Euronoise 2018
Crete

A simple model to predict the cognitive performance in distracting background speech

Tobias Renz
Institute for Acoustics and Building Physics, University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany.

Philip Leistner
Fraunhofer Institute for Building Physics, Nobelstrasse 12, 70569 Stuttgart, Germany.

Andreas Liebl
Fraunhofer Institute for Building Physics, Nobelstrasse 12, 70569 Stuttgart, Germany.

Summary
Background speech in open-plan offices causes frequently high distraction and dissatisfaction. Particularly the short-term memory is impaired by irrelevant speech sounds. A good acoustical design is required to achieve acceptable acoustical privacy in open-plan offices. In addition to sound absorbers and sound screens, sound masking is an efficient measure to control the background noise level and to mask disturbing speech sounds. The existing algorithmic approaches to estimate the performance impairment when people are subjected to irrelevant background speech require considerable computational effort. A simple model is presented to predict the short-term memory performance in distracting background speech. The results of various laboratory experiments with masked speech sounds at both, low and high speech intelligibility, were considered. The fitted model can be used to optimise the design of open-plan offices and to predict the effectiveness of a masking sound.

PACS no. 43.55.Hy, 43.66.Dc, 43.72.Dv, 43.50.Fe, 43.66.Ba

1. Introduction

As of today, the acoustical design of open-plan offices remains challenging because it requires sufficient treatment of two conflicting aims, maintaining a sound environment that enables conversations over short distances but providing acoustical privacy that enables workers to concentrate at work at the same time. By means of current room acoustic simulation tools acoustical consultants can predict the effect of different acoustical products in open-plan offices on various acoustical parameters. The international standard ISO 3382-3 [1] suggests the use of reverberation time, spatial decay of sound pressure level (SPL) and parameters that are based on Speech Transmission Index (STI) to evaluate the acoustical design of open-plan offices. Recently, Haapakangas et al. [2] have shown that STI correlates with noise disturbance in open-plan offices. More than ten years ago, Hongisto [3] presented a model that predicts the impact of distracting background speech on work performance.

It is well known that auditory background with changing-state features deteriorates working memory performance. This effect of auditory distraction is especially pronounced in memory for order of visually or auditorily presented digits. The auditory distraction due to irrelevant background noise is known as Irrelevant Sound Effect [4]. Performance decrements occur when the required cognitive processes conflict with those that are involved in processing the auditory stimuli, known as the interference-by-process principle [5, 6]. According to this explanation, any sound with sufficient temporal-spectral variability is expected to deteriorate the serial memory. Recent findings suggest that spectral fluctuations affect short-term memory more than level changes [7]. For an overview of recent studies in this field see e.g. [8].

Testing the mean error rates of subjects in a digit span task in laboratory conditions is a well-established method to determine the auditory distraction by background sounds. Also more complex tasks like proofreading [9, 10], reading comprehension [11, 12], and mental arithmetic [13, 14] are disrupted by irrelevant noise. Showing performance effects in field studies is challenging because either only self-reported performance in the current state is analysed (e.g. [15]) or the acoustical conditions need to be carefully varied while no other essential changes take place (e.g. varying the artificial masking sound [16]). The results of field studies that analyse the subjective responses of office workers before and after a renovation may be distorted by other effects than acoustical changes. Hence, it seemed reasonable to limit the further analysis of the effect of distracting background speech on working memory performance to results obtained in laboratory experiments at the first.

The mentioned model [3] predicts the decrease of performance (DP) due to speech of varying intelligibility. The model was based on three experimental studies that measured the performance in number recall and proofreading, respectively. Additionally, one field study was included that analysed the self-reported daily waste of working time before and after an office relocation. Since its introduction in 2005, this model has been applied to the acoustical design and evaluation of open-plan offices (e.g. [1]). Various researchers (e.g. [17, 18]) have reviewed this model coming to different conclusions. The first study [17] tested the performance in five tasks during sound conditions with three different STI values (0.10, 0.35, and 0.65). There were no significant performance differences between the two lowest STI values indicating that the cognitive performance may start to deteriorate at STI values higher than 0.2, as suggested by Hongisto [3]. The latter study [18] tested the performance in four tasks during six acoustic conditions and concluded that the steepest slope of overall performance decrease occurred between STI values of 0.23 and 0.34 while the differences between 0.34 and 0.71 were negligible. The results also indicate that the performance curve varies between different cognitive tasks. The results of these two studies are inconsistent suggesting that there are some uncertainties involved in the determination of STI and its correlation with cognitive performance. Limiting the model to results with the same experimental design (e.g. monaural speech recording with masking sound) and task (e.g. digit span) would have reduced the scope of application but may have provided more consistent results.

From a practical point of view, it is difficult to measure or simulate the STI because the occurring background noise levels during a usual workday are often unknown and the STI is not defined for situations with fluctuating background noise levels. When multiple persons speak at different locations, the distracting impact of background speech is sustained [19, 20]. However, Hongisto's model [3] cannot account for multi-talker environments. As the model was fitted to data that contained one talker under steady-state background noise, the STI was able to predict the speech intelligibility. One may assume that this DP model estimates the disturbing impact of speech intelligibility but not of temporal-spectral variability. The STI does not model temporal-spectral variability, and hence cannot account for aspects like different speech tempo and intonation or fluctuating background noise, for instance, when background babble is present. Since temporal-spectral variability is expected to correlate with speech intelligibility in sound conditions with one talker under steady-state background noise, the model correlated well with the DP.

In contrast to Finland's National Building Code that sets STI limits, German standards such as VDI 2569 [21] or DIN 45645-2 [22] avoid the use of the STI. The noise rating level L_r represents the statutory basis for the assessment of office noise in office buildings in Germany, Austria and Switzerland. The aim of this study was to extend the SPL measurements in an occupied office to a new metric that correlates with the work performance. Hence, this paper presents a new approach to predict the cognitive performance in distracting background speech. By using global level statistics the SPLs of fluctuating speech sounds and rather stationary background noise are estimated. Subtracting these two levels provides a ratio of the estimated SPL of fluctuating sounds to stationary sounds at the receiver position and may enable a DP prediction. Using global level statistics seemed appropriate because both, the signal-to-noise ratio (SNR) and temporal-spectral variability, can be considered to a certain extent. Contrary to Hongisto's model [3], this model does not require a loudspeaker to simulate a speech source, and hence it can be applied in occupied open-plan offices. It can also consider multi-talker sound environments. The overall sound condition is analysed at receiver points.

The model was fitted with data points that contained one speech sound similar to Hongisto's approach [3]. Only short-term memory performance was considered as response variable and the DP was normalised to the maximum DP during clear speech to compensate differences in the mean working memory capacity between different subject groups. Kaarlela-Tuomaala et al. [23] analysed similar level statistics in a study that compared the subjective perception of an acoustic environment during a relocation from private office rooms to an open-plan office and concluded that the variability of SPL was not related to the self-rated disturbance caused by noise. However, the study measured only two percentile levels $L_{A,1\%}$ and $L_{A,99\%}$ and did not distinguish between speech from the person working at the workplace and background speech from colleagues.

This paper provides first a description of the used studies and the procedure (see Section 2). In Section 3 six model fits with different percentile level metrics are compared. This is followed by a validation that analyses the prediction quality of the suggested models. In the concluding Section 4, the potentials and shortcomings of the presented model are outlined.

2. Methods and materials

2.1. Sound conditions

Five experimental studies were considered. The first dataset consisted of twelve sound conditions testing the effect of steady-state noise and reversed speech maskers at SNRs between –12 and –3 dB [24]. The second study was based on two laboratory experiments with twelve and ten sound conditions that evaluated the effect of babble as compared to stationary masking sounds at SNRs between –12 and –3 dB [25]. The third dataset included eight sound conditions with distracting speech at different SNRs of –6, –3 and 0 dB, as well as steady-state and variable speech-like noise [26]. The fourth study analysed the effect of two different ventilation sounds on cognitive performance at different SPLs between 25 and 45 dB(A) with and without distracting speech background, covering twelve sound conditions in total [27]. The last study analysed the effect of ten different masking sounds at –2.5 dB SNR on cognitive performance aside from two control conditions with silence and clear speech [28]. One sound condition was excluded from the further analysis because the participants were subjected to their favourite music, and hence the sound condition was different for each subject. The studies that are described in Refs. [26, 27] contained seven sound conditions with noise but without speech. These seven sound conditions were not considered. In total, 58 data points remained. The study in Ref. [26] contained a control condition with pink noise at 25 dB(A) SPL. This sound condition was included in the analysis. Table I provides an overview of the different studies.

Table I. Overview of the included studies.

Source	Primary comparison
[24]	Reversed speech and stationary noise
[25]	Babble and stationary sounds
[26]	Steady and variable speech-like noise
[27]	Different ventilation sounds
[28]	Various masking signals

2.2. Design and procedure

All experiments except from the experiment described in Ref. [26] were performed in the High Performance Indoor Environment Laboratory at Fraunhofer Institute for Building Physics with a volume of 132 m^3 under consistent room temperature, volumetric flow rate, relative humidity, and illuminance. The study in Ref. [26] was conducted at Catholic University of Eichstätt-Ingolstadt. All sound conditions were presented at both ears using Sennheiser HD 280 PRO or Sennheiser HD 600 headphones, respectively (Sennheiser electronics GmbH & Co. KG, Wedemark, Germany).

The experimental design of all experiments was a one-way repeated measures design with ten to twelve levels according to the tested sound conditions, i.e., all subjects performed the test during the same sound conditions. Silence and unmasked speech were included as control conditions, except from the study in Ref. [26] that used pink noise at 25 dB(A) as comparable control condition. All experiments tested 24 to 30 subjects. The participants received a small payment.

During each sound condition the subjects performed a serial recall task where subjects had to memorise a sequence of the numbers from 1 to 9 and recall it in the exact order of presentation after a retention interval of 8 to 10 s. Each digit that was not recalled correctly at the presented serial position was counted as an error. Subjective ratings were collected directly after the serial recall task by means of a questionnaire. Recall was carried out by clicking numbers in the same order on a 3 x 3 array on the screen. The percentage of incorrectly recalled digit positions of all tested sequences of one condition (mean error rate) was determined.

2.3. Predictor variables

Zuydervliet et al. [29] suggested that background levels can be described by the ninetieth percentile $L_{AF,90\%}$ and activity levels by the tenth percentile $L_{AF,10\%}$ while reducing the difference between them is the objective of sound masking. L'Espérance et al. [30] refined the proposed algorithm concluding that the difference between the tenth percentile $L_{AF,10\%}$ and ninety-ninth percentile $L_{AF,99\%}$ is an appropriate metric to evaluate the level of disturbing noise in offices.

On the basis of these proposed models the same percentile levels $L_{A,1\%}$, $L_{A,90\%}$, and $L_{A,99\%}$ were taken into consideration but the common time constants impulse, fast, and slow were considered. The values were calculated with the software ArtemiS version 12.05.1512 (HEAD acoustics GmbH, Herzogenrath, Germany). In binaural listening conditions the maximum value at both ears was used. Since the sounds were presented from frontal direction there were only minor differences between both ears present.

A first analysis showed that $L_{A,10\%}$–$L_{A,99\%}$ as well as $L_{A,10\%}$–$L_{A,90\%}$ are sensitive to SNR changes between −12 and 0 dB. Use of shorter time constants exhibited a higher sensitivity whereas SPL fast measured led to much higher level differences in sound conditions with clear speech as compared to SPL impulse or slow measured.

2.4. Model fitting

The measured mean error rates were referred to the silent baseline condition by subtracting the measured mean error rate during silence. In a second step, the DP values were normalised by dividing all values by the DP value that was measured at the control condition with clear speech to account for differences between the tested subject groups. The relationship between the respective predictor variable and the response variable DP was modelled by two models, a logistic function and a Boltzmann sigmoid function. The models were calculated in RStudio Version 1.1.383 (RStudio, Inc., Boston, MA, USA). The models were compared by using Spearman's rank correlation coefficient r_S as well as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [31].

2.5. Model validation

The measured DP values in short-term memory of the studies of Jahncke et al. [18], Ebissou et al. [32], and Ellermeier and Hellbrück [33] were used to assess the prediction quality. The studies provided 21 data points. This data was not included in the first model fit to enable a validation whether the model can also be used for data obtained in different environments than used at Fraunhofer Institute for Building Physics. The DP of the study of Ebissou et al. [32] was referred to the values of subject group 1 that showed a distinct effect of the sound conditions on mean error rates in serial recall. The data of Experiment 2A in Ref. [33] was referred to noise alone instead of a silent sound condition while the data of Experiment 2B was referred to masked speech at –12 dB SNR that resulted in similar mean error rates than pink noise in Experiment 2A. The normalisation of the DP values of Experiment 2B was performed by dividing all DP values by the DP during masked speech at +4 dB SNR because unmasked speech was not included in this experiment and speech at +4 dB SNR produced similar mean error rates as clear speech in Experiment 2A.

3. Results

The model with Boltzmann sigmoid function with four parameters did not increase the prediction quality as compared to the model with logistic function with three parameters. Thus, the analysis was limited to the logistic model. Table II provides an overview of the quality of the fitted models. Based on r_S, AIC, and BIC, the predictor variables $L_{AS,10\%}$–$L_{AS,90\%}$, $L_{AF,10\%}$–$L_{AF,90\%}$, and $L_{AI,10\%}$–$L_{AI,99\%}$ were favoured. $L_{AS,10\%}$–$L_{AS,90\%}$ was excluded from further analysis because the determined model had a very steep slope within a small range of values, and hence it did not seem to result in a robust prediction model.

The respective models are depicted in Figures 1 and 2. The graphs show the normalised DP values

Table II. Quality of the fitted models.

Predictor variable	AIC	BIC	r_S
$L_{AS,10\%}$–$L_{AS,99\%}$	13	21	0.65
$L_{AS,10\%}$–$L_{AS,90\%}$	11	19	0.67
$L_{AF,10\%}$–$L_{AF,99\%}$	15	23	0.67
$L_{AF,10\%}$–$L_{AF,90\%}$	10	18	0.66
$L_{AI,10\%}$–$L_{AI,99\%}$	14	22	0.59
$L_{AI,10\%}$–$L_{AI,90\%}$	19	28	0.49

of the six considered laboratory experiments over the respective predictor variable. The curve shows the determined model. The followings Equations 1 and 2 describe the formula of these models

$$DP\left(L_{AF,10\%} - L_{AF,90\%}\right) = \frac{1.0}{1 + exp(2.9 - 1.4 \times (L_{AF,10\%} - L_{AF,90\%}))}, \quad (1)$$

$$DP\left(L_{AI,10\%} - L_{AI,99\%}\right) = \frac{1.0}{1 + exp(5.5 - 1.5 \times (L_{AI,10\%} - L_{AI,99\%}))}. \quad (2)$$

Figures 3 and 4 depict the results of the validation of both favoured models with the data of four additional experiments. The curves represent the fitted models that were determined with the first dataset. The mean squared error (MSE) between the data points and the respective model was used to assess the quality of both prediction models, $DP(L_{AF,10\%} - L_{AF,90\%})$ and $DP(L_{AI,10\%}-L_{AI,99\%})$. The MSE was 0.039 and 0.043, respectively.

4. Discussion

The presented models that are based on the predictor variables $L_{AF,10\%}$–$L_{AF,90\%}$ and $L_{AI,10\%}$–$L_{AI,99\%}$ can be used to estimate the resulting cognitive performance during an acoustic environment. According to the determined MSE values, the model $DP(L_{AF,10\%}$–$L_{AF,90\%})$ appears to perform better than $DP(L_{AI,10\%}$–$L_{AI,99\%})$. The suggested metrics are easy to measure in-situ. Since occupational safety and health in Germany requires the measurement of noise rating levels L_r during representative tasks as described by the standard DIN 45645-2 [22], the metrics could be determined as part of such measurements without additional cost.

In a subsequent step these models could be used for the acoustical design of open-plan offices when adequate information about the background noise is available (e.g. when a sound masking system is used).

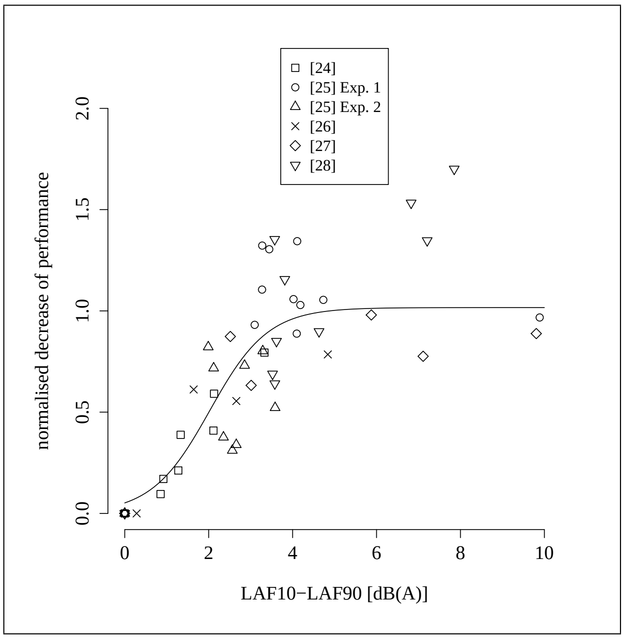

Figure 1. Normalised DP over $L_{AF,10\%}$–$L_{AF,90\%}$; the solid line shows the fitted model. Please note that some values of $L_{AF,10\%}$–$L_{AF,90\%}$ were above 10, and hence are not plotted in this figure.

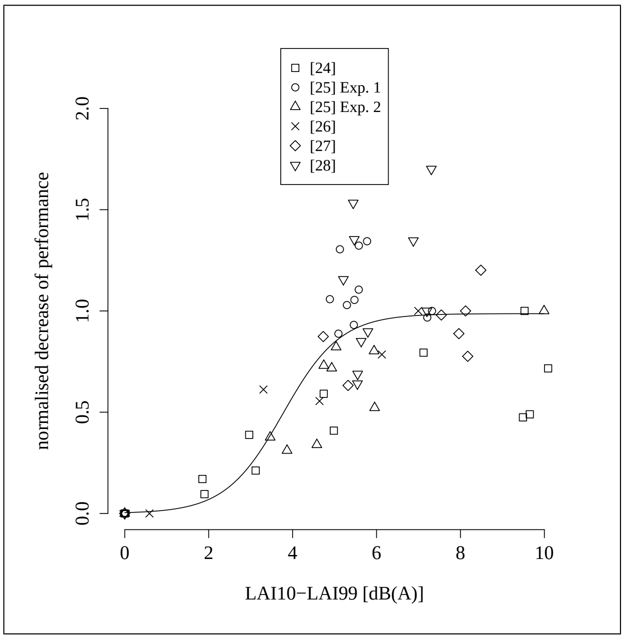

Figure 2. Normalised DP over $L_{AI,10\%}$–$L_{AI,99\%}$; the solid line shows the fitted model.

There are obvious shortcomings that may limit the applicability of the presented models. First, the considered data consisted of a single voice that was masked by various masking sounds. It remains unclear whether the models can be applied to more complex sound environments, e.g., with multiple talkers that may be spatially separated. Second, the predictor variables estimate an SPL ratio between speech and steady background noise but they take hardly

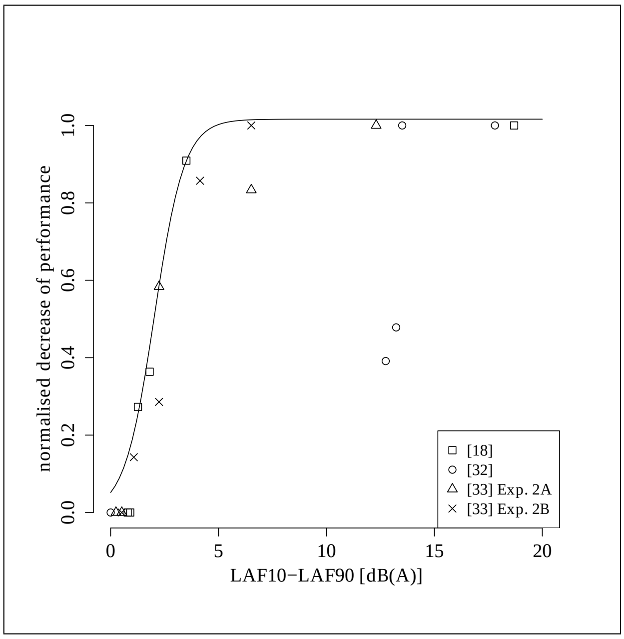

Figure 3. Validation with normalised DP over $L_{AF,10\%}$–$L_{AF,90\%}$; the solid line shows the respective model. The studies in Refs. [18, 32, 33] were considered for the validation.

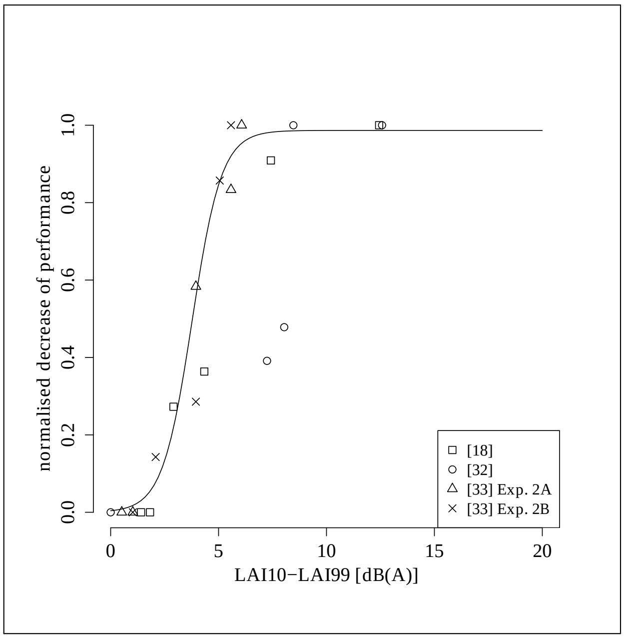

Figure 4. Validation with normalised DP over $L_{AI,10\%}$–$L_{AI,99\%}$; the solid line shows the respective model.

temporal-spectral variability into account. Further analysis is necessary to verify whether these models are suitable to asses work performance during irrelevant background speech in practical applications.

Acknowledgement

The authors would like to thank Helena Jahncke and Patrick Chevret for providing audio footage of former

research studies, which enabled a first validation of the model.

References

[1] ISO 3382-3:2012. Acoustics – Measurement of room acoustic parameters – Part 3: Open plan offices. International Organization for Standardization, Geneva, Switzerland, 2012.

[2] A. Haapakangas, V. Hongisto, M. Eerola, T. Kuusisto: Distraction distance and perceived disturbance by noise–An analysis of 21 open-plan offices. J. Acoust. Soc. Am. 141 (2017) 127-136.

[3] V. Hongisto: A model predicting the effect of speech of varying intelligibility on work performance. Indoor Air 15 (2005) 458-468.

[4] C. P. Beaman, D. M. Jones: Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. J. Exp. Psychol. Learn. Mem. Cogn. 23 (1997) 459-471.

[5] J. E. Marsh, D. M. Jones: Cross-modal distraction by background speech: What role for meaning? Noise Health 12 (2010) 210-216.

[6] D. M. Jones, S. Tremblay: Interference in memory by process or content? A reply to Neath (2000). Psychon. Bull. Rev. 7 (2000) 550-558.

[7] W. Ellermeier, K. Zimmer: The psychoacoustics of the irrelevant sound effect. Acoust. Sci. Technol. 35 (2014) 10-16.

[8] A. Liebl, H. Jahncke: Review of research on the effects of noise on cognitive performance 2014-2017. 12th ICBEN, Zurich, Switzerland, 2017

[9] N. Halin, J. E. Marsh, A. Haga, M. Holmgren, P. Sörqvist: Effects of speech on proofreading: Can task-engagement manipulations shield against distraction? J. Exp. Psychol. Appl. 20 (2014) 69-80.

[10] N. Venetjoki, A. Kaarlela-Tuomaala, E. Keskinen, V. Hongisto: The effect of speech and speech intelligibility on task performance. Ergonomics 49 (2006) 1068-1091.

[11] N. Halin, J. E. Marsh, A. Hellman, I. Hellström, P. Sörqvist: A shield against distraction. J. Appl. Res. Mem. Cogn. 3 (2014) 31-36.

[12] R. C. Martin, M. S. Wogalter, J. G. Forlano: Reading comprehension in the presence of unattended speech and music. J. Mem. Lang. 27 (1988) 382-398.

[13] S. J. Schlittmeier, J. Hellbrück, R. Thaden, M. Vorländer: The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance. Ergonomics 51 (2008) 719-736.

[14] S. P. Banbury, D. C. Berry: Disruption of office-related tasks by speech and office noise. Br. J. Psychol. 89 (1998) 499-517.

[15] A. Haapakangas, R. Helenius, E. Keskinen, V. Hongisto: Perceived acoustic environment, work performance and well-being - survey results from Finnish offices. 9th ICBEN, Foxwoods, CT, USA, 2008.

[16] V. Hongisto, J. Varjo, D. Oliva, A. Haapakangas, E. Benway: Perception of water-based masking sounds – long-term experiment in an open-plan office. Front. Psychol. 8 (2017) 1-14.

[17] M. Haka, A. Haapakangas, J. KerÃďnen, J. Hakala, E. Keskinen, V. Hongisto: Performance effects and subjective disturbance of speech in acoustically different office types – a laboratory experiment. Indoor Air 19 (2009) 454-467.

[18] H. Jahncke, V. Hongisto, P. Virjonen: Cognitive performance during irrelevant speech: Effects of speech intelligibility and office-task characteristics. Appl. Acoust. 74 (2013) 307-316.

[19] M. Yadav, J. Kim, D. Cabrera, R. de Dear: Auditory distraction in open-plan office environments: The effect of multi-talker acoustics. Appl. Acoust. 126 (2017) 68-80.

[20] M. Zaglauer, H. Drotleff, A. Liebl: Background babble in open-plan offices: A natural masker of disruptive speech? Appl. Acoust. 118 (2017) 1-7.

[21] Draft of VDI 2569:2016-02. Schallschutz und akustische Gestaltung im Büro. Verein Deutscher Ingenieure e.V., Düsseldorf, Germany, 2016.

[22] DIN 45645-2:2012-09. Ermittlung von Beurteilungspegeln aus Messungen – Teil 2: Ermittlung des Beurteilungspegels am Arbeitsplatz bei Tätigkeiten unterhalb des Pegelbereiches der Gehörgefährdung. DIN Deutsches Institut für Normung e.V., Berlin, Germany, 2009.

[23] A. Kaarlela-Tuomaala, R. Helenius, E. Keskinen, V. Hongisto: Effects of acoustic environment on work in private office rooms and open-plan offices – longitudinal study during relocation. Ergonomics 52 (2009) 1423-1444.

[24] T. Renz, P. Leistner, A. Liebl: Auditory distraction by speech: Sound masking with speech-shaped stationary noise outperforms –5 dB per octave shaped noise. J. Acoust. Soc. Am. 143 (2018) EL212-EL217.

[25] T. Renz, P. Leistner, A. Liebl: Auditory distraction by speech: Can a babble masker restore working memory performance and subjective perception to baseline? Appl. Acoust. 137 (2018) 151-160.

[26] A. Liebl, A. Assfalg, S. J. Schlittmeier: The effects of speech intelligibility and temporal-spectral variability on performance and annoyance ratings. Appl. Acoust. 110 (2016) 170-175.

[27] A. Liebl, M. Zaglauer, N. Martin: More is less – Positive effects of higher volume ventilation sound on cognitive performance and acoustic comfort in offices. Proc. INTER-NOISE 2016, 3527-3535.

[28] N. Martin, A. Liebl: Sound masking in office environments – Trade-off between masking effect and user acceptance. Proc. Acoust. 2017, 3598-3598.

[29] R. Zuydervliet, R. Chanaud, A. L'Espérance: Adaptive masking in open offices. Acoust., Paris, France, 2008.

[30] A. L'Espérance, A. Boudreau, L.-A. Boudreault, F. Gariépy, R. Mackenzie: Adaptive volume control for sound masking. NOISE-CON, Grand Rapids, MI, USA, 2017.

[31] K. Aho, D. Derryberry, T. Peterson: Model selection for ecologists: the worldview of AIC and BIC. Ecology 95 (2014) 631-636.

[32] A. Ebissou, E. Parizet, P. Chevret: Use of the speech transmission index for the assessment of sound annoyance in open-plan offices. Appl. Acoust. 88 (2015) 90-95.

[33] W. Ellermeier, J. Hellbrück: Is level irrelevant in "Irrelevant Speech"? Effects of loudness, signal-to-noise ratio, and binaural unmasking. J. Exp. Psychol. Hum. Percept. Perform. 24 (1998) 1406-1414.