Listening test design for synchronous acquisition of physiological data and cognitive performance in disturbing noise

Christian Laufs[1,2], Andreas Herweg[1], Christoph Hoog Antink[2]
[1] *HEAD acoustics GmbH, 52134 Herzogenrath, Germany*
[2] *KIS*MED (AI-Systems in Medicine), Technische Universität Darmstadt, 64283 Darmstadt, Germany*
Corresponding author: christian.laufs@head-acoustics.com

Introduction

In addition to the aural noise effects that have been studied in detail, sound events that are perceived as noise often result in so-called extra-aural noise effects [4]. They describe negative impacts to the entire body, psychological or physiological, in response to noise. Examples of extra-aural noise effects include decreased cognitive performance or bodily stress responses. The former can be assessed in laboratory studies using established cognitive tests. Numerous studies have demonstrated the negative effect of various noises on cognitive performance by comparing performance indicators of cognitive tests between processing the test in silence and with presentation of the noise as described in a review by Schlittmeier [8]. She concludes that a serial recall test is used most frequently. Here, the participants have to remember combinations of numbers or letters in order to be able to reproduce them as accurately as possible. In addition, the so-called N-Back-Test is used in some of these studies to demonstrate the negative effects of noise on cognitive performance.

Since extra-aural noise effects usually affect several dimensions, the physiological response to noise is being studied increasingly [5]. For the stress response mentioned above, well-studied mechanisms operate in the body. In such a reaction, noise perceived as a stressor activates the sympathetic nervous system, which puts the body into a state of alarm in order to be able to react to the stressor [9]. In this process, various parameters change in the body, such as an increased heart rate or a decreased heart rate variability. Increased sweat production in the palms of the hands also represents a reaction to a stressor. Physiological sensors make it possible to record these changes in the body and thus allow conclusions to be drawn about the participant's stress response in reaction to sounds. In this field, the measurement of heart rate variability by means of an ECG system and the measurement of skin conductance as a value for sweat production in the palms of the hands have already proven to be reliable indicators of a stress response [3]. In the context of listening tests, these parameters have already been used to identify stress responses. In addition to the self-reported questioning of the participants about perceived stress responses, the physiological measurements open up the recording of a further dimension, which also allows the detection of unconscious stress responses [9]. However, since the various processes in the body during a stress response are often associated with certain latencies or require certain measurement durations, the addition of physiological measurements imposes a variety of requirements on the experimental design.

The study conducted here therefore represents a pre-study for an investigation with physiological measurements. Physiological measurements are not yet used in this pre-study, but the experimental design has been adapted to the requirements of physiological measurements. The use of a cognitive test is primarily aimed at obtaining an equal and controllable activity for all participants, while the physiological measurements will take place in the main study. However, since a cognitive test can also provide information about another dimension of extra-aural noise effects, this pre-study will investigate whether a measurement of influences on cognitive performance is also possible in this experimental design, which has already been adapted to physiological measurements.

Methods

Participants

A total of 15 participants (age: 24-36 years; Mean = 27.6 years, SD = 3.4 years, 4 female) took part in the experiment. Before the experiment started, informed consent was obtained from each participant for the whole procedure of the experiment. All participants were native German speakers, as German speech signals were used as stimuli.

Cognitive Test

During the experiment, the participants performed a cognitive test. The engagement in cognitive work is intended to create a more realistic environment with everyday tasks that occur, for example, in the workplace. In addition, the test can measure losses in cognitive performance as an extra-aural effect. A cognitive test that allows simultaneous measurement of HRV and skin conductance is the N-Back-Test. In this test, participants are presented with individual letters on the screen one after the other. For each letter, participants must press one of two possible buttons. One button must be pressed for letters that are the same as the N-th letter before and thus represent a target. For all other letters where this is not the case, the other button must be pressed. As a performance indicator, the percentage of missed targets is used in the N-Back-Test. In addition, the reaction time can be used as an indicator. Letters were presented for 500 ms each, followed by a blank screen for the duration of an inter-stimulus interval (ISI) until the next letter was

presented. Only easily distinguishable letters were taken for the letter sequences. In addition, the number of targets as well as their positions was balanced between each test phase of the N-Back-Test. For example, the number of targets in direct succession or targets with one letter distance was balanced over the individual test phases. Furthermore, it was ensured that the number of lures, i.e. identical letters at positions N-1 and N+1, was balanced across the test phases.
In a small pretest, a suitable difficulty of the N-Back-Test was to be determined by comparing different variants of the test with each other. The variables of the test that could be changed were the duration of the ISI and the variable N. In this pretest, ten participants (age: 26-30 years; Mean = 28.3 years, SD = 2.1 years, 3 female) completed four different variants, each with three repetitions of the individual variants. The four variants were

- 2-Back Test (1500 ms ISI),
- 2-Back Test (1000 ms ISI),
- 2-Back Test (700 ms ISI),
- 3-Back Test (3250 ms ISI).

The aim of the comparison was to achieve relative high error rate on missed targets during processing in silence, comparable to error rates from other cognitive tests that have already demonstrated losses in cognitive performance. For the individual variants, this resulted in the error rates in Figure 1.

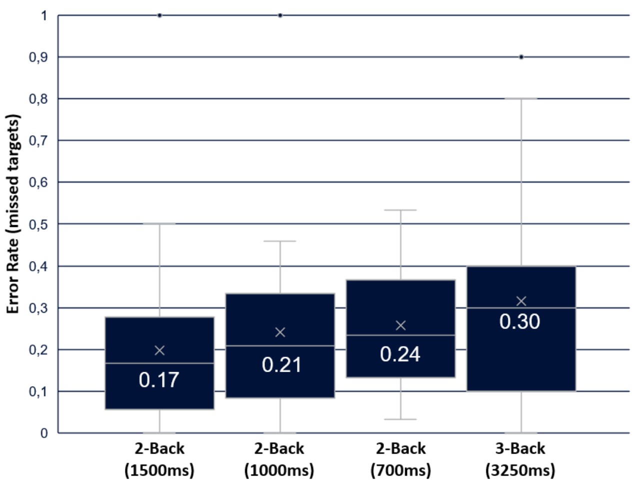

Figure 1: Error rate of missed targets within the four different variants of the N-Back Test

The 2-Back-Test with 700 ms ISI was chosen since the mean error rate of 24 % is comparable to error rates in silence from studies with the serial recall task [6]. Although the 3-Back-Test produced an even higher error rate, the participants reported that they used different strategies to complete the test due to the relatively long ISI. It is reasonable to assume that different strategies also show different sensitivity to interference by noise.

Acoustic Stimuli

In order to validate whether losses in cognitive performance could also be detected in this experimental design, acoustic stimuli were selected that had already significantly reduced cognitive performance in other studies. First, irrelevant speech was chosen, using a continuous speech signal (65 dB(A)) by a German female speaker. In several studies, speech signals have decreased cognitive performance [6] [7]. In addition to the speech signal, traffic noise (65 dB(A)) was used as a second stimulus, as this category of noise is one of the most significant categories in terms of impact on health [10].

Experimental Design

For the construction of an experimental design, the requirements of the physiological measurements for the design should first be discussed in more detail. To analyze HRV, features in time- and frequency-domain are first calculated from the ECG signal. To obtain meaningful features, a sufficient number of heartbeats must be recorded within the measurement interval. For features in time domain, measurement intervals of at least 60 s are required [2]. For the calculation of features in frequency domain, a minimum duration of 90 s is required. In order to be able to make valid statements based on the features in the future main study, the duration of a test phase and thus the duration of the presentation of individual stimuli was therefore set to 120 s. Also for the measurement of skin conductance, this duration leads to a calculation of valid slowly changing (tonic) features [3]. These tonic features of skin conductance are the Skin Conductance Level (SCL) and the frequency of Skin Conductance Responses (fSCRs). These SCRs in particular, which represent individual peaks in skin conductance, require a time of at least 20 s to return to a resting value [3]. This time specification determines the choice of the duration of rest periods between the individual test phases, which was set at 60 s. As a final experimental design, the processing time of the 2-Back-Test was 120 s, which was followed by a rest period of 60 s, during which the participants were asked to remain seated.
The different test phases took place under three different conditions: the processing of the 2-Back-Test with simultaneous presentation of the speech signal or traffic noise and the processing of the test without presentation of an acoustic stimulus. Each of the conditions was repeated once, resulting in six sessions of 120 s of the 2-Back-Test. The order of conditions was counterbalanced using a Latin Square between all participants. Two identical conditions never followed each other directly. The experiment started with a training session of the 2-Back-Test.
After each session of the 2-Back-Test, participants were asked via a Visual Analog Scale (VAS) how disturbed they felt during the previous test. The VAS allows the experimental procedure to be minimally disturbed by the questioning and enables a subsequent comparison between the self-reported VAS-data and the cognitive performance.

Hypothesis

From the independent variables (presence and absence of irrelevant speech or traffic noise) and the dependent variables (measures of cognitive performance; error rate and reaction time) and self-reported disturbance, the following hypotheses emerge based on the literature: H1: The presentation of irrelevant speech or traffic noise reduces cognitive performance in comparison to silence. H2: The presentation of irrelevant speech or traffic noise increases self-reported interference in comparison to silence.

Experimental Environment

The experiment took place in the listening studio of HEAD acoustics with a constant room temperature. The acoustic stimuli were presented using calibrated and equalized hardware from HEAD acoustics via Sennheiser HD 650 headphones. Like the N-Back-Test, the complete experiment was created in HEAD acoustics' SQala software utilizing the SQala Extension API.

Results

Cognitive Performance

For the cognitive test results, the error rate on missed targets was first evaluated as a score for cognitive performance. As described in Figure 2, this resulted in an average error rate of 22.6 % (SD = 12.2 %) for processing the 2-Back-Test in silence. Under the auditory stimulus of irrelevant speech, the average error rate increased to 29.3 % (SD = 10.0 %), and for the presentation of traffic noise there was an increase to 24.5 % (SD=12.6 %). After verification using the Shapiro Wilk Test ($p > .05$ in all cases), a normal distribution of the data was found for all conditions. Thus, the one-way repeated measures ANOVA could be performed, that showed no significant differences of the mean values ($F(2, 87) = 1.79$, $p = .17$, $\eta_g^2 = 0.04$).

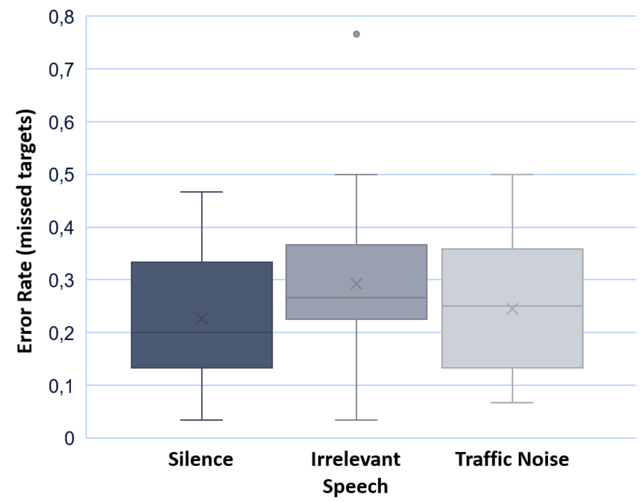

Figure 2: Results of the error rates on missed targets for the individual sessions of the N-Back-Test in silence and with the acoustic stimuli of irrelevant speech and traffic noise.

As a second score for cognitive performance, the reaction time of the participants for answering the 2-Back-Test when a target was presented was analyzed. As described in Figure 3, this resulted in an average reaction time of 535 ms (SD = 62 ms) for processing the test in silence. When irrelevant speech was presented, the average reaction time decreased slightly to 531 ms (SD = 66 ms), and when traffic noise was presented, the average reaction time increased to 557 ms (SD = 61 ms). After verification using the Shapiro Wilk Test ($p > .05$ in all cases), a normal distribution of the data was found for all conditions. Thus, the one-way repeated measures ANOVA could be performed, that showed no significant differences of the mean values ($F(2, 87) = 1.3$, $p = .18$, $\eta_g^2 = 0.03$).

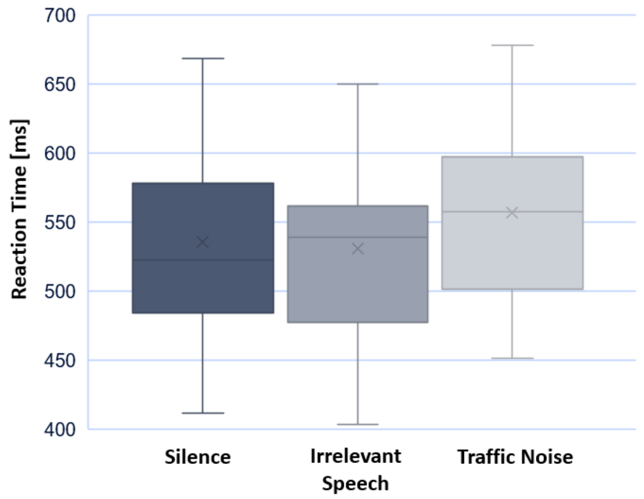

Figure 3: Results of the reaction time for pressing the button on targets for the individual sessions of the N-Back-Test in silence and with the acoustic stimuli of irrelevant speech and traffic noise.

Self-Reported Disturbance

The score for the self-reported disturbance were the responses of the VAS for the individual conditions, which could take values from 0 to 100. As described in Figure 4 this resulted in an average VAS rating of 8.5 (SD = 18.6) for processing the test in silence. When processing the test during the presence of irrelevant speech, the average VAS rating increased to 69.6 (SD = 27.5), and when presenting traffic noise, the VAS rating increased to an average value of 57.8 (SD = 23.2). A normal distribution was not present for any of the conditions according to the Shapiro Wilk test ($p < .05$ in all cases). The Friedman test was therefore used for statistical analysis, that indicated significant differences ($\chi^2(2) = 34.85$, $p < .0001$). In the course of the Wilcoxon Test with Bonferroni Correction as a subsequent post-hoc test, there was a significant increase in the VAS rating between silence and irrelevant speech ($p < .0001$) and between silence and traffic noise ($p < .0001$), as it is visualized in Figure 4. Analyzing the conditions with acoustic stimuli, there was a significant higher VAS rating in the condition with irrelevant speech compared to the condition with traffic noise ($p = .017$).

Discussion

The evaluation of the collected data shows that the expected losses in cognitive performance under presentation of acoustic stimuli were not significant, contrary to hypothesis H1. Nevertheless, the observed tendencies in the increase of the mean error rate from 22.6 % to 29.3 % during presentation of irrelevant speech and in the increase of the mean reaction time from 535 ms to 557 ms during presentation of traffic noise may indicate the presence of extra-aural effects. The fact that the error rate was not significantly higher in the presentation of speech, as it was found by Liebl et al. [6] for example, could be

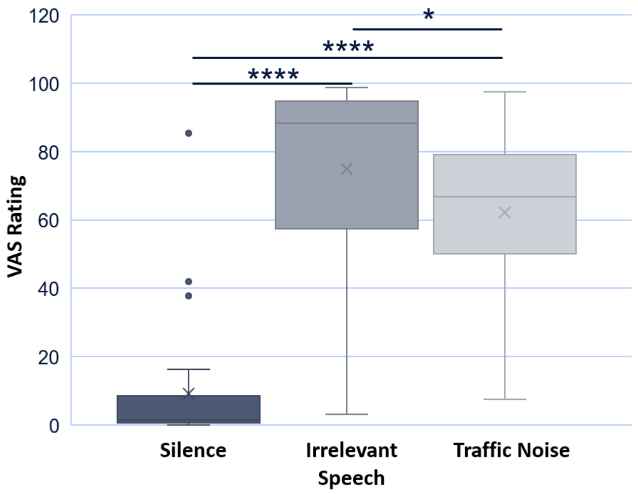

Figure 4: Results of the VAS ratings for the individual sessions of the N-Back-Test in silence and with the acoustic stimuli of irrelevant speech and traffic noise.

due to the different cognitive tests used. In their study they applied the frequently used serial recall test, which mainly requires short-term memory. The influence of irrelevant speech on short-term memory has already been demonstrated several times [8]. In the N-Back-Test, on the other hand, it is the working memory that is more involved. The choice of cognitive test for this pre-study fell on the N-Back-Test, since it is considered to be more suitable for the simultaneous use of physiological measurements. This is because, compared to Serial Recall, the N-Back test produces a constant activity for the participants and thus a more constant underlying basic stress level. On the other hand, it should be noted that the number of participants for this pre-study was relatively small and the tendencies might have become significant with a larger number of participants.

The higher error rate in the presentation of speech compared to the presentation of traffic noise has already been found in other studies [7]. Likewise, the influence of traffic noise on the reaction time of the participants has already been observed, which coincides with the tendencies detected in this study [1]. Regarding the recording of the reaction time, it should be mentioned that it was recorded via time stamps of the computer. Thereby, the computer induced delay is often determined. In this study this was not tested, since no considerable computations in the program occur during the recording. The error rate in silence was at 22.6 %. In the pretest with different versions of the N-Back-Test, an comparable error rate of 24.0 % was determined when processing the 2-Back-Test in silence. The reproducibility is therefore suggested for this version of the N-Back-Test.

Regarding self-reported VAS ratings, hypothesis H2 was confirmed with significant increases in VAS ratings in the presence of irrelevant speech and traffic noise. The increases are consistent with observed tendencies in cognitive performance. However, the large increases from 8.5 to 69.6 respectively 57.8 suggest that there are other extra-aural effects besides the decreases in cognitive performance visible only in tendencies. These further extra-aural effects could be e.g. stress responses triggered by the background sounds [5].

Conclusion

The suspected extra-aural effects were significantly reflected only in the self-reported data of the participants but not in the cognitive performance. This pre-study thus shows that the experimental design adapted to physiological measurements does not clearly reveal cognitive effects of the disturbing noises, but that tendencies of these effects can be observed. In subsequent studies with this design, where the focus will be on the analysis of physiological data, the analysis of concurrent cognitive effects must therefore be handled with caution. Nevertheless, the N-Back-Test promises to be suitable as a cognitive test for the simultaneous measurement of physiological data due to its constant load and simplicity of execution.

References

[1] Alimohammadi, I. and Zokaei, M. and Sandrock, S.: *The effect of road traffic noise on reaction time.* Health promotion perspectives, 2015

[2] Baek, H. and Cho, C. and Cho, J. and Woo, J.: *Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability.* Telemedicine and e-Health, 2015

[3] Boucsein, W.: *Electrodermal Activity.* Springer Science & Business Media, 2012

[4] Genuit, K. and Fiebig, A.: *Die Psychoakustik im Bereich der Lärmwirkungsforschung.* Praktische Arbeitsmedizin 9, 2007

[5] Laufs, C. and Herweg, A. and Loh, K. and Fels, J.: *Relationship between the psychoacoustic parameter sharpness and the physiological parameter skin conductance for the assessment of extra-aural noise effects.* Jahrestagung für Akustik DAGA, 2021

[6] Liebl, A. and Assfalg, A. and Schlittmeier, S.: *The effects of speech intelligibility and temporal–spectral variability on performance and annoyance ratings.* Applied Acoustics, 2016

[7] Schlittmeier, S. and Feil, A. and Liebl, A. and Hellbrück, J.: *The impact of road traffic noise on cognitive performance in attention-based tasks depends on noise level even within moderate-level ranges.* Noise & health, 2015

[8] Schlittmeier, S.: *Review of research on the effects of noise on cognitive performance 2017-2021.* Deutsche Nationalbibliothek, 2021

[9] Spreng, M.: *Central nervous system activation by noise.* Noise and health, 2000

[10] World Health Organization (Europe): *Environmental noise guidelines for the European Region.*(2018)