3rd International Workshop on Perceptual Quality of Systems (PQS 2010)
6-8 September 2010, Bautzen, Germany

Neural correlates of speech degradation – Subjective ratings and brain activation in case of signal-correlated noise

Jan-Niklas Antons[1], *Robert Schleicher*[1], *Ingo Wolf*[1], *Anne K. Porbadnigk*[2], *Benjamin Blankertz*[2,3], *Sebastian Möller*[1] *and Gabriel Curio*[4]

[1] Deutsche Telekom Laboratories, Berlin Institute of Technology, Berlin, Germany
[2] Machine Learning Laboratory, Berlin Institute of Technology, Berlin, Germany
[3] Fraunhofer FIRST, Intelligent Data Analysis Group, Berlin, Germany
[4] Neurophysics Group, Charité-University Medicine, Campus Benjamin Franklin, Berlin, Germany
Jan-Niklas.Antons@Telekom.de

Abstract

Recent studies with magnetoencephalography (MEG) have shown that the human auditory cortex is particularly sensitive to reduction in sound and speech quality. In this paper, we examine whether this sensitivity is also visible in the electroencephalogram (EEG) and whether it is possible to detect subconscious processes which can then be used to improve the behavioral assessment of speech quality. In order to compare the physiological results with behavioral measurements, we degraded a speech stimulus (vowel /a/) in a scalable way and asked for a pair comparison (PC) and a comparison category rating (CCR) of the degraded stimuli. In addition, the brain activity of eleven healthy subjects was measured with EEG, focusing on event-related potentials (ERPs). We found that the threshold as set by the Modulated Noise Reference Unit (MNRU) for the PC and CCR are on a similar signal-to-noise level. We trained classifiers, which were found capable of distinguishing between events which are seemingly similar at the behavioral level (i.e., no button press). Converging evidence suggests that the classifier results could reflect subconscious cortical sensitivity to sound degradations.

Index Terms: Electroencephalography, Speech, Transmission Quality, Subconscious Processing, Shrinkage LDA

1. Introduction

The pair comparison (PC) and comparison category rating (CCR) are commonly used and well-defined subjective test procedures for speech and audio quality assessment in the field of telecommunications [1]. The PC method implies that the test sequences are presented in pairs. Subjects are asked to rate if the quality of the second stimulus is better or worse than that of the first stimulus. The CCR method compares different test conditions with a fixed reference of high quality. The listeners get presented a pair of speech samples for each trial and have to rate the quality of the second compared to the quality of the first, e.g., from excellent (100) to bad (0) [2]. Unfortunately, these approaches do not provide information about possible subconscious processes which could prime for slowly growing dissatisfaction with an audio transmission.
Recent studies in neuroscience showed the promising application of neuro physiological methods for speech quality evaluation by measuring pre-/ sub conscious brain activity [3]. It has been proved that the auditory cortex is particularly sensitive to a reduction in sound quality as visible in the magnetoencephalogram.
A promising solution for quick and mobile measurement of brain activation is the EEG. Coles and Rugg [4] define the electroencephalogram as a voltage variation over time, between a pair of electrodes which are attached to the surface of the human scalp. While recording the EEG, a stimulus can be presented to the subjects. Voltage changes may occur with a fixed temporal relationship to that auditory stimulus. The voltage changes in epochs that are related to the brain's response to the stimulus constitute the event-related potential (ERP). Two well-known components are the mismatch negativity (MMN) and the P300.
ERPs are commonly elicited in the so-called oddball paradigm, in which a random sequence of stimuli is presented. The stimuli can be classified as belonging to one of two categories, one stimuli ('standard', 'non-target' (NT)) occurring frequently (e.g., p = .80), and the other ('deviant', 'target' (T)) occurring infrequently (p = .20). The task of the participants is then to classify the stimuli, either by counting or by pressing a button when a target is presented.
The mismatch negativity (MMN) is elicited by any change in auditory stimulation. This component is thought to reflect a pre-attentive process that detects a difference between an incoming stimulus and the sensory memory trace of preceding stimuli, based on the standard. An MMN can be elicited even in the absence of the participant's attention. For instance, Sculthorpe et al. could elicit MMN even during sleep [5]. The mismatching stimuli can differ on any discriminable auditory dimension, such as pitch, duration, intensity or location [6-8].
The P300 is a large, positive component in the ERP that typically peaks 300 ms or later after onset of a deviant. Stimuli defined as deviants in a task generally elicit larger P300 amplitudes than standards, even when they are equal in probability [6]. In contrast to the MMN, stimuli that would normally elicit neural responses, do not result in a P300 component when they are ignored or when attention is directed away from them. Furthermore, the P300 is elicited only after the stimulus has been evaluated and categorized.
The more complex the stimulus is, the longer the latency of the P300, which can vary from approximately 250 ms up to 1000 ms [6].
Especially for stimuli of high quality, brain signals might reflect subconsciously perceived differences which are not expressed consciously, i.e., via verbal ratings. Schubert et al. could show that the conscious processing of suprathreshold stimuli differs significantly from subconscious processing [9]. In order to investigate the supplementary impact of the mentioned methods, we compared PC, CCR and components of the EEG in this study.

10.21437/PQS.2010-19

2. Material and Methods

Stimulus material was a recording of the vowel /a/ which was recorded in an unechoic chamber from a male speaker. The stimuli were degraded by a Modulated Noise Reference Unit (MNRU) according to ITU-T Rec. P.810 [10] in a controlled and scalable way. For this purpose, signal-correlated noise was added to the original signal at the following signal-to-noise ratios: 5, 10, 14, 16, 18 dB, and 20 to 35 dB in 1 dB steps. As a control condition for the ERPs, an additional recording of the vowel /i/ was used. In order to allow for individual differences, we conducted a pre-test before the actual experiment for selecting appropriate stimuli with regard to signal-to-noise ratio in dB for each subject. The selected noise levels should be recognized with a probability of T1:100%, T2:75%, T3:25% and T4:0%. The average signal to-noise ratios (SNR) for the deviant stimuli were T1: 5, T2:21, T3:24 and T4:28 dB. In the experiment, an average recognition rate of T1:99%, T2:46%, T3:22% and T4:7% were reached. Stimuli that were correctly classified by a subject are termed as 'hits' (true positives, true negatives) and the others as 'misses' (false positives, false negatives). Eleven German students and university staff of TU Berlin, Germany (mean age 25 years, 4 male) participated in the EEG study; none of them reported any hearing impairment.

2.1. EEG recordings

A 64-channel EEG system (Brain Products) was used for the recordings. Acoustic stimuli were presented via an in-ear headphone (Sennheiser) binaurally, at an individual preferred listening level. Per subject, 8 to 12 blocks were recorded, resulting in a total of 107 blocks. During each block, 300 auditory stimuli were presented, each with a duration of 160 ms. An oddball paradigm was used with the undisturbed phoneme /a/ as the standard (p=.70) and the phoneme /a/ disturbed with four varying degrees of signal-correlated noise (p=.06 each) as deviants. An additional 6% of stimuli were the phoneme /i/, which was used as control stimulus. The task of the subjects was to press a button whenever they detected one of the deviants or the control stimulus (identification task).

2.2. PC task

During the pair comparison (PC), listeners were presented with a pair of speech samples on each trial. One stimulus of each pair was the reference. The order of the degraded and reference samples was chosen random for each trial. Subjects were asked to rate if they can detect differences between the two presented stimuli. The behavioral measurements were carried out in a second separate session after the EEG recording. Four subjects out of the 11 that took part in the EEG study also participated in the behavioral measurements, resulting in a total of 18 subjects.

2.3. CCR task

Again listeners were presented with pairs of speech samples, containing in a random order, always the reference and one randomly selected degraded stimuli. The subjects judged the quality of the second sample relative to that of the first and used the scale from excellent (100) to bad (0). Per subject, two blocks (one PC and one CCR) were recorded. During each block, 82 auditory stimuli were presented.

3. Results

3.1. PC

The results of the PC revealed a perception threshold of 21 dB signal-to-noise ratio, as introduced by the MNRU. The probability of detecting the degradation is above 50% at that level (Fig. 1).

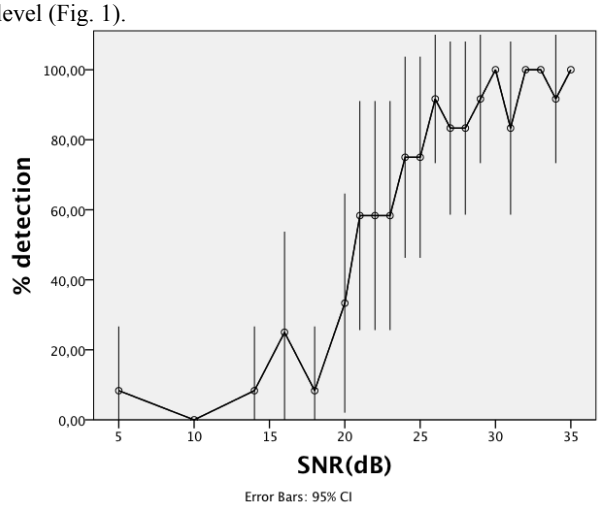

Figure 1: Detection rate (in percent) of the degradation for all SNR(dB) levels used over all subjects.

3.2. CCR

The analysis of variance (ANOVA) with degradation intensity as the independent variable and the mean opinion score (MOS) as the dependent variable on the CCR data revealed a main effect on the factor Stimulus (strength of degradation). The post-hoc test (Scheffé adjustment for pairwise comparisons) was significant at a level of 21 dB ($p<.05$), the quality was rated significantly lower in comparison with the reference (Fig. 2).

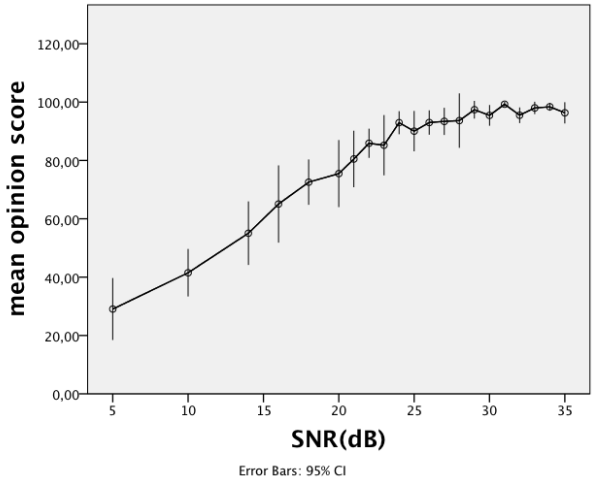

Figure 2: Mean opinion score for all SNR(dB) levels used.

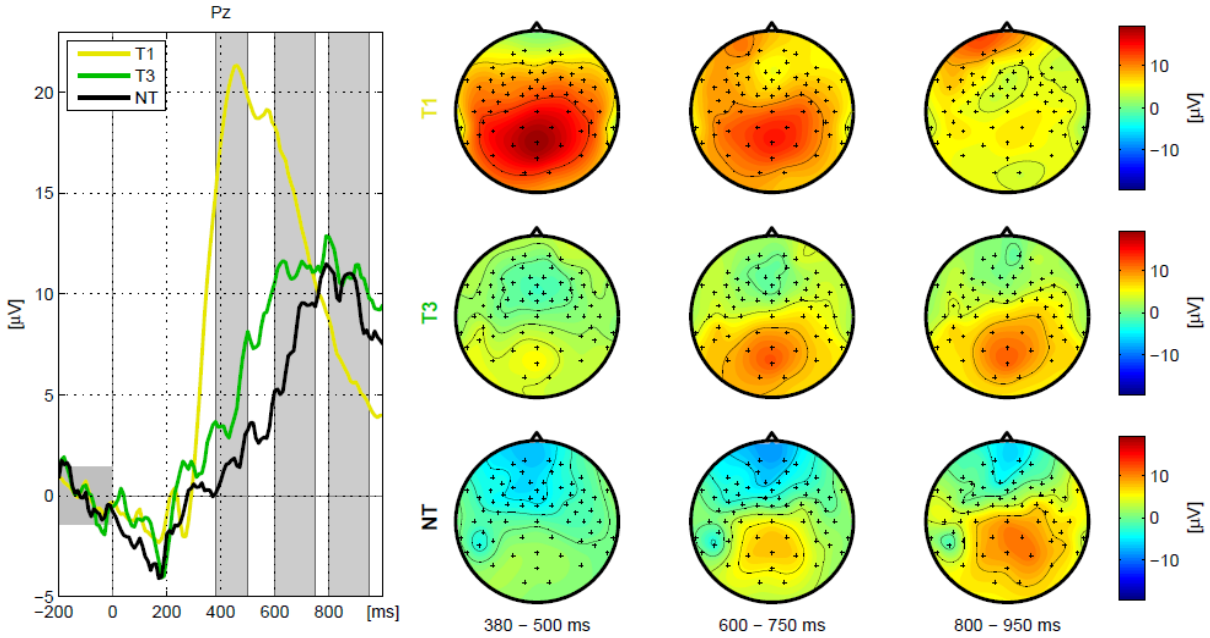

Figure 3: Grand average ERPs for the conditions T1, T3 and non -targets (NT). Left: Time course of ERPs in the time interval -200 to 1000 ms with t = 0 as the time point of stimulus onset. Right: Scalp distribution, time intervals 380 – 500, 600–750, 800–950 ms (marked in grey in the time course on the left) . Positive activation is indicated in red. The maps show the head as seen from the top, with the nose pointing upwards. For NT only false positive were used.

3.3. EEG

The disturbed audio stimuli elicited a characteristic pattern, with an early negativity (M MN pattern about 300ms post -stimulus, frontotemporal, Fig. 5) followed by a P300 . The amplitudes of the P300 component are higher if the stimulus is noisier (see Fig. 3, left). In addition the latency of the P300 component varies with the noise level. Increased demands, the 'neuronal eff ort' for the detection of sound-quality degradation result in a higher latency in the EEG (see Fig. 3, right). For two out of the 11 subjects (VPcad, VPcae) , misses of deviant class T2 resulted in an ERP pattern that was cle arly similar to that elicited by hits of the same deviant class. This can be seen exemplarily in the topographies of running t-test values (explanation of the running t-test [11]) for subject VPcad in Figure 5 (hits versus NT, misses versus NT).
In order to further explore this finding, we analyzed the similarities between the ERP patterns elicited by T2 hits and T2 misses. For this purpose, a classifier based on shrinkage LDA (linear discriminant analysis) [12] was trained to distinguish between hits and one half of the non -targets. The classifier was then tested on the misses of the deviant class and the other half of the non -targets. For both subjects , the classifier was well able to distinguish between misses and non-targets, due to the similarity of the activation patterns of hits and misses. Classification resulted in an area under the ROC (receiver operating characteristic) curve (AUC) of 0.7617 and 0.6659 for subjects VPcad (see Fig. 4) and VPcae , respectively. The receiver operating characteristic (R OC) can be represented by plotting the true positive rate versus the false positive rate. ROC analysis allows selecting optimal models independently from the class distribution. The ROC AUC statistic is commonly used for model comparison and can be interpreted as the probability for a classifier to assign a higher score to the positive instance when randomly one positive and one negative instance is selected [13]. For both VPcag and VPcah, an AUC value of 0.62 was reached. Details on classification results can be found in [14] . Thus, even though the behavioral data suggests that a stimulus is not perceived as being degraded, the correspondin g neural activation does indicate for a certain percentage of the trials that the noise is processed subconsciously.

4. Discussion

The result of the two conscious behavioral judgements revealed a threshold at the same noise level. For the PC, the

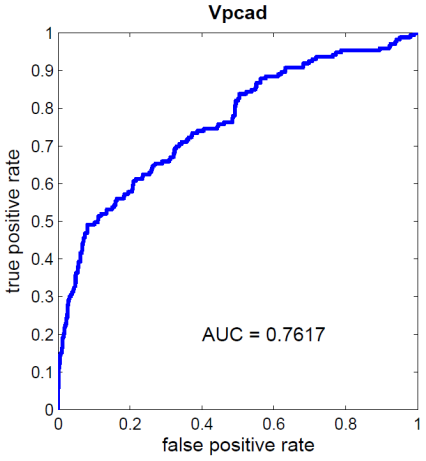

Figure 4: ROC curve for subject VPcad.

level of 50% detection rate of the degradation is reached at 21dB. For the quality judgment (CC R), we can show a significant drop-off of quality at 21dB. Additionally we gained

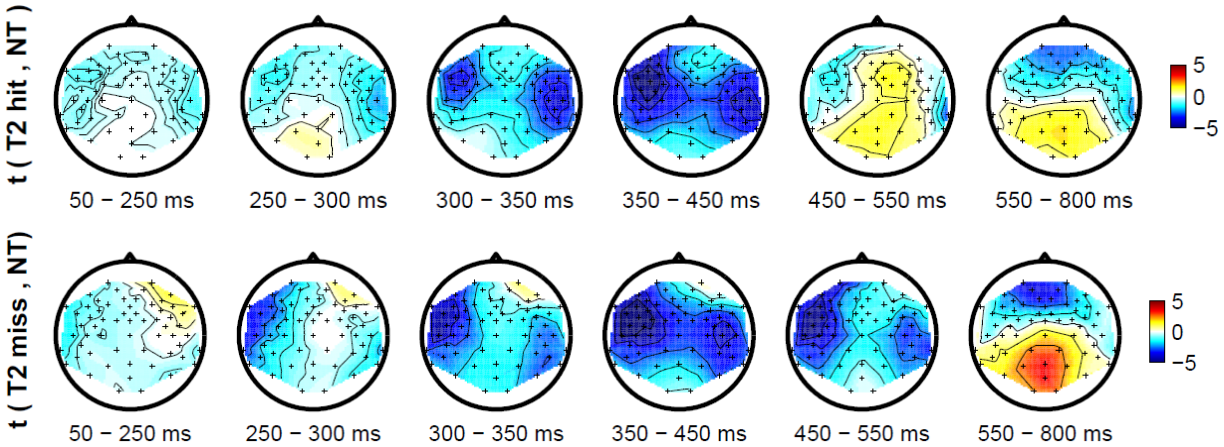

Figure 4: Topographies of running t-test values for subject VPcad and the intervals 50 -250, 25-300, 300-350, 350-450, 450-500, 500-550, 550-800 ms. Top: t-values for the comparison hits in deviant class T2 with the standard stimuli. Bottom: t-values for the comparison misses in deviant class T2 with the standard stimuli.

information about the conscious stimulus processing by analyzing the P300. The strength of the noise level has an impact on the amplitudes and latency of the P300 component. Furthermore, we present classification results which indicate that even though noise is missed on a conscious level, it might still be processed on a subconscious level. Our classifier (based on shrinkage LDA) that was trained on hit s is able to discriminate between missed T2 stimuli and correctly recognized non-targets for four subjects.

The main result of this study is that ERPs have the potential to be used successfully as a quantitative measure for the assessment of auditory quality. For disturbed stimuli for which the noise level is below/next to the threshold, in our study was the T2 stimulus detected with an average of 46 % (21 dB), we show evidence for four subjects that the noise is processed subconsciously for a certain percentage of the trials.

5. Conclusions

We could show that the result of the behavioral methods PC and CCR are on a similar signal -to-noise threshold, as set by the MNRU. Additionally, we could show that a typical ERP activation pattern is still identifiable, eve n when subjective tests are not able to reveal sufficiently that noise in the signal is processed by the subjects. EEG-based classifiers can identify speech samples which are rated high qualitative consciously - neurally, however, noise contamination is detected, possibly affecting the long -term contentment with the transmission quality.

Using EEG data for quality research is a possibility to detect minimal differences in audio signals of high quality. In a follow-up study, we will investigate whether ERPs still provide an adequate measure for the detection of disturbances for audio stimuli that are longer than phonemes, such as words or sentences. Moreover, other classes of degradation (e.g., band pass filters) could be tested and used for integration in existing frameworks of quality prediction [15].

6. Acknowledgements

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

7. References

[1] ITU–T Rec. P.910, “Subjective video quality assessment methods for multimedia applicat ions”, Int. Telecomm. Union, Geneva, 2008.

[2] ITU–T Rec. P.800, “Methods for subjective determination of transmission quality”, Int. Telecomm. Union, Geneva, 1996.

[3] Miettinen, I., Tiitinen, H., Alku, P., May, P., “Sensitivity of the human auditory cortex to acoustic degradation of speech and non-speech sound”, BMC Neuroscience, 11(24), 1471-2202, 2010.

[4] S Coles, M., Rugg, M., “Event-related brain potentials: an introduction”, in Rugg, M., Coles, M., (Eds.), “Electrophysiology of Mind: Event-Related Brain Potentials and Cognition” Oxford University Press, 1995.

[5] Sculthorpe, L., Ouellet, D., Campbell, K., “MMN elicitation during natural sleep to violations of an auditory pattern”, Brain Reaearch, 1390:52-62, 2009.

[6] Duncan, C., Barry, R., Connolly, J., Fischer, C., Michie, P., Näätänen, R., Polich, J., Reinvang, I., Petten, C., “ Event-related potentials in clinical research: Guidelines for eliciting ,recording, and quantifying mismatch negativity, P300, and N400”, Clinical Neurophysiology, 120:1883-1903, 2009.

[7] Garrido, M., Kilner, J., Stephan, K., Friston, K., “The mismatch negativity: A review of underlying mechanisms”, Clinical Neurophysiology, 120:453-463, 2009.

[8] Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., “The mismatch negativity (MMN) in basic research of central auditory processing: A review”, Clinical Neurophysiology, 118:2544-2590, 2007.

[9] Schubert, R., Blankenburg, F., Lemm, S., Villringer, A., Curio, G., "Now you feel it--now you don't: ERP correlates of somatosensory awareness", Psychophysiology, 43(1):31-40, 2006.

[10] ITU-T Rec. P.810, ”Modulated noise reference unit (MNRU)”, Int. Telecomm. Union, Geneva, 1996.

[11] McGee, T., Kraus, N., Nicol, T., „Is it really mismatch negativity? An assessment of methods for determining response validity in individual subjects”, Electroencephalography and clinical Neurophysiology, 104:359-368, 1997.

[12] Blankertz, B., Lemm, S., Treder, S., Haufe, S., Müller, K.-R., “Single-trial analysis and classification of erp components – a tutorial”, in press NeuroImage, 2010.

[13] Fawcett, T., “An introduction to ROC analysis”, Pattern Recognition Letters, 27:861-874, 2006.

[14] Porbadnigk A.K., Antons J.-N., Blankertz B., Treder M.S., Schleicher R., Moeller S. and Curio G., “Using ERPs for Assessing the (Sub)Conscious Perception of Noise”, Proc. of the 32nd Int’l Conf. of the IEEE Engineering in Medicine and Biology Society, 2010.

[15] Wältermann, M., Scholz, K., Möller, S., “An Instrumental Measure for End-to-end Speech Transmission Quality Based on Perceptual Dimensions: Framework and Realization”, Interspeech 2008.