Using ERPs for Assessing the (Sub)Conscious Perception of Noise

Anne K. Porbadnigk, Jan-N. Antons, Benjamin Blankertz, Matthias S. Treder, Robert Schleicher, Sebastian Möller, Gabriel Curio

***Abstract*— In this paper, we investigate the use of event-related potentials (ERPs) as a quantitative measure for quality assessment of disturbed audio signals. For this purpose, we ran an EEG study (N=11) using an oddball paradigm, during which subjects were presented with the phoneme /a/, superimposed with varying degrees of signal-correlated noise. Based on this data set, we address the question to which degree the degradation of the auditory stimuli is reflected on a neural level, even if the disturbance is below the threshold of conscious perception. For those stimuli that are consciously recognized as being disturbed, we suggest the use of the amplitude and latency of the P300 component for assessing the level of disturbance. For disturbed stimuli for which the noise is not perceived consciously, we show for two subjects that a classifier based on shrinkage LDA can be applied successfully to single out stimuli, for which the noise was presumably processed subconsciously.**

I. INTRODUCTION

A. Motivation

The quality of auditory signals has a huge impact on the joy of use when operating a phone or listening to recorded music. In the field of telecommunications, the quality of audio signals is typically assessed based on subjective test procedures that rely on behavioral data, such as the Absolute Category Rating (ACR, [1]). However, some differences in quality might be too subtle to be detected consciously and thus cannot be revealed on the behavioral level. Nonetheless, these differences might be reflected in the neural correlates of auditory perception. Subconscious processing of noise might affect the long term contentment of users, potentially leading to a growing dissatisfaction over time. As shown in a recent MEG study [2], neuro-physiological measures have the potential for being used as an objective and accurate measure for auditory quality assessment. These measures have the potential to complement behavioral approaches, while at the same time giving deeper insights into the cerebral mechanisms of noise perception. In this paper, we present initial steps towards such a methodology for investigating the neural correlates of (sub)conscious perception of noise based on EEG recordings, employing methods typically used for Brain Computer Interfaces (BCIs, [3]).

This work was supported by the German Federal Ministry of Education and Research within the Bernstein Focus Neurotechnology.

A.K. Porbadnigk, B. Blankertz and M.S. Treder are with the Machine Learning Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany anne.k.porbadnigk@tu-berlin.de

B. Blankertz is also with the Intelligent Data Analysis Group, Fraunhofer FIRST, 12489 Berlin, Germany

J.-N. Antons, R. Schleicher and S. Möller are with the Quality and Usability Lab, Telekom Laboratories, 10587 Berlin, Germany

G. Curio is with the Department of Neurology and Clinical Neurophysiology, Charite, 12200 Berlin, Germany

B. Related Work

Even though auditory paradigms have only very recently gained more interest for use in BCI settings, first studies on using spatially distributed auditory cues for EEG-based BCIs showed promising results [4]. In a recent EEG study, presumably subconscious processing steps of certain features of musical chords were found that were not reflected on the behavioral level [5]. Furthermore, it was recently shown in an MEG study that the auditory cortex is sensitive to the degradation of different types of acoustic stimuli [2]. In contrast to these approaches, we investigate how the perception of phonemes, superimposed with different levels of noise, is reflected in phasic EEG components.

For the analysis of event-related potentials (ERPs), we focus on the P300 component in this paper. The P300 is elicited as a reaction of the brain to deviating stimuli in an oddball paradigm [6]. It can be based on the match between mental representation in working memory of a target stimulus and an incoming stimulus, for instance. P300 latency is thought to index classification speed, which is proportional to the time required to detect and evaluate a target stimulus [7].

With respect to ERP classification, conventional Linear Discriminant Analysis (LDA, [8]) has proven to be the most successful linear method. However, it has been shown that shrinkage LDA outperforms classical LDA approaches by far when it comes to the classification of single-trial ERPs [9]. This method is based on using shrinkage for the regularization of the empirical covariance matrix that needs to be estimated (see [10] for details on parameter estimation). The receiver operating characteristic (ROC, [11]) is frequently used to characterize the performance of classifiers, oftentimes plotting the true positive rate against the false positive rate in a so-called ROC curve [12]. The area under the ROC curve (AUC) is then commonly calculated in order to condense classification performance to a single value [13].

II. MATERIAL AND METHODS

For the EEG study presented here, N=11 participants were recorded (mean age 25), using a 64 channel EEG system by Brain Products. The auditory stimuli were presented binaurally, using in-ear headphones by Sennheiser. Per subject, 8 to 12 blocks were recorded, resulting in a total of 107 blocks. During each block, 300 auditory stimuli were presented, each of which had a duration of 160 ms (1000 ms stimulus onset asynchrony).

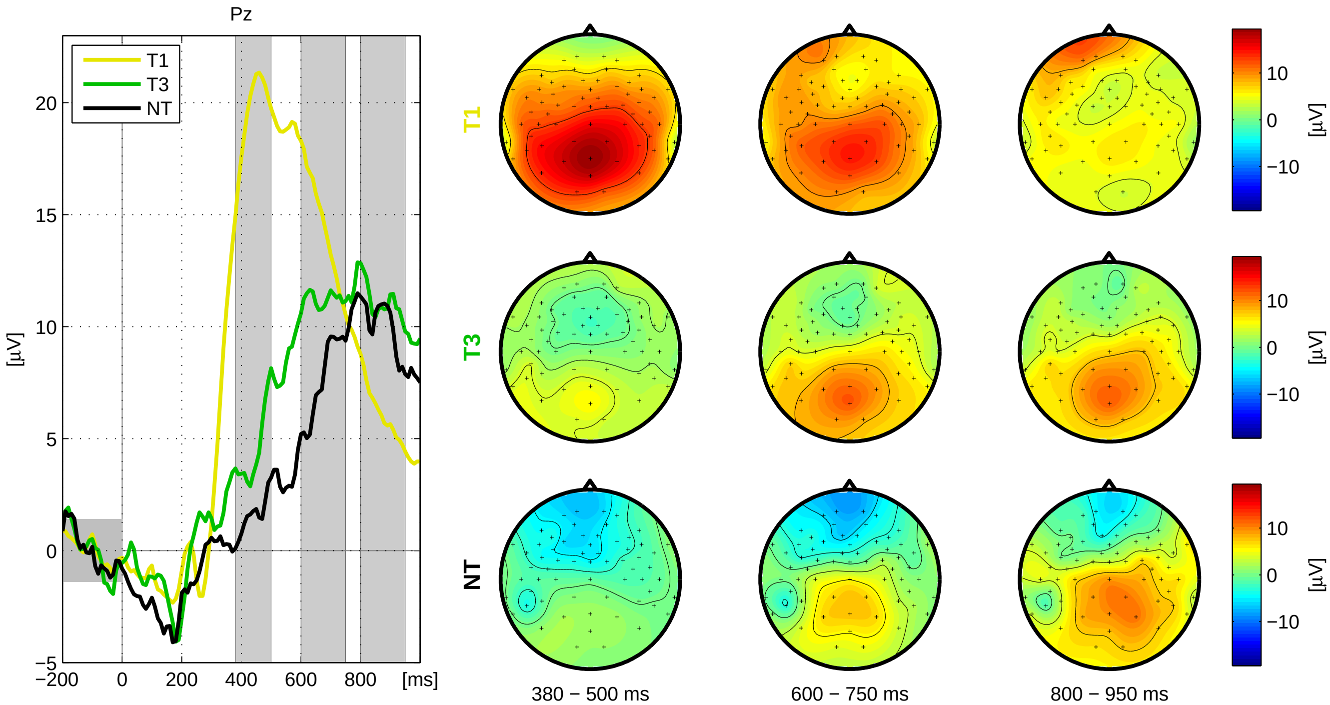

Fig. 1. Grand average ERPs, exemplarily for the two target conditions T1 and T3 as well for non-targets (NT). *Left:* Time course of ERPs at Pz in the time interval -200 to 1000 ms with $t = 0$ being the time point of stimulus onset. *Right:* Scalp distribution for the three conditions in the time intervals 380–500, 600–750, 800–950 ms (marked in gray in the time course). The maps represent a top view on the head with nose pointing upwards. For calculating the ERPs, only those trials were selected that were classified as targets by the subject (true positives for T1 and T3, false positives for non-targets).

A. *Paradigm*

An oddball paradigm was used during which an audio signal (phoneme /a/) was presented to the subjects that was either undisturbed (70% of stimuli; non-targets NT) or superimposed with four varying degrees of signal-correlated noise (6% per class; targets). An additional 6% of the stimuli was made up of the phoneme /i/ as control stimulus (target). The task of the subjects was to press a button whenever they detected one of the deviant stimuli (targets).

B. *Stimulus Selection*

In order to account for individual differences, a pre-test was run before the actual experiment in order to select appropriate stimuli for each subject. This test aimed at finding four individual noise levels Ti ($i \in \{1, 2, 3, 4\}$) that would be recognized with a probability of 100%, 75%, 25% and 0%, respectively. The resulting signal-to-noise ratios (SNR) for the deviant stimuli were set to 5, 21, 24 and 28 dB on average, resulting in average recognition rates of 99%, 46%, 22% and 7% in the experiment. A Modulated Noise Reference Unit (MNRU) was used for creating the modified signals [14]. In the remainder of this paper, we refer to those stimuli that were correctly classified by a subject as 'hits' (true positives, true negatives) and to the others as 'misses' (false positives, false negatives). It should be noted that the correct classification of deviants involved a button press (true positive), whereas it involved the absence of a button press for non-targets (true negative).

C. *Classifier*

For training and testing, the correctly classified non-target trials (true negatives) of a given subject were split into a training set and a test set of the same size. The classifier was then trained to distinguish between non-targets of the test set and the hits of a deviant stimulus class. Before training and testing the classifier, seven intervals (within [100ms, 650ms]) were determined for each subject that were most discriminative for each of the two classes in the training set (based on ROC values). Subsequently, for both the training and the test set, the mean was calculated within these seven intervals for each trial, which was then used as input for the classifier [9]. During each of these steps, data from all 64 recorded electrodes was used.

III. RESULTS

A. *General Pattern of Activation*

The conscious perception of disturbed audio stimuli (true positives) results in a typical activation pattern, specifically an early temporal negativity (130–230 ms post-stimulus), as well as a P300 component (250–500 ms post-stimulus for T1). Generally speaking, the relatively difficult detection task causes the ERP components to occur later than would usually be expected.

Fig. 1 shows the ERPs for three exemplarily selected stimulus classes, averaged over all 11 subjects (left: time course at position Pz, right: scalp maps). Only those trials are

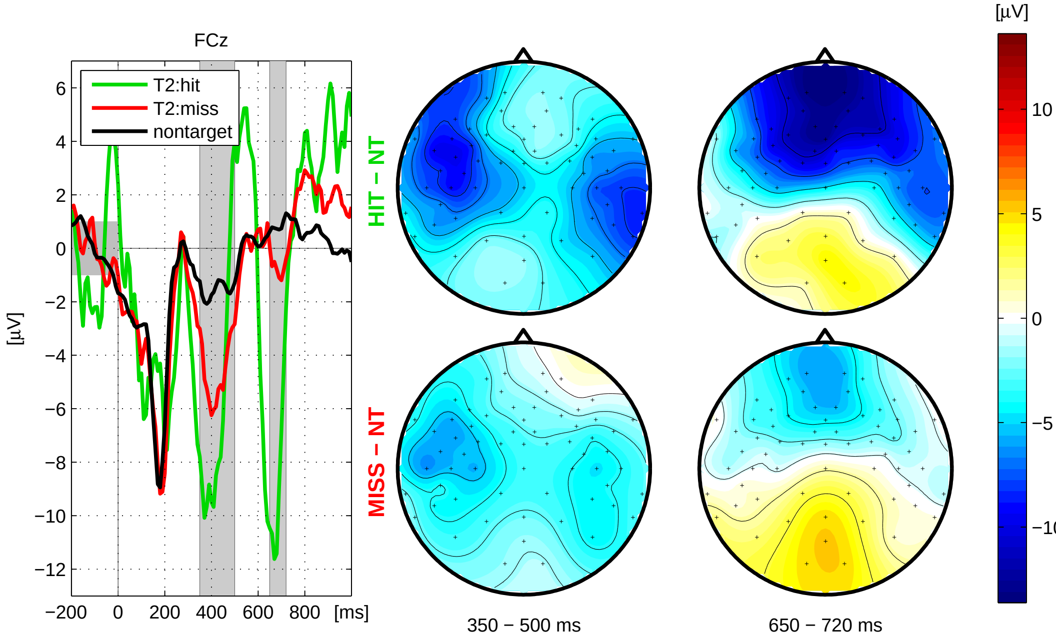

Fig. 2. Time course and difference topographies displaying the similarity of hits and misses for subject VPcad. *Left:* ERPs at electrode FCz for hits and misses of target class T2 and for non-targets. *Right:* Scalp topographies of the difference of hits and misses minus the non-target class in the time intervals 350–500 ms and 650–720 ms.

considered that were classified as targets by the subjects (button press response). This means that for the deviant classes T1 and T3, only true positives are considered, whereas only false positives are taken into account for the non-targets.

As can be seen, the harder it is to detect a target, the lower the amplitude and the higher the latency of the P300 component, possibly reflecting the 'neuronal effort' involved. This causal relationship can be found in all but one subject, despite considerable inter-subject differences in amplitude and latency of the P300 (T1: 11–35 μV, 430–700 ms; T2: 7–23 μV, 400–890 ms). Interestingly, the false positives also elicit a subdued pattern of activation, that bears some similarity with that of the true positives. In contrast to the P300, the latency of the early negativity remains almost the same for all conditions, probably reflecting sensory processing of the audio stimuli.

B. Classifying the Threshold of Noise Perception

In general, the grand averages of the ERPs for the hits and misses of deviants are clearly distinguishable from each other, with missed deviants eliciting a similar activation as correctly classified non-targets (no noise perceived). However, for two out of the 11 subjects (VPcad, VPcae), misses of deviant class T2 result in an ERP pattern that is clearly different from that of non-targets and shows a striking resemblance with hits of the same deviant class. This is shown exemplarily for subject VPcad in Fig. 2. As can be seen in the difference topographies (right panel), the ERP pattern of misses is subdued compared to that of hits, but shows nonetheless a similar morphology. The first two negative components (N1 at 120–220 ms, N2 at 350–500 ms) presumably reflect early sensory perception. While N1 is similar in attenuation for all conditions, component N2 is subdued for misses compared to hits and even more so for non-targets, showing differences in perception already on an early sensory level. These differences in amplitude are more pronounced in component N3 (650–720 ms) that can probably be attributed to cognitive processes.

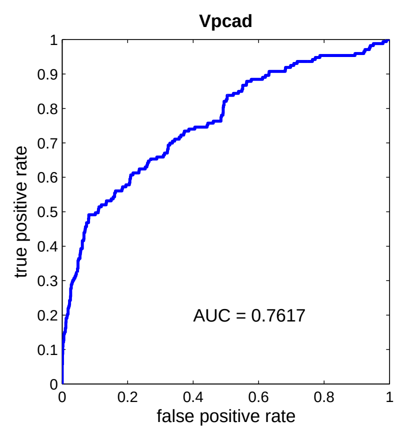

Fig. 3. ROC curve resulting from a classifier based on shrinkage LDA. The classifier was trained to distinguish between T2 hits and one half of the NT hits for subject VPcad and then tested on T2 misses and the other half of the NT hits (same subject).

In order to explore these findings further, we used shrinkage LDA to assess the similarities between the ERP patterns elicited by T2 hits, T2 misses and correctly classified non-targets. For this purpose, a classifier based on shrinkage LDA was trained to distinguish, for a given subject and condition, between the ERP patterns elicited by correctly recognized deviants and by half of the correctly recognized non-targets. Subsequently, the classifier was tested on the misses of the deviant class and the other half of the non-targets.

As can be seen exemplarily in the ROC curve for subject VPcad (Fig. 3), the classifier is well able to distinguish between the two classes. This is the case for both subjects, with an area under the ROC curve (AUC) of 0.7617 and 0.6659 for VPcad and VPcae, respectively. Thus, even though the behavioral data suggests that the two different stimulus types were perceived in the same way, the corresponding neural activation does differ. We conjecture that the noise might be processed subconsciously during these trials, as it may be below the threshold of conscious perception.

For both subject VPcad and VPcae, T2 is the first degradation level for which the subjects show a substantial amount of misses, with a ratio of hits:misses of 30:173 and 75:103, respectively. Even though we cannot yet achieve a comparable classification performance for the other subjects for comparable deviant classes, the classifier can separate false negatives and true negatives up to a certain degree (AUC=0.62 for both subject VPcag and VPcah for T2).

IV. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

1) General Pattern of Activation: For true positives, the ERP analysis reveals a 'neuronal effort' involved in detecting noise: the harder it is to detect noise in a stimulus, the higher the latency and the lower the amplitude of the P300 component. Seemingly, the more the disturbance is pressing to obtain neural resources, the earlier and more pronounced the P300. It is important to note that all of these trials are hits and thus indistinguishable on a behavioral level. The latency and amplitude of the P300 could therefore be used as a measure to assess the degree to which a subject is disturbed by a noisy stimulus.

2) Classification: We trained a classifier based on shrinkage LDA that is able to discriminate between missed T2 stimuli and correctly recognized non-targets for two subjects, even though the behavioral data suggests that subjects did not perceive noise in either of these stimulus classes. As the classifier was trained to discriminate between T2 hits and non-targets, we conjecture that it singles out trials where noise is missed on a conscious level, but still processed on a subconscious level.

It needs to be taken into account that the true class labels are unknown to us, as we cannot judge whether the noise in a signal was truly not perceived at all, processed subconsciously or whether the subject forgot to press the button. However, the significantly higher perception rates for the T1 stimuli suggest that trials in the T2 condition were mostly missed because they were below the perception threshold (first two cases). The higher the number of actually missed deviant trials (likely to be classified as non-targets), the more the performance of the classifier can be expected to be negatively affected. This assumption is supported by the fact that the AUC values of the classifiers get worse, the harder it becomes to detect the noise in a stimulus.

3) Summary: As a main result of this study, we conclude that ERPs have the potential to be used successfully as a quantitative measure for the assessment of auditory quality, providing complementary information to conventional behavioral methods. Our results show this exemplarily for the (sub)conscious detection of signal-correlated noise in phonemes. For those stimuli that are consciously recognized as being disturbed (hits), we suggest that the latency and amplitude of the P300 component ('neuronal effort') might be used to assess the level of disturbance. For disturbed stimuli for which the noise is below the threshold of conscious perception, we show for two subjects that a classifier based on shrinkage LDA can be applied successfully to single out stimuli for which the noise is presumably processed subconsciously.

B. Future Work

In a follow-up study, it could be investigated, whether ERPs still provide an adequate measure for the detection of disturbances in audio stimuli that are longer than phonemes, such as words or sentences. Moreover, degradation levels could be used that are finer-grained and focus on the threshold between conscious and subconscious processing.

V. ACKNOWLEDGMENTS

This work was supported by the Bundesministerium für Bildung und Forschung (BMBF), FKZ 01GQ0850.

REFERENCES

[1] *ITUR Rec. BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, Int. Telecomm. Union, Geneva Std., 1997.

[2] I. Miettinen, H. Tiitinen, P. Alku, and P. May, "Sensitivity of the human auditory cortex to acoustic degradation of speech and non-speech sounds," *BMC Neuroscience*, vol. 11, p. 24, 2010.

[3] G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, Eds., *Toward Brain-Computer Interfacing*. Cambridge, MA: MIT Press, 2007.

[4] 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, p. e9813, 2010.

[5] I. Sturm, "Machine learning methods for investigating music perception: Detecting neural correlates of perceived consonance in single-trial ERPs," Master's thesis, University of Potsdam, Department of Computer Science, 2010.

[6] S. Sutton, M. Braren, J. Zubin, and E. John, "Evoked-potential correlates of stimulus uncertainty," *Science*, vol. 150(3700), p. 1187 1188, 1965.

[7] J. Polich, "Updating P300: an integrative theory of P3a and P3b," *Clin Neurophysiol*, vol. 118(10), pp. 2128–48, 2007.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley & Sons, 2001.

[9] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components - a tutorial," *Neuroimage*, in press, 2010.

[10] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Stat Appl Genet Mol Biol*, vol. 4, Article32, 2005.

[11] M. D. Green and J. A. Swets, *Signal detection theory and psychophysics*. Huntington, NY: Krieger, 1966.

[12] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 88, pp. 861–874, 2006.

[13] A. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30 (7), pp. 1145–1159, 1997.

[14] *ITU-T Rec. P.810: Modulated noise reference unit (MNRU)*, Int. Telecomm. Union, Geneva Std., 1996.