier outputs more often. We find that the introduced method can accelerate a multiclass decision while slightly loosing accuracy. We computed the ITR [12] for both methods, finding that using the unbalanced method we can slightly increase the bit rate (Fig. 4).

IV. CONCLUSIONS

This study presents a novel paradigm for an auditory BCI speller with two-dimensional stimuli and a predictive text system. Subjects spelled two sentences with 16 resp. 36 characters in an online experiment. We find that 10 of 12 subjects are able to successfully use the system with a high accuracy.

We also approach a new method for subtrial selection, taking online binary classification results into account. This method aims to accelerate a decision by presenting specific cues more frequent if they initiated more significant
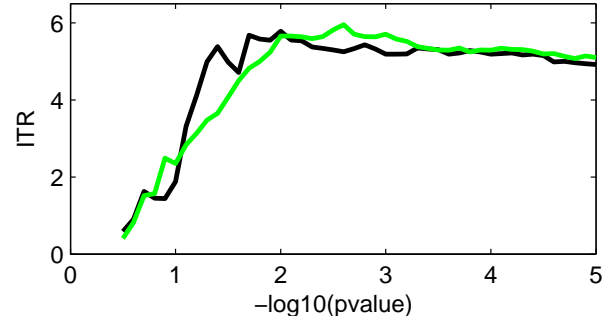

Fig. 4.   Grand-averaged Information Transfer Rate (ITR) in bits/minute as a function of stopping threshold. The standard sequence-wise paradigm (maximum value: 5.78 for $p_{crit} = 0.01$) is represented in black, the unbalanced paradigm (maximum value: 5.95 for $p_{crit} = 0.0025$) for subtrial selection is marked green.

classifier outputs than others before. We find a slightly increased bitrate for the unbalanced method compared to the standard sequence-wise for subtrial selection.

REFERENCES

[1] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials." *Electroencephalogr Clin Neurophysiol*, vol. 70, no. 6, pp. 510–523, Dec 1988.

[2] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced p300 speller performance." *J Neurosci Methods*, vol. 167, no. 1, pp. 15–21, Jan 2008.

[3] M. Salvaris and F. Sepulveda, "Visual modifications on the P300 speller BCI paradigm." *J Neural Eng*, vol. 6, no. 4, p. 046011, 2009.

[4] F. Nijboer, E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N. Birbaumer, and A. Kbler, "A P300-based brain-computer interface for people with amyotrophic lateral sclerosis." *Clin Neurophysiol*, vol. 119, no. 8, pp. 1909–1916, 2008.

[5] D. S. Klobassa, T. M. Vaughan, P. Brunner, N. E. Schwartz, J. R. Wolpaw, C. Neuper, and E. W. Sellers, "Toward a high-throughput auditory p300-based brain-computer interface." *Clin Neurophysiol*, vol. 120, no. 7, pp. 1252–1261, Jul 2009.

[6] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, "A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients." *Ann N Y Acad Sci*, vol. 1157, pp. 90–100, 2009.

[7] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, "An auditory oddball (p300) spelling system for brain-computer interfaces." *Psychophysiology*, vol. 46, no. 3, pp. 617–625, May 2009.

[8] M. Schreuder, B. Blankertz, and M. Tangermann, "A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue." *PLoS One*, vol. 5, no. 4, 2010.

[9] B. Venthur and B. Blankertz, "A platform-independent open-source feedback framework for BCI systems," in *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course 2008*. Verlag der Technischen Universität Graz, 2008, pp. 385–389.

[10] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365 – 411, 2004.

[11] G. D. Ruxton, "The unequal variance t-test is an underused alternative to student's t-test and the mann–whitney u test," *Behavioral Ecology*, vol. 17, no. 4, pp. 688–690, July 2006.

[12] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control." *Clin Neurophysiol*, vol. 113, no. 6, pp. 767–791, Jun 2002.