Audio Engineering Society

Convention Paper

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Subjective Listening Tests and Neural Correlates of Speech Degradation in Case of Signal-correlated Noise

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

[1] Deutsche Telekom Laboratories, Berlin Institute of Technology, Berlin, Germany

[2] Machine Learning Laboratory, Berlin Institute of Technology, Berlin, Germany

[3] Department of Neurology and Clinical Neurophysiology, Charité, Berlin, Germany.

ABSTRACT

In this paper, we examine whether particularly sensitivity of the human cortex to reduction in speech quality is visible in the electroencephalogram (EEG) and whether these measures can be used to improve the behavioral assessment of speech quality. We degraded a speech stimulus (vowel /a/) in a scalable way and asked for a behavioural rating. In addition, the brain activity was measured with EEG. We trained classifiers, which were found capable of distinguishing between events which are seemingly similar at the behavioral level (i.e., no button press), neurally, however, noise contamination is detected, possibly affecting the long-term contentment with the transmission quality.

1. INTRUDUCTION

Subjective test procedures like the pair comparison (PC) and comparison category rating (CCR) are well-defined and used for speech and audio quality assessment in telecommunication research [1]. During the PC method test stimuli 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. If intelligibility is in the focus of interest late components of the event-related potentials are inspected. Recent studies in neuroscience showed the promising application of neuro physiological methods for speech quality evaluation by measuring pre-/ sub conscious brain activity in earlier components [3]. 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. The P300 is a large, positive component in the ERP that typically peaks 300 ms or later after onset of a deviant. 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 [5-8]. In order to investigate the supplementary impact of the mentioned methods, we compared PC, CCR and components of the EEG in this study.

2. MATERIAL AND METHODS

Stimulus material was a recording of the vowel /a/. The stimuli were degraded by a Modulated Noise Reference Unit (MNRU) according to ITU-T Rec. P.810 [9] in a controlled and scalable way. The average signal to-noise ratios (SNR) for the deviant stimuli were: Target 1 (T1) = 5, T2 = 21, T3 = 24 and T4 = 28 dB. 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. The undisturbed phoneme /a/ was used as the standard (p=.70) and the phoneme /a/ disturbed with four varying degrees of signal-correlated noise (p=.06 each) was used as deviants. An additional 6% of stimuli were the phoneme /i/, which was used as control stimulus. A 64-channel EEG system was used for the recordings. Acoustic stimuli were presented via an in-ear headphone 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. The task of the subjects was to press a button whenever they detected one of the deviants or the control stimulus (identification task). During the behavioral assessment which was recorded in a separate session, listeners were presented with pairs 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 during the pair comparison (PC). For the CCR, listeners had to judge the quality of the second sample compared with the first and use the scale from excellent (100) to bad (0).

3. RESULTS

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).

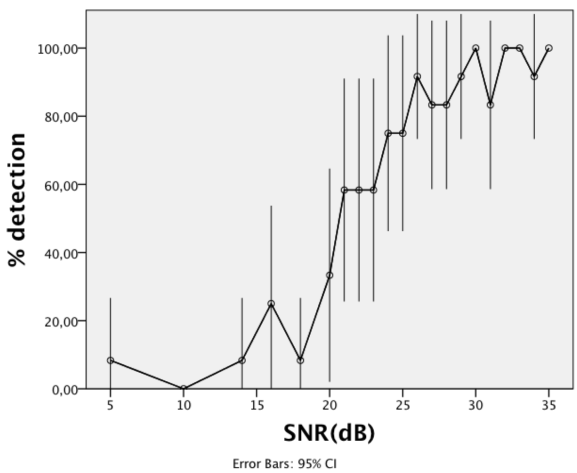

Figure 1 Detection rate (in percent) of the degradation for all SNR(dB) levels used over all subjects.

For the 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).

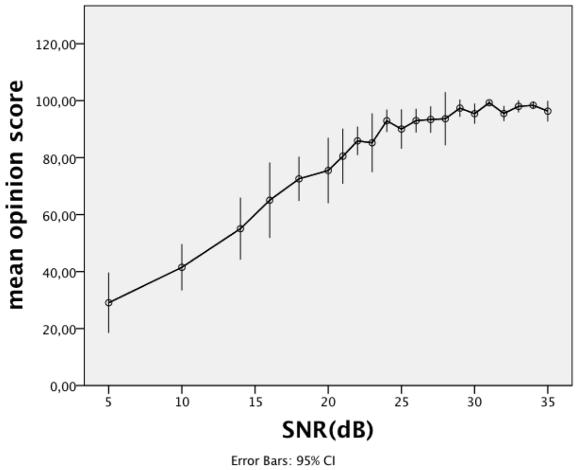

Figure 2 Mean opinion score for all SNR(dB) levels used.

The disturbed audio stimuli elicited a characteristic pattern, with an early negativity (MMN pattern about 300ms post-stimulus) followed by a P300. The amplitudes for the P300 component are higher if the stimulus is noisier. In addition, increased demands that is to say the 'neuronal effort' for the detection of sound-quality degradation results in a higher latency in the EEG (Fig. 3). We trained a classifier based on shrinkage LDA (linear discriminant analysis), to distinguish between hits and one half of the non-targets. Area under the ROC (receiver operating characteristic) curve (AUC) for subject VPcad = 0.7617, VPcae = 0.6659, VPcag and VPcah, = 0.62 was reached [10-12].

4. DISCUSSION

The result of the two conscious behavioral judgements revealed a threshold at the same noise level (21 dB). 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 not detected on a conscious level, it might still be processed on a subconscious level. In our study, the disturbed stimuli for which the noise level is below but close to the threshold was 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.

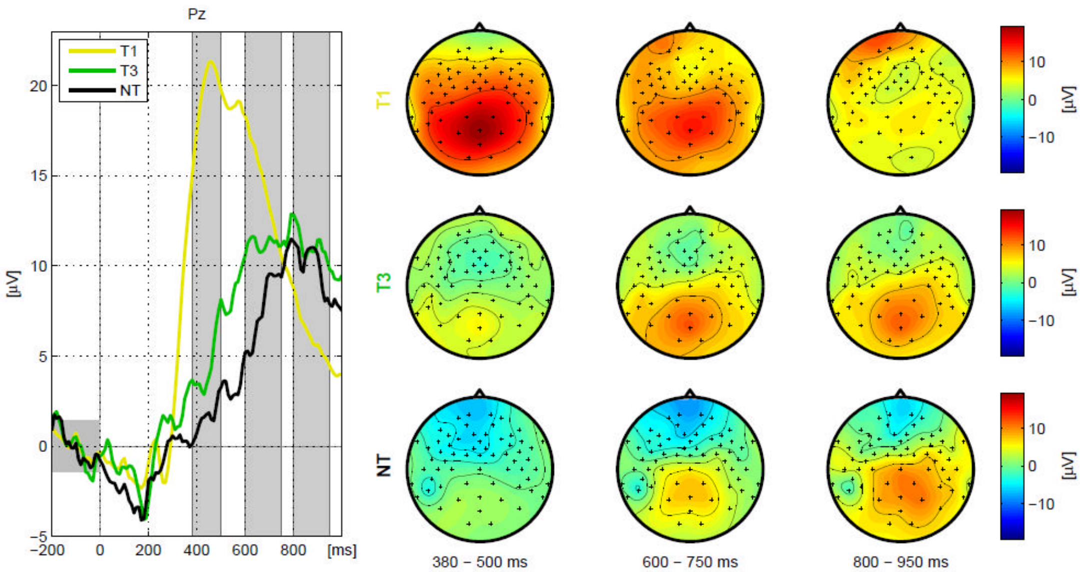

Figure 3: Grand average ERPs for the conditions T1, T3 and non-targets (NT). Left: Time course of ERPs in the time interval -200 to 1000 ms with t = 0 as the time point of stimulus onset. Right: Scalp distribution, time intervals 380–500, 600–750, 800–950 ms (marked in grey in the time course on the 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.

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, even when subjective tests do not to reveal that noise in the signal is actually processed by the subjects. Thus, even though the behavioral data suggests that a stimulus is not perceived as being degraded, the corresponding neural activation does indicate for a certain percentage of the trials that the noise is processed subconsciously. Using EEG data for quality research is a possibility to detect minimal differences in audio signals of high quality.

6. ACKNOWLEDGEMENTS

This work was supported by the Federal Ministry of Education and Research (BMBF) FKZ 01GQ0850.

7. REFERENCES

[1] ITU–T Rec. P.910, "Subjective video quality assessment methods for multimedia applications", 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 Potentialsand Cognition" Oxford University Press, 1995.

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

[6] Duncan, C., Barry, R., Connolly, J., Fischer, C., Michie, P.,Näätänen, R., Polich, J., Reinvang, I., Petten, C., " Eventrelated potentials in clinical research: Guidelines for eliciting ,recording, and quantifying mismatch negativity, P300, andN400", 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] ITU-T Rec. P.810, "Modulated noise reference unit (MNRU)", Int. Telecomm. Union, Geneva, 1996.

[10] 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.

[11] Fawcett, T., "An introduction to ROC analysis", PatternRecognition Letters, 27:861-874, 2006.

[12] 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.