EFFECTS OF THE ACOUSTICAL WORK ENVIRONMENT ON READING PERFORMANCE IN EMPLOYEES: A LABORATORY STUDY ON THE EVALUATION OF A READING TASK

Helga Sukowski

*Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Dortmund, Germany*

*e-mail: sukowski.helga@baua.bund.de*

Reading is a typical task in everyday life of employees. Many employees have to get information or instructions from written material in the frame of their occupation. Reading was therefore chosen as the main task in a research project on noise effects on cognitive performance at the work place. In order to get valid results on the possible effects, it is important to use a task that is sensitive to noise, and that is evaluated with regard to training effects and also to the equivalence of different test versions, when more than one version is used.

The project has several work packages. Currently an experimental study with 51 participants has been finished, where the participants worked on a reading task that was developed by this author. The task has two different versions. Participants were asked to find mistakes in sentences that were presented on a computer screen. Since the focus of this study was mainly on the task itself, participants worked on the task either twice in a silent condition (group 1) or twice in a condition with speech as background sound (group 2). Each person worked once on test version A and once on test version B.

The main questions in this study were: (1) Is there a difference between the mean results in the first and the second run, and if so, is this different for group 1 and group 2? (2) Are there differences in the performance between both test versions? The results are presented and discussed in this contribution. They will finally be used to optimise the reading task and the design of the test runs, in order to get a solid basis for a prospected work package in which the effects of different typical sounds from the work place will be investigated.

Keywords: effects of noise, work place, reading performance
1. Introduction

Unwanted noise in the work place is well known as one factor that can cause detrimental effects on employees. These effects can become manifest in many different ways like hearing impairments, other effects on health, effects on performance, effects on experienced effort or annoyance. Therefore noise is one important factor that has to be considered in the risk assessment which has to be carried out by employers. Noise should also be part of the psycho-social risk assessment [see e.g. 1]. There are also regulations in order to protect employees from detrimental effects of noise with high sound pressure levels [2] and of noise with lower sound pressure levels [3].

Effects of noise on reading, which are in the focus of the current contribution, belong to the general research field of effects of noise on cognitive performance. This is one component of the “non-auditory effects of noise”. Although there is a number of studies dealing with this topic [see e.g. 4] many questions are still open - and this holds especially for the aspects that are relevant in the research dealing with effects on employees at the work place. Therefore, the author of the current contribution already stated in previous publications that this topic needs to be treated in a more systematic way in future, because the demand to get information from written material (=reading) is a frequent and essential part of everyday work for many employees in many different occupations [5].

Reading performance can be measured in different ways, with different tasks and, according to the respective research question, in different research settings like field studies, laboratory studies, different research designs, e.g. including a silent condition as a reference or not, or realizing many different variations in the applied sound conditions.

One type of reading task is the so called proofreading task. This task can be realized as a task finding typing errors or other errors not depending on the context (non-contextual errors) e.g. 6, and it can also be realised as a task, where the participants have to understand the context in order to find a mistake (contextual errors; studies using both kinds of mistakes, see e.g. [7, 8]), or in which they have to revert to their knowledge in order to recognize that an information cannot be true in the given item. The errors requiring more than focusing on one isolated word seem to be more likely to be missed when reading in a noise condition or in a supposed more disturbing situation with background noise, respectively.

In order to pursue the goal to investigate the possible effects of noise on reading performance in a more systematic way, it seems necessary as a first step to invest in the development or selection of a suitable task or suitable test items. Especially in studies where participants have to work twice on a task it is important to have valid information about the equality of the two test versions applied and the possible training effects that are likely to occur in studies on reading performance.

The long term aim of the research project delineated below is to investigate the effects of typical sounds at the work place on reading performance. Due to the described background the first step in this project is a study that focuses on the mentioned aspects in a self-developed reading task, in order to optimise the task for further applications in studies on noise effects on employees at work.

2. From a pilot study to the current research project

In 2016 a pilot study was carried out in the context of research on effects of noise on cognitive performance at the work place. In that pilot study 24 participants worked on two versions (A and B) of a self-developed reading task in silence or in a condition with speech as background sound. The aims were to investigate the suitability of the task itself, in which participants had to find mistakes in written sentences, and to test whether the task is in principle appropriate to detect effects on performance between both acoustical conditions. The pilot study revealed no significant difference between the results

---

[1]: Author.
[2]: Reference.
[3]: Reference.
[4]: Reference.
[5]: Reference.
[6]: Reference.
[7]: Reference.
[8]: Reference.
ascertained with the two different test versions. A significant difference was found between the performance in the different acoustical conditions (sub-group, n=12) [5].

Although the first versions of the proofreading task turned out to be practicable and useful to detect effects of a speech sound on reading performance, further evaluations were carried out with regard to the difficulty of the single test items and the category of constructed mistakes, respectively, as well as on the equality of both test versions on item level [9], in order to optimise the task. In addition to the input for modifying the test material, the pilot study was also a good basis to collect experience for the entire experiment. Changes in the test material include replacements of a few pairs of items (pair = parallel items in both versions), modifications of one item in a pair and modifications of the time given to work on a particular item.

Based on the results and conclusions from the pilot study and based on previous experience in this topic [e.g. 10], a new research project titled “Effects of the acoustic work environment on reading performance and well-being in employees. An experimental investigation in the laboratory” is now on the way [11]. Regarding the reading performance the project has two main work packages with one empirical study each.

The aim of the first study is to get further information about the self-developed proofreading task by testing a larger group of participants. The data collection in this study is finished. The second study will be carried out with the aim to investigate the effects of various realistic scenarios of occupational noise on reading performance. Regarding the sample, one important aspect in both laboratory studies is that the participants should be employees aged between 30 and 60 years. This is a group that so far has rarely been included in this kind of laboratory studies. For the entire research project a positive vote was granted by the ethics committee of the Federal Institute for Occupational Safety and Health (BAuA). The following sections in this contribution deal with the first study of this research project and focus on the results in reading performance.

The main questions in this study were:

1. Is there a difference between the mean results in the first and the second run, when participants work on the two test versions in the same acoustical condition, and if so,
2. is this difference in performance different for a group working twice in a silent condition and a group working twice in a condition with speech as background sound?
3. Are there differences in the performance between both test versions?

At this stage, this study was not designed to carry out an analysis regarding a direct comparison between the total performances in the two groups with different acoustical conditions. As explained above this will be part of a prospected future study.

### 3. Methods

#### 3.1 Participants

In total 51 persons worked on the reading task in both test sessions. The results from 45 participants (female: 16, male: 29) were included in the data analysis. The results from six participants were excluded, because of unexpected disturbances while working on the reading task in one test run (e.g. a phone ring on a participants’ phone) or because participants mentioned that they missed to follow particular aspects of the instruction while working on the task (e.g. forgotten to confirm the decisions for each item with a “finish” button as instructed). There are some indications that in two further cases the participants also might have had problems with the instruction. Since in these cases the problems were not explicitly mentioned or not as obvious these cases were not excluded from the sample at the current stage. The age of the participants ranged from 30 to 59 years (mean = 47.8 years; median = 52 years). All participants were employees from different local companies, institutions or authorities. According to the self-assessment along an initial questionnaire (see section 3.6), 43 of the 45 participants reported
their hearing status as normal, whilst one participant reported a slight hearing loss and one participant did not respond to this question. Since a repeated measurement design was employed, slight differences in the individual hearing status are not relevant for this study.

### 3.2 Reading test procedure

As test material the modified versions of the self-developed reading task were used. Again 52 items (one or two sentences) were applied in each version. The task for the participants was to find mistakes in the presented items. The kind of mistakes varied. There were grammatical mistakes, orthographic mistakes or there were elements in the items that did not make sense. Some items did not contain a mistake. The number of these was not known by the participants. The items were presented one by one on a computer screen for a predefined time slot. The duration of the time slots varied between the items. In the time while one item was present the participants could read it and make their decision. Participants worked on the items via touchscreen. They could mark the relevant element or fields with their fingers or with a special pen.

The instruction to work on the items was in short: Read the item, mark the word being or containing the mistake or mark the field “everything correct” and finish your decision by marking the field “finished”; then move on to the next item. The instruction was given in written form and with some standardised verbal supplements by the investigator. Then two training runs followed. The first training run consisted of three items which were presented without time limit. Participants worked on these items under guidance of the investigator and they had the chance to try the different functions (e.g. changing from one chosen word to another one). The second training run contained six items with time limit, so that participants could get familiar with the procedure in the real test run. The test runs itself took a maximum time of 14-15 minutes in each session. The results presented below refer to the number of correctly finished items in each test run or each test version, respectively.

### 3.3 Acoustical conditions

The two acoustical conditions were a silence condition (“silence”) and a condition with speech as background sound (“speech”). Because the experiment was carried out in a large sound insulated laboratory (see below) the silence condition was real silence without any background sound. Participants wore closed supra-aural headphones (Sennheiser HD 25) in both acoustical conditions in order to standardise the test conditions as far as possible in all groups. In the speech condition a speech sound was presented stereophonically via these headphones while working on the reading task. The speech sound was the same as in the pilot study (several sections from an audio drama). The sound pressure levels measured with a dummy head for the maximum time of a test run at the two arranged work places (see below) were:

- Work place 1: left $L_{eq} = 57.5$ dB(A), right $L_{eq} = 60.4$ dB(A)
- Work place 2: left $L_{eq} = 57.9$ dB(A), right $L_{eq} = 61.1$ dB(A).

The sound was played back at each work place from a laptop computer (HP-Envy X360, 15.4” Touch-Display) via an external sound device (Roland UA-55 Quad Capture Audio Interface).

### 3.4 Study design

With respect to the aim of this study, each participant worked on both test versions (once on version A and once on version B) in the same acoustical condition, that means either both times in the silent condition (group 1) or both times in the condition with background speech (group 2). The first point of measurement is called T1 throughout this contribution, the second point is called T2. The desired time window between the two points of measurement (two test sessions) was a minimum of one week and a maximum of 10 days. In addition, it was intended that both appointments for one participant were nearly at the same time of day. With very few exceptions these goals regarding the scheduling and the time
of day were achieved. The sequence of the two test versions was balanced in both groups. Therefore, there were finally four different test sets. These four sets and the number of participants in each set are listed in Table 1.

Table 1: Number of participants in all four test sets

<table>
<thead>
<tr>
<th>Sequence of test versions</th>
<th>Test conditions</th>
<th>Group 1: Silence, Silence</th>
<th>Group 2: Speech, Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td></td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>BA</td>
<td></td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>22</td>
<td>23</td>
</tr>
</tbody>
</table>

3.5 Laboratory and work places

The laboratory was a large sound insulated laboratory (l: 8.4 m; w: 6.4 m; h: 4.2 m) with low reverberation times (<125 Hz: 0.2-0.26 sec.; >125 Hz: 0.1 +- 0.02 sec.). In that large laboratory two work places for the participants were set up. Each work place consisted of a desk, an office chair, a laptop computer, headphones, pens and a special pen for use on the touchscreen. The desk was of comfortable size and both the desk and the chair were adjustable in height, so that it was possible to find the most comfortable position for each participant. The two work places were opposite of each other, but separated by a large non-transparent screen. When two persons participated simultaneously they could not observe each other during task processing.

3.6 The test runs

The investigators conducted the experimental sessions along a written manual. Therefore a high amount of standardisation was realised. At the beginning of the first test run, before starting data collection, all potential attendees received written and verbal information about the study and data protection, and they gave informed consent to participate in the study. All participants were then asked to fill in a short questionnaire with some questions on personal data, current occupation, experience with participation in other experimental studies, experiences with working on a computer screen, hearing status and the general feeling of disturbance by noise. In addition to the reading task each test run also included some questions on mood, tiredness, effort, disturbance, self-assessed performance and concentration. The investigator was present during the entire test sessions, but sitting quietly in approx. 3 m distance to the experimental work places while the participants worked on the reading task.

4. Results

Analyses were carried out for the number of “correctly finished items”. At this stage for each item it was distinguished between the outcome “the item was completely finished correctly” and “there was some mistake”, regardless of which kind the mistake was. Therefore, the possible range for the results from one test run was between 0 (no item finished correctly) and 52 (all items finished correctly).

4.1 Ranges of correctly finished items

The ranges of the number of correctly finished items are as follows:

- Entire group, N=45: 16 to 49 items
- Group 1 (silence, silence), n=22: 23 to 49 items
- Group 2 (speech, speech), n=23: 16 to 48 items
The results in the group working on the task twice with speech as a background sound had a slightly larger range than in the group working twice in silence.

### 4.2 Ranges of the differences between the first and the second test run

The range of the differences between both test runs was calculated by the number of correctly finished items in the second run (T2) minus the number of correctly finished items in the first run (T1). The ranges were:
- Entire group, N=45: -5 to 13 items
- Group 1 (silence, silence), n=22: -5 to 10 items
- Group 2 (speech, speech), n=23: -4 to 13 items

### 4.3 Comparison of the results in the first test run (T1) and the second test run (T2)

The statistical analysis (paired t-test) revealed a significant difference between the results ascertained in both test runs ($t(44) = -4.294, p < 0.05$); Mean ($M_{T1} = 32.93; M_{T2} = 35.89$). The same analysis was carried out for both test conditions (group 1 and group 2) separately. The following results were found: group 1 $t(21) = -2.419, \ p = 0.025$; $M_{T1} = 34.73; M_{T2} = 37.09$; group 2 $t(22) = -3.609, \ p = 0.002$; $M_{T1} = 31.22; M_{T2} = 34.74$). That means participants finished a significantly larger number of items correctly in the second test run than in the first run and this finding holds for both groups (see Figure 1, results are illustrated as percentage of correctly finished items). The difference between mean values is slightly larger in group 2 than in group 1.

![Figure 1: Mean percentages of correctly finished items in the first and the second test run, separately shown for the test conditions "silence, silence" and "speech, speech" (* = the difference is significant). The total numbers of correctly finished items are also given in brackets in each column.](image)

### 4.4 Comparison of the results measured with different test versions (A and B)

A paired t-test considering all participants (N=45) revealed a significant difference between the results ascertained with test version A and test version B $t(44) = 2.262, \ p = 0.029$. A higher number of items was finished correctly in test version A than in test version B ($M_A = 35.29; M_B = 33.53$). The difference in the mean number of correctly finished items between both test versions is less than two items. Analysis carried out for each test condition (group 1 and group 2) separately revealed the following results: group 1 $t(21) = 0.581, \ p = 0.568$; $M_A = 36.23; M_B = 35.59$; group 2 $t(22) = 2.632, \ p = 0.015$; $M_A = 34.39; M_B = 31.57$. These results show that the significant difference is only apparent in
the test condition “speech, speech”. No significant difference between the results ascertained with different test versions was found in the group “silence, silence”. Both participants who might have had problems with the instruction (see 3.1) participated in the test condition “speech, speech”. A paired t-test that was carried out without these two participants did not show a significant difference between the results for both test versions ([t(20) = 2.064, p = 0.052]; $M_A = 33.76; M_B = 31.62$).

5. Discussion and outlook

The results on the number of correctly finished items showed that also the revised versions of the test can cover a wide range of individually different performance in proofreading. The finding that no participant detected all mistakes and also no participant detected none of the mistakes means that the task all in all is of an adequate difficulty for the proposed group of participants.

The finding that some participants missed to follow particular aspects of the instruction while working on the task and had therefore finally to be excluded from further analysis is a new observation and needs further attention. The instruction was standardised and the training runs were carried out with the same number of items for all participants. The initial training run without time limit was introduced as a new element in order to give the chance for the participants to get familiar with the test procedure with help of the investigator, before starting the training run with time limit. That means, many aspects were taken into account to arrange an optimal preparation phase. Nevertheless it seems useful to implement further modifications during this preparation phase in order to achieve a similar status of familiarity with the procedure. In this context all aspects that might be relevant will be taken into account whether they are associated with the problems explicitly mentioned by the participants and also possibly with problems that were not mentioned but become obvious in a more detailed analysis, e.g. when the results for the first five or ten items are inspected in each run. This will provide information whether a substantial number of participants would have needed more training items or more time to get familiar.

As a long-term solution it will be taken into consideration to apply an adaptive training run that ends, as soon as a defined number of items could be finished correctly.

The findings with respect to the first and the second test run confirm the results from the pilot study, and also the magnitude of the mean training effect is similar. On average participants performed better in the second run than in the first run. This effect was found in both test conditions (group 1 and group 2). This effect might be reduced by a more extensive preparation phase as described above, or by a longer interval between both test runs. This should of course be reflected on for the layout of future measurements in the laboratory. However, the results from the current study are still valuable, because it is important to know that there is a training effect when the task is applied twice, and also how big this effect approximately is, in order to include this knowledge about the task when results regarding effects of noise are interpreted. Figure 1 also suggests a detrimental effect of the background speech on the overall performance in the reading task. The current study, however, was not designed yet to analyse this effect. The focus was rather on the different test versions and the training effect between both test runs in the different acoustical conditions.

The analyses carried out with respect to the two different test versions revealed a significant difference between the results ascertained with version A and version B. Regarding the entire group, test version B turned out to be more difficult than test version A. This significant difference only holds for the test condition “speech, speech”. This result also needs further attention. It might be that version B includes items that are more “sensitive” to noise, so that differences in the difficulty become more obvious in the sound condition. This will be investigated by a detailed analysis on item level. As done for the results from the pilot study the difficulty for each item will be inspected as well as the kind of processing error.

In this contribution the first results of the study and the entire research project on noise effects on reading at the workplace were reported. Since the focus of this first empirical study in this project was
mainly on the task itself, the results cannot be used for practical advices for the work place yet. After the pilot study, this study was the next essential step on the way to treat this topic in a more systematic way. Further detailed analyses are planned in order to optimise the task and to optimise the entire procedure with the aim to make the task useful for the next empirical study, in which the effects of different typical background sounds on reading at the work place will be investigated.

6. Acknowledgments

Many thanks to all volunteers for their cooperation, participation and their interest in this study and to the employers who allowed to call the attention of their employees for this study by our flyers or notices. Also many thanks to all colleagues of the laboratory unit being involved in the project, mainly to Ulrich Hold for programming the computer-based reading test procedure and for arranging the technical setup in the laboratory as well as to Ilka Arendt for her support on organisational aspects and data collection. Very helpful support was also given by the colleagues in the information centre of the BAuA who scheduled the participants, and by our student assistant Nina Ahrweiler who worked on the data entry and data transfer.

REFERENCES


