PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Remembering landmarks in a virtual maze: Does the disturbance impact of background speech depend on the spatial information inherent in the speech signal?

Sabine J. SCHLITTMEIER[1]; Edina FINTOR[2]

[1, 2] RWTH Aachen University, Germany

ABSTRACT

Whether a particular background sound impairs cognitive task performance depends on both sound and task characteristics. Although this so-called interference-by-process principle (e.g., 1) is widely accepted, it has not yet been fully empirically tested. In recent years, research on cognitive noise effects has focused on the detrimental influence of background speech's intelligibility, semantics, and temporal-spectral variability on performance in verbal tasks (e.g., verbal short-term memory, reading, writing). For visual-spatial tasks, however, the spatial characteristics of background noise should induce a performance decrement.

We present a series of experiments in which short-term memory for visual-spatial information was explored during background speech conditions in which a talker's location either varied or remained stable. Considering wayfinding and orientation as everyday visual-spatial tasks, we set up virtual mazes with landmarks (e.g., a bakery, a park). Participants had to find their way through these virtual mazes and afterward recall the spatial position of the landmarks. The effects of the different background speech conditions – including or excluding varied spatial information – on visual-spatial memory is discussed in terms of cognitive psychological implications as well as applied contexts.

Keywords: binaural sound, visual-spatial cognition, sound effects

1. INTRODUCTION

Background sounds do not impair cognitive performance per se. Whether a given sound disturbs cognitive performance depends on a kind of "fit" between background sound and task characteristics.

A performance decrement occurs when volitional task processing and the obligatory pre-attentive processing of background sound burden the same cognitive processes or functions. This so-called "interference-by-process" hypothesis (e.g., 1) has been tested in recent years predominantly with respect to verbal tasks and the influence of background speech. Here, for example, background speech's semantic content has been shown to be pivotal in impairing performance on tasks that rely on semantic processing, e.g., reading comprehension or text production by writing (2, 3). With respect to visual-spatial tasks, it seems reasonable to assume that the automated processing of spatial information inherent in a background sound signal (e.g., overheard speech produced by a conversation of several talkers in the acoustic background of an open-plan office) affects performance on tasks that rely on the voluntary processing of spatial information (e.g., construction task, design tasks). The present studies aimed to test this assumption considering performance on a visual-spatial task. Using wayfinding and orientation as everyday examples of such tasks, we set up virtual mazes with visual landmarks (e.g., a bakery, a park). Participants had to navigate through these spatial configurations and afterward recall the spatial positions of the landmarks. Our studies aimed to test whether background speech that incorporates varied spatial information (i.e., the speaker´s location changed

[1] Sabine.Schlittmeier@psych.rwth-aachen.de

[2] Edina.Fintor@psych.rwth-aachen.de

after each sentence) disturbs visual-spatial recall performance more compared to speech with constant spatial information (i.e., the speaker remained at one position).

In fact, little evidence on visual-spatial cognitive performances with respect to background sound effects is available, and the existing evidence is, furthermore, not consistent. Jones et al. (4) and Kvetnaja (5) both used the so-called Corsi block task to measure visual-spatial short-term memory capacity. In this task, a series of 7-9 positions are presented on a screen and participants are asked to recall the positions by clicking the exact presentation order. Although Jones et al. (4; Exp. 4) found a detrimental effect of background speech on visual-spatial short-term memory, Kvetnaja (5) did not replicate this finding despite increased sample size (and thus increased statistical power).

It is important to note that in both studies (4, 5), the tested background-speech signal did not include varying spatial information. This is due to the fact that both studies aimed to test the so-called changing-state hypothesis (6), which can be considered a specification of the later proposed interference-by-process hypothesis. The changing-state hypothesis assumes that deliberate processing of task-relevant serial information (here serial recall of visual-spatial positions) interferes with the serial processing of the auditory-perceptive tokens of the background sounds. Serial information is assumed to be inherent in background sound, consisting of distinct auditory-perceptual tokens varying successively, as is typical for background speech. Do note, however, that in contrast to the sparse and contradictory empirical results with respect to visual-spatial short-term memory, the disturbance effect of background speech on *verbal* short-term memory (measured via serial recall of a series of digits, consonants, or words) is an empirically robust finding replicated in a large amount of studies (e.g., 7). Furthermore, for this verbal version of serial recall task, two published studies have shown that spatial information of background speech contributes to its disturbance effect; yet due to the experimental set-up, these empirical results have to be discussed as spatial release from masking (8, 9) rather than interference of cognitive processes.

The intent of the present studies was to test whether background speech containing varied spatial information affects visual-spatial cognitive performance. To this end, Study 1 was designed to use a novel wayfinding and landmark recall task. Participants worked on this complex visual-spatial task either in silence (baseline condition) or while hearing background speech with the background speaker´s location changing after each spoken sentence. Furthermore, the Corsi block task was used as a standard measure of visual-spatial short-term memory capacity. By doing so, the capacity of each participant could be related to his/her landmark recall performance, allowing us to statistically test whether performance on the novel maze task relies also on this basic cognitive function.

Study 2 tested whether background speech per se disturbs landmark recall performance or whether the varied spatial information in the background speech signal is the decisive characteristic of a performance decrement. For this purpose, participants worked on the wayfinding and landmark recall task in silence; in the presence of background speech with varying spatial information (i.e. the speaker`s position varied over time); and in the presence of a corresponding speech signal with constant spatial information (i.e., the speaker remained in one position).

2. STUDY 1

2.1 Methods

2.7.1 Participants

Twenty-eight students from the RWTH Aachen University (32 female) aged between 18 and 55 years (*Md* = 24.0 years) participated in Study 1. They responded to a notice calling for participation and reported normal hearing. All participants provided written informed consent. Participants received no incentives or credit points for their participation. All participants were able to differentiate and correctly indicate the five auralized talker positions in a localization test.

2.7.2 Apparatus and Stimuli

Mazes were created with the free Software Maze Suite (10, 11), which was also used to administer the navigation task. All mazes were of equal size and shared the following characteristics: dark floor, brick walls, and the sky above. A starting position and a goal region were defined in a non-trivial manner. Panel a) of Figure 1 displays a sample screenshot of a maze and Panel b) the corresponding map.

Eight landmarks were shown in each maze: a fruit stand, a cash dispenser (ATM), a petrol station, a bus stop, a church, a park, a playground, and a bakery. Reaching the goal of the maze required the participants to pass by all landmarks. Consequently, no landmarks were placed in dead ends to guarantee that a participant who reached the target point of the respective maze saw also all the pictures. For landmark recall, a sketch map of the maze – created from a bird's eye view – was given in which starting position (head symbol) and goal region (colored area) were given without indicating landmarks (cp. Panel b) of Figure 2). In several studies, visual-spatial short-term memory in a wayfinding tasks was already successfully measured using sketch maps (e.g., 12, 13).

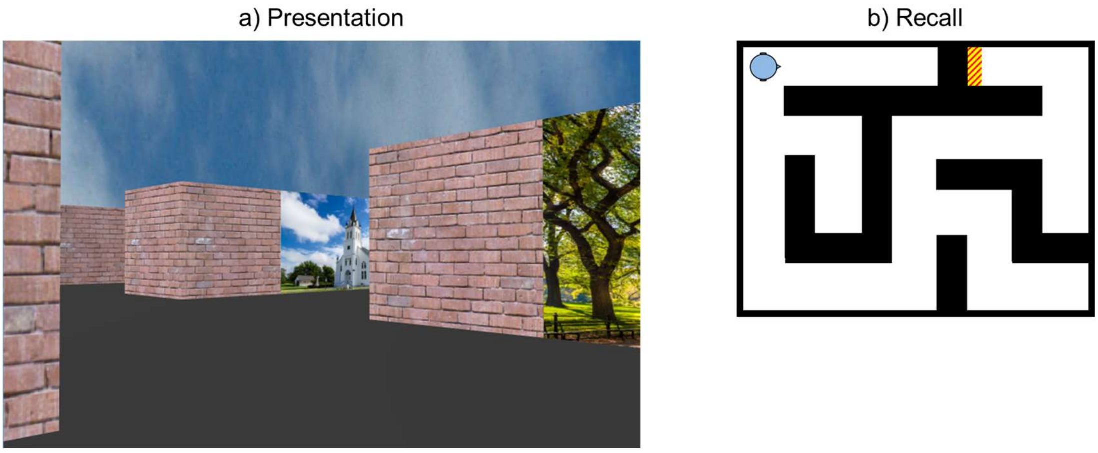

Figure 1 – Panel a) shows a sample screenshot of a maze while navigating through it during presentation period; panel b) shows a sketch map for landmark recall.

The background speech signal was derived from the OLSA (Oldenburger Satztest; Oldenburg Sentence Test; 14). The sentences spoken by a male speaker were processed to derive a binaural signal. Every single sentence was syntactically correct but semantically relatively meaningless, following the rule: subject + verb + numeral + adjective + noun, e.g., "Stefan wins twelve green flowers," "Britta sees eighteen beautiful rings." Sentences were auralized in such a manner that the speaker was heard from one out of five positions on a semicircle in front of the listener/participant (-90°, -45°, 0°, +45°, +90°) at a distance of 2 m (15). Successive sentences always differed randomly with respect to talker position. Pauses between sentences were 75 ms, 100 ms, 125 ms, 150 ms, 175 ms, or 200 ms long generated randomly. Altogether, 30 min background speech was generated with the material being repeated after 15 min.

The Corsi block task was conducted with the experimental software PsyScope X B57 (16) on an Apple MacBook Pro 2013 with a 15'' non-glare screen. In the presentation period, seven squares ("blocks") marking seven positions were displayed on the screen. An X was moving through all seven positions in random order. On the recall screen, all seven positions were marked with X's (Figure 2).

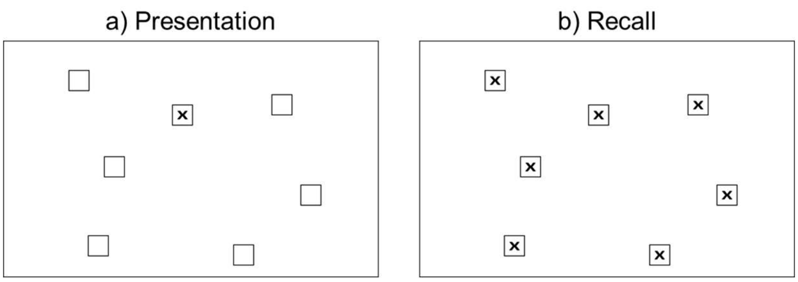

Figure 2 – Screenshots taken during presentation period (Panel a) and recall period (Panel b) of the Corsi block task

The Questionnaire on Spatial Strategies (Fragebogen räumlicher Strategien, FRS; 17; for an English version, cf. 18) was given to participants but the corresponding results are not subject of the present proceedings.

2.7.3 Procedure

Participants were tested in single sessions in a soundproof chamber at the Institute of Psychology (IfP) at the RWTH Aachen University. Testing took about 1.25 hours, and testing sessions were structured as follows: initial interview, localization test, maze task, questionnaire on spatial strategies (FRS), and Corsi block task.

In the initial interview, participants were questioned about individual vision and hearing status as well as experience with VR and gaming. Afterward, a localization test was administered to ensure that all participants were able to differentiate the five auralized talker positions. Therefore, a single sentence of the OLSA was played back twice (300 ms pause) from one position and the participant was asked to indicate the position in the figure, as shown in Panel b) of Figure 4. Each position was tested 5 times in random order.

Subsequently, the maze task started. The participant´s task was to navigate through a maze and to memorize the localization and the identity of the landmarks that he/she passed. After participants navigated through the maze, they were asked to mark the landmarks (position/wall section and identity indicated by a number) on a sketch map (landmark recall), as given in Panel b) Figure 1. Therefore, participants received a sheet of paper showing all landmarks with assigned numbers. First, a practice maze without recall was given to participants to allow them to become familiar with navigating through the maze and with the landmarks. Subsequently, the second practice maze was given, including recall with a sketch map, followed by the two experimental blocks, one block solved in silence, the other block with background speech, each encompassing four mazes. The sequence of these two sound conditions was counterbalanced across participants. Time to navigate through one maze was restricted to 2 min (time out) to avoid giving participants additional time to learn landmark positions.

After the maze task, the participants completed the questionnaire on spatial strategies (FRS) during a short break of several minutes. At the end of the test session, the Corsi block task was administered to measure individual visual-spatial short-term memory capacity. Here, two practice trials were followed by nine testing trials. A trial started with the rectangles decreasing in size before an X moved through all seven positions in random order. The participants` task was to remember the sequence of positions. After the X had passed through all seven positions, there was a short pause of 10 s (retention interval). Then the recall screen appeared (cp. Figure 2). The participant was asked to click the positions in exactly the same order in which the X moved through the positions. A block that was clicked disappeared from the screen. Correcting errors was not possible.

2.2 Results

One point was given for correctly indicating the spatial position of a landmark on a sketch map and one more point for correctly indicating the landmark's identity. Accordingly, a maximum of 16 points (4 mazes x 2 points) could be achieved in silence and 16 further points during background speech. Panel a) of Figure 3 depicts mean landmark recall performance during the two sound conditions. A t-test for dependent samples revealed a highly significant sound effect, $t(27) = 3.06$, $p < .01$, two-tailed, Cohen's $d = 0.41$. Thus, background speech of varying spatial characteristics significantly reduced wayfinding and landmark recall performance compared to the silent control condition.

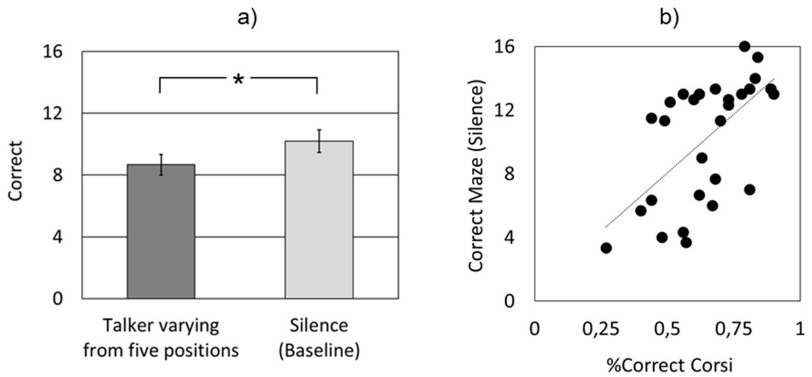

Figure 3 – Panel a) depicts mean recall performance for landmarks in the maze task with standard errors (n = 28), Panel b) shows a scatter plot of Corsi block performance by performance on the wayfinding and landmark recall task during silence (baseline condition).

In the Corsi block task, each block that was not recalled exactly in the serial position in which it was presented counted as an error. The mean %correct was assumed to be an indicator of individual visual-spatial short-term memory capacity. Panel b) of Figure 3 depicts the scatter plot for %correct in the Corsi block task by individual landmark recall performance during silence in the maze task. Both performance measures correlated significantly, $r_S = .61$, $p < .001$. This correlation indicates that performance on these two tasks may be administered by the same cognitive function, i.e., visual-spatial short-term memory (cp. also 19).

3. STUDY 2

In Study 1, background speech was used in which the talker´s position varied. Study 2 was conducted to test whether the varying spatial characteristics of the background speech signal significantly contribute to a performance decrement in the wayfinding and landmark recall task. Here, performance in the maze task was measured during three different background sound conditions: silence, background speech from a talker who remained in one position, and the same speech condition as in Study 1 with a talker speaking from varying positions.

3.1 Methods

3.1.1 Participants

32 students (24 female) of the RWTH Aachen University participated in Study 2. Participants were aged between 18 and 33 years (Md = 22.5 years) and had responded to a notice seeking participants. All participants provided written informed consent, reported normal hearing and received credit points for participation. All participants were able to differentiate and correctly indicate the five auralized talker positions in a localization test.

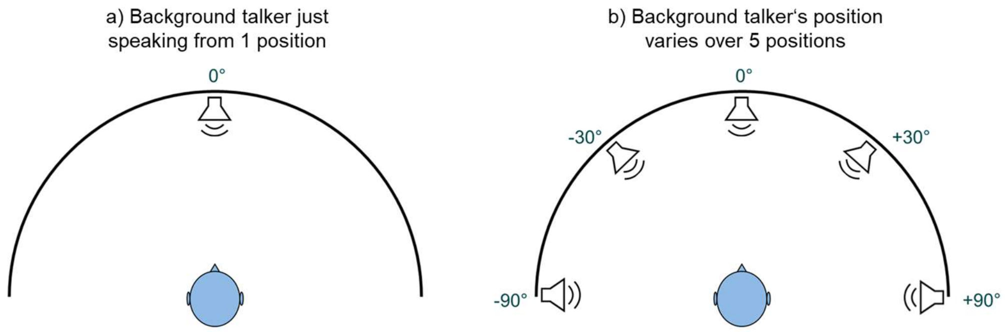

Figure 4 – Background speech was auralized either as being spoken by a talker remaining on one spatial position (Panel a) or with the talker´s position varying over 5 different spatial positions (Panel b).

3.1.2 Apparatus and Stimuli

The same experimental set-up was used as in Study 1 with the following two exceptions:
First, wayfinding and landmark recall performance was measured during three sound conditions: silence and two irrelevant background speech conditions. One speech signal involved a talker speaking successively from five spatially different positions, as in Study 1. The second speech signal was derived by binaurally auralizing the sentences spoken by the OLSA's male speaker from one spatial position in front of the listener (i.e., the participant) (cp. Panel a) of Figure 4). Second, the FRS questionnaire and the Corsi block task were not administered in Study 2.

3.1.3 Procedure

The procedure was the same as in Study 1 with the following two exceptions. First, three experimental blocks, each consisting of 4 mazes, were administered due to the testing of three sound conditions: silence, speaker located in one position, speaker switching across five positions. The sequence of sound conditions was counterbalanced over participants. Second, the FRS questionnaire, and the Corsi block task were not administered in Study 2.

3.2 Results

Landmark recall performance was analyzed as in Study 1: One point was given for each correct position of a landmark on a sketch map (i.e. certain position on a certain wall), and a further point for the correctly identity of the landmark (e.g. a bakery). Thus, 16 points (4 mazes x 2 points) could be achieved at maximum in each of the three sound conditions. Figure 5 visualizes group means of recall performance during the different background sound conditions.

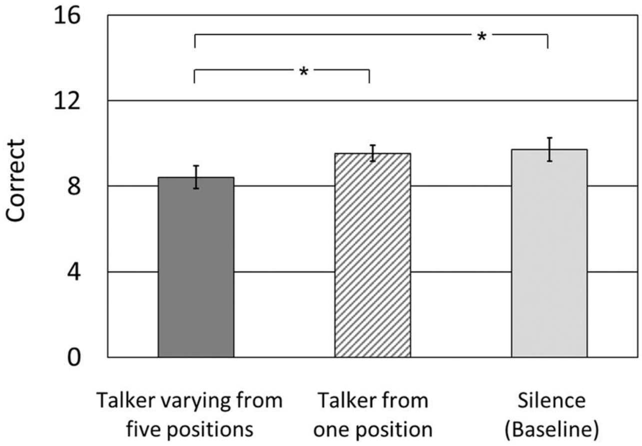

Figure 5 – Mean landmark recall performance with standard errors in Study 2 (n=32)

An ANOVA (analysis of variance) on the within-subject factor *sound* (talker from five positions, talker from one position, silence) verifies a significant effect on this factor, $F(2,62)= 3.35$, $p = .04$, $\eta = .10$. Paired t-tests reveal that this significant sound effect is due to significantly enhanced error rates during background speech with varied spatial information (i.e. talker successively speaking from five different positions) compared to silence (baseline condition), $t(31) = 2.03$, $p = .05$, two-tailed, Cohen's $d = 0.43$, and also compared to background speech with constant spatial information (i.e. talker speaks always from one position), $t(31) = 2.38$, $p = .02$, two-tailed, Cohen's $d = 0.43$. Performance during the latter two conditions does not differ, $t(31) = 0.35$, $p = .73$, two-tailed, Cohen's $d = 0.07$. Thus, background speech inheriting varying spatial characteristics significantly reduced wayfinding and landmark recall performance compared to a silent control condition but also compared to background speech with constant spatial information.

4. DISCUSSION AND CONCLUSIONS

The research intent of the present study was to test whether background speech containing variable spatial information affects visual-spatial cognitive performance. Therefore, a novel wayfinding and landmark recall task was used. Study 1 verified that background speech with varying spatial characteristics (i.e., with the speaker´s location changing after each spoken sentence) significantly disturbs wayfinding and landmark recall performance compared to a silent control condition. This finding was replicated and extended in Study 2. Study 2 also applied the wayfinding and landmark recall tasks but tested whether background speech in and of itself disturbs landmark recall performance or whether the varying spatial information in the background speech signal is necessary for a performance decrement to occur. In this study, background speech with varying spatial information (i.e., the speaker's location changed after each sentence) exclusively and significantly reduced landmark recall performance, whereas a speech signal with constant spatial information (i.e., the speaker remained at one position) did not affect complex visual-spatial cognition. Conclusively, Study 2 replicated Study 1 regarding the disturbance effect of a speech signal with distinct and spatial varying characteristics and extended the previous findings by verifying that the spatial characteristics – and not the speech nature of the background signal – are necessary for disturbance to occur.

From the cognitive psychological perspective, the two results of the present studies are of special importance. First, individual performance in the wayfinding and landmark recall task correlated significantly with individual performance in the Corsi block task (Study 1). Since the latter is considered a standard measure of visual-spatial short-term memory capacity, it can be assumed that performance on the complex visual-spatial task of wayfinding and landmark recall also relies, among others, on this basic cognitive function. Thus, the present study might also clarify the existing contradictory findings: Whereas Jones et al. (1995) reported a significant disturbance effect of background speech with constant spatial characteristic on visual-spatial short-term memory performance, Kvetnaja (2018) was not able to replicate this finding despite increased sample size. The results of the present study support Kvetnaja's results and her interpretation of Jones et al.'s finding as a random effect, which became statistically significant by chance and due to an alpha error in statistical testing, respectively.

Second, the detrimental effect of background speech derived from a spatially varying talker supports automatic and obligatory processing of spatial audio information even if the background speech is irrelevant to the task at hand and should be ignored. It seems reasonable to assume that the automated processing of spatial information inherent in the background speech signal (namely the speaker's location) affects performance on tasks that rely on the voluntary processing of spatial information. Future research might further test this assumption using other tasks that rely on the cognitive processing of spatial information (e.g., construction task, design tasks). With this, our results are interesting, and they can potentially stimulate further research from an applied perspective. Background sound in daily life is as a standard binaural sound. And as such, it is characterized by a defined spatial characteristic. For example, overheard speech produced by several talkers in the acoustic background of an open-plan office has per se a defined spatial characteristic, which, however, has only sparsely been considered in studies on detrimental effects of background speech (but cp. Jones & Macken, 1995; Renz et al., 2018).

Finally, the fact that the effect of a given background sound on cognitive performance depends on both sound and task characteristics poses a particular challenge to research. On the one hand, basic research using highly controlled stimulus and task characteristics is necessary to reveal the basic cognitive functions and processes underlying task performance, sound processing, and a potential cross-talk between the two. On the other hand, more complex tasks and more plausible background sounds are desired to validate the obtained empirical results in real life situations. Audiovisual and interactive virtual reality (VR) technologies might be able to provide a tool for bridging between basic and applied research questions and corresponding experimental set-ups. For example, in an interactive audiovisual VR, the present navigation task and background talkers from spatially fixed as well as moving positions ("listening while walking to an also moving talker") might be further tested and evaluated.

Recent theories of auditory cognition and extant empirical findings were predominantly derived from studies with "simple" audio reproduction techniques (such as mono/dichotic representation without spatial cues). However, research projects which investigate the extent to which these empirical findings and the corresponding cognitive psychological models (e.g., on attention, short-term memory, communication, and scene analysis) hold true and/or can be validated, modified, or extended in close(r)-to-real-life audiovisual virtual environments are now realizable. Hence, research that will benefit from or require the use of interactive VR in experiments on auditory cognition can be identified.

ACKNOWLEDGMENTS

The authors would like to thank the M.Sc. psychology students Cornelia Plein, Özge Benk, Elena Halstenberg from the RWTH Aachen University for data collection in Study 1 and the B.Sc. psychology student Lena Hintzen for starting with data collection in Study 2. This work was financially supported by the Head Genuit Stiftung. We are also grateful to Lukas Aspöck and Florian Pausch from the Institut of Technical Acoustics (ITA), RWTH Aachen University, for assisting us in deriving the auralized binaural speech signal.

REFERENCES

1. Marsh JE, Hughes RW, Jones DM. Interference by process, not content, determines semantic auditory distraction. Cognition. 2009; 110:23–38. doi: 10.1016/j.cognition.2008.08.003.
2. Vasilev MR, Kirkby JA, Angele B. Auditory Distraction During Reading: A Bayesian Meta-Analysis of a Continuing Controversy. Perspect Psychol Sci. 2018 A.D.; 13:567–597. doi: 10.1177/1745691617747398.
3. Sörqvist P, Nöstl A, Halin N. Disruption of writing processes by the semanticity of background speech. Scand J Psychol. 2012 A.D.; 53:97–102. doi: 10.1111/j.1467-9450.2011.00936.x.

4. Jones D, Farrand P, Stuart G, Morris N. Functional equivalence of verbal and spatial information in serial short-term memory. J Exp Psychol Learn Mem Cogn. 1995; 21:1008–1018.
5. Kvetnaya T. Registered Replication Report: Testing disruptive effects of irrelevant speech on visual-spatial working memory. Journal of European Psychology Students. 2018; 2018:10–15.
6. Jones D, Madden C, Miles C. Privileged access by irrelevant speech to short-term memory: The role of changing state. Q J Exp Psychol A. 1992 A.D.; 44:645–669. doi: 10.1080/14640749208401304.
7. Schlittmeier SJ, Weissgerber T, Kerber S, Fastl H, Hellbruck J. Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. Atten Percept Psychophys. 2012 A.D.; 74:194–203. doi: 10.3758/s13414-011-0230-7.
8. Jones DM, Macken WJ. Auditory babble and cognitive efficiency: Role of number of voices and their location. Journal of Experimental Psychology: Applied. 1995; 1:216–226. doi: 10.1037/1076-898X.1.3.216.
9. Renz T, Leistner P, Liebl A. The effect of spatial separation of sound masking and distracting speech sounds on working memory performance and annoyance. Acta Acustica united with Acustica. 2018; 104:611–622. doi: 10.3813/AAA.919201.
10. Ayaz H, Allen SL, Platek SM, Onaral B. Maze Suite 1.0: A complete set of tools to prepare, present, and analyze navigational and spatial cognitive neuroscience experiments. Behav Res. 2008; 40:353–359. doi: 10.3758/BRM.40.1.353.
11. Ayaz H, Shewokis PA, Curtin A, Izzetoglu M, Izzetoglu K, Onaral B. Using MazeSuite and functional near infrared spectroscopy to study learning in spatial navigation. J Vis Exp. Epub ahead of print. doi: 10.3791/3443.
12. Garden S, Cornoldi C, Logie RH. Visuo-spatial working memory in navigation. Applied Cognitive Psychology. 2002; 16:35–50. doi: 10.1002/acp.746.
13. Knight MJ, Tlauka M. Map learning and working memory: Multimodal learning strategies. Q J Exp Psychol (Hove). 2018; 71:1405–1418. doi: 10.1080/17470218.2017.1326954.
14. HörTech (2011). OLSA – Oldenburger Satztest. Adaptive Sprachaudiometrie mit Sätzen in Ruhe und im Störgeräusch. Oldenburg: HörTech.
15. Schröder D, Vorländer M. RAVEN: A real-time framework for the auralization of interactive virtual environments. Forum Acusticum; Dansk Akustik Selskab; European Acoustics Association; European Congress on Acoustics. Proceedings of Forum Acusticum 2011: 27 June - 01 July, Aalborg, Denmark. Madrid: Spanish Acoustical Society; 2011. [published in Acta acustica united with Acustica]
16. Cohen J, Macwhinney B, Flatt M, Provost J. PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. Behavior Research Methods, Instruments, & Computers. 1993; 25:257–271. doi: 10.3758/BF03204507.
17. Münzer S, Fehringer BCOF, Kühl T. Standardized norm data for three self-report scales on egocentric and allocentric environmental spatial strategies. Data Brief. 2016; 8:803–811. doi: 10.1016/j.dib.2016.06.039.
18. Münzer S, Hölscher C. Entwicklung und Validierung eines Fragebogens zu räumlichen Strategien. Diagnostica. 2011; 57:111–125. doi: 10.1026/0012-1924/a000040
19. Hund AM. Visuospatial working memory facilitates indoor wayfinding and direction giving. Journal of Environmental Psychology. 2016; 45:233–238. doi: 10.1016/j.jenvp.2016.01.008.