Noise effects on reading and attention: Investigating the role of the chosen test procedure

H. Sukowski[1], S. van de Par[2]

University of Oldenburg, Acoustics Group, 26111 Oldenburg, GERMANY
[1]helga.sukowski@uni-oldenburg.de, [2]steven.van.de.par@uni-oldenburg.de

ABSTRACT

In order to quantify the influence of noise on human cognitive performance different kinds of tasks can be used to measure effects on a specific skill. Some of the tests usually employed are diagnostic tests which were originally developed to describe the current state of a person, e.g. with respect to their reading performance, arithmetic performance, their memory capacity or their attention. One reason why different studies sometimes show different results concerning the effects of noise might also be related to the fact that the chosen methods have different "noise sensitivity".

In the current study we investigate the effects of speech noise on a reading task and on two different attention tests. The reading task was primarily developed for children and could detect effects of different traffic noises on the performance in a previous study. Here it will be investigated if the test principles in this procedure are also suitable for studies on noise effects with adults. The two attention tests differ in their complexity. While one of them measures sustained attention based on finding visually presented identical patterns the other one is based on an arithmetic task which involves memory and decision demands. The tests are carried out in a group setting, and with some participants also in a single test setting. The results for various comparisons (between groups; between skills; between attention tests) are also shown and discussed in the light of the role a particular test procedure plays for the final results and conclusions drawn with respect to noise effects.

INTRODUCTION

In studies on effects of noise on cognitive performance it is one typical approach to investigate the effects on reading or attention. This holds for studies where the effects of *chronic* noise exposure are in the focus (e.g. Clark et al. 2006, Sanz et al. 1993) as well as for studies where the effects of *acute* noise exposure are under investigation (e.g. Banbury & Berry 1998, Martin et al. 1988). Both, reading and attention are important abilities for a successful task performance at school, in many professional settings, and also in many occasions in everyday life.

In studies on effects of *chronic* noise exposure the essential research question is, whether there are differences in a specific skill between people living in different areas with different noise exposure. Sometimes this approach is extended to the question, whether there is a change in the performance when noise exposure changes over time (Hygge et al. 2002), and in some studies a further aim is to compile dose-response curves which show reading performance in relation to the exposition to environmental noise (Green et al. 1982, Clark et al. 2006, 2013).

A respective national standardized test is usually employed to uncover noise effects in studies on chronic noise exposure. These tests might be different in the way how

the measurement of a specific skill is operationalized. However, it is common that these standardized tests come with normalized performance values for the particular skill. In addition to the differences between different noise-exposure groups it can therefore also be concluded whether the performance in general is within the expected normal range for a defined (age) group.

For studies that are specifically aimed at *chronic* noise exposure, it is not essential whether the test procedure can capture noise effects in an acute noise situation, because in these studies something like a long-term, “cumulative effect” on the respective skill is quantified, which can in principle be mediated by different factors.

The demand for a test procedure used in studies on *acute* noise effects, however, is different. When investigating effects of *acute* noise on a specific skill, the chosen task should of course really capture the respective ability in general. It is, however, also essential that the task is sufficiently sensitive to noise. “Noise sensitivity” here means that the task should be capable to uncover whether the presentation of sound affects the specific performance. In the ideal case the task can also distinguish between the effects of different kinds of experimental noise conditions. This demand is not directly fulfilled for every test procedure, because a conventional reading test or attention test is typically developed with the aim to be a diagnostic instrument, which is not automatically the same as being an instrument to measure effects of noise.

Reading is a complex skill. From various studies it is known that not all components which are typically used to characterize the reading performance (e.g. speed, accuracy, comprehension) are affected by acute noise in the same way (e.g. Weinstein 1974). Even for similar tasks like proof reading it was demonstrated that the detection of different types of experimental mistakes is affected by noise in a different way (Weinstein 1974, Liebl 2006). In summary this means, if a performance task only includes items that are largely unaffected by noise or if the task only covers specific components like e.g. working speed the results might suggest that the noise does not harm the specific ability, although detrimental effects on that skill in general are existent. With this background the current study was carried out with the aim to investigate the role of the specific test procedures with respect to their capability to uncover effects of noise. One reading task and two different attention tests were applied in a silent condition and in different speech sound conditions.

The employed reading task was originally developed several years ago by the first author (Sukowski 2009) for use in a large study on the effects of traffic noise on cognitive performance in children (see BMBF research network “Leiser Verkehr” (Low noise traffic), sub-project 2222 [www.fv-leiserverkehr.de]). This reading task is named OKiLeLe for *Oldenburger Kinder-Leise-Leseaufgabe*. While developing the reading task the aim was to create a procedure that includes the current knowledge about noise effects on reading, and that was appropriate in the given research context. The task was applied in different traffic noise conditions. It revealed differences in reading performance with respect to the specific sound conditions (Sukowski & Schick 2007, Sukowski 2009). Of course, the procedure needs to be tested in further sound conditions in order to be able to generalize the results from this specific study and to optimize the item set. But the overall conclusion of the previous study was that the new test procedure is in principle suitable for studies on noise effects in children.

The OKiLeLe task was now applied in a study with adult participants. The main aim was to evaluate whether the realized test principles in this procedure would in

general be appropriate to capture noise effects also in this group of more experienced readers. Based on these findings one additional, prospective aim is to create a modified procedure that is suitable for future studies with adults.

Concerning the attention tests our main aim was to compare the "noise sensitivity" of two different test procedures. Both attention tests applied in this study differ in the material the participants have to work on, the instruction, and also with respect to the very important aspect to what extent working memory is involved to work on the task correctly. With the background knowledge that effects on working memory are well documented the hypothesis concerning this part of the study was that the task including a stronger demand on working memory will be affected more by the experimental sound conditions than the task that implies less demand on working memory.

METHODS

The current explorative study was carried out in two parts, a group setting (n = 33; maximum of 8 persons in one group) and a single test setting (n = 10). Overall three different tests were applied, the described reading test and two different attention tests. In this contribution the results referring to the group setting are presented.

Participants

In this part of the study 11 men and 22 women participated (age: 21 to 37 years). All participants had German as their first language. For the reading test two participants were excluded from further analysis because they did not work on the task according to the instruction.

Room and time

The measurements were carried out in a typical small lecture room in the evening time between 18:30 h and 19:30 h. Eight desks were positioned in two rows so that there was a maximum of four participants in one row. The tests were allocated to two sessions one week apart. Each session lasted for about one hour.

Test procedures

(A) *Reading task* (Sukowski 2009). In this test the task for the participants is to find mistakes in written sentences and to mark the mistakes in a defined way. The procedure has two parallel versions with 48 items each (12 items on one page). The items consist of one or two sentences. 36 out of the 48 items contain one mistake, 12 have no mistake. Half of the mistakes are realized as grammatical and half as semantic mistakes. There was a time limit of 50 s per page. This instruction is different from the procedure used in the previous study with children, who had 10 minutes to work on the entire test.

(B) *Attention test: Frankfurter Aufmerksamkeitsinventar 2* (FAIR-2, Moosbrugger et al. 2011). In this attention test visually similar patterns have to be found and marked in a specific way. Participants had to work on two sheets with 16 lines each (20 symbols per line). The time limit was three minutes for each sheet. This test has two parallel versions.

(C) *Attention test: Konzentrations-Leistungstest (*KLT-R*,* Düker et al. 2001*)*: In this test the participants have to work on arithmetic exercises heavily involving short term memory. For each item, they are asked to perform two simple calculations separately and then add or subtract the two results. Whether they have to add or subtract depends on the comparison of the results from both calculations. That means, participants have to keep in mind the general instruction, and while they work on a

single item they have to remember two numbers, decide which rule to apply, and then carry out an additional calculation. The original version of this test consists of nine columns with 20 items each. In the current study it was shortened to four columns only. The time for each column is two minutes. This test procedure also has two parallel versions.

Sound conditions

In total there were three different sound conditions in this study, one control condition (A) and two experimental sound conditions (B and C). Since the strongest effects of noise on cognitive performance often are shown for speech sounds, and especially in the case of reading for *comprehensible* speech, two speech sounds in German language were chosen as experimental sound conditions.

(A) *Control condition*. This condition was fairly quiet. It was the “natural” background sound in a seminar room occupied by up to eight people during evening hours. No additional masking sound was used. In all test sessions there were no unexpected sound events inside or outside the room.

(B) *Speech sound 1 (noise-1)*. This sound was part of an audio drama in German language, with different speakers in different room conditions. The sound was presented at a sound level of 62 dB(A) Leq.

(C) *Speech sound 2 (noise-2).* This sound was a stream of announcements with phrases that are typical for a public address system in a long distance train or on a platform, e.g. information about possible and missed connecting trains. This sound was presented at 65 dB(A) Leq.

The speech sounds were presented via two loudspeakers which were positioned at the front of the room. The sound pressure levels listed above were measured at a position in the middle of the first row.

Test conditions - experimental design

The reading test was carried out in the control condition and the sound condition noise-1. The FAIR-2 procedure was carried out in all sound conditions. The KLT-R was done in the control condition and the sound condition noise-2. All participants worked on all named combinations of tasks and sounds. In total there were four different presentation orders concerning the kind of the test and the used test version (“test conditions”, TC). In all conditions, the reading test was carried out first in both sessions.

RESULTS

For all tests the effect on the *number of correctly completed items* (“correct”) was analysed. In addition, the *percent error value* (“percent error”) was considered for the reading task and the KLT-R.

Reading task

Analyses of variance with repeated measurements were carried out with the inner-subject factor “sound condition” (2 steps) and the between-subject factor “test condition” (4 steps). For the *number of correctly completed items* a significant main effect for the factor “sound condition” [$F(1/27) = 13.98$, $p < 0.05$] (Mean values: control: 35.61; noise-1: 33.23) and a significant interaction [$F(3/27) = 7.50$, $p < 0.05$] between both factors were revealed. No main effect for the factor “test condition” was observed. Post-hoc tests carried out for each test condition separately showed

significant effects in two out of the four test conditions, with higher *numbers of correctly completed items* in the control condition than in noise. In both conditions the participants worked first in noise and then in the control condition. For the *percent error value* a significant main effect for the factor "sound condition" was revealed [$F(1/27) = 5.69$, $p < 0.05$]. The mean values were higher in noise than in the control condition (see Figure 1). That means, the performance was better in the control condition. No significant effect for the factor "test condition" and no interaction between both factors were observed.

A further analysis of the *percent error values* also considering the factor "experimental mistake" (semantic, grammatical) showed: (1) It was easier to find the semantic mistakes than the grammatical mistakes. (2) A tendency effect for the difference between the results in both sound conditions was only given for the grammatical mistakes ($t(30) = -1.87$, $p = 0.072$).

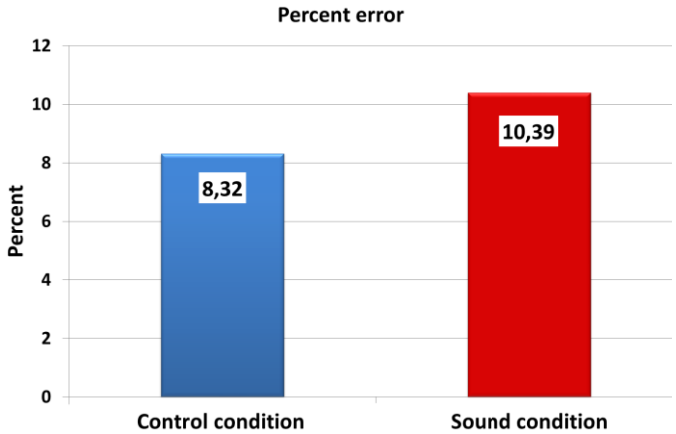

Figure 1: Percent error in the control and the sound condition in the reading test ($n = 31$).

Attention test FAIR-2

In this test on average a very low number of mistakes was observed. Therefore the analysis focused on the *number of correctly completed items*. In the analysis of variance with repeated measurements for the factor "sound condition" all three steps were taken into account.

The analysis revealed a significant main effect for the factor "sound condition" [$F(2/58) = 14.85$, $p < 0.05$], a tendency effect for the factor "test condition" [$F(3/29) = 2.29$, $p < 0.1$] and a significant interaction between both factors [$F(6/58) = 18.27$, $p < 0.05$]. The post-hoc test (pairwise comparisons) for "sound condition" showed significant differences between the results in the conditions "control" and "noise-2" as well as between the conditions "noise-1" and "noise-2". In both cases there were more *correctly completed items* in the sound condition "noise-2" than in the other conditions. The post-hoc test for "test condition" did not reveal significant differences between single test conditions.

Further analyses of variance carried out for each test condition separately showed significant main effects for the factor "sound condition" in each of the four groups (each: $p < 0.05$) but the pattern of the mean results varied: TC_01: control < noise-1 < noise-2; TC_02 and TC_04: noise-1 < control < noise-2; TC_03: noise-2 < control < noise-1.

Attention test KLT-R

For the KLT-R an analysis of variance like for the reading test was carried out. The analysis revealed a significant main effect for the "sound condition" on the *number of correctly completed items* [$F(1/29) = 12.65$, $p < 0.05$]. Participants achieved a larger number in the control condition than in the speech sound condition (mean values: control: 33.33; noise-2: 29.94). No significant effect was found for the factor "test condition".

Since a significant interaction between the factors "sound condition" and "test condition" was found [$F(3/29) = 9.15$, $p < 0.05$] post-hoc tests were carried out to investigate the effects of the sound for each test condition separately. These tests showed a significant difference between "control" and "noise-2" in those two conditions where participants worked first on the KLT-R under noise and then in the control condition (TC_01 and TC_04). In the control condition they achieved a higher *number of correctly completed items*. In both conditions with the reverse order no significant effect was observed (see Figure 2).

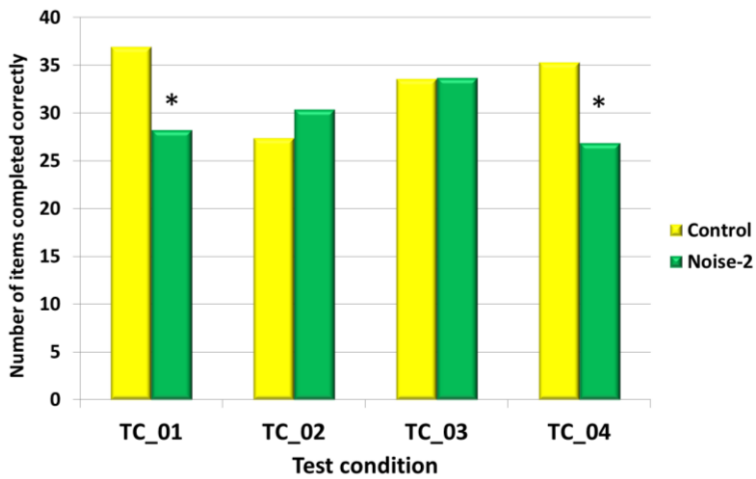

Figure 2: KLT-R: Number of correctly completed items in the control and in the noise-2-condition, separately for each test condition ($n = 33$ in total). Significant differences are marked with an asterisk.

For the variable *percent error* no main effect for "sound condition" or "test condition" was observed. Since a significant interaction was revealed also in this case, likewise post-hoc tests were carried out for each test condition. These tests revealed a tendency effect in one of the four test conditions ($p = 0.089$) between the results in the control and the speech sound condition, with a higher *percent error value* in the control condition (control: 13.77%; noise-2: 9.19%). In this respective test condition the participants worked first on the KLT-R in the control condition and then under noise-2.

SUMMARY AND DISCUSSION

In the current study one reading task and two different attention tests were applied in different sound conditions. In the reading task significant effects of speech noise on performance were revealed. In this noise condition detrimental effects on the *number of correctly completed items* and the *percent error value* were found. Since for the *correctly completed items* the sound effect was only significant when participants worked first in the noise condition and then in the control condition, it seems that there is a training effect from one run to the other. Presumably the training effect is added to the noise effect in these conditions, while in the other conditions with the

reversed order, the training effect reduced a possible noise effect to a non-significant difference. To test this assumption a larger number of participants needs to be tested and additional test conditions should be applied (two runs in a control condition and two runs in a sound condition). The variable *percent error* turned out to be sensitive to noise and to be largely unaffected by a training effect. This result is in line with the observation in the previous studies with children (Sukowski 2009). This variable is therefore very useful in order to detect and to quantify effects of noise in this reading task.

Although these results suggest that even the original version of this procedure can in principle be used in studies with adult participants to detect effects of noise, it is still an aim to create a similar procedure along the new results which is optimized for studies with adults. The results observed for semantic and grammatical mistakes separately, which revealed stronger effects on the detection of the grammatical mistakes, provide a good basis for an initial modification. The solution, however, will not be to remove all items with semantic mistakes and to keep only the items with grammatical mistakes, because the very mixture of different types of items makes the procedure “lively”. One possible alternative will be to realize semantic mistakes in more complex sentences and to add further kinds of mistakes. In addition, a more detailed analysis of the current data including the effects of sentence length will need to be carried out to get to more hints for the optimization of the procedure. All in all, the results from the current study emphasize that it is important to keep an eye on the concrete reading test procedure or, respectively, on the concrete test items employed, in order to avoid misleading interpretations of results concerning detrimental effects of noise.

According to the expectations the two different attention tests showed different “noise sensitivity”. The test procedure with a low demand on working memory (FAIR-2) showed a significant improvement in the *number of correctly completed items* in one of the noise conditions in comparison to the other conditions. In the test procedure including stronger demands on working memory (KLT-R), considering all participants a significant detrimental noise effect on the *number of correctly completed items* was observed.

As for the reading test this result in the KLT-R only holds for that test sequence when the participants worked first in the noise condition and then in the control condition. Therefore a similar explanation as mentioned above can be applied. In contrast to the reading task no main effect of the sound condition on the *percent error value* was observed for the KLT-R. That means, an effect on a variable which is largely unaffected by training is missing in this test. But as mentioned before, this test was applied in a shortened form in this study. It is possible that the suggested training effect on the *number of correctly completed items* would diminish when the test would be applied in its full length.

Concerning the FAIR-2 procedure the results should be treated very carefully. It seems that this test procedure was not an ideal instrument for use in our study. This preliminary conclusion is not drawn based on the finding that the procedure did not show detrimental noise effects, but rather on the observation that there was apparently no optimal matching of test procedure and experimental design. There are also problems due to the constraints given by the procedure itself. The variations in the results revealed for the four different test conditions indicate that there might be further factors influencing the results. One further factor might be the double use of one test version. As mentioned above, the test has two different versions. Since it

was carried out three times in this study, participants consequently had to work twice on one version in our design. A further problem is that there was a very low total number of mistakes so that the mistakes itself do not provide a sufficient basis for a detailed analysis. Finally, a further problem was that in this study more participants finished the entire test than expected from the description in the test manual. Some participants finished up to 30 seconds before the given time limit was over. Therefore, it seems that this procedure was too easy for the current group of participants.

Although the results in the attention tests turned out to match the assumptions, some problems with the chosen attention test procedures became obvious in this study. Concerning some of the aspects further studies with a larger number of participants and an extended experimental design might help to clarify the addressed issues, but other problems will not diminish even when testing a larger group of participants. Therefore, if the employed attention tests are to be applied in further studies the mentioned problems need to be considered carefully, in order to actually learn as much as possible about noise effects on humans.

Acknowledgements

Many thanks to all who participated in this study at that late time of day, and to Annika Urbschat, who contributed greatly to this study as a student research assistant.

REFERENCES

Banbury S, Berry DC (1998). Disruption of office related tasks by speech and office noise. British Journal of Psychology: Applied 89: 499-517.

Clark C, Martin R, van Kempen E, Alfred T, Head J, Davis HW, Haines MM, Barrio IL, Matheson M, Stansfeld SA (2006). Exposure-effect relations between aircraft and road traffic noise exposure at school and reading comprehension. American Journal of Epidemiology 163(1): 27-37.

Clark C, Head J, Stansfeld SA (2013). Longitudinal effects of aircraft noise exposure on children's health and cognition: a six-year follow-up of the UK RANCH cohort. Journal of Environmental Psychology 35: 1-9.

Düker H, Lienert GA, Lukesch H, Mayrhofer S (2001). Konzentrations-Leistungs-Test: KLT-R. Revidierte Fassung. Göttingen: Hogrefe.

Green K, Pasternack B, Shore R (1982). Effects of aircraft noise on reading ability in school-aged children. Archives of Environmental Health 37: 24-31.

Hygge S, Evans GW, Bullinger M (2002). A prospective study of some effects of aircraft noise on cognitive performance in schoolchildren. Psychological Science 13: 469-474.

Liebl A (2006). Auswirkungen von Hintergrundschall auf das Lesen und Verstehen von Texten. Dissertation, Katholische Universität Eichstätt-Ingolstadt.

Martin RC, Wogalter MS, Forlano JG (1988). Reading comprehension in the presence of unattended speech and music. Journal of Memory and Language 27: 382-398.

Moosbrugger H, Oehlschlägel J, Steinwascher M (2011). Frankfurter Aufmerksamkeits-Inventar 2 (FAIR-2). Bern: Huber.

Sanz SA, García AM, García A (1993). Road traffic noise around schools: a risk for pupil's performance? International Archives of Occupational and Environmental Health 65: 205-207.

Sukowski H (2009). Untersuchungen zur Lärmwirkung auf die Leseleistung von Grundschulkindern. Dissertation, Universität Oldenburg.

Sukowski H, Schick A (2007). Influence of traffic sound parameters on the reading performance in elementary school children. Proc. 14th Intern. Congress on sound and vibration - ICSV14, Cairns, Australien, July 2007.

Weinstein ND (1974). Effect of noise on intellectual performance. Journal of Applied Psychology 59: 548-554.