34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

# Exploring the Use of Tactile Feedback in an ERP-Based Auditory BCI

Martijn Schreuder[1], Marieke E. Thurlings[1,2,3], Anne-Marie Brouwer[2],
Jan B.F. Van Erp[2] and Michael Tangermann[1]

*Abstract*— **Giving direct, continuous feedback on a brain state is common practice in motor imagery based brain-computer interfaces (BCI), but has not been reported for BCIs based on event-related potentials (ERP), where feedback is only given once after a sequence of stimuli. Potentially, direct feedback could allow the user to adjust his strategy during a running trial to obtain the required response.**

**In order to test the usefulness of such feedback, directionally congruent vibrotactile feedback was given during an online auditory BCI experiment. Users received either no feedback, short feedback pulses or continuous feedback. The feedback conditions showed reduced performance both on a behavioral task and in terms of classification accuracy. Several explanations are discussed that give interesting starting points for further research on this topic.**

## I. INTRODUCTION

Brain-computer interface (BCI) systems allow a person to control a device without the use of the brain's normal efferent pathways. Several types of BCI paradigms have extensively been described in literature, the most prominent ones being those based on motor imagery (MI) induced event-related desynchronization (ERD) [8] and those based on event-related potentials (ERP) [6]. They both allow the user to convey his intention to a machine by voluntarily inducing a specific brain state.

We previously introduced an auditory ERP based BCI paradigm where the spatial direction of auditory cues is a discriminative property [11], [9], [10]. The principle behind this so-called AMUSE paradigm is that a sequence of different tones is played, each tone always from the same unique location around the subject. Though all tones elicit an ERP in the brain, the response that results from the direction that the subject is paying attention to contains enhanced components; generally these are the N2 and the P3. Such enhancements can be detected from the ongoing EEG signal by machine learning methods and are used to decode the user's intention. They can, however, also be used to give the user direct feedback on his task performance, even during a running trial. This way, the user can potentially improve his strategy for inducing the intended brain state, and it thus truly closes the loop in ERP based BCI.

Though it is common practice to provide the user of an MI based BCI with such direct feedback [2], this is not the case for ERP based BCIs. One study introduced a stimulus sequence generation paradigm where the more likely stimuli got presented more often [7]. This can be considered as an implicit form of feedback. To our knowledge, explicit and direct feedback has thus far not been reported for ERP based BCIs. Though ERP based BCIs generally already have higher performances than their MI siblings, the interesting question remains if feedback can benefit the user of an ERP BCI and further improve performance. In other words, can the user integrate such feedback and optimize his strategy accordingly? A first step towards answering that question is presented here.

The AMUSE paradigm was adapted to include direct feedback during ongoing trials. In order not to interfere with the stimulation, this feedback was given through a vibrotactile vest developed at TNO [13], which has already been used successfully for driving a BCI [4], [12]. Experiments took place at TNO Soesterberg, the Netherlands.

Results indicate that in this setting, concurrent stimulation and feedback does not improve, but rather degrades performance. This was found for classification performance as well as for behavioral scores (counting the number of target tones). Some limitations of the current approach are discussed, and recommendations for further research given.

## II. METHODS

### A. Subjects

Recruited subjects were ten healthy volunteers (six female, mean age 29.5, range 19-52) with no reported current or prior neurological disorder and normal hearing. The latter was not formally tested. Subjects were naïve to auditory BCI and were compensated for their time. Approval of the experiment was acquired from the local ethical committee. All subjects provided verbal and written informed consent and subsequent analysis and presentation of data was anonymized.

### B. Setup

Subjects were seated in a comfortable chair facing a screen at approximately 1m distance, and were surrounded by six speakers. The speakers were placed at ear height in a circle, equidistant from the subject's head (at ∼65 cm) and

*This work is supported by the European ICT Programme Project FP7-224631, the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science, and the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie). This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

[1]M. Schreuder, M.E. Thurlings and M. Tangermann are with the BBCI group of the Machine Learning Department, Berlin Institute of Technology, Berlin, Germany.

[2]M.E. Thurlings, A.-M. Brouwer and J.B.F. Van Erp are with the Department of Perceptual and Cognitive Systems TNO, Soesterberg, the Netherlands.

[3]M.E. Thurlings is furthermore with the Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands.

Correspondence: `schreuder@tu-berlin.de`

with 60° circular displacement [9], [10]. Furthermore, the subjects were fitted with a belt with six tactors [13], one for each speaker direction. Tactor and speaker had the same angular displacement, thus providing congruent auditory-tactile mapping. Each tactor could be actuated individually.

As the tactors produced sound, minimal pink noise was played through two separate speakers, placed on the floor left and right of the subject. For consistency it was played during all conditions, even those where no tactile feedback was given.

Using a BrainAmp amplifier (Brain Products, Germany), 14 EEG channels were recorded and referenced against linked mastoid electrodes. The signals were filtered by a hardware analog bandpass filter between 0.1 and 250 Hz before being sampled at 1 kHz and stored for offline analyses. For online use, the signal was low-pass filtered below 40 Hz, down sampled to 100 Hz and streamed directly to the online Berlin BCI system. The auditory and tactile stimulation, online BCI toolbox and offline analyses were all implemented in Matlab (Mathworks), making use of the Psychophysics Toolbox [3].

### C. *Stimuli*

Auditory stimuli consisted of a complex of band-pass filtered white noise and a sinusoidal tone overlay. Both the direction and the tone/noise complex conveyed the same information; each tone always came from the same location. Stimuli where constructed to provide features for frequency ranges that are optimal for both inter-aural timing differences ($<$3 kHz) and inter-aural level differences ($>$3 kHz). The stimuli were played at 60 dB. As the noise that was played to mask the tactor sound partially masked the stimulus noise as well, the task was more difficult than previously described [9], [10]. This reduced the ceiling effect on performance and allowed for both negative and positive changes.

### D. *Terminology*

As any ERP based BCI, AMUSE relies on block-structured, repetitive stimulation. A single stimulus is called an epoch, and each epoch is classified individually. With a six class BCI, six unique consecutive epochs are called an iteration. Presentation order within an iteration is pseudo-random, with some neighboring constraints to prevent repetitions in the next iteration. A single BCI class-decision, called trial, consists of several of such iterations. Here, a run refers to a set of consecutive trials belonging to the same condition.

### E. *Experimental protocol*

The experiment existed of two phases: calibration and online. The following three conditions were tested: *NoFB* (no tactile feedback was given), *200 ms* (a short, 200 ms tactile feedback burst was given after each iteration) and *Cont* (continuous feedback was given, which was updated after each iteration). Tactile feedback was only started after the fourth iteration of each trial, to avoid the uncertain decisions due to lack of evidence. During the calibration phase, the tactile feedback was operated pseudo-randomly with a bias

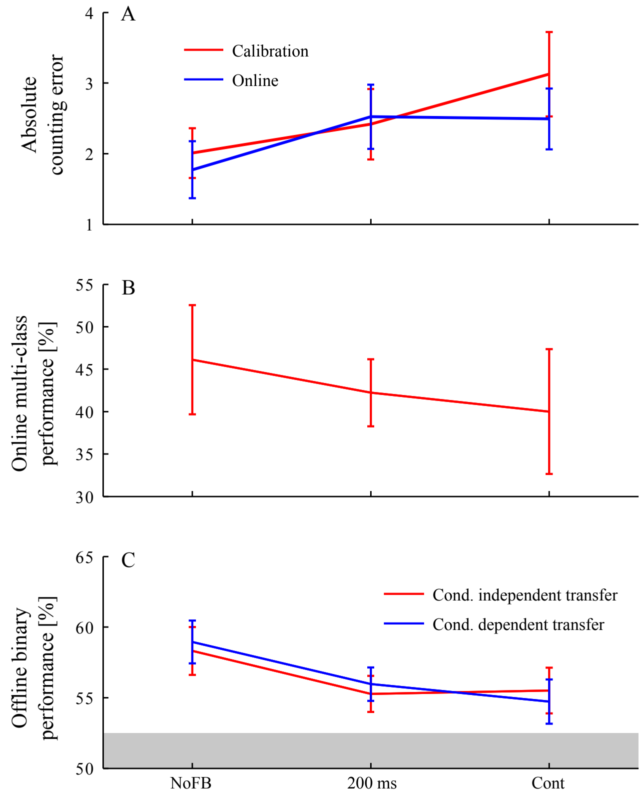


Fig. 1. Behavioral and classification performance. A) Subjects were asked to report the counted number of targets after each trial. The average of the absolute deviation from the correct number is reported for the calibration (red) and online (blue) phase. B) Online BCI performance in the multi-class setting ("one out of six" decision). C) Offline reclassification using condition independent (red) and dependent (blue) classifiers. The shaded area in represents the chance interval. For all plots, the errorbars represent standard error of mean.

towards correct feedback; subjects were informed that they had no control over it.

Four calibration runs were performed, each consisting of 12 trials. The first run was always without feedback; the second to fourth run contained each of the above conditions once and the order was randomized between subjects. The subject's task was to focus his attention to stimuli from a given target direction and count the number of target occurrences. After each trial, a screen prompted the subject to enter the number of counted targets. During calibration, the number of iterations per trial was varied between 15 and 18; only the last 15 of which were used for analyses. During the online phase this was fixed to 15 iterations.

A shrinkage regularized linear classifier [1] was trained on all four data sets combined, with the goal to create a condition independent classifier. For feature extraction, the data was cut into epochs (interval: -150 ms to 800 ms relative to stimulus onset). The single epochs were baselined (-150 prestimulus to stimulus onset) and data in three hand-picked intervals was averaged. The trained classifier was used during all consecutive online trials to drive the tactile feedback and the BCI.

Three blocks of three online runs (one for each condition) were performed, each run consisting of 12 trials. Condition order was randomized in the first block, and this order was kept in the second and third block. The task and interface were exactly the same as in the calibration phase, with the exception that subjects now had control over the feedback and at the end of the trial the subjects were informed of the BCI decision in two ways: first, one second after the last stimulus, the tone from the winning direction was played and, second, a green tick mark or a red cross was shown in the center of the screen for two seconds, indicating a right or wrong decision, respectively.

After the session, subjects were asked to indicate their favorite condition. They were asked to do so, disregarding the performance obtained in each condition. The latter can however not be guaranteed, and results are thus used as qualitative information only.

### F. Offline analyses

The online multi class performance is the BCI relevant metric. However, as it is based on accumulated evidence, it could lack some of the granularity necessary for finding condition related differences. Therefore, the results of several offline analyses are also reported here. For all offline analyses the features were slightly different: the data were down-sampled using a-priori knowledge to set a higher sampling density over early components (30-350 ms: average over 30 ms bins), than late components (360-800 ms: average over 60 ms bins). Also, baselining was replaced by a high-pass filter with pass frequency .5 Hz (.1 cut off).

Two analyses were performed. First, an individual classifier was trained for each condition, using one of the training runs (even when two were available). This classifier was then applied to the respective online session and the binary classification performance is reported. For comparison, a fourth classifier was trained on a mix of data from the three conditions, with the total number of calibration trials equal to that of the other classifiers. It simulates the approach of a condition independent classifier, as taken during the online phase, but with the same number of training trials. This comparison is necessary to investigate the influence of the condition independent classifier, which may effect each condition differently.

Second, the differences in ERP component contribution to classification performance are investigated. Multiple cross-validations were performed in sliding window intervals (width 50 ms, overlap 40 ms) on the online data for all three conditions individually. Samples in an interval were averaged, and the resulting feature vector contained a single temporal feature per channel. Each 50 ms timewindow thus gave a crossvalidation score, which roughly indicates the contribution of that window to the overall performance.

All reported binary performances were class-wised normalized, to account for the highly unbalanced data (1 target for every 5 non-targets). Condition comparisons were done using paired t-tests with Bonferroni correction for multiple comparisons. Chance levels were estimated by applying 500 cross validations on the data with randomized labels. A

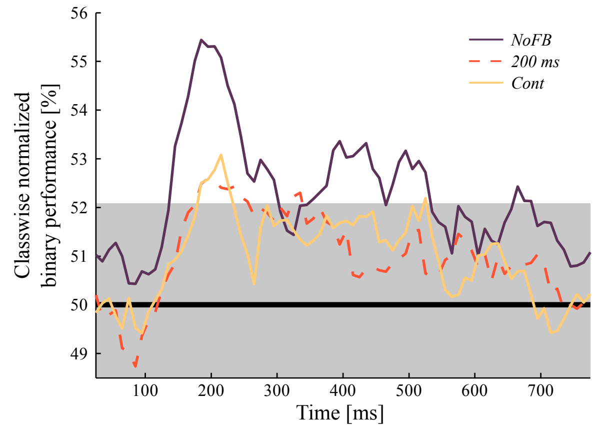


Fig. 2. Grand average sliding window classification performance. Contributions to the overall classification stem mainly from N2 and P3 in condition *NoFB*. The separability of these components is highly reduced, or negated for the other two conditions. The gray area refers to chance level classification performance.

classification score that is outside of the 5 percentile of these random scores is considered significantly above chance.

## III. RESULTS

### A. Online performance and counting

Mean absolute online counting error was 1.78, 2.5 and 2.5 for conditions *NoFB*, *200 ms* and *Cont*, respectively (see blue line in Fig. 1-a). Significance was confirmed by a paired t-test (*NoFB* vs *200 ms*: $p<0.05$; *NoFB* vs *Cont*: $p<.01$). Feedback thus shows a deteriorating effect on the counting performance in the online phase. Condition dependent counting differences in the training phase were not statistically significant.

With a chance level at 16%, the mean online multi-class classification performance for the three conditions *NoFB*, *200 ms* and *Cont* was 46.1%, 42.2% and 40%, respectively (see Fig. 1-b). This trend was not significant, which is likely due to the partial error correction by the accumulation of evidence.

### B. Offline analyses

Fig. 1-c shows the result of the offline reclassification. The red line indicates the binary transfer performance of the condition independent classifier, thus representing the online protocol. For conditions *NoFB*, *200 ms* and *Cont*, the mean binary transfer performance is 58.3%, 55.3% and 55.5%, respectively. Significance was confirmed by a paired t-test (*NoFB* vs *200 ms*: $p<0.05$; *NoFB* vs *Cont*: $p<.05$).

The blue line represents the performance of three individually trained classifiers, which are thus condition dependent. For conditions *NoFB*, *200 ms* and *Cont*, the mean binary transfer performance is 59.0%, 56.0% and 54.7%, respectively. Significance was confirmed by a paired t-test (*NoFB* vs *200 ms*: $p<0.05$; *NoFB* vs *Cont*: $p<.01$).

Results of the sliding window analysis can be found in Fig. 2. Under normal circumstances (*NoFB*), contributions to the classification performance stem from an early component around 200 ms (N200) and a later component around 400 ms (P3), both of which are individually classifiable above chance level. For both feedback conditions, the separability is reduced to almost around or below chance level, especially for the later P3 component.

### C. Favorite condition

Five of ten subjects reported that the *NoFB* condition was best for them, indicating that the feedback was annoying (physically and mentally), seemed to jump or that it was only helpful when it matched the target. One subject gave preference to the *200 ms* condition, and two preferred the *Cont* condition. Two subjects did not give preference to any condition. Six of eight indicated preferred conditions coincided with the best performing (online) condition.

## IV. CONCLUSIONS

It is generally assumed that direct feedback helps the subject reach a proper brainstate in MI BCIs; an assumption that has recently been challenged [5]. The current study reports on the effect of explicit and direct feedback on the BCI performance in an ERP based BCI. Using auditory stimuli and vibrotactile feedback that were designed to be congruent in direction, a truly closed-loop and multimodal ERP based BCI system was realized.

The online and offline classification performance are considerably lower than reported in previous work [9], [10], which is likely due to the increased task difficulty resulting from the background noise. On top of this, feedback reduced the subjects' ability to perform the BCI task. Behaviorally, the errors made in the counting task increased for conditions with feedback, as opposed to the control condition without feedback. This was significant for the online data only, which may be due to the fact that the tactile feedback was psuedo-random during calibration and biased towards correct feedback. In line with the behavioral results, the majority of subjects preferred the control condition, indicating that the feedback was distracting them from the task at hand.

A similar negative influence of feedback was found for the classification performance. Though the online multi-class performance seemed to be effected, this was not found to be significant. However, when looking at the finer grained binary decisions, the effect becomes clearer. Both feedback conditions had a similarly and significantly deteriorated performance when a condition independent classifier was used, resembling the online protocol. To verify that these effects are not due to the choice for an independent classifier, an individual classifier was used for each condition separately. Using these dependent classifiers led to a similar and significant deteriorating effect of feedback on binary classification performance. Thus, the choice for an independent classifier does not explain these results.

An alternative explanation for the found effects may be that the localization of the auditory stimuli was less precise due to the played background noise. This not only increases the task difficulty in general, but may also compromise the directional congruency between auditory stimulus and tactile feedback. Furthermore, the auditory stimuli and tactile feedback were congruent in direction but not in location. Therefore, correct tactile feedback may still have drawn spatial attention away from the auditory stimuli. This reduced attention on the primary task may result in less pronounced target ERP waveforms, explaining the reduced class discriminative components in the data. Lastly, as erroneous tactile feedback is per definition not directionally congruent with the focused auditory stimulus, it may be particularly effective in drawing away spatial attention. This may have added to the reported distracting effect of the feedback, and resulted in a negative reinforcement effect.

This study for the first time investigates the effect of direct feedback on an ERP-based BCI. Although a positive effect of feedback was not found in this particular setting, several explanations have been discussed as to why this might be the case. What happens when these are properly addressed remains in interesting open research question.

## ACKNOWLEDGMENT

The authors would like to thank Antoon Wennemers for support with the tactile vest.

## REFERENCES

[1] Benjamin Blankertz, Steven Lemm, Matthias Sebastian Treder, Stefan Haufe, and Klaus-Robert Müller. Single-trial analysis and classification of ERP components – a tutorial. *NeuroImage*, 56:814–825, 2011.

[2] Benjamin Blankertz, Michael Tangermann, Florin Popescu, Matthias Krauledat, Siamac Fazli, Márton Danóczy, Gabriel Curio, and Klaus-Robert Müller. The Berlin Brain-Computer Interface. In Jacek M. Zurada, Gary G. Yen, and Jun Wang, editors, *WCCI 2008 Plenary/Invited Lectures*, volume 5050 of *LNCS*, pages 79–101. Springer, Berlin Heidelberg, 2008.

[3] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10(4):433–436, 1997.

[4] Anne-Marie Brouwer and J. B. F. van Erp. A tactile P300 brain-computer interface. *Frontiers in Neuroscience*, 4(19):036003, 2010.

[5] Jan Conradi, Benjamin Blankertz, Michael Tangermann, Volker Kunzmann, and Gabriel Curio. Brain-computer interfacing in tetraplegic patients with high spinal cord injury. *International Journal of Bioelectromagnetism*, 11(2):65–68, 2009.

[6] 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*, 70:510–523, 1988.

[7] Johannes Höhne, Martijn Schreuder, Benjamin Blankertz, and Michael Tangermann. Two-dimensional auditory P300 speller with predictive text system. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2010, pages 4185–4188, 2010.

[8] Gert Pfurtscheller and F. H. Lopes da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 110(11):1842–1857, Nov 1999.

[9] Martijn Schreuder, Benjamin Blankertz, and Michael Tangermann. A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue. *PLoS ONE*, 5(4):e9813, 2010.

[10] Martijn Schreuder, Thomas Rost, and Michael Tangermann. Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. *Frontiers in Neuroscience*, 5:112, 2011.

[11] Martijn Schreuder, Michael Tangermann, and Benjamin Blankertz. Initial results of a high-speed spatial auditory BCI. *International Journal of Bioelectromagnetism*, 11(2):105–109, 2009.

[12] Marieke E. Thurlings, Jan B.F. van Erp, Anne-Marie Brouwer, Benjamin Blankertz, and Peter Werkhoven. Control-display mapping in brain-computer interfaces. *Ergonomics*, 55(5):564–580, 2012.

[13] Jan BF Van Erp. Presenting directions with a vibrotactile torso display. *Ergonomics*, 48(3):302–313, 2005.