PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Comparison of models to predict the effect of background speech on work performance in open-plan offices

Tobias RENZ

Renz Solutions GmbH, Germany

Abstract

There are different models to estimate the relationships between measurable acoustical parameters and office workers' perception, well-being and performance. These models are based on speech transmission index (STI), fluctuation strength (FS) or percentile level statistics. Whilst STI requires a loudspeaker and is measured in unoccupied conditions, other metrics, such as FS or percentile level statistics, can be determined in situ during usual working hours. Nonetheless, the established models have some shortcomings. STI can estimate the effect of one simulated speech source on short-term memory performance but it cannot assess the effect of office noise exposure in occupied offices. FS correlates with the impact of temporal-spectral variability of a background sound on short-term memory performance but it depends on the sound pressure level which does not have an effect on the short-term memory performance. Percentile level statistics correlate with the speech-to-noise ratio: higher differences between the 10th and 90th percentile levels measured with fast time weighting lead to lower number recall performances. As part of this study 110 sound conditions under which subjects have to complete a number recall task are evaluated with respect to their relationships with these three acoustical parameters. A cross-validation reveals comparable prediction qualities of the models.

Keywords: Open-plan office, Work performance, Irrelevant sound effect

1 INTRODUCTION

1.1 Effects of Noise on Office Workers

Background speech in open-plan offices can deteriorate the subjective well-being and work performance (1). Speech from colleagues' conversations or telephone calls is the most distracting noise source (2).

Auditory background with changing-state features deteriorates working memory performance. This effect of auditory distraction is especially pronounced in memory for order information, which is known as the irrelevant sound effect (3). Performance decrements occur when the required cognitive processes conflict with those that are involved in processing the auditory stimuli, according to the interference-by-process principle (4). Hence, any sound with sufficient temporal-spectral variability is expected to deteriorate the serial memory. Testing the mean error rates of subjects in a serial recall task in laboratory conditions is a well-established method to determine the auditory distraction by background sounds. A serial recall test consists of a series of digits or letters that are presented sequentially on a computer screen (auditory presentation is also possible) and have to be recalled at the end of the sequence presentation after a short delay (retention interval).

Room acoustics simulations enable acoustical consultants to predict the effects of acoustical materials and products on various parameters. The standard ISO 3382-3 suggests the use of reverberation time, spatial decay of sound pressure level (SPL) and parameters based on the speech transmission index (STI) to evaluate the acoustical design of open-plan offices. STI, which is a metric to predict the speech intelligibility, is suggested to predict both subjective noise disturbance and work performance (e.g. Refs. (1, 2, 5, 6)).

1.2 Review of Existing Prediction Models

Hongisto's model (5) based on the STI predicts the decrease of performance (DP) due to speech of varying intelligibility. The model is based on three laboratory studies that analyse the performance in number recall or proofreading. Additionally, one field study is included that evaluates the self-reported daily waste of working time before and after an office relocation (2). Since its introduction in 2005 this model has been applied to the acoustical design and evaluation of open-plan offices (e.g. in the standard ISO 3382-3) and has become a well-established tool for the acoustical assessment of open-plan offices.

In practice, it is difficult to measure or simulate the STI because the occurring background noise levels during a usual workday are often unknown. The STI is not defined for fluctuating background noise, as observed in occupied open-plan offices. In addition, the standard DIN 60268-16 mentions that STI measurements are not suitable to assess systems that produce speech privacy through sound masking because the STI is not verified for such conditions. Moreover, the selection of transmission paths remains challenging because it is not clear which talkers have to be considered as speech sources and which talkers vanish in the background noise. Hongisto's model (5) cannot account for multi-talker environments. As the model is fitted to data that contains one talker under steady-state background noise, the STI is able to predict the speech intelligibility. The DP model may estimate the disturbing impact of speech intelligibility but not of temporal-spectral variability. Since temporal-spectral variability correlates with speech intelligibility in sound conditions with one talker under steady-state background noise, the model correlates highly with the DP.

Schlittmeier *et al.* (7) model the irrelevant sound effect by using the hearing sensation fluctuation strength (FS). FS is a predictor of the hearing sensation of fluctuations that are caused by low amplitude or frequency modulations below approx. 20 Hz modulation frequency. FS peaks at about 2 Hz to 5 Hz which is equivalent to the typical syllabic rate of speech (8). The model applies to speech as well as non-speech sounds. The FS correlates well with the sensation of amplitude or frequency modulations but not with the DP due to temporal-spectral variability of sounds. FS is not very sensitive within the range of typical office sounds consisting of speech and non-speech sounds. The FS can vary by decades between masked and unmasked speech while changes in the speech-to-noise ratio result in rather small changes in FS. In the following, the speech-to-noise ratio denotes the level difference between the A-weighted SPL of a distracting speech sound and the A-weighted SPL of a masking or background sound.

Liebl *et al.* (9) review both models, concluding that Hongisto's model (5) cannot account for distracting non-speech sounds such as office noise or background music while Schlittmeier's model (7) cannot account for the additional performance decrement of speech as compared to speech-like noise. The main finding of the study can be outlined by the mean error rates that are observed during variable speech-like noise: STI cannot account for the disturbing impact of the temporal-spectral variability while FS overestimates the DP because the FS is similar to a speech signal that causes a much higher DP. Liebl *et al.* (9) mention that the approach based on the STI is often misunderstood by facility managers and office owners which equate the STI with speech intelligibility and semantic content of background speech. However, the irrelevant sound effect denotes a phenomenon which is rather based on temporal-spectral characteristics of the acoustic background regardless of whether subjects understand the background speech.

In contrast to Finland's National Building Code that sets STI limits, German standards (such as VDI 2569 or DIN 45645-2) avoid the use of the STI. The rating level L_r, that is based on A-weighted energy-equivalent SPL measurements and the consideration of penalties for impulsiveness, tonality and informational content, provides the basis for the assessment of noise immissions in office buildings in Germany. The set values that are described in the standard guideline VDI 2058-3 are not expected to correlate with the subjective perception or working memory performance of office workers that are subjected to office noise (10). Hence, there is a need to improve the assessment of this parameter or to identify a different metric that enables in situ measurements and correlates with the effect of office noise on workers.

1.3 Aim of This Study

This paper compares a new approach to predict the working memory performance under distracting background speech based on percentile level statistics with the established models that are based on STI and FS. Based on practical experiences in German open-plan offices, the new model can be applied during the everyday work of acoustical consultants and occupational safety and health institutes. By using percentile levels (SPL that is exceeded for a defined percentage of time during the measurement time), the SPL of fluctuating speech peaks and the SPL of stationary background noise are estimated. Zuydervliet *et al.* (11) suggest that background levels can be described by the 90th percentile $L_{AF,90\%}$ and that activity levels can be described by the 10th percentile $L_{AF,10\%}$. The difference of these two levels correlates with the speech-to-noise ratio at the receiver position. Kaarlela-Tuomaala *et al.* (2) analyse similar level statistics in a study that compares the subjective perception of an acoustic environment during a relocation from private office rooms to an open-plan office and conclude that the variability of SPL is not related to the self-rated disturbance caused by noise. However, the study considers only the percentile levels $L_{A,1\%}$ and $L_{A,99\%}$ and the measurements do not distinguish between speech sounds from the person working at the respective workstation and distracting background speech from colleagues.

In contrast to Hongisto's model (5), the suggested model does not require a loudspeaker to simulate a speech source, and hence it can be applied to occupied open-plan offices and it can consider multi-talker sound environments. Similar to Schlittmeier's model (7), the overall sound condition is analysed at receiver points instead of the STI on distinct transmission paths. Contrary to the STI, the model does not take reverberation into account. Reverberation times in open-plan offices are usually between 0.3 s and 0.7 s, and hence do not have a notable impact on the resulting STI values. Compared to the FS, the presented predictor is easier to determine because the computational cost are notably lower allowing for measurements with conventional sound level meters. The model is developed with data points that contain one speech sound similar to Hongisto's study (5).

2 METHODS AND MATERIALS

2.1 Sound Conditions

Twelve laboratory experiments from ten different studies are considered. Table 1 provides an overview of the studies. All studies comprise two control conditions: the first control condition is a distracting speech signal and the second control condition consists of silence or stationary noise at moderate SPL (25 dB). The other sound conditions contain speech with different speech intelligibility. The speech intelligibility is reduced by adding a masking sound to the signal. Different masking sounds are used, such as stationary noise, babble, reversed speech, nature sounds, and ventilation sounds. The speech-to-noise ratio varies between –12 dB and +4 dB. One sound condition from the study by Martin and Liebl (12) is excluded from the further analyses because the participants are subjected to their favourite music, and hence the sound condition is different for each subject.

Table 1 – Overview of the included studies

No.	Authors (reference)	Primary comparison	Presentation
1	Renz *et al.* (13, 14)	reversed speech and stationary noise	headphones
2	Renz *et al.* (15)	stationary and babble masker	headphones
3	Liebl *et al.* (9)	steady and variable speech-like noise	headphones
4	Liebl *et al.* (16)	different ventilation sounds	headphones
5	Martin and Liebl (12)	various masking signals	headphones
6	Renz *et al.* (17)	spatial masking release	headphones
7	Renz *et al.* (18)	local and conventional masking	loudspeakers
8	Ebissou *et al.* (19)	use of STI for annoyance assessment	loudspeaker
9	Ellermeier and Hellbrück (20)	effect of speech-to-noise ratio	headphones
10	Jahncke *et al.* (6)	effect of speech intelligibility	headphones

The experimental design of all experiments is a one-way repeated measures design with five to twelve levels according to the tested sound conditions, i.e. all subjects perform the test during the same sound conditions. During each sound condition the subjects perform a serial recall task where they have to memorise a sequence of nine digits and recall it in the exact order of presentation after a short retention interval. Each item that is not recalled correctly at the presented serial position is counted as an error. The retention interval varies between the studies (0–10 s). The study by Jahncke *et al.* (6) does not include a serial recall task but uses free recall by visual presentation of lists of ten words. For the studies with serial recall the percentage of incorrectly recalled digit positions of all tested sequences of one condition (mean error rate) is determined. In the study by Ellermeier and Hellbrück (20) subjects can leave blanks empty. In this case a random integer between 1 and 9 is chosen for this serial position. Results that are only depicted in graphs are estimated (studies by Ellermeier and Hellbrück (20) and Jahncke *et al.* (6)). The sound signals that are used by Ellermeier and Hellbrück (20) are generated manually, and thus they may differ from the original sound conditions.

The absolute performance change, the DP, is calculated as the difference in the mean performance between the silent control condition and the analysed sound condition or as the difference in the mean error rates between the sound condition and the silent control condition:

$$DP = P_{silence} - P_{condition}. \tag{1}$$

In some studies the silent control condition comprises stationary noise at moderate levels which does not have an effect on the working memory performance (e.g. Ref. (20)).

2.2 Predictor Variables

The percentile level difference $L_{AF,10\%}$–$L_{AF,90\%}$, the STI, and the FS are considered as predictor. SPL and FS values are computed with the software ArtemiS version 12.05.1512 (HEAD acoustics GmbH, Herzogenrath, Germany). The STI values are reported in most studies or otherwise they are determined by taking the SPLs of the speech and masking signals into account (21). In binaural listening conditions the maximum values at both ears are used because prior analyses showed that the serial recall performance correlates with the STI and FS values at the advantaged ear (cf. Ref. (17)) and because the standard DIN 60268-16 suggests the use of the higher STI value if binaural recordings are performed with a dummy head.

2.3 Model Fitting

A three-parameter logistic model based on the logistic function is considered:

$$f(x) = \frac{L}{1+exp\left(-k(x-x_0)\right)}, \tag{2}$$

with the curve's maximum value L, the steepness of the curve k, and the x-value of the curve's midpoint x_0.

The use of a logistic function seems appropriate because it can model the S-shaped curve that is commonly observed for sentence intelligibility and DP under speech with varying speech-to-noise ratios (cf. Refs. (5, 22)). The models are determined in R (23) with the software RStudio Version 1.1.383 (RStudio, Inc., Boston, MA, USA). The model fitting is based on two steps of a cross-validation. Firstly, the 89 DP values of all studies that are performed at the Fraunhofer Institute for Building Physics (first seven studies in Table 1) are considered as response variable to determine and validate the models. Secondly, the 21 DP values of the three studies that are performed at other research groups with different experimental conditions (last three studies in Table 1) are considered to validate the determined models.

A repeated eleven-fold cross-validation with ten repetitions is performed to determine the models. Repeating a k-fold cross-validation can be used to effectively increase the precision of the estimates while maintaining a small bias (24). The model fitting is based on the following four steps, as described by Kuhn and Johnson (25):

1. Ten randomly ordered lists are created that contain the 89 data points of the predictor and the DP values.

2. For each of these 10 lists, 8 train sets with 78 data points and 8 test sets with the remaining 11 data points are created, while always considering 11 different data points.

3. The respective model fits for each of the 80 train sets are determined.

4. The average values of the model parameters are calculated.

The validation procedure is as follows:

1. For each of the 80 test sets the root-mean-square error (RMSE), which describes the standard deviation of the differences between the predicted values and the observed values of the test set, is computed.
2. The average values of all 80 determined RMSE values are calculated and compared between the considered models.

The RMSE is determined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}, \tag{3}$$

with the predicted values $\hat{y}_i$ and the observed values y_i over the $n = 11$ data points.

Besides the RMSE values, the Pearson correlation coefficients r_{xy} and Spearman's rank correlation coefficients r_S are determined to compare the different models. The Pearson correlation coefficient is not suitable to compare non-linear relationships. The values are still computed because correlation coefficients are reported in similar studies (e.g. Refs. (1, 7)).

In the second step, the RMSE, r_{xy} and r_S values are calculated across the 21 additional data points.

3 RESULTS

The results of the cross-validation are outlined in Table 2. STI results in the lowest vlues of r_S and RMSE. STI, FS and $L_{AF,10\%}$–$L_{AF,90\%}$ reveal comparable results with respect to r_S and RMSE. $L_{AF,10\%}$–$L_{AF,90\%}$ appears to be suitable to predict the working memory performance under distracting background speech.

Figures 1 and 2 illustrate the determined models in step 1 and step 2, respectively. The DP over the STI and $L_{AF,10\%}$–$L_{AF,90\%}$ seems to follow an S-shaped curve. It should be noted that FS and $L_{AF,10\%}$–$L_{AF,90\%}$ show a wide range of values because they are not limited to values between 0 and 1. The FS values range from 0 to 0.37, but the change in DP occurs over a much smaller range (0–0.01).

Table 2 – Cross-validation results of the working memory performance models: Pearson correlation coefficient r_{xy}, Spearman's rank correlation coefficient r_S, and the RMSE averaged over all 80 test sets. *, **, *** Represent significance at the $p < .05$, .01, and .001 levels, respectively

Predictor variable	Response variable	$r_{xy,1}$	$r_{S,1}$	$RMSE_1$	$r_{xy,2}$	$r_{S,2}$	$RMSE_2$
STI	DP	0.74^{***}	0.82^{***}	2.5	0.88^{***}	0.90^{***}	3.9
FS	DP	0.36^{***}	0.74^{***}	2.6	0.79^{***}	0.84^{***}	4.9
$L_{AF,10\%}$–$L_{AF,90\%}$	DP	0.53^{***}	0.75^{***}	2.7	0.47^{***}	0.73^{***}	4.0

4 DISCUSSION

STI, FS and $L_{AF,10\%}$–$L_{AF,90\%}$ show comparable prediction accuracy of working memory performance under distracting background speech. The presented model based on $L_{AF,10\%}$–$L_{AF,90\%}$ can be used to estimate the resulting DP in occupied offices. The metric is easy to measure in situ. Since occupational safety and health in Germany requires the measurement of rating levels L_r in occupied offices the suggested metric can be determined without additional cost. Whilst the STI is determined on one distinct transmission path, the presented model can consider multi-talker environments.

Some shortcomings remain that may limit the applicability of the presented model and require further research. First, all considered data includes only a single voice that is masked by multiple sounds; multi-talker

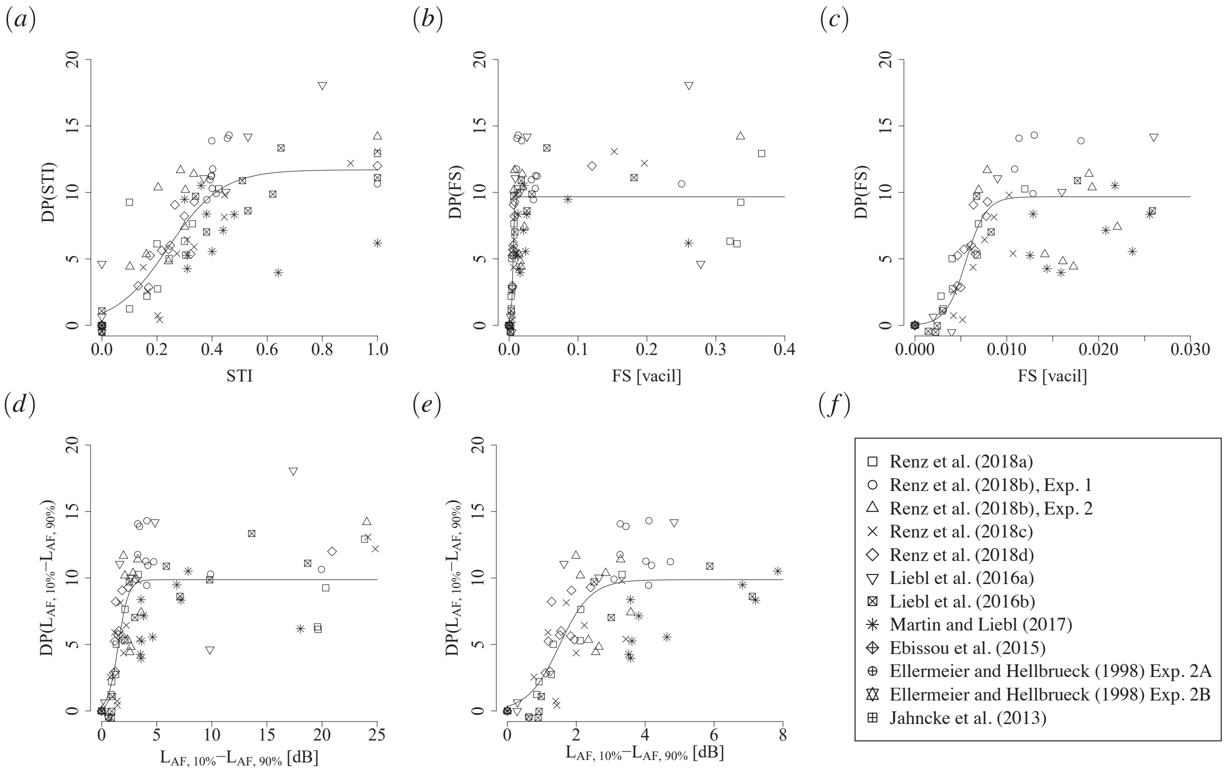

Figure 1 – Plots of the fitted and cross-validated working memory performance models of the response variable DP over the predictor variables STI (a), FS (b), zoomed area of FS (c), $L_{AF,10\%}$–$L_{AF,90\%}$ (d), and zoomed area of $L_{AF,10\%}$–$L_{AF,90\%}$ (e). The solid line shows the determined model. A legend is depicted in (f)

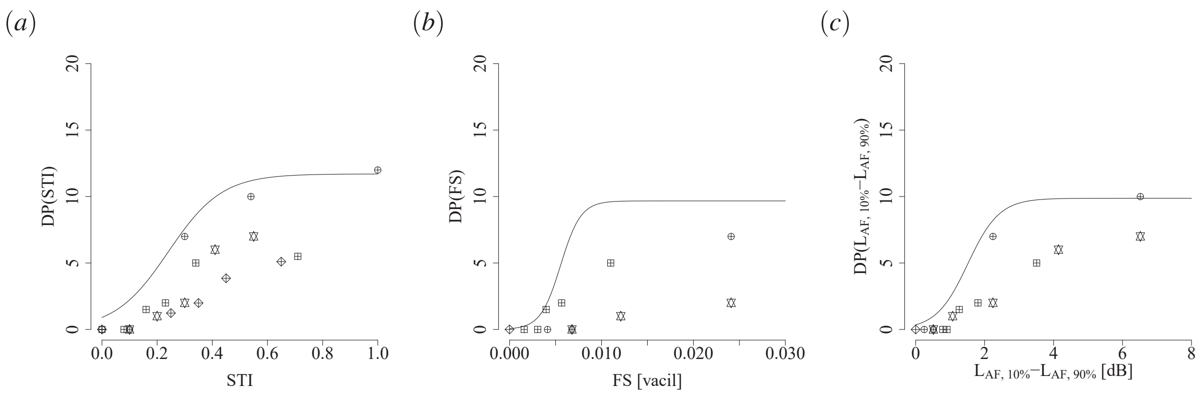

Figure 2 – Plots of the additionally considered data points of the response variable DP over the predictor variables STI (a), FS (b), and $L_{AF,10\%}$–$L_{AF,90\%}$ (c). The solid line shows the determined model in step 1

environments are not included. It remains unclear if the model can be applied to complex binaural environments, for instance, an office environment with multiple talkers that are spatially separated. Second, $L_{AF,10\%}$–$L_{AF,90\%}$ estimates a ratio between the speech level peaks and the ambient noise level but it may not sufficiently account for the effects of spectral fluctuations and semantic content. Further analyses are necessary to verify whether this model can be applied to in situ measurements to assess the acoustic satisfaction and work performance under distracting background speech.

5 CONCLUSIONS

Hongisto's model (5) based on the STI is an established model for the acoustical design of open-plan offices. According to the performed analyses, the maximum DP occurs at STI values above 0.5, and hence occurring STI values should be below 0.5. Nevertheless, the STI cannot be used to measure the office workers' exposure to noise and to assess its impact on the individual's subjective perception and performance impairment, which is commonly done by occupational safety and health practitioners by means of SPL measurements.

Schlittmeier's model (7) based on the FS is difficult to apply in practice because the FS depends on the absolute SPL and there are only a few software tools available to compute the FS. In contrast to the FS, the percentile level difference $L_{AF,10\%}$–$L_{AF,90\%}$ can be determined in situ as part of SPL measurements with conventional sound level meters.

ACKNOWLEDGEMENTS

I would like to thank the Institute for Acoustics and Building Physics at the University of Stuttgart, and particularly Philip Leistner who supervised and supported my PhD project. Philip Leistner and Andreas Liebl provided valuable advice on the design of the laboratory experiments (13–15, 17, 18).

Helena Jahncke, Noemi Martin, Patrick Chevret, and Andreas Liebl provided audio material that enabled these analyses. The author greatly appreciates their help. Without their cooperation these analyses would not have been possible to the same extent.

REFERENCES

[1] Haapakangas A, Hongisto V, Eerola M, Kuusisto T. Distraction distance and perceived disturbance by noise—An analysis of 21 open-plan offices. J Acoust Soc Am. 2017;141(1):127–136.

[2] Kaarlela-Tuomaala A, Helenius R, Keskinen E, Hongisto V. Effects of acoustic environment on work in private office rooms and open-plan offices – longitudinal study during relocation. Ergonomics. 2009;52(11):1423–1444.

[3] Beaman CP, Jones DM. Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. J Exp Psychol Learn Mem Cogn. 1997;23(2):459–471.

[4] Marsh JE, Jones DM. Cross-modal distraction by background speech: What role for meaning? Noise Heal. 2010;12(49):210–216.

[5] Hongisto V. A model predicting the effect of speech of varying intelligibility on work performance. Indoor Air. 2005;15(6):458–468.

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

[7] Schlittmeier SJ, Weissgerber T, Kerber S, Fastl H, Hellbrück J. Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. Attention, Perception, Psychophys. 2012;74(1):194–203.

[8] Edwards E, Chang EF. Syllabic (˜2–5 Hz) and fluctuation (˜1–10 Hz) ranges in speech and auditory processing. Hear Res. 2013;305:113–134.

[9] Liebl A, Assfalg A, Schlittmeier SJ. The effects of speech intelligibility and temporal-spectral variability on performance and annoyance ratings. Appl Acoust. 2016;110:170–175.

[10] Renz T, Leistner P, Liebl A. Use of energy-equivalent sound pressure levels and percentile level differences to assess the impact of speech on cognitive performance and annoyance perception. Appl Acoust. 2019;153:71–77.

[11] Zuydervliet R, Chanaud R, L'Espérance A. Adaptive masking in open offices. In: Acoustics. Paris (France); 2008. .

[12] Martin ND, Liebl A. Sound masking in office environments – Trade-off between masking effect and user acceptance. J Acoust Soc Am. 2017;141(5):3598–3598.

[13] Renz T, Leistner P, Liebl A. Auditory distraction by speech: Sound masking with speech-shaped stationary noise outperforms –5 dB per octave shaped noise. J Acoust Soc Am. 2018;143(3):EL212–EL217.

[14] Renz T, Leistner P, Liebl A. Auditory distraction by speech: Comparison of fluctuating and steady speech-like masking sounds. J Acoust Soc Am. 2018;144(2):EL83–EL88.

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

[16] Liebl A, Zaglauer M, Martin ND. More is less – Positive effects of higher volume ventilation sound on cognitive performance and acoustic comfort in offices. In: INTER-NOISE NOISE-CON Congr. Conf. Proc. Hamburg (Germany); 2016. p. 3527–3535.

[17] 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 Acust united with Acust. 2018;104(4):611–622.

[18] Renz T, Leistner P, Liebl A. Effects of the location of sound masking loudspeakers on cognitive performance in open-plan offices: Local sound masking is as efficient as conventional sound masking. Appl Acoust. 2018;139:24–33.

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

[20] Ellermeier W, Hellbrück J. Is level irrelevant in "Irrelevant Speech"? Effects of loudness, signal-to-noise ratio, and binaural unmasking. J Exp Psychol Hum Percept Perform. 1998;24(5):1406–1414.

[21] Houtgast T, Steeneken HJM, Plomp R. Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. Acta Acust united with Acust. 1980;46(1):60–72.

[22] Plomp R, Mimpen AM. Improving the reliability of testing the speech reception threshold for sentences. Audiology. 1979;18(1):43–52.

[23] R Core Team. R: A Language and Environment for Statistical Computing. Vienna (Austria); 2016.

[24] Kim JH. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Comput Stat Data Anal. 2009;53(11):3735–3745.

[25] Kuhn M, Johnson K. Applied Predictive Modeling. New York, NY (USA): Springer New York; 2013.